APUNTES OPOSICIÓN CUERPO DE DIPLOMADOS COMERCIALES DEL ESTADO

Tercer Ejercicio

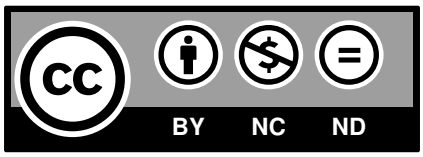
Parte A) Microeconomía y Economía del Sector Público

Manuel Cabado García

Última revisión: 12/10/2026

© 2026 Manuel Cabado García

Esta obra se distribuye bajo licencia Creative Commons Atribución/Reconocimiento-NoComercial-SinDerivados 4.0 Internacional (CC BY-NC-ND 4.0)
(https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es)

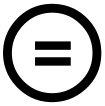




Atribución: Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial: Usted no puede hacer uso del material con propósitos comerciales.



SinDerivadas: Si remezcla, transforma o crea a partir del material, no podrá distribuir el material modificado.

No hay restricciones adicionales: No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Estos apuntes reelaboran materiales de distintas fuentes, que se citan en la bibliografía de cada tema. El contenido de terceros conserva los derechos de sus autores y no queda cubierto por esta licencia.

Los enlaces a sitios web de terceros se proporcionan de buena fe y solo a título informativo. El autor declina toda responsabilidad por los materiales contenidos en cualquier sitio web de terceros al que se haga referencia en este trabajo.

La última versión de este libro, gratuita, se encuentra siempre en la siguiente dirección:
https://manuelcabadogarcia.es

Contenido

Índice de figuras **10**

Índice de cuadros **15**

1 Pensamiento económico (I): los economistas clásicos y neoclásicos. Referencia a Marx. **17**

1.1 Introducción 17
1.2 Los economistas clásicos 18
1.3 Los predecesores de la revolución marginalista 33
1.4 La revolución marginalista 36
1.5 Los economistas neoclásicos 43
1.6 Conclusiones 53
Apéndice 55
Bibliografía 70

2 Pensamiento económico (II): de Keynes a la actualidad. Referencia a la Síntesis Neoclásica, al Monetarismo, a la Nueva Macroeconomía Clásica y a los modelos Neokeynesianos de primera y segunda generación. **73**

2.1 Introducción 73
2.2 John Maynard Keynes 74
2.3 Síntesis Neoclásica 81
2.4 Monetaristas 85
2.5 Nueva Macroeconomía Clásica 89
2.6 Nueva Economía Keynesiana 93
2.7 Otros desarrollos 98
2.8 Conclusiones 101
Apéndice 102
Bibliografía 121

3 Teoría de la demanda del consumidor. Axiomas sobre las preferencias, función de utilidad y función de demanda marshalliana. Otros desarrollos. **125**

3.1 Introducción . . . 125
3.2 Enfoque cardinal de la teoría neoclásica . . . 126
3.3 Enfoque ordinal de la teoría neoclásica . . . 127
3.4 Estática comparativa . . . 139
3.5 Condiciones de un sistema de demanda completo . . . 146
3.6 Demanda de mercado . . . 147
3.7 Conclusiones . . . 148
Apéndice . . . 150
Bibliografía . . . 162

4 Desarrollos de la teoría de la demanda: dualidad. Referencia a la integrabilidad de las preferencias y medidas de cambio en el bienestar. Elección en situaciones de riesgo. Otros desarrollos recientes. 163
4.1 Introducción . . . 163
4.2 Teoría de la dualidad . . . 164
4.3 Elección en situaciones de riesgo . . . 184
4.4 Conclusiones . . . 198
Apéndice . . . 200
Bibliografía . . . 212

5 Teoría de la producción. Caracterización de la tecnología de la empresa en el corto y el largo plazo. El conjunto de posibilidades de producción. La función de producción. Rendimientos locales y globales a escala. Elasticidad de sustitución. Producción conjunta. 217
5.1 Introducción . . . 217
5.2 Construcción de la función de producción . . . 218
5.3 Análisis de la función de producción . . . 223
5.4 Producción multiproducto . . . 233
5.5 Conclusiones . . . 236
Apéndice . . . 237
Bibliografía . . . 242

6 Teoría de los costes. Análisis de la dualidad en el ámbito de la empresa. Análisis temporal. Otros desarrollos. 243
6.1 Introducción . . . 243
6.2 Teoría neoclásica de los costes . . . 244
6.3 Críticas a la teoría tradicional . . . 258
6.4 Otros desarrollos . . . 262
6.5 Conclusiones . . . 266
Apéndice . . . 268

 © 2026 Manuel Cabado García

Bibliografía 277

7 Análisis de mercados (I). El modelo de competencia perfecta. Análisis de equilibrio parcial: corto y largo plazo, dinámicas de ajuste y estabilidad del equilibrio. Análisis de eficiencia y bienestar. 279

7.1 Introducción 279
7.2 Caracterización del modelo de competencia perfecta 280
7.3 Equilibrio de la empresa y de la industria 282
7.4 Equilibrio competitivo 291
7.5 Estática comparativa 295
7.6 Relajación de supuestos 297
7.7 Conclusiones 303
Apéndice 304
Bibliografía 309

8 Análisis de mercados (II). Teoría del monopolio. Discriminación de precios. Monopolio natural. Análisis de eficiencia y bienestar. La regulación del monopolio. 311

8.1 Introducción 311
8.2 Caracterización del monopolio 312
8.3 Equilibrio del mercado 313
8.4 Tipos de monopolio 318
8.5 Regulación y control 324
8.6 Conclusiones 336
Apéndice 337
Bibliografía 352

9 Análisis de mercados (III). Teoría del oligopolio: soluciones no cooperativas y soluciones cooperativas. Análisis mediante la aplicación de la teoría de juegos. 355

9.1 Introducción 355
9.2 Caracterización del oligopolio 356
9.3 Oligopolios no cooperativos 357
9.4 Oligopolios cooperativos 364
9.5 Sostenibilidad del oligopolio 371
9.6 Conclusiones 376
Apéndice 377
Bibliografía 386

10 Análisis de mercados (IV). Diferenciación de productos. La competencia monopolística y otros desarrollos. 389
10.1 Introducción . . . 389
10.2 Justificación de la diferenciación de productos . . . 390
10.3 Diferenciación horizontal espacial . . . 391
10.4 Diferenciación horizontal no espacial . . . 397
10.5 Diferenciación vertical . . . 410
10.6 Conclusiones . . . 412
Apéndice . . . 414
Bibliografía . . . 422

11 Teorías microeconómicas del mercado de trabajo (I). La teoría neoclásica del mercado de trabajo. Análisis intertemporal de la oferta de trabajo. Teoría del capital humano. 425
11.1 Introducción . . . 425
11.2 Teoría neoclásica . . . 426
11.3 Análisis intertemporal . . . 442
11.4 Capital humano . . . 446
11.5 Conclusiones . . . 452
Apéndice . . . 453
Bibliografía . . . 463

12 Teorías microeconómicas del mercado de trabajo (II). La curva de Beveridge, el modelo de búsqueda y emparejamiento y costes de ajuste. Mención a los modelos de negociación, salarios de eficiencia y contratos implícitos. 469
12.1 Introducción . . . 469
12.2 Beveridge y búsqueda y emparejamiento . . . 470
12.3 Costes de ajuste . . . 477
12.4 Negociación, salarios de eficiencia y contratos . . . 480
12.5 Conclusiones . . . 486
Apéndice . . . 487
Bibliografía . . . 494

13 Economía de la información y teoría de la agencia: selección adversa y riesgo moral. 497
13.1 Introducción . . . 497
13.2 Selección Adversa . . . 498
13.3 Riesgo Moral . . . 506
13.4 Selección adversa y riesgo moral a la vez . . . 513

 © 2026 Manuel Cabado García

13.5 Conclusiones . . . 515
Apéndice . . . 517
Bibliografía . . . 526

14 Economía del bienestar (I). Óptimo económico y equilibrio general competitivo. Los teoremas fundamentales del bienestar. Teoría del "second best". 529
14.1 Introducción . . . 529
14.2 Óptimo económico . . . 530
14.3 Equilibrio competitivo y Teoremas del Bienestar . . . 541
14.4 Los criterios de compensación . . . 545
14.5 Teoría del Second Best . . . 547
14.6 Conclusiones . . . 550
Apéndice . . . 551
Bibliografía . . . 563

15 Economía del bienestar (II). Los fallos del mercado: externalidades y bienes públicos. Intervención y fallos del sector público. 567
15.1 Introducción . . . 567
15.2 Concepto de Eficiencia Económica . . . 568
15.3 Optimalidad de la competencia perfecta . . . 570
15.4 Fallos de mercado: bienes públicos y externalidades . . . 571
15.5 Otros fallos de mercado . . . 586
15.6 Fallos del sector público . . . 587
15.7 Conclusiones . . . 589
Apéndice . . . 590
Bibliografía . . . 605

16 Economía del bienestar (III): teoría de la elección colectiva. El teorema de imposibilidad de Arrow. Funciones de bienestar social, reglas de votación y otros desarrollos posteriores. 609
16.1 Introducción . . . 609
16.2 El problema de la Elección Colectiva . . . 610
16.3 Criterios que debe cumplir la Regla de la Elección Colectiva . . . 612
16.4 El Teorema de Imposibilidad de Arrow . . . 612
16.5 Elección mediante votación . . . 613
16.6 Conclusiones . . . 629
Apéndice . . . 630
Bibliografía . . . 639

17 El gasto público. Razones de su crecimiento. Teoría de la burocracia y constitución fiscal. El debate sobre el tamaño del Sector Público, en lo referente al gasto. 643
17.1 Introducción . . . 643
17.2 Teorías que explican las causas del aumento del Sector Público . . . 644
17.3 Debate sobre el tamaño del Sector Público . . . 653
17.4 La Constitución Fiscal . . . 657
17.5 Conclusiones . . . 662
Apéndice . . . 663
Bibliografía . . . 668

18 Poder de mercado y regulación óptima: instrumentos, efectos y limitaciones. Desarrollos en presencia de información asimétrica. Aplicaciones prácticas. 671
18.1 Introducción . . . 671
18.2 Información completa . . . 672
18.3 Información asimétrica . . . 681
18.4 Varias empresas . . . 686
18.5 Aplicaciones prácticas . . . 688
18.6 Conclusiones . . . 693
Apéndice . . . 694
Bibliografía . . . 700

19 El Estado del Bienestar y los instrumentos de redistribución y protección social. Medición e indicadores del impacto redistributivo. 705
19.1 Introducción . . . 705
19.2 Justificación . . . 706
19.3 Prestaciones económicas . . . 707
19.4 Provisión de bienes preferentes . . . 717
19.5 Impacto redistributivo . . . 722
19.6 Conclusiones . . . 726
Apéndice . . . 727
Bibliografía . . . 737

20 Teoría de la imposición: traslación e incidencia de los impuestos. Enfoques de equilibrio parcial y general. Efecto renta y efecto sustitución de los impuestos. Concepto y medición del exceso de gravamen. 741
20.1 Introducción . . . 741
20.2 Exceso de gravamen: concepto y medición . . . 742
20.3 Incidencia y traslación de impuestos . . . 750

 © 2026 Manuel Cabado García

20.4 Conclusiones 765
Apéndice 767
Bibliografía 779

21 Imposición y oferta. Efectos incentivos de los impuestos. Medidas de progresividad impositiva. El compromiso entre eficiencia y equidad de la política tributaria. 781
21.1 Introducción 781
21.2 Principios impositivos 782
21.3 Eficiencia económica 782
21.4 Equidad 793
21.5 Compromiso entre eficiencia y equidad 797
21.6 Conclusiones 799
Apéndice 800
Bibliografía 814

22 La teoría del federalismo fiscal: la asignación óptima de las funciones entre los distintos niveles de gobierno para alcanzar el bienestar máximo. La financiación de las haciendas territoriales: impuestos, transferencias y deuda. 817
22.1 Introducción 817
22.2 Justificación económica de la descentralización 818
22.3 Tamaño de la jurisdicción y el reparto de competencias 825
22.4 Financiación de las haciendas territoriales 829
22.5 Conclusiones 841
Apéndice 843
Bibliografía 853

Glosario 855

Índice de figuras

2.1 Modelo IS-LM 82
2.2 Curva de Phillips 85
2.3 Curva de Phillips monetarista 89

3.1 Función de utilidad en 3 dimensiones 132
3.2 Mapa de curvas de indiferencia 133
3.3 Tipos de funciones de utilidad 135
3.4 Restricción presupuestaria 136
3.5 Equilibrio del consumidor 138
3.6 Obtención de la curva de demanda 140
3.7 Descomposición de efecto renta y sustitución 142
3.8 Efecto de las variaciones en la renta 145
3.9 Curvas de Engel 145
3.10 Obtención de la curva de demanda de mercado 148
3.11 Axioma débil de preferencia revelada 154
3.12 Axioma fuerte de preferencia revelada 155
3.13 Preferencia revelada: efecto renta y efecto sustitución 156

4.1 Problema de Maximización de Utilidad 167
4.2 Problema de Minimización de Costes 170
4.3 Descomposición efecto renta y sustitución 176
4.4 Variación compensatoria y equivalente 180
4.5 Actitudes frente al riesgo 190
4.6 Equivalente cierto y prima de riesgo 191
4.7 Decisión de aseguramiento 196
4.8 Función de valor de la Teoría de las Perspectivas 198

5.1 Conjunto de producción y frontera de posibilidades de producción 220
5.2 Función de producción en tres dimensiones 223
5.3 Producción con un factor variable 225
5.4 Mapa de curvas isocuantas 227
5.5 Región económica de producción 228

5.6 Efectos del progreso técnico . . . 232
5.7 Definiciones de progreso técnico . . . 233
5.8 Caja de Edgeworth y Frontera de Posibilidades de Producción . . . 234
5.9 Tipos de funciones de producción . . . 238
5.10 Axiomas y características de escala . . . 241

6.1 Curvas isocuantas y rectas isocostes . . . 246
6.2 Senda de expansión . . . 248
6.3 Curvas de costes de una empresa a corto plazo . . . 253
6.4 Curvas de costes de una empresa a largo plazo . . . 255
6.5 Relación entre el largo plazo y el corto plazo: Teorema de Le Châtelier-Samuelson . . . 256
6.6 Curvas de coste sin rendimientos mixtos . . . 259
6.7 Teoría de costes de Stigler a corto plazo . . . 260
6.8 Teoría de costes de Stigler a largo plazo . . . 261
6.9 Economías de alcance . . . 263
6.10 Subaditividad de costes . . . 265
6.11 Curva de aprendizaje . . . 266
6.12 Las economías de escala frente al aprendizaje . . . 267

7.1 Equilibrio competitivo a muy corto plazo . . . 285
7.2 Equilibrio competitivo de la empresa a corto plazo . . . 287
7.3 Equilibrio competitivo a corto plazo . . . 288
7.4 Equilibrio competitivo a largo plazo . . . 290
7.5 Eficiencia asignativa de la competencia perfecta . . . 293
7.6 Aumento de los costes fijos . . . 296
7.7 Aumento de los costes variables . . . 296
7.8 Rendimientos a escala . . . 299
7.9 Oferta a largo plazo de una industria con coste creciente . . . 301
7.10 Oferta a largo plazo de una industria con coste decreciente . . . 301
7.11 Modelo de la telaraña con fluctuaciones convergentes . . . 303

8.1 Equilibrio del monopolio . . . 315
8.2 Ineficiencia asignativa del monopolio . . . 317
8.3 Discriminación de precios de segundo grado . . . 321
8.4 Sostenibilidad del monopolio natural . . . 323
8.5 Regulación fiscal del monopolio . . . 325
8.6 Efecto Averch-Johnson . . . 332

9.1 Obtención gráfica de la curva de reacción . . . 358
9.2 Equilibrio en el modelo de Cournot . . . 359

9.3 Equilibrio en el modelo de Stackelberg . . . 361
9.4 Cártel como monopolio multiplanta . . . 365
9.5 Curva de colusión . . . 366
9.6 No cumplimiento del acuerdo de colusión . . . 367
9.7 Modelo del precio límite . . . 373

10.1 Modelo Lancaster simplificado . . . 391
10.2 Ciudad lineal . . . 393
10.3 Demanda Quebrada . . . 400
10.4 Modelo de Chamberlin: equilibrio en el corto plazo . . . 401
10.5 Modelo de Chamberlin: equilibrio en el largo plazo . . . 402
10.6 Teorema exceso de capacidad . . . 403
10.7 Diferenciación vertical . . . 411
10.8 Modelo de Lancaster . . . 416
10.9 Modelo de Lancaster: fidelidad a las marcas . . . 417
10.10Modelo de Lancaster: desarrollo de nuevos productos . . . 418

11.1 Determinación de la oferta de trabajo . . . 428
11.2 Curva de oferta de trabajo . . . 430
11.3 Oferta de trabajo a la que se enfrenta la empresa . . . 431
11.4 Demanda de trabajo en el corto plazo . . . 432
11.5 Demanda de trabajo a largo plazo . . . 434
11.6 Demanda agregada de trabajo a corto plazo . . . 435
11.7 Equilibrio del mercado de trabajo neoclásico . . . 436
11.8 Poder de mercado de la empresa: monopolio y monopsonio . . . 438
11.9 Poder de monopolio del trabajador . . . 441
11.10Modelo de capital humano: formación general . . . 448
11.11Monopolio bilateral en el mercado de trabajo . . . 454

12.1 Curva de Beveridge . . . 471
12.2 Equilibrio del modelo DMP . . . 476
12.3 Banda de inacción con costes de ajuste lineales y asimétricos . . . 479
12.4 Modelos de derecho a gestionar y de contrato eficiente . . . 481
12.5 Condición de Solow . . . 484
12.6 Modelo de Shapiro y Stiglitz . . . 485

13.1 Modelo del mercado de limones de Akerlof . . . 500
13.2 Modelo de señalización de Spence . . . 503
13.3 Modelo de autoselección . . . 506
13.4 Modelo de racionamiento del crédito de Stiglitz y Weiss . . . 514

 © 2026 Manuel Cabado García

14.1 Caja de Edgeworth de producción . . . 535
14.2 Frontera de posibilidades de producción . . . 536
14.3 Caja de Edgeworth de consumo . . . 538
14.4 Frontera de posibilidades de utilidad . . . 539
14.5 Óptimo global . . . 540
14.6 Gran frontera de posibilidades de utilidad . . . 541
14.7 Criterio de la función de bienestar social . . . 542
14.8 Criterios de compensación . . . 548
14.9 GFPU con restricción en la economía . . . 549

15.1 Frontera de posibilidades de producción . . . 569
15.2 Condición de Samuelson . . . 574
15.3 Infraprovisión del bien público . . . 576
15.4 Externalidad negativa en la producción . . . 582
15.5 Corrección de la externalidad: Impuestos pigouvianos . . . 585
15.6 Tipos de externalidades . . . 598

16.1 Tipos de Función de Bienestar Social . . . 611
16.2 Regla de la mayoría óptima en cada elección . . . 616
16.3 Paradoja de Condorcet . . . 617
16.4 Preferencias unimodales . . . 618
16.5 Elección del bien público por parte de los agentes . . . 620
16.6 Votación sobre el nivel de gasto . . . 620
16.7 Nivel de provisión ineficiente de gasto público . . . 622
16.8 Resultado del modelo Downs-Hotelling . . . 625

17.1 Teoría de la burocracia de Niskanen . . . 650
17.2 Tamaño del Sector Público y crecimiento económico: el modelo de Barro (1990) . . . 657

18.1 Reglas de precio en el monopolio natural . . . 676
18.2 Precios de Ramsey-Boiteux con dos mercados . . . 679
18.3 Efecto Averch-Johnson en el plano capital-trabajo . . . 680

19.1 Evolución de la riqueza a lo largo del ciclo vital . . . 711
19.2 Impuesto negativo sobre la renta . . . 716
19.3 Índice de Gini . . . 723
19.4 Valor actual neto fiscal de un inmigrante adicional según la edad de llegada y la cualificación . . . 731
19.5 Contribución neta media a las finanzas públicas por edad y origen, Dinamarca, 2018 . . . 732

20.1 Exceso de gravamen 745
20.2 Exceso de gravamen con más de un impuesto 748
20.3 Incidencia de un impuesto unitario al consumo en equilibrio parcial bajo competencia perfecta 751
20.4 Incidencia de un impuesto unitario a la producción en equilibrio parcial bajo monopolio 753
20.5 Incidencia de un impuesto pero con monopolio bajo demanda isoelástica . 754
20.6 Impuesto al capital en un país pequeño 764
20.7 Pérdida de eficiencia por precio máximo y mínimo 770
20.8 Exceso de gravamen de un subsidio 772

21.1 Efecto de impuestos sobre la oferta de trabajo 788
21.2 Efectos de impuestos sobre el ahorro 790
21.3 Impuestos sobre el rendimiento en cartera 791
21.4 Índice de Gini y curva de Lorenz 797
21.5 Compromiso entre eficiencia y equidad 798
21.6 La curva de Laffer 806

22.1 Tamaño óptimo de la jurisdicción 827
22.2 Transferencia incondicional 834
22.3 Transferencia condicional no compensatoria 835
22.4 Transferencia condicional compensatoria 837

 © 2026 Manuel Cabado García

Índice de cuadros

2.1 Ganadores del Premio Nobel de Economía 102

4.1 Resumen de la teoría de la dualidad 173
4.2 Orden de magnitud de medidas de bienestar 183

7.1 Resumen de la estática comparativa 297

8.1 Fórmulas básicas 348

9.1 Resultados de oligopolio 364
9.2 El dilema del prisionero 368

11.1 Resumen de resultados 439

15.1 Tipos de bienes 572

18.1 Efectos y limitaciones de los instrumentos de regulación 689

20.1 Equivalencias impositivas 760

22.1 Dilema del prisionero en competencia fiscal 831

 © 2026 Manuel Cabado García

Tema 1

Pensamiento económico (I): los economistas clásicos y neoclásicos. Referencia a Marx.

1.1. Introducción ⏳ 1'

La teoría del pensamiento económico consiste en el análisis de las diferentes doctrinas económicas que han surgido a lo largo de la historia. Convencionalmente comienza con los clásicos aunque el proceso económico ya había sido objeto de estudio por escuelas anteriores:

- Mercantilistas: siglos XVI y XVII en Inglaterra, abogaban por la acumulación de metales preciosos y políticas proteccionistas.
- Fisiócratas: siglo XVIII en Francia, proponían la acumulación de excedente generado por la agricultura como vía de crecimiento económico.

En 1776 el foco se traslada a Inglaterra con la publicación de la obra magna de Adam Smith, que marca el comienzo de la Economía clásica, cuya doctrina será criticada por la escuela marxista o la revolución marginalista, que marcará el inicio de los neoclásicos.

Ambas tradiciones pueden entenderse, de forma simplificada, a partir del problema económico al que se enfrentan: los clásicos razonan como si operasen en un mundo de recursos prácticamente ilimitados, por lo que su preocupación central es la distribución del producto entre las clases sociales. Los neoclásicos, en cambio, sitúan la escasez en el centro del análisis, lo que explica que su instrumento distintivo sea la utilidad marginal.

El objetivo de esta exposición es el análisis de la economía clásica y neoclásica, el cual es fundamental ya que:

1. Marca las bases del pensamiento económico moderno.
2. Por primera vez se plantean conjuntamente varios problemas económicos.

1.2. Los economistas clásicos ⏳ 10'

Los autores clásicos comparten, en mayor o menor medida, el campo de estudio: los determinantes de la riqueza, la teoría del valor, la teoría de la distribución de la renta, el papel del comercio internacional, etc. Así, la mayoría de ellos se centrarán en analizar tres cuestiones fundamentales:

1. La Teoría del Valor: Qué factores son los que determinan el valor de un bien.
2. La Teoría de la Distribución de la Renta: Cómo se distribuye el valor de dicho bien entre los diferentes agentes de la sociedad.
3. La Teoría de la Acumulación de la Riqueza: Qué factores contribuyen al crecimiento económico y la acumulación de la riqueza hasta la llegada a un estado estacionario.

Adam Smith es considerado como el fundador de la ciencia económica.

Adam Smith (1723-1790)

Nace en Escocia y es un académico, a diferencia de otros economistas clásicos, como David Ricardo o Stuart Mill[1]. Tradicionalmente, se le ha considerado como el fundador de la ciencia económica, y su libro *Una investigación sobre la naturaleza y causas de la riqueza de las naciones* o, simplemente, *La riqueza de las naciones* (Smith, 1776), como el punto de partida de esta disciplina. Sin embargo, muchas de las ideas de Smith no eran totalmente originales:

1. Su idea del laissez-faire ya había sido defendida por los fisiócratas en su concepto del "orden natural".
2. Hume (Hume, 1752) ya había advertido del peligro de las practicas mercantilistas.

El verdadero valor de Smith reside en cómo supo recopilar todas estas ideas y construir con ellas un modelo completo del funcionamiento de la economía. Smith sigue un método deductivo, combinado en ocasiones con el inductivo, e histórico.

Determinantes de la riqueza

En su obra *La riqueza de las naciones* (Smith, 1776), Smith se plantea una gran pregunta: ¿en qué consiste la riqueza de las naciones? Para Smith, la respuesta es evidente: la riqueza de una nación deriva de su trabajo. Concretamente, la riqueza de un país depende de dos circunstancias: la productividad del trabajo y la proporción de la población ocupada en

[1] Nace en Kirkcaldy, cerca de Edimburgo, y ocupa la cátedra de Filosofía Moral en la Universidad de Glasgow, donde en 1759 publica *La teoría de los sentimientos morales*. Durante su viaje a Francia como preceptor del duque de Buccleuch entra en contacto directo con Quesnay y Turgot, cuyas ideas terminarán de fraguar en *La riqueza de las naciones* (1776).

 © 2026 Manuel Cabado García

trabajo productivo, definido como aquel que añade valor al objeto sobre el que se aplica.[2] De las dos, Smith considera decisiva la primera, por lo que la pregunta es ¿cuál es el determinante de la productividad?

Para Smith, la división del trabajo, es decir, la especialización de cada trabajador en una fase concreta del proceso productivo, es el principal determinante de la productividad, hecho que ilustró con la fábrica de alfileres[3]. A su vez, la división del trabajo depende de dos factores:

1. Extensión del mercado: cuanto más grande sea el mercado, mayor será el volumen que pueda venderse y, por tanto, mayor será la oportunidad de división del trabajo.
2. Acumulación del capital: la producción requiere tiempo, y el capitalista debe adelantar el coste de los salarios y de los demás insumos hasta que la producción se venda. Este capital adelantado es el capital circulante, y la parte destinada a pagar salarios se conoce como "fondo de salarios".

Por lo tanto, de la división del trabajo se derivan dos implicaciones. La primera es la aparición del dinero, pues ahora el productor no tendrá bienes, sino "bienes incompletos", cuyo trueque es más difícil. La segunda es la necesidad del comercio internacional para expandir el mercado objetivo.

Teoría del valor

En cuanto a la teoría del valor, Smith retoma una distinción conocida desde la antigüedad entre el valor de uso, que es la capacidad de un bien para satisfacer una necesidad, y el "valor de cambio" o precio, que es su capacidad para comprar otros bienes.

Sin embargo, la "paradoja del agua y los diamantes" enfatiza la incoherencia entre los dos términos anteriores: mientras que el valor de uso del agua es mucho mayor que el de los diamantes, su valor de cambio es mucho menor. Esto lo solucionarán los marginalistas más adelante.

Por consiguiente, Smith se centra en estudiar qué es lo que determina el valor de cambio o precio de un bien.

[2] La intuición es que un criado doméstico puede trabajar tantas horas como un obrero de una manufactura, pero su servicio se consume en el mismo momento en que se presta y no deja un objeto vendible que añada valor ni permita acumular capital. El ejemplo es del propio Smith (*La riqueza de las naciones*, Libro II, cap. III).

[3] Cuando un trabajador realiza todas las operaciones necesarias para producir un alfiler, apenas consigue fabricar uno al día y, desde luego, no llega a veinte. Si, en cambio, el proceso se divide en unas dieciocho operaciones y cada trabajador se especializa en una o dos de ellas, un taller de diez personas fabrica unos 48.000 alfileres diarios, es decir, 4.800 por trabajador. No obstante, tiene costes sociales: los trabajadores deben realizar tareas repetitivas y monótonas. Los seres humanos se convierten en máquinas atadas a un proceso de producción y deshumanizadas por las simples, repetitivas y aburridas tareas que realizan. Pero Smith no tenía duda de que la división del trabajo aumentaba en general el bienestar humano.

Para determinar el valor de cambio o precio de un bien, Smith distingue entre precio de mercado, a corto plazo, y precio natural, a largo plazo:

- El precio de mercado sería el precio de equilibrio a corto plazo, en competencia perfecta, que viene determinado por las fluctuaciones de la oferta y la demanda.
- El precio natural sería el precio de equilibrio del mercado a largo plazo, también en competencia perfecta, que viene determinado por los costes de producción, es decir, por el pago a los factores productivos

Por consiguiente, el precio natural es un precio central en torno al cual gravitan constantemente los precios de todas las mercancías. La importancia que Smith atribuye al coste de producción como determinante del precio natural demuestra que su teoría está enfocada al lado de la oferta.

Teoría de la distribución

Smith no realiza un análisis de los determinantes de la distribución de la renta entre las distintas clases sociales (cosa que sí hará David Ricardo), pero sí describe individualmente cada una de las remuneraciones existentes: salarios, beneficios y rentas, cuya suma determina el precio a largo plazo, precio natural, de las cosas.

- Salarios, al igual que otros clásicos como Stuart Mill, Smith defiende la doctrina del fondo de salarios, según la cual el salario medio es el cociente entre el fondo de capital que los capitalistas destinan a pagar trabajo y el número de trabajadores.[4] De esta manera, Smith concibe el siguiente proceso de formación de salarios:
  - Corto plazo: tanto el fondo de salarios como la población están dados, por lo que los salarios se determinan mediante la negociación[5]
  - Largo plazo: en principio podría pensarse que el continuo crecimiento de la población reducirá el salario hasta el nivel mínimo de subsistencia[6], pues si el salario fuera mayor, la población aumentaría y, con ella, la oferta de trabajo, hasta devolver el salario a ese nivel. La curva de oferta de trabajo sería entonces horizontal al salario de subsistencia y el salario, exógeno.
    Sin embargo, Smith argumenta que esto no siempre se cumple, pues la demanda

[4] La intuición es la de una tarta de tamaño fijo: si el fondo es de 1.000 unidades y hay 100 trabajadores, cada uno cobra 10, y si la población trabajadora crece hasta 125, el salario cae a 8 sin que nadie lo haya decidido. Mill abandonó esta doctrina en 1869.

[5] Smith ya anticipa una idea cercana al moderno "salario de eficiencia" (Smith, 1776) (Libro I, cap. VIII): los individuos responden a los incentivos, de modo que a mayor salario firmado, mayor productividad. Esta conclusión difiere de la teoría de los mercantilistas, que defienden salarios de subsistencia, y de la teoría neoclásica, en la que la productividad determina el salario, y no al revés.

[6] Ley de hierro de los salarios: tendencia del salario a situarse a largo plazo en el nivel de subsistencia, porque todo salario superior eleva la población y la oferta de trabajo hasta devolverlo a ese nivel. La expresión la populariza Ferdinand Lassalle en 1863, pero la idea procede de Malthus y Ricardo.

 © 2026 Manuel Cabado García

de trabajo puede crecer más deprisa que la población, y distingue así entre "estado progresivo" y "estado estacionario"[7].

- Beneficios, Smith concibe el beneficio como la remuneración al empresario capitalista y no distingue entre "capitalista" y "empresario" (debido a que los mercados de capitales en el siglo XVIII estaban sin desarrollar). El beneficio es la remuneración del empresario-capitalista por proveer capital[8] y por el riesgo asumido. Este beneficio, que Smith considera tremendamente incierto[9], dependerá fundamentalmente de la tasa de interés del capital que, a su vez, depende de la oferta y la demanda de bienes de capital (de las oportunidades de inversión que dependerán de si hay más o menos competencia)[10].
- Rentas, considera las rentas en términos residuales, es decir, lo que queda después de remunerar el resto de factores. Así, el economista escocés considera las rentas como parte del valor de la producción que se paga por el uso de la tierra, por lo que básicamente se interpreta como un coste de arrendamiento.

Esta teoría de la distribución, denominada por Sraffa la teoría de la suma, al ser sus elementos independientes será considerada por David Ricardo como insatisfactoria al no realizar un análisis conjunto y sin considerar su evolución temporal. A esto se suma un problema de circularidad: en el capítulo 6 de *La riqueza de las naciones* Smith explica el precio en función de la renta, pero en el capítulo 11 invierte la relación: es la renta la que se explica en función del precio. Smith no fue consciente de esta circularidad, y será precisamente David Ricardo quien la resuelva.

Comercio internacional

Smith rompe con la filosofía proteccionista de los mercantilistas[11], pues argumenta que el libre comercio es un juego de suma positiva. Para ello, se basa de nuevo en la teoría de la división del trabajo y la especialización.

[7] De esta manera, Smith consideró que las colonias se encontraban en un "estado progresivo", mientras que el imperio chino había alcanzado el "estado estacionario" y, por consiguiente, los salarios se situaban allí en el nivel de subsistencia.

[8] En los clásicos el capital no se asemeja a un factor productivo como tal, esta escuela se centra en el factor trabajo considerado como valor de todas las cosas, también de los bienes de capital, y como determinante último del proceso productivo. El capital se entenderá como: i) factor (fondo de salarios o capital circulante) ii) Bien (como trabajo acumulado en forma de bienes de capital físicos, es decir, maquinaria, que se mide por horas de trabajo necesarias para su producción)

[9] El beneficio es volátil y no se puede conocer, por lo que se aproxima por el tipo de interés (i + compensación por el riesgo incurrido en el proceso productivo) que se mueve en la misma dirección.

[10] Estas oportunidades de inversión irían disminuyendo conforme aumenta la competencia (caída secular de la tasa de beneficios). Por el efecto de la competencia en los diferentes mercados (trabajo, inversión,...) que reducirían tanto los salarios como los precios, y agotaría las oportunidades derivadas de nuevos mercados.

[11] "Es premisa de todo padre de familia prudente no producir en casa lo que resulta más barato en el exterior."

Según Smith, cada país debe especializarse en la producción de aquel bien en el que posea ventaja absoluta en coste, es decir, aquel bien cuyo coste de producción en el país sea menor al coste de producción en cualquier otro país: es la conocida como teoría de la ventaja absoluta.

Sin embargo, el mayor problema de esta teoría, que resolverá David Ricardo, consiste en establecer el patrón de comercio cuando un país es el productor más eficiente de todos los bienes.

Política económica

Smith considera que cada individuo es el mejor juez de su propio interés, por lo que ha de tener libertad a la hora de satisfacerlo. Si se le otorga dicha libertad, el individuo no solo alcanzará su propio interés, sino que, movido por la mano invisible, contribuirá al bienestar colectivo[12]. Estas ideas constituyen la base del liberalismo económico, que se traduce en propuestas concretas como la defensa del laissez-faire o la supresión de los gremios. Así, Adam Smith es considerado como el padre de esta doctrina[13].

Además, Smith considera que el papel del Estado debe ser limitado, atendiendo solamente a tres objetivos fundamentales: justicia, seguridad y obras públicas (Smith, 1776) (Libro V, cap. 1). Según Smith, todas estas actividades suponen un gasto que debe financiarse con impuestos —que deben cumplir los cánones de igualdad, certidumbre, comodidad de pago y economía en la recaudación— en lugar de con deuda, ya que la deuda convierte el trabajo productivo en improductivo[14]. Además, defiende el equilibrio presupuestario, es decir, $T = G$.

Robert Malthus (1766-1834)

Robert Malthus fue un clérigo inglés que estudió matemáticas y que realizó importantes aportaciones a la economía, principalmente en dos aspectos: los problemas del crecimiento de la población y los problemas de la insuficiencia de la demanda. A diferencia del optimismo de Smith sobre el progreso ilimitado, Malthus advirtió que la población crece exponencialmente mientras los recursos alimenticios lo hacen aritméticamente.

[12] "No es la benevolencia del panadero lo que nos procura el alimento, sino la consideración de su propio interés".

[13] No obstante, es erróneo considerar a Smith como un defensor del capitalismo salvaje que, de hecho, fue profesor de moral. En su obra *La teoría de los sentimientos morales* (Smith, 1759) considera que el hombre no es un individuo egoísta, sino que se interesa por el bienestar de otros y de hecho la felicidad de quienes le rodean le resulta necesaria.

[14] Esto se debe a que dicho capital, en lugar de dedicarse a la inversión y a la producción, se presta al Estado a cambio de un tipo de interés. Por eso defenderá el uso de los impuestos frente a la deuda y, especialmente, aquellos que recaigan sobre la tierra, ya que considera que no afectan directamente a la producción.

 © 2026 Manuel Cabado García

Problemas del crecimiento de la población

Esta teoría se deriva de su obra *Ensayo sobre el Principio de la Población* (Malthus, 1798). La idea es que el alimento crecerá en una progresión aritmética, mientras que la población lo hará en progresión geométrica, por lo que será imposible la elevación permanente del nivel de vida de la clase trabajadora por encima de las puras necesidades fisiológicas.

Esta brecha solo puede cerrarse por dos vías[15]:

1. Frenos positivos, que son aquellos que tienden a aumentar la tasa de mortalidad como la hambruna, las guerras y las epidemias.
2. Frenos preventivos o normativos, que son aquellos que tienden a reducir la tasa de fertilidad mediante mecanismos de control de la natalidad, como por ejemplo el retraso o abstención del matrimonio.

Partiendo de este análisis, Malthus concluye que es la condición humana la que limita el crecimiento económico, dado que un aumento del salario por encima de su nivel de subsistencia da lugar a un aumento de la población. Sin embargo, los trágicos presagios maltusianos no se han cumplido, debido, entre otros motivos, a que:

- El aumento de la educación y la riqueza suele venir acompañado de un descenso de la natalidad, rompiendo la progresión geométrica de la población.
- Los avances tecnológicos pueden acelerar la producción de alimentos.

No obstante, sigue siendo una teoría que, sorprendentemente, es de tiempo en tiempo resucitada por algunos grupos. Ejemplos destacados del siglo XX son *La bomba demográfica* de Paul Ehrlich (Ehrlich, 1968), que auguraba hambrunas masivas por el crecimiento poblacional, y el informe *Los límites del crecimiento* del Club de Roma (Meadows et al., 1972), que extendió la lógica maltusiana a otros recursos finitos —petróleo, metales— además de los alimentos. Como el propio Malthus, ambas predicciones envejecieron mal: la Revolución Verde disparó la producción agrícola y el mecanismo de precios incentivó la sustitución de recursos escasos, dos vías de escape que ninguno de los dos modelos anticipó.

[15] ¿Hay innovaciones que pudiesen haber puesto freno a las teorías de Malthus?:

- Revolución agrícola británica s.XVIII-XIX
  - Enclosure acts s.XVII (1604-1914). Expropiación de open fields
  - Rotación de cultivos
  - Selección de semillas
  - Aparición del arado de hierro, maquinas sembradoras
- Invención de la vacuna s.XVIII. Edward Jenner (1749.1823). vacuna contra la viruela.
- En el siglo XIX segunda generación de vacunas. Louis Pasteur vacuna contra la cólera, rabia. Se llama vacuna por la viruela bovina de Jenner.

Problemas de la insuficiencia de la demanda

Malthus concibe dos tipos de bienes: los básicos, de los que nunca habrá sobreproducción, y el resto, sobre los que sí puede haber una brecha entre la demanda y la oferta. Malthus niega la ley de Say, según la cual toda oferta crea su propia demanda ya que todos los ingresos del empresario se reparten vía rentas, que pueden destinarse al consumo o al ahorro, y como todo el ahorro se destina a inversión, al final todo el valor de los bienes producidos se dedica a demanda interna.

Sin embargo, Malthus cree que no todo el ahorro se gasta, pues:

- Los trabajadores gastan su parte.
- Los capitalistas son ahorradores y piensan consumir en el futuro más que en el presente. Por lo tanto, el dinero no solo es un medio de cambio, sino también un depósito de valor.
- Así, es necesario que los terratenientes y el Estado consuman, o de lo contrario habrá insuficiencia de demanda y habrá producción que quedará sin venderse.

De este último punto se derivan las siguientes recomendaciones de política económica de Malthus:

- Favorecer a los terratenientes en lugar de a los pobres, que apenas contribuyen al aumento de la demanda.
- Medidas proteccionistas que defiendan la demanda interna (desincentivar las compras de terratenientes al exterior).
- Preferencia por el gasto público frente a las rebajas impositivas, ya que el gasto público estimula directamente la demanda interna, mientras que las rebajas impositivas se pueden retener en forma de ahorro.
- Otras medidas: No auxilio a los pobres (aumenta la demanda de alimentos, no elimina la pobreza y reduce los incentivos al trabajo, justificará las leyes de granos, inculcar a las clases pobres los principios de la clase media (educación y sufragio universal).

El error de Malthus fue que no consideró la inversión como demanda efectiva. Malthus no distingue en su estudio entre la demanda de inversión (que realizan los capitalistas) de la demanda de consumo de los terratenientes, ni tampoco el efecto que cada tipo de demanda tendría sobre el crecimiento de los países. Si el ahorro se traduce en acumulación de inventarios y no en FBKF, se produce efectivamente déficit de demanda. Sin embargo, no considera que el ahorro de los terratenientes pueda ir a inversión, y por tanto mayor demanda, evitando así la sobreproducción.

 © 2026 Manuel Cabado García

David Ricardo (1772-1823)

Nace en Londres y, a diferencia de Smith, no solo no era académico, sino que ni siquiera tenía estudios universitarios, siendo su ocupación la de empresario financiero[16]. Su obra principal fue *Principios de Economía Política y Tributación* (Ricardo, 1817). Ricardo sigue un método puramente deductivo, hasta el punto de que es considerado un pionero en la construcción de modelos —lo que Schumpeter criticaría más tarde como el "vicio ricardiano"[17].

Ricardo desarrolla sus trabajos en un contexto particular: las guerras napoleónicas, al cerrar el comercio, habían forzado el incremento de la producción agrícola en Gran Bretaña. Los terratenientes, alarmados ante la caída de los precios agrícolas que empezó a registrarse en 1812, sostenían que esa protección que habían recibido durante la guerra debía mantenerse en tiempos de paz, y lograron así la aprobación de las leyes de Granos, que estarían vigentes hasta 1846 pese a las protestas de muchos economistas, entre ellos Ricardo[18].

El economista inglés trata de elaborar una teoría que respalde sus argumentos. Para ello, se centrará en la teoría de la distribución, para lo que reformuló previamente la teoría del valor de Smith. A continuación, se analizarán ambas teorías de Ricardo.

Teoría del valor

A corto plazo, Ricardo coincide con Smith en que el precio de los bienes dependerá de las fluctuaciones de la oferta y la demanda. Por otro lado, la principal novedad de la teoría del valor de Ricardo es que este no considera la tierra como un determinante del precio de largo plazo de los bienes.

En su "teoría de la renta diferencial" (Ricardo, 1817) (cap. 2), Ricardo señala que la renta de la tierra no es un coste de producción, sino un excedente que surge porque la tierra fértil es escasa y de calidad desigual. Por consiguiente, según Ricardo, el valor de cambio o precio de un bien vendrá determinado por el coste de producción derivado del uso de trabajo y de capital. Además, puesto que el capital puede expresarse en última instancia en términos de trabajo, el principal determinante del precio natural es el trabajo incorporado.

No obstante, Ricardo admite que el precio también dependerá de la escasez para ciertos

[16] Concretamente, fue agente de bolsa de origen sefardí que amasó una fortuna joven y se retiró de los negocios en 1814 para dedicarse a la economía política y, más tarde, a la política parlamentaria.

[17] Consiste en extraer conclusiones de política económica muy concretas y taxativas a partir de un modelo extremadamente simplificado y abstracto, sin contrastarlo empíricamente: salta directamente de la teoría pura a la prescripción de política, sin el paso intermedio de comprobar si los supuestos del modelo se sostienen en la realidad.

[18] Un ejemplo empírico moderno de este episodio es Juhász (2018), que muestra cómo la protección temporal del bloqueo continental permitió a las regiones francesas más aisladas del comercio británico desarrollar capacidad en el hilado de algodón mecanizado, con efectos persistentes sobre su industrialización hasta bien entrado el siglo XIX.

bienes, como por ejemplo las obras de arte no reproducibles, y de la proporción de capital fijo usado[19].

Teoría de la distribución

Una vez introducida la teoría del valor, se procederá a analizar la teoría de la distribución, que constituye una de las principales aportaciones de Ricardo a la teoría económica y que, como se muestra a continuación, tiene una aplicación directa como crítica a las leyes de granos. Concretamente, la lógica de la teoría de la renta diferencial es la siguiente:

1. Conforme crece la población, hay que obtener más cereal, bien cultivando tierras cada vez menos fértiles (margen extensivo), bien aplicando más trabajo y capital a las ya cultivadas, con rendimientos decrecientes (ley de Turgot (Turgot, 1767), margen intensivo). En ambos casos aumentan los costes de producción. Por consiguiente, el precio del cereal aumenta.
2. La renta diferencial es el excedente que obtienen las tierras más fértiles porque el precio del cereal lo fija el coste de producción en la tierra menos fértil en cultivo. En esa tierra marginal la renta es nula, pues el precio apenas cubre sus costes, y en cada una de las demás la renta es igual a la diferencia entre ese precio y su propio coste, menor[20][21].
3. Por su parte, los salarios se mantienen fijados al nivel de subsistencia, pero como los precios agrícolas han aumentado, el nivel de subsistencia también aumenta, llevando a un alza de los salarios.
4. El aumento de las rentas y de los salarios conduce a una disminución progresiva del beneficio, que irá tendiendo a cero con el paso del tiempo.
5. Como los beneficios son el motor de la acumulación del capital, y dicha acumulación permite el crecimiento de la economía, al desaparecer los beneficios, la economía tiende a un estado estacionario, es decir, aquel en el que los beneficios son nulos, el salario sigue fijado en el nivel de subsistencia y el resto del ingreso corresponde a la renta de la tierra.

Formalmente, distinguiendo el margen intensivo —más trabajo por unidad de tierra ya cultivada— del margen extensivo —incorporación de tierras cada vez menos fértiles—, la

[19] La razón es que el capital fijo tarda más tiempo en recuperarse. Por lo tanto, debe existir alguna compensación por el tiempo que pasa hasta que una inversión en capital fijo da sus frutos. De lo contrario, nadie invertiría en este, y se preferiría invertir en mercancías fácilmente vendibles.

[20] La intuición es que, si producir una tonelada de trigo cuesta 100 en la peor tierra cultivada y 70 en la mejor, el trigo se vende a 100 y el propietario de la mejor tierra cobra una renta de 30 por tonelada.

[21] Igual que Smith, en el modelo de Ricardo los salarios tienden al nivel de subsistencia (Ley de Hierro de los salarios), sin embargo, los mismos pueden situarse en niveles superiores dependiendo de las fuerzas de oferta y de demanda. No obstante, Ricardo trató el mecanismo de los salarios de subsistencia como si operasen a corto plazo, lo cual resulta incorrecto el considerar los salarios a nivel constante en el estado estacionario

 © 2026 Manuel Cabado García

renta se expresa como

$$\text{Renta} = (PMe_L - PMg_L) \cdot L \tag{1.1}$$

, es decir, el exceso de la productividad media del trabajo sobre su productividad marginal. En la tierra marginal ambas se igualan y la renta se anula. El beneficio, por el principio del excedente, es lo que queda tras remunerar a trabajo y tierra:

$$\text{Beneficio} = (PMg_L - w) \cdot L \tag{1.2}$$

, siendo w el salario de subsistencia.

Para Ricardo, los factores que permiten frenar la llegada al estado estacionario son dos:

1. Progreso técnico
2. Libre comercio, ya que la llegada de trigo barato del exterior permite abandonar las tierras menos fértiles y cultivar solo las más fértiles, de manera que la renta de la tierra disminuye y el beneficio aumenta. De ahí que Ricardo defendiera activamente la supresión de la Ley de Granos.

De este análisis se desprenden las siguientes conclusiones:

- La caída secular de la tasa de beneficio, al igual que concluía Smith pero, a diferencia de éste en el que el beneficio caía con motivo de la competencia, para Ricardo el beneficio desaparece como consecuencia de la caída de la productividad por unidad de tierra explotada.
- Existencia de un Estado Estacionario que se caracteriza por unos beneficios nulos, un salario de subsistencia y unas altas rentas para los terratenientes. Ante este pesimista futuro, Ricardo defenderá el Comercio Internacional, ya que evita el uso del margen intensivo y extensivo de la agricultura[22].

En definitiva, la teoría de la distribución de Ricardo representa el primer modelo coherente que explica el reparto del ingreso entre las distintas clases de propietarios a lo largo del tiempo, mientras que la teoría de la distribución de Smith se limitaba únicamente a describir las remuneraciones. Además, Ricardo elabora una teoría de la distribución conflictiva, en el sentido de que muestra la oposición de intereses de diferentes clases sociales, lo que tendrá cierta influencia sobre Karl Marx[23].

[22] Este pesimismo, compartido con Malthus, es el origen popularmente atribuido al apodo de la economía clásica como "ciencia lúgubre". Sin embargo, el origen real de la expresión es bien distinto: la acuñó el ensayista escocés Thomas Carlyle en 1849 (Carlyle, 1849), no como crítica al pesimismo malthusiano o ricardiano, sino como ataque a John Stuart Mill y a otros economistas que, apoyándose en el principio de que todos los individuos son productivamente equivalentes con independencia de su raza, defendían la abolición de la esclavitud.

[23] Además, presenta una serie de inconvenientes: no se considera el sector industrial o manufacturero

Comercio internacional

Ricardo comparte con Smith la defensa de las ganancias derivadas del comercio, pero propondrá que los países no se especialicen según su ventaja absoluta, sino según su ventaja comparativa. Un país tendrá ventaja comparativa en aquel bien que produzca con menor coste relativo (coste de oportunidad) tal y como explicó, en el capítulo 7 de sus *Principios*, en su ejemplo del paño y el vino entre Inglaterra y Portugal[24].

Por lo tanto, que la industria de un país sea tecnológicamente superior a la del resto de países no es una condición suficiente ni necesaria para su supervivencia, ya que la clave está en la diferencia relativa de los costes, es decir, en la ventaja comparativa. Por la misma razón, las industrias de países tecnológicamente menos avanzados podrían competir con la de los países más desarrollados si dichas industrias tienen ventaja comparativa.

Política económica

Al igual que Smith, Ricardo defenderá las virtudes de un sistema autorregulado, laissez-faire, y la financiación de los bienes públicos a través de los impuestos, a la tierra y a los bienes de lujo[25], y no de deuda. No obstante, en el capítulo XVII de su obra *Principios de Economía Política y Tributación* (Ricardo, 1817), Ricardo planteará la posibilidad teórica de que impuestos y deuda pueden acabar siendo sustitutivos. Más de 150 años

en la producción (es decir, una vez decae la tasa de beneficio en la agricultura, el capital se mueve al sector manufacturero, aunque en última instancia también tendería a un estado estacionario), los salarios son íntegramente gastados en trigo, no considera el progreso tecnológico (que podría compensar los rendimientos decrecientes) , etc.

El análisis de Ricardo respecto a la nueva tecnología es ambiguo en sus conclusiones. Por un lado, una mejor tecnología aumentaría la productividad del trabajo y en consecuencia la producción total para un nivel de trabajo dado. No obstante, esa nueva tecnología usaría más capital fijo en maquinaria, que se detrae de la cantidad disponible como fondo de salarios, y por tanto, las posibilidades de contratar más trabajadores. En cualquier caso, Ricardo subestimó la capacidad de PT en la agricultura, que podría compensar los rendimientos decrecientes.

[24] Estas serían las críticas a la teoría de la ventaja comparativa de Ricardo:

1. Especialización poco realista: El modelo asume coeficientes fijos en la producción, lo que implica una especialización completa que no corresponde a la realidad.
2. Falta de análisis redistributivo: Al considerar solo el trabajo como factor relevante, se ignoran los efectos redistributivos del comercio internacional, que sí abordan los modelos neoclásicos posteriores.
3. Ignorancia de demanda y costes crecientes: Ricardo asumió costes constantes y desestimó la demanda, lo que excluye posibles beneficios del comercio entre países con estructuras de costes similares.
4. Indeterminación de la relación de intercambio: Ricardo no aborda los precios a los que se intercambian las mercancías, dejando incompleto su análisis. Este vacío lo llena John Stuart Mill mediante las demandas recíprocas y Marshall con instrumentos gráficos.
5. Tamaño relativo de los países: En casos donde un país es demasiado pequeño, podría no abastecer al otro incluso con especialización completa.

[25] Un impuesto sobre la renta recae sobre los terratenientes sin posibilidad de traslación, mientras que los impuestos sobre salarios recaen sobre los beneficios (dado que los salarios son exógenos, fijados por el nivel de subsistencia). Por tanto, recomendará que se minimicen los gravámenes sobre beneficios y salarios, y que se apliquen los mayores impuestos a los terratenientes y a los bienes de lujo.

 © 2026 Manuel Cabado García

después, esta idea fue revivida por Robert Barro en su artículo "Are Government Bonds Net Wealth?" (Barro, 1974), bajo el nombre de "equivalencia ricardiana".

Respecto a la teoría monetaria, Ricardo también reconoce las virtudes del papel-moneda que se estaba comenzando a emitir en su tiempo, pero a la vez reconoce los peligros que acarrea: si los gobiernos imprimen demasiado papel moneda, esto puede dar lugar a un aumento del nivel de precios[26]. En este sentido, sugiere que una solución podría ser vincular el stock de oferta monetaria a algo material, como los lingotes de oro. De este modo, se le puede considerar un precursor del patrón oro.

John Stuart Mill (1806-1873)

John Stuart Mill fue un filósofo, economista y político británico del siglo XIX[27], conocido por su defensa del utilitarismo, el liberalismo clásico y los derechos individuales. Su obra *Principios de Economía Política* (Mill, 1848) culmina y sintetiza el pensamiento clásico, hasta el punto de sustituir a *La riqueza de las naciones* como manual de referencia durante las siguientes cuatro décadas, hasta que lo desplazaron los *Principios de Economía* de Alfred Marshall en 1890. Precisamente por eso se le considera el nexo entre la escuela clásica y la neoclásica: es también el primer autor en articular explícitamente, en un ensayo de 1836 (Mill, 1836), la idea del "homo economicus", el individuo racional que persigue su interés maximizando su riqueza. De esta forma, buscó refinar y humanizar el pensamiento clásico. Compartió la fe en el libre mercado, pero abogó por una mayor intervención estatal para la justicia social. Consideró que la producción seguía leyes naturales, pero la distribución de la riqueza podía modificarse a través de políticas, tendiendo puentes hacia ideas más reformistas.

Teoría del valor

Mill es defensor de la teoría clásica del valor, esto es, que los precios están determinados por los costes de producción, pero deja un margen para la influencia de la demanda en el

[26] Ricardo fue un defensor de la teoría de Locke y Hume, que defendía la existencia de un nexo entre masa monetaria y nivel de precios (bullionismo). El bullionismo es el nombre que se da a los economistas que estaban en contra de la política monetaria discrecional. En concreto, en 1808, el aumento de la emisión de billetes en Inglaterra empezó a sentirse a través de un aumento de los precios y un menor tipo de interés. Esto provocó la creación de un comité que investigase el elevado precio de los metales preciosos, el Bullion Committee, cuyo informe, el *Bullion Report* (1810), establecía que la vuelta a la convertibilidad era la única manera de restaurar la confianza en el valor del dinero.

Ricardo participará en el *Bullion Report* en el que se recuperan las ideas de Hume sobre la ecuación cuantitativa del dinero y sobre la conexión entre la masa monetaria y el nivel de precios. Por tanto, defenderá el control del crecimiento de la masa monetaria y el patrón oro como mecanismo para controlar el nivel general de precios.

[27] Fue un niño prodigio educado por su padre, James Mill —economista y amigo íntimo de Ricardo—, bajo una disciplina extraordinariamente exigente: a los ocho años ya leía en griego original a los clásicos, y antes de cumplir veinte debatía con Jeremy Bentham. El precio de esa exigencia fue una depresión nerviosa en su juventud, de la que se recuperó gracias, entre otras influencias, a Harriet Taylor, después su esposa, que le acercó a las ideas socialistas humanistas de la época.

precio al distinguir diferentes tipos de bienes[28]:

- Bienes de oferta inelástica, como las obras de arte, y en los que el precio viene determinado íntegramente por la demanda.
- Bienes de oferta totalmente elástica, como las manufacturas, y en los que el precio sí viene determinado por la oferta, esto es, por los costes de producción.
- Bienes de oferta con pendiente positiva, es decir, aquellos cuya producción puede aumentar, pero a un coste creciente.

El análisis de Mill de la teoría del valor es muy relevante al servir de base para dos cuestiones importantes:

1. La crítica marginalista de que solo en el caso del segundo tipo de bienes el precio viene determinado íntegramente por la producción.
2. La concepción de las curvas de oferta y demanda como funciones del precio.

Teoría de la distribución

En su estudio de la distribución, destaca lo siguiente:

- Beneficio: Sigue la teoría de Ricardo.
- Renta: Extiende el concepto no sólo a la tierra sino a todos los factores que tengan demanda inelástica.
- Trabajadores: Reconoce el poder sindical.

Así, en su análisis de la teoría de la distribución Mill introduce el papel de los sindicatos, que, en la medida en que tengan poder suficiente, podrán hacer aumentar los salarios, no a costa de un menor empleo, sino a costa de un menor beneficio de los empresarios, lo que, a su vez, podría conducir a un estado estacionario.

Sin embargo, a diferencia de sus predecesores, Mill considera que el estado estacionario no tiene por qué ser desagradable. En efecto, el autor tiene una visión del estado estacionario como un estado pacífico, ilustrado, en el que los agentes pueden aprovechar "el tiempo libre para estrechar los lazos sociales y desarrollar su propia capacidad intelectual".

Además, el autor distingue entre las leyes que rigen la producción, que son inmutables, como las leyes físicas, y las que rigen la distribución, que sí pueden cambiar porque dependen de las personas y de las instituciones. Según Mill, modificando las leyes de la distribución se puede variar el estado estacionario, por lo que éste no es inamovible y exógeno.

[28] En la época de Mill aún no se había desarrollado el concepto de elasticidad (es marshalliano), por lo que lo que decía Mill (y en general los clásicos) se expresa con lenguaje de hoy en día (interpretaciones de los distintos autores).

 © 2026 Manuel Cabado García

Concretamente, Mill proponía la educación de la clase trabajadora para que ésta controlase su natalidad y permitir, así, un salario superior al de subsistencia. Además, entendía que, si el salario era superior al de subsistencia, la inversión no tenía por qué caer ya que el ahorro de los trabajadores podía destinarse a la acumulación de capital.

Comercio internacional

Ni Smith ni Ricardo determinaron la relación de intercambio, esto es, los precios a los que se intercambian internacionalmente los bienes. De hecho, la ventaja comparativa es una condición necesaria pero no suficiente para que tenga lugar el comercio internacional. En efecto, es necesario también que la relación real de intercambio se sitúe entre los precios relativos de autarquía de ambos países, pues solo así cada uno obtiene, a cambio de su exportación, más de lo que obtendría produciendo el otro bien en casa.[29]

La determinación de la relación real de intercambio la lleva a cabo Mill a través de las llamadas demandas recíprocas, que son la demanda de cada país de los productos importados y que, al mismo tiempo, determina la oferta de productos exportados.

La relación real de intercambio será aquella que haga que el valor que cualquier país reciba por sus exportaciones sea exactamente igual al que tenga que pagar por sus importaciones. En términos gráficos, sería donde se cortasen las curvas de oferta neta definidas posteriormente por Marshall. Mill completó además este análisis con el estudio de los efectos de los aranceles sobre la relación de intercambio y con su teoría de la oferta conjunta —aplicable a bienes que se producen necesariamente juntos, como la lana y la carne de una misma oveja—, dos aportaciones metodológicas menos conocidas pero influyentes.

Economía del bienestar

Mill realiza aportaciones pioneras en el análisis de la intervención pública en la economía que tendrán gran influencia incluso hasta la actualidad. Concretamente, Mill defenderá que las autoridades deben intervenir en respuesta a lo que él llamó "defectos de mercado", que no son más que lo que hoy se denomina fallos de mercado.

Además, el autor distingue entre las funciones necesarias y las funciones facultativas del sector público, incluyendo dentro de las funciones necesarias lo que más tarde será el Estado del Bienestar como la regulación del trabajo infantil, determinadas prestaciones, etc. En cuanto a la tributación, Mill formula algunos de los primeros principios modernos de la Hacienda Pública —capacidad de pago, neutralidad impositiva, equilibrio presupuestario y existencia de un mínimo exento— y, en línea con ellos, aboga por impuestos

[29] La intuición es que, en el ejemplo de Ricardo, un barril de vino cuesta en Inglaterra lo mismo que 1,2 piezas de paño y en Portugal lo mismo que 0,89. Cualquier relación comprendida entre 0,89 y 1,2 piezas de paño por barril beneficia a los dos países, y las demandas recíprocas determinan en qué punto de ese intervalo se fija. Las cifras son las de Ricardo (*Principios de economía política y tributación*, cap. VII): en Inglaterra el paño cuesta 100 horas de trabajo y el vino 120, y en Portugal, 90 y 80, respectivamente.

progresivos frente a los proporcionales.

Karl Marx (1818-1883)

Karl Marx nace en Prusia, en el seno de una familia judía, y estudia Derecho, Filosofía e Historia en Bonn y Berlín[30]. Su relación con los economistas clásicos[31], especialmente con Adam Smith y David Ricardo, es una de doble filo: bebió de sus fuentes —sobre todo de la Teoría del Valor-Trabajo de Ricardo— para después criticarlas ferozmente, sobre todo en su obra *El Capital* (Marx, 1867). Mientras los clásicos empleaban la teoría del valor-trabajo para explicar la generación de riqueza, Marx la transformó en una herramienta para desvelar la explotación inherente al capitalismo.

Marx parte del método dialéctico de Hegel —de quien se consideraba discípulo—, pero invierte sus términos: mientras que para Hegel es el Espíritu Absoluto el que se autorrealiza a través de la historia, para Marx son las condiciones materiales de vida las que determinan la superestructura.

Para entender esta crítica, es crucial comprender el concepto marxista de estructura y superestructura: la estructura económica de una sociedad —las relaciones de producción y las fuerzas productivas— determina la superestructura —las instituciones políticas, legales y las formas de conciencia social, como la religión o la ideología—. De este modo, la explotación económica de los trabajadores no es un simple problema de salarios, sino la base material que sostiene todo el sistema social y justifica la dominación de la clase capitalista[32].

Su idea central es la plusvalía: los trabajadores producen más valor del que reciben en salarios, y esa diferencia, apropiada por los capitalistas, es la esencia de la explotación y la fuente de sus ganancias.

A diferencia de los clásicos, que veían el capitalismo como un sistema natural tendente al equilibrio, Marx lo consideraba intrínsecamente inestable, marcado por contradicciones internas y un inevitable conflicto de clases entre capitalistas y proletarios. Predijo que estas tensiones desembocarían en una revolución que derrocaría el sistema, abriendo paso a una sociedad sin clases: el comunismo. En esencia, Marx utilizó las herramientas analíticas de los clásicos para llegar a conclusiones radicalmente opuestas: no fue un continuador, sino un crítico revolucionario que sentó las bases de una nueva escuela de pensamiento.

[30] Frecuenta los círculos de la izquierda hegeliana y es expulsado de Alemania, Francia y Bélgica por el radicalismo de sus ideas, hasta instalarse definitivamente en Londres, donde vive el resto de su vida entre penurias económicas, sostenido en buena parte por la ayuda de su compatriota y colaborador Friedrich Engels.

[31] Fue, de hecho, el propio Marx quien acuñó el término "economistas clásicos" para referirse a la economía ricardiana.

[32] "La tradición de todas las generaciones muertas oprime como una pesadilla el cerebro de los vivos" (Marx, 1852). La frase resume bien su idea de que el peso de las estructuras heredadas condiciona, incluso atrapa, la acción de los vivos.

 © 2026 Manuel Cabado García

Su estudio es vital hoy para entender las dinámicas de poder, las desigualdades económicas y las críticas fundamentales al sistema capitalista. Concretamente, algunas de las tendencias secundarias que Marx apuntó se han ido cumpliendo desde una perspectiva histórica: la limitación de los horarios, el incremento de la composición orgánica del capital, la reducción de costes, la tendencia a la concentración en ciertas industrias o la recurrencia de crisis cíclicas.

No obstante, el núcleo duro de las predicciones no ha tenido lugar: los salarios han ido aumentando a la par que la productividad y las condiciones de las clases más bajas han mejorado de forma progresiva[33]. Al igual que el resto de los autores clásicos, Marx infravaloró el progreso tecnológico.

Clásicos y neoclásicos

Los economistas clásicos plantearon y respondieron algunas de las principales cuestiones dentro de la teoría económica, como las causas del crecimiento de las naciones, el desarrollo de una teoría del valor y de una teoría de la distribución de la renta e incluso analizaron las virtudes del comercio internacional.

No obstante, sus trabajos no presentaban una metodología sistemática y la ausencia de instrumental matemático les impedía resolver algunas de las paradojas que persistían en el seno de la teoría económica, como fue la paradoja del "agua y los diamantes de Smith". Además, su análisis se centró predominantemente en el lado de la oferta a la hora de determinar el valor de los bienes.

Será con la revolución marginalista, que posteriormente se consolida en la Escuela Neoclásica con la que se producirá una sistematización del método económico y se asentarán las bases de lo que hoy constituye la Microeconomía moderna, además de introducir principios fundamentales como es el principio del margen. Con la escuela neoclásica se consagra la Economía como ciencia.

1.3. Los predecesores de la revolución marginalista ⏳ 1'30"

La revolución marginalista es el resultado de los trabajos de varios autores, entre los que destacan William Stanley Jevons, Carl Menger, y Léon Walras. Se trata de trabajos simultáneos en el tiempo, pero independientes entre sí. De hecho, cada uno de estos

[33] Bartels et al. (2023) contrastan empíricamente estas predicciones con datos de panel de regiones prusianas del siglo XIX: la acumulación de capital sí elevó la participación del capital en la renta y la desigualdad, como predecía Marx, pero no llevó a mayor concentración del capital ni a más apoyo electoral al socialismo. Fueron, en cambio, los sindicatos y la actividad huelguística los que limitaron la desigualdad y fomentaron el apoyo político al socialismo —más en línea con los revisionistas de Eduard Bernstein que con la ortodoxia marxista.

autores vivió en un país diferente: Jevons en Reino Unido, Menger en Austria, y Walras en Francia. En 1871, Jevons publica en inglés *La teoría de la economía política* (Jevons, 1871). Ese mismo año, Menger publica en alemán sus *Principios de Economía* (Menger, 1871), y en 1874 Walras, profesor en Suiza, publica los *Elementos de Economía Política Pura* (Walras, 1874). Por ello, se suele considerar 1871 como el inicio de la revolución marginalista, germen de lo que Thorstein Veblen bautizaría en 1900 como "economía neoclásica"[34].

Sin embargo, pese a su nombre, la revolución marginalista no es verdaderamente una revolución, por dos motivos principales:

1. No se trata de un acontecimiento puntual, sino que es más bien el resultado de un proceso.
2. Las ideas que introduce no son totalmente revolucionarias, sino que se pueden identificar muchos precursores. En este apartado se analizan brevemente los precursores más importantes[35].

La verdadera revolución fue, en realidad, metodológica: el paso al individualismo metodológico —frente al análisis clásico por clases sociales— y la introducción sistemática del cálculo diferencial, con la excepción parcial de la escuela austríaca.

A su vez, algunos elementos de los clásicos influirán notablemente en la obra de los marginalistas, como los rendimientos decrecientes que mencionó Ricardo o la consideración de unos mercados en libre competencia. Concretamente, hay una serie de autores que, con anterioridad a los marginalistas, ya hablaban de algún modo de la utilidad marginal, si bien sus teorías en este ámbito serán incompletas o imprecisas:

- Daniel Bernoulli: este autor resolvió la "paradoja de San Petersburgo" en 1738, suponiendo una utilidad marginal decreciente.

[34] Veblen acuña el término en su artículo "Preconceptions of Economic Science" (Veblen, 1900), en el que relaciona a los marginalistas de la tradición de Marshall con los de la escuela austríaca. Alfred Marshall no forma parte del grupo fundador porque publica sus obras en la década de 1890 y, aunque argumentó que sus ideas fueron contemporáneas a las del resto de autores mencionados, esto no ha sido clarificado.

[35] El desarrollo de esta revolución se realiza de manera más o menos independiente a través de diversas escuelas de pensamiento.

- Escuela británica (Jevons, Edgeworth, Marshall, Pigou, Wicksteed): Son los primeros en incluir la oferta y la demanda en la determinación de los precios.
- Escuela austríaca (Menger, Böhm-Bawerk, Wieser): defiende el método deductivo frente al inductivismo de la Escuela Histórica alemana, rechaza el uso de las matemáticas y sostiene que existen leyes económicas exactas que se derivan de la conducta individual. Se dedican al estudio de las regularidades en los procesos de evolución social. Hay que decir que, si bien esta escuela influye en las teorías neoclásicas, se separa posteriormente de esta corriente para estudiar la economía desde un enfoque diferente. Hayek, Schumpeter, Von Mises (austriacos de segunda generación, s.XX).
- Escuela de Lausana (Walras, Pareto, Leontieff): usan profusamente las matemáticas. Se busca dotar a la economía de rasgos científicos.
- Escuela Americana (Clark, Irving Fisher)
- Escuela Sueca (Wicksell, Ohlin, Myrdal)

 © 2026 Manuel Cabado García

- William Forster Lloyd: en su *Lección sobre la noción de valor* (Lloyd, 1834), es uno de los primeros que distingue explícitamente entre utilidad total y marginal, pero no lo aplica a problemas económicos.
- Jeremy Bentham: Bentham también es pionero en definir el carácter decreciente de la utilidad marginal.

Sin embargo, serán otros autores los precursores más cercanos del marginalismo.

Precursores más cercanos del marginalismo

Estos serían los precursores más cercanos a los marginalistas:

- Antoine Augustin Cournot (1801-1877) fue un matemático y economista francés que, en sus *Recherches sur les principes mathématiques de la théorie des richesses* (Cournot, 1838), fue pionero en usar un diagrama para explicar cómo se forman los precios mediante la intersección de una curva de demanda con pendiente negativa y una curva de oferta con pendiente positiva. De este modo, Cournot ya estaba defendiendo la ley de la demanda, según la cual la cantidad demandada disminuye cuando aumenta el precio. Actualmente, se conoce que la cantidad demandada también depende de otros factores como la renta, los precios de otros bienes, etc., pero éstos se toman como exógenos, de modo que el análisis de Cournot es correcto desde esta perspectiva.
- Sin embargo, Cournot rechazó el concepto de utilidad como fundamento de la función de demanda y en cambio optó por un enfoque empírico, es decir, construyó la curva de demanda a través del cálculo de relaciones cantidad-precio basadas en la observación de datos.
- Jules Dupuit (1804-1866) fue un ingeniero, matemático y economista francés que, en su artículo de 1844 "De la mesure de l'utilité des travaux publics" (Dupuit, 1844), al igual que Cournot, señala que existe una relación inversa entre cantidad demandada y precio. Sin embargo, a diferencia de Cournot, Dupuit ya trata de relacionar la demanda con su utilidad. En concreto, Dupuit argumenta que la demanda tiene pendiente negativa porque la utilidad marginal es normalmente decreciente. Es más, va más lejos aún y afirma que la curva de demanda es en sí misma la curva de utilidad. Llega a la conclusión de que, como la demanda es medible cardinalmente, la utilidad también lo es, por lo que es posible realizar comparaciones interpersonales de utilidad, argumento que más adelante será desechado por la teoría económica. Dupuit introduce además el germen del concepto de excedente del consumidor, que Marshall desarrollará décadas después.
- Johann Heinrich von Thünen (1783-1850) fue un economista alemán que, en su obra *Der isolierte Staat* (von Thünen, 1826), fue de los primeros en señalar que, desde

una perspectiva de optimización, los factores productivos deberán ser remunerados de acuerdo a su productividad marginal.

- Hermann Heinrich Gossen (1810-1858) fue un economista alemán. De sus obras se derivan tres enunciados relativos a la utilidad marginal que serán bautizados posteriormente por otros autores como las "leyes de Gossen"[36]:
  - Primera Ley de Gossen: postula que "cada acto sucesivo de consumo produce un placer cada vez menor, hasta llegarse a un punto de saciedad". Por tanto, esta ley señala que la utilidad marginal será decreciente. No obstante, esta aportación no es original de Gossen, pues, como se explicó antes, autores como Bernoulli o Bentham ya habían introducido esta idea.
  - Segunda Ley de Gossen: postula que el individuo deberá distribuir su renta entre los bienes de modo que la utilidad marginal de cada uno, dividida por su precio, sea la misma para todos. De este modo, el último euro gastado en cada bien reporta la misma utilidad.
  - Tercera Ley de Gossen (ley de la escasez): postula que un bien solo tiene valor cuando su demanda excede a su oferta.

1.4. La revolución marginalista ⏳ 2'

La teoría del valor trabajo era el gran defecto de los economistas clásicos. Esta teoría carecía de generalidad, puesto que solo era válida para uno de los tres tipos de bienes que Mill había defendido en su análisis.

La preocupación de los marginalistas era demostrar que es la demanda, y no la oferta, la que determina el precio. Para ello van a conceder un protagonismo crucial a la utilidad y, más en concreto, a la utilidad marginal.

Así, los marginalistas, especialmente Menger y Jevons y, tras Menger, todos los miembros de la escuela austríaca —desarrollada con más detalle en el Apéndice 1.A.3—, van a defender una teoría subjetiva del valor, de manera que, para ellos, el valor de las cosas no depende de su naturaleza objetiva, sino de una estimación subjetiva de los seres humanos, que es lo que representa la utilidad:

- Por utilidad se entiende la capacidad de ciertas mercancías para satisfacer las necesidades de los individuos, lo cual provoca que quieran comprarlas.
- Por lo tanto, el valor (precio) de los bienes desde el punto de vista clásico proviene del pasado (coste de producción), mientras que desde el punto de vista marginalista, el valor proviene del futuro, es decir, la utilidad que reportará su consumo.

[36] Cuando Jevons, Menger y Walras publicaron sus obras en la década de 1870, desconocían la de Gossen(Gossen, 1854). Fue el propio Jevons quien, al descubrirla después, bautizó estas leyes en su honor.

 © 2026 Manuel Cabado García

Sin embargo, esta misma insistencia en la demanda contiene ya el germen de su propio error: al pasarse al extremo opuesto de los clásicos, la teoría marginalista no podrá explicar la formación del precio de los bienes con oferta totalmente elástica. La solución a este nuevo desequilibrio llegará más tarde, con la síntesis de la escuela neoclásica.

En suma, Menger, Jevons, y Walras compartían las siguientes ideas:

1. La utilidad marginal, ligada a la escasez, es el elemento clave de la teoría del valor.
2. La utilidad es medible cardinalmente, es decir, puede expresarse en unidades, de modo que tiene sentido decir cuánto más satisface un bien que otro, y no solo cuál se prefiere.[37]
3. El individuo es capaz de realizar comparaciones interpersonales de utilidad[38].
4. Aceptan las dos primeras leyes de Gossen.
5. La función de utilidad es separable y aditiva $\left(U\left(x_1,\ldots,x_N\right)=\sum_{i=1}^{N}u_i\left(x_i\right)\right)$, por lo que no se contempla la influencia de bienes sustitutivos y complementarios.

De esta forma, mientras que los clásicos se preocupaban por las causas y las consecuencias a largo plazo del crecimiento económico, los marginalistas tienen unas preocupaciones de más corto plazo.

- En este análisis de corto plazo, asumirán que ciertos elementos como la población, el stock de capital y de recursos naturales, o la tecnología vienen dados.
- Esto es importante, porque este énfasis en un entorno en el que los factores productivos vienen dados supondrá que la oferta es completamente inelástica, por lo que el precio depende íntegramente de la demanda.
  La implicación más importante de esto último es que, como la remuneración de los factores productivos depende del precio del producto y éste depende, a su vez, de la demanda, que a su vez depende de la utilidad, la remuneración de los factores productivos viene determinada por la utilidad.

Primera generación: Menger y Jevons

Menger y Jevons son, junto con Walras, los tres autores que descubren de forma simultánea e independiente el principio marginal a comienzos de la década de 1870. A diferencia de Walras, cuya aportación principal —el equilibrio general— se tratará en un epígrafe propio más adelante, Menger y Jevons comparten un mismo objetivo: fundamentar la teoría del valor en la utilidad marginal del consumidor, aunque llegarán a él por caminos metodológicos radicalmente distintos, como se verá a continuación.

[37] La intuición es la diferencia entre decir que un plato gusta "el doble" que otro, lo que exige una escala cardinal, y decir solo que gusta "más", que basta en un enfoque ordinal.

[38] Con la excepción matizada de Jevons, que las consideraba imposibles por la distinta distribución de la renta entre individuos.

Carl Menger

Carl Menger (1840-1921) fue un economista y jurista austriaco, autor de *Principios de Economía Política* (Menger, 1871). Las ideas de Menger harán que este, desde su cátedra de economía en Viena, se convierta en líder de la escuela austríaca, cuyos principales exponentes serán Friedrich Hayek, Friedrich von Wieser, Eugen Böhm von Bawerk o Joseph Alois Schumpeter. Menger y los demás miembros de la escuela austríaca emplean un método deductivo, que Ludwig von Mises sistematizará en el siglo XX bajo el nombre de praxeología[39] y, por lo general, rechazan el uso del método matemático.

Teoría del valor Menger fue quien resolvió de manera más clara la paradoja del agua y los diamantes de Adam Smith:

- La paradoja del agua y los diamantes enfatiza la incoherencia entre dos términos: mientras que el valor de uso o utilidad del agua es mucho mayor que el de los diamantes, su valor de cambio o precio es mucho menor.
- La clave de esta paradoja se encuentra en la confusión entre utilidad total y marginal, siendo esta el incremento de la utilidad debido a la última unidad consumida del bien. De esta forma, es la escasez de los diamantes y la abundancia de agua lo que explica la diferencia entre el precio y la utilidad del agua y los diamantes.

Por otro lado, hay que destacar que Menger y sus discípulos de la escuela austríaca llevarán al extremo su teoría subjetiva del valor, de modo que:

- Rechazarán las comparaciones interpersonales de utilidad.
- Negarán cualquier papel del coste de producción en la determinación de los precios.

Teoría de la distribución Por otro lado, Menger distingue entre "bienes de orden inferior", es decir, los que son capaces de satisfacer necesidades humanas directamente y "bienes de orden superior", es decir, factores de producción que satisfacen necesidades humanas de forma indirecta, contribuyendo a la producción de bienes de orden inferior.

Respecto a los bienes de orden superior, Menger señala que se caracterizan por su complementariedad y su imputación:

[39] La praxeología es el método que deduce toda la teoría económica, por pura lógica, a partir del axioma de que el ser humano actúa con un propósito. El término lo acuña el francés Louis Bourdeau en 1882 y le da su sentido moderno el filósofo Alfred Espinas en 1890, de quien lo toma Mises, discípulo de Böhm-Bawerk, para convertirlo en el fundamento de la economía austriaca en su obra *Human Action* (Mises, 1949), a partir de un único axioma —el "axioma de la acción": el ser humano actúa, es decir, emprende conductas conscientes orientadas a fines elegidos. Para Mises este axioma es autoevidente, en el sentido de que cualquier intento de refutarlo presupone, de hecho, una acción (la de argumentar), confirmándolo. A partir de él, la praxeología deriva por deducción lógica pura —sin necesidad de contrastación empírica— todo el cuerpo teórico de la economía: fenómenos como el intercambio o la formación de precios se siguen necesariamente de la elección individual, y no de regularidades estadísticas observadas.

 © 2026 Manuel Cabado García

- Complementariedad: los bienes de orden superior no tienen carácter de "bien" por sí solos, sino que necesitan complementarse con otros bienes de orden superior (pues si el productor no cuenta con un factor productivo que sea imprescindible en el proceso, el resto de factores productivos pierde su utilidad).
- Imputación: idea similar a la anterior, pero en este caso se refiere a que los bienes de orden superior perderán su condición de bienes si los bienes de orden inferior que producen pierden también su condición de bienes (por ejemplo, si la gente deja de demandar el bien de orden inferior que producen).

En definitiva, se rompe de nuevo con la idea de los clásicos de que el trabajo incorporado, bienes de orden superior, determina el valor del bien de orden inferior. Según Menger, es la utilidad marginal la que determina el valor de un bien de orden inferior. En este sentido, Menger va más allá y estudia cómo se determina el valor de los factores, es decir, bienes de orden superior. Para ello, va a utilizar un concepto clave en Microeconomía, pero que hasta entonces había sido poco aplicado al análisis económico: el coste de oportunidad.

Aplicando este principio a la producción, concluirá que el valor de una unidad de factor de producción es igual al valor del producto que dejaría de producirse si se perdiera dicha unidad. Esto no es otra cosa que una teoría de la valoración de factores según su productividad marginal y su coste de oportunidad.

Además, Menger también señala que, dado que la transformación de bienes de orden superior en bienes de orden inferior lleva su tiempo, y está sujeta a incertidumbre, los bienes de orden inferior tendrán un valor algo más elevado que los de orden superior. De este modo, Menger introduce en su análisis, implícitamente, el concepto de preferencia temporal, que corresponde originalmente a Böhm von Bawerk (Böhm-Bawerk, 1884), quien señala que los bienes presentes son técnicamente superiores a los bienes futuros.

Otros aspectos Para Menger, al contrario que para otros descubridores de la utilidad marginal, los mercados no tienen por qué estar en equilibrio cuando los individuos maximizan su utilidad. Esto se debe a que Menger reconoce que existen problemas de información imperfecta. De ahí la importancia que Menger, y la escuela austríaca en general, otorgan a la figura del empresario, como un agente que busca información y aprovecha oportunidades de beneficio, de lo que se deriva un concepto de competencia como un proceso dinámico, a diferencia del concepto estático de Jevons y Walras. Menger es, además, de los primeros en analizar el dinero como una institución que surge espontáneamente, "orgánicamente", para reducir los costes de transacción del trueque.

William Stanley Jevons

William Stanley Jevons (1835-1882) fue un economista inglés. Un aspecto importante de la obra de Jevons es su idea de que el método apropiado para la economía debía ser el

matemático. Esto se manifestó principalmente en dos aspectos:

- Utilidad marginal. Jevons expresó la utilidad marginal como la variación de la utilidad al variar la cantidad consumida de un bien: dU/dX.
- Relación Marginal de Sustitución. Jevons también aplica el análisis matemático a la teoría del intercambio para concluir que el intercambio entre dos individuos se produce hasta que los cocientes de utilidades marginales de ambos se igualan, es decir, hasta que las RMS de cada individuo se igualen.

Desarrolla la teoría del intercambio basada en la utilidad y en la ley de la indiferencia. Según esta ley, en un mismo mercado y en un mismo momento todas las unidades de una mercancía homogénea se intercambian al mismo precio, porque al comprador le resulta indiferente cuál de ellas recibe.[40]

Por otro lado, la cadena de Jevons es un concepto clave en su teoría del valor. Según Jevons, el valor de los bienes no está determinado directamente por el coste de producción, como sostenían los economistas clásicos, sino por el grado final de utilidad. La cadena se estructura así (Martín Martín, 2010):

- El coste de producción determina la oferta.
- La oferta determina el grado final de utilidad.
- El grado final de utilidad determina el valor.

Esta formulación permitió a Jevons romper con la teoría clásica del coste de producción y fundamentar su análisis en la utilidad marginal decreciente. Además, la ley de utilidad marginal decreciente le permitió formular la ecuación del intercambio, según la cual, en equilibrio, la razón a la que se intercambian dos bienes es la inversa de la razón entre sus grados finales de utilidad.

Este enfoque revolucionó la economía al introducir una perspectiva más subjetiva y psicológica en la determinación del valor. Sin embargo, Jevons asumió erróneamente que este análisis conducía a un único equilibrio. Sería Edgeworth quien demostrara, mediante las curvas de indiferencia, que existen infinitos equilibrios posibles según la dotación inicial[41].

Finalmente, cabe señalar que fue un precursor de la economía ambiental, al poner de manifiesto la preocupación por el agotamiento de los combustibles fósiles (carbón) y su relación con la preeminencia de la economía británica.

[40] La intuición es que nadie paga 2 euros por una barra de pan en un puesto si en el de al lado venden una idéntica por 1,50 euros, de modo que en un mercado bien comunicado todas las barras acaban vendiéndose al mismo precio.

[41] Jevons también extendió el análisis marginal al mercado de trabajo (el trabajador ofrece su trabajo hasta que la desutilidad de la última unidad iguala la utilidad marginal del salario real) y anticipó la idea de los "equilibrios de manchas solares" (sunspot equilibria), que resurgiría en el siglo XX en los modelos de equilibrios múltiples. Su "cadena" de determinación del valor fue, además, criticada por Marshall, Edgeworth y Walras, que mostraron que, con oferta variable, los precios se determinan de forma simultánea, no secuencial.

 © 2026 Manuel Cabado García

En conjunto, el análisis de esta primera generación invierte por completo la causalidad clásica —el valor de un factor depende de la utilidad que reportan los bienes finales que ayuda a producir, no al revés— y se aplica casi en exclusiva al lado de la demanda, asumiendo implícitamente una oferta rígida a corto plazo.

Segunda generación: Wicksteed y Wicksell

Menger, Jevons y Walras aplicaron el análisis marginal casi en exclusiva a la demanda, ignorando por completo el lado de la oferta. Resulta paradójico que la teoría de la productividad marginal no se desarrollara junto con la de la utilidad marginal, sino que hubiera que esperar veinte años, hasta 1890-1894[42], para que una segunda generación de marginalistas como Wicksteed, Wicksell y Clark ampliara el análisis marginal, aplicándolo al lado de la oferta y desarrollando la teoría de la productividad marginal. Esta establece que, al igual que el valor de los bienes depende de su utilidad marginal, la remuneración de los factores productivos se basa en su productividad marginal. Los puntos clave incluyen:

- Producción y distribución: La producción sigue una función continua con productividades marginales decrecientes. El uso óptimo de los factores se alcanza cuando la última unidad asignada a cada uno iguala su productividad marginal (principio de equimarginalidad). Cada factor es remunerado según su productividad marginal, asegurando que el producto total se distribuya completamente:
  - Wicksteed (1894): Para agotar el producto, se requiere competencia perfecta y funciones de producción homogéneas de grado 1 (rendimientos constantes a escala)[43].
  - Wicksell (1900): Sostiene que basta con competencia perfecta y operación en el mínimo de la curva de costes medios a largo plazo.

- Diferencias con los clásicos:
  - Alcance: aplican el principio del margen a todos los sectores productivos, no solo a la agricultura, como hizo Ricardo con la renta de la tierra.
  - Tierra: mientras que los clásicos cuestionaban la legitimidad de remunerar este factor, para los marginalistas es un factor más que debe ser remunerado.

[42] Dos factores explican este retraso: el aumento generalizado de los salarios por encima de los niveles de subsistencia durante la Segunda Revolución Industrial, que contradecía las predicciones pesimistas de Ricardo y Malthus, y el auge de la teoría marxista, que exigía una justificación ética alternativa a la forma en que se distribuye la renta.

[43] Por el Teorema de Euler, si $Y = f(K, L)$ es homogénea de grado uno, entonces $Y = \frac{\partial Y}{\partial K} \cdot K + \frac{\partial Y}{\partial L} \cdot L$. En competencia perfecta cada factor se remunera según su productividad marginal ($r = \frac{\partial Y}{\partial K}$, $w = \frac{\partial Y}{\partial L}$), por lo que sustituyendo se obtiene $Y = r \cdot K + w \cdot L$: toda la producción se destina a remunerar a los factores y el beneficio residual es nulo.

- Capital: Los marginalistas lo ven como un factor productivo que debe ser remunerado, en contraste con los clásicos, que lo consideran un producto ya retribuido.
- Beneficios: Los marginalistas diferencian entre la remuneración del capitalista (interés) y del empresario (beneficio), planteando al empresario como un cuarto factor de producción —una distinción que los clásicos no necesitaban, pues en su época las sociedades por acciones apenas estaban desarrolladas.

Así, la teoría marginalista de la distribución refuerza el sistema capitalista, justificando que cada factor sea remunerado conforme a su aportación. John Bates Clark, en *La distribución de la riqueza* (Clark, 1899), lleva esta idea más allá y sostiene que dicha remuneración es éticamente correcta: cada factor recibe exactamente aquello que aporta a la sociedad, por lo que rechaza expresamente la idea de explotación central en la obra de Marx.

Si el producto se agota exactamente entre los factores, queda pendiente una pregunta: ¿qué justifica entonces la existencia de un beneficio para el empresario? Y, junto a él, ¿por qué el capital recibe una remuneración adicional en forma de interés? La segunda generación marginalista ofrece varias respuestas a ambas cuestiones.

- Teoría del beneficio: conviven tres respuestas. Von Thünen entiende el beneficio como la recompensa por incurrir en riesgos no asegurables: es el residuo que queda tras remunerar a los factores productivos y cubrir los riesgos asegurables, lo que le lleva a plantear al empresario como un cuarto factor productivo. Clark, en cambio, considera que la remuneración del empresario como gestor no debe llamarse beneficio, sino salario. El beneficio genuino solo puede explicarse como el resultado del desequilibrio transitorio que genera el avance tecnológico en la transición al largo plazo, es decir, una renta puramente temporal. Schumpeter lleva esta idea a su extremo (Schumpeter, 1911): si no hay cambio técnico no hay incertidumbre y, por tanto, la tasa de beneficio debe ser nula. Solo la innovación tecnológica genera un beneficio positivo, que además se disipa rápidamente por la imitación de la competencia.
- Teoría del interés: también conviven varias explicaciones. Para Jevons, el capital permite alargar el proceso productivo, funcionando como un adelanto a los trabajadores y una renuncia temporal del capitalista, de modo que el tipo de interés remunera ese tiempo de espera. Para Böhm-Bawerk y Menger, el tipo de interés es el precio del tiempo, la prima por disfrutar de bienes presentes en lugar de esperar a los futuros. Irving Fisher (Fisher, 1907) rompe con ambos enfoques: no concibe el interés como la participación del capital en la renta, sino como una forma general de examinar los flujos de renta de cualquier clase, que se descuentan al tipo de interés corriente para obtener su valor presente. Su nivel depende de fuerzas subjetivas (la preferencia de consumo presente frente a futuro) y objetivas (las oportunidades de

 © 2026 Manuel Cabado García

inversión y la productividad de los factores). El llamado "teorema de la separación de Fisher" —que en un mercado de capitales perfecto las decisiones de inversión y de financiación de la empresa son independientes— precede directamente al célebre teorema de Modigliani-Miller.

En definitiva, la segunda generación marginalista culmina ofreciendo su propia teoría de la distribución: la tasa de beneficio del capital es decreciente, no porque sea residual, como sostenían los clásicos, sino porque la acumulación de capital reduce su productividad marginal —una conclusión que, como se verá en el Apéndice 1.A.4, la Controversia de las Dos Cambridge pondrá en cuestión.

1.5. Los economistas neoclásicos ⏳ 5'

Tanto la teoría clásica del valor-trabajo como la teoría subjetiva del valor de los marginalistas presentan el mismo problema: su incapacidad para ofrecer una explicación general al proceso de determinación de los precios. Marshall pone fin a este problema, al explicar que la oferta y la demanda contribuyen simultáneamente a determinar el equilibrio.

Para ello, recurre al célebre ejemplo de las tijeras, en el Libro V, capítulo III de sus *Principios de Economía* (Marshall, 1890): "Discutir acerca de si el valor está determinado por la utilidad o por el coste de producción sería lo mismo que discutir acerca de si es la lámina inferior de unas tijeras o la superior la que corta un trozo de papel".

Surge así la escuela neoclásica, cuyos padres intelectuales serán Marshall y Walras. Ambos autores comparten esta visión del proceso de formación de precios como resultado de la interacción de la oferta y la demanda, aunque con diferencias:

- Marshall se centra en el análisis de un mercado, haciendo uso de un enfoque de equilibrio parcial. Se apoya, pues, en el principio ceteris paribus, que le permite sostener que una parte de la economía pueda estudiarse de forma aislada, pues se supone que su efecto sobre el resto de la economía será insignificante.
- Walras tiene en cuenta la interacción entre diferentes mercados, recurriendo a un enfoque de equilibrio general.

Alfred Marshall (1842-1924)

Su obra principal es *Principios de economía política* (Marshall, 1890)[44]. Se puede decir que con Marshall el marginalismo se convierte en la escuela neoclásica, en un intento de reconciliar la escuela clásica con la marginalista. Marshall será el líder de la escuela

[44]En esta obra, Marshall emplea un método riguroso, pero su preocupación por ser accesible y la prioridad que daba a la resolución de problemas de la realidad cotidiana, le llevaron a relegar la parte formal de su análisis a notas a pie de página y anexos. Además, Marshall introduce el análisis gráfico en la Economía.

de Cambridge, dentro de la cual se inscriben otros economistas como Arthur Pigou o su alumno John Maynard Keynes[45].

A continuación se presentan las aportaciones de Marshall en la teoría de la demanda, excedente del consumidor, teoría de la oferta, cuasi-rentas y equilibrio

Teoría de la demanda

Marshall coincide con los marginalistas en varios aspectos relativos a la teoría de la demanda:

1. La demanda depende de la utilidad marginal.
2. La función de utilidad es separable y aditiva, es decir, la utilidad de cada bien es independiente de las restantes, por lo que se ignoran las relaciones de sustitución y complementariedad, lo que constituye una de las principales críticas de la demanda Marshalliana.
3. La utilidad es cardinal.
4. Acepta las dos primeras leyes de Gossen. Así:
   a) Primera ley: utilidad marginal, UMg, decreciente.
   b) Segunda ley: igualdad de las utilidades marginales ponderadas:

$$\frac{UMg_1}{P_1} = \frac{UMg_2}{P_2} = \ldots = UMg_{\text{renta}} = \lambda \tag{1.3}$$

   de donde se obtiene que

$$UMg_i = P_i \cdot UMg_{\text{renta}} \tag{1.4}$$

[45] En 1924 el propio Keynes escribió un obituario para Marshall, donde reflexiona sobre la vida de su maestro y esboza lo que para él debería ser el economista ideal:

> *"In another respect the diversity of his nature was pure advantage. The study of economics does not seem to require any specialised gifts of an unusually high order. Is it not, intellectually regarded, a very easy subject compared with the higher branches of philosophy and pure science? Yet good, or even competent, economists are the rarest of birds. An easy subject, at which very few excel! The paradox finds its explanation perhaps, in that the master-economist must possess a rare combination of gifts. He must reach a high standard in several different directions and must combine talents not often found together. He must be mathematician, historian, statesman, philosopher-in some degree. He must understand symbols and speak in words. He must contemplate the particular in terms of the general, and touch abstract and concrete in the same flight of thought. He must study the present in the light of the past for the purposes of the future. No part of man's nature or his institutions must lie entirely outside his regard. He must be purposeful and disinterested in a simultaneous mood; as aloof and incorruptible as an artist, yet sometimes as near the earth as a politician. Much, but not all, of this ideal many-sidedness Marshall possessed. But chiefly his mixed training and divided nature furnished him with the most essential and fundamental of the economist's necessary gifts-he was conspicuously historian and mathematician, a dealer in the particular and the general, the temporal and the eternal, at the same time" (Keynes, 1924).*

 © 2026 Manuel Cabado García

donde el precio del bien i, P_i, representa la función inversa de demanda de dicho bien. Así:

- Por tanto, la demanda tiene pendiente negativa porque la utilidad marginal es decreciente: el individuo está dispuesto a pagar más por las primeras unidades que por las siguientes.
- Más aún: suponiendo que la utilidad marginal de la renta sea constante, la demanda (P_i) podría medir perfectamente las variaciones de utilidad de los individuos (UMg_i).

A su vez, establece una relación directa entre la demanda y la utilidad marginal. Para ello, comienza con una función de utilidad de la siguiente forma:

$$UT = \sum_{t=0}^{\infty} u(x_i) + M \tag{1.5}$$

por lo que la función de utilidad de Marshall es cardinalista (medida en unidades monetarias), aditivamente separable y cuasilineal en la renta (por lo que la utilidad marginal de la renta o dinero, M, será constante)[46].

Así pues, el análisis anterior llevará a Marshall a formular la famosa ley general de la demanda, según la cual la cantidad demandada aumentará cuando el precio disminuye y disminuirá cuando el precio aumenta[47][48].

Excedente del consumidor

Marshall inicia el estudio del bienestar social a través del concepto de excedente del consumidor, concepto acuñado por Dupuit y extendido por Marshall.

[46] Una diferencia importante entre Marshall y los marginalistas es que el primero consideró que las funciones de utilidad de los individuos sí eran agregables (en contraste con la versión más estricta de la teoría subjetiva del valor), de manera que se puede obtener una función de demanda agregada como suma horizontal de las individuales (sin llegar a analizar los problemas inherentes a la agregación, distintas preferencias, distintas rentas...). Ignora el ER (función cuasi-lineal respecto al dinero) y el ES entre los bienes.

[47] Para que esta ley sea cierta, Marshall supuso que el ER era despreciable, de modo que la Umg del dinero permanece constante en los cambios pequeños de precio de cualquier mercancía. De esta forma, solo funciona el ES ante cambios en el precio. Además, como no existen relaciones de complementariedad ni de sustitubilidad no hace falta diferenciar el ES bruto del neto.

[48] Finalmente, cabe señalar que Marshall también realiza otras aportaciones importantes a la teoría de la demanda:

- ER y ES. Es consciente de la existencia (aunque los obvia), ante variaciones en los precios, de un efecto sustitución y de un efecto renta. En base a estos dos efectos, Marshall distingue entre bienes normales, inferiores y Giffen.
- Elasticidad-precio. También se asocia a Marshall con el concepto de elasticidad-precio de la demanda.
- Movimientos EN y DE la curva. Distingue entre movimientos EN la curva de demanda y desplazamientos DE la curva.
- Excedente del consumidor

Como la utilidad marginal del consumidor es decreciente, éste recibe más utilidad por las primeras unidades consumidas que por las últimas, pero paga por todas ellas el precio que paga por la última, por lo que paga una cantidad total inferior a la valoración subjetiva que hace de dicho consumo[49].

Teoría de la oferta

En su análisis de la oferta, en primer lugar Marshall introducirá dos conceptos relevantes: el de industria y el de empresa representativa:

- Industria: es un conjunto de empresas que producen la misma mercancía.
- Empresa representativa: no es una empresa media, en términos estadísticos, sino aquella que tiene acceso normal a la tecnología disponible dentro de su industria.

A continuación, Marshall introduce en su análisis el concepto de tiempo y distingue así entre cuatro períodos de tiempo distintos:

1. Muy corto plazo o "período de mercado". La oferta es fija y, por tanto, la demanda determina íntegramente el precio de mercado.
2. Corto plazo. Se trata de una situación en la que la tecnología no varía y la dotación de algunos factores productivos es fija. Es decir, hay algún factor productivo cuya dotación se puede modificar, es decir, es posible variar el volumen de producción, pero no el tamaño de planta. En este análisis, Marshall distinguirá entre costes fijos y variables, mostrando que una empresa seguirá operando en el corto plazo siempre que cubra sus costes variables (incluso si incurre en pérdidas). La curva de oferta a corto plazo de una industria perfectamente competitiva es por tanto aquel tramo de la curva de coste marginal que está por encima del mínimo de la curva de coste variable medio. El precio se determina mediante la interacción de la oferta y la demanda.
3. Largo plazo. Se trata de una situación en la que la tecnología no varía, pero la dotación de todos los factores productivos es variable (es decir, el productor puede adecuar el tamaño de planta). En este escenario de largo plazo, Marshall señala que la curva de oferta puede ser horizontal, creciente o decreciente, en función de si se opera con costes constantes, crecientes, o decrecientes, respectivamente. Profundizando en este análisis, Marshall se pregunta qué motivos pueden estar detrás de cada uno de los tres escenarios de costes que se acaban de plantear. En este sentido, resulta de particular interés el análisis que realiza del caso de los costes decrecientes, señalando que este tipo de costes puede deberse a dos tipos de fuerzas:

[49] El principal problema del análisis marshalliano del EC es que lo expresó en términos monetarios, bajo el supuesto de un ER despreciable. Debido a que la UMg de la renta es decreciente a medida que ésta aumenta, la demanda marshalliana sobreestimará el EC si se trata de un bien normal, y lo infraestimará si se trata de un bien inferior.

 © 2026 Manuel Cabado García

   a) Economías internas a la empresa: los costes medios decrecen con la producción corriente de la propia empresa.

   b) Economías externas a la empresa pero internas a la industria: los costes medios de cada empresa decrecen con la producción agregada de la industria[50].

4. Muy largo plazo o "período secular". Permite variar, además de las dotaciones de factores productivos, la tecnología.

La conclusión general que obtiene Marshall de este análisis de los distintos horizontes temporales es que, cuanto más largo sea el período considerado, mayor influencia tendrá el coste de producción sobre el precio y, cuanto más corto, mayor será la influencia de la demanda

Cuasi-rentas

La aplicación del concepto de tiempo al análisis de la oferta también permite a Marshall obtener el concepto de cuasi-renta[51], que es la remuneración que obtienen los factores fijos en el corto plazo:

- En el corto plazo, la empresa, para producir, se preocupa principalmente de cubrir sus costes variables, y todo lo que obtenga por encima de dicha cantidad será una remuneración a los factores fijos, una cuasi-renta. "Cuasi" porque es una renta que desaparece en el largo plazo, en el que no hay factores fijos.
- Si, además, los ingresos a corto plazo superan lo necesario para remunerar los factores fijos, ese exceso es un beneficio extraordinario, que la entrada de competidores eliminará a largo plazo.

El concepto de cuasi-renta permite aclarar la controversia entre los economistas clásicos y los marginalistas respecto a la relación entre precio y coste:

- Mientras que los clásicos afirman que las remuneraciones a los factores productivos determinan los precios de los bienes.

[50] Nótese que la existencia de costes decrecientes entra en contradicción con el supuesto de competencia perfecta: si la curva de coste marginal siempre decrece, no se puede maximizar el beneficio ni obtener la curva de oferta de la industria.

[51] La cuasi-renta es el exceso de los ingresos sobre los costes variables que obtienen a corto plazo los factores cuya oferta no puede aumentarse de inmediato. Es decir, los ingresos adicionales que recibe un factor productivo cuando en el momento la oferta no puede aumentarse (c/p).

Además, cabe hacer referencia a su teoría de la distribución: Marshall se desmarca de la teoría de la distribución por clases sociales. A través del empleo de la oferta y la demanda de cada mercado se puede determinar el precio de los factores productivos en cada mercado. Sin embargo, desconfiará de la teoría de la distribución basada en la productividad marginal ya que, por lo general, la tecnología requiere que, al aumentar la cantidad de un factor se aumente también la del resto. Para solucionar el problema de medición de las Pmg ideó el concepto de producto neto en el margen: si un trabajador requiere un martillo, el producto neto del trabajo será el producto del trabajador menos el coste del martillo.

- Los marginalistas afirman que los precios de los bienes determinan las remuneraciones a los factores productivos.

Sin embargo, Marshall señala que el que se dé una relación u otra depende del tiempo que se considere:

- En el corto y muy corto plazo, la oferta de los factores productivos es prácticamente fija, por lo que es el precio del bien el que determina las remuneraciones de los factores y no al revés: si aumenta la demanda, el precio aumenta, pero las cantidades de los factores no pueden aumentar, por lo que su remuneración aumenta.
- En el largo y muy largo plazo, las dotaciones de los factores productivos son variables, por lo que son las remuneraciones a los factores productivos las que determinan el precio del bien.

Equilibrio

La interacción de la oferta y la demanda tiene lugar en el "punto de cruce Marshalliano" (aspa marshalliana) de ambas curvas, que determinará el equilibrio del mercado. De esta manera, reconcilia la postura clásica, donde la oferta es la que determina el precio, y la postura marginalista, que opina que es la demanda la que determina el precio. Un rasgo distintivo del enfoque marshalliano, frente al walrasiano, es que Marshall considera la cantidad como la variable independiente y el precio como la dependiente: para una cantidad dada, la oferta indica el precio mínimo al que las empresas están dispuestas a vender, y la demanda el precio máximo al que los consumidores están dispuestos a comprar.

Respecto a la estabilidad del equilibrio, Marshall considera que, ante una situación que no sea la de equilibrio, los ajustes necesarios se producirán vía cantidades y no por tanteo del precio de manera que si, por ejemplo, el precio de demanda excede al de oferta, entra en funcionamiento una fuerza que induce a los productores a aumentar su cantidad ofrecida, y a los consumidores a reducir sus compras.

Otras aportaciones

Otras aportaciones del economista británico son:

- Teoría del bienestar: A finales del s.XIX el movimiento socialista se encontraba en auge, por lo que los economistas se planteaban el problema de cómo fundamentar la adecuada intervención del Estado. En este contexto, surge la Economía del Bienestar. Marshall distinguirá la riqueza de la utilidad o bienestar social. A través del EC se puede medir la utilidad en términos de dinero. Defenderá que el sistema de libre mercado permite maximizar el bienestar de la sociedad, por lo que se mostró cauteloso frente a la intervención. Fue su discípulo y sucesor en la cátedra de Cambridge,

 © 2026 Manuel Cabado García

Arthur Cecil Pigou, quien fundó propiamente la economía del bienestar moderna: introdujo el concepto de externalidad —la diferencia entre el producto social marginal neto y el privado— y propuso los impuestos y subvenciones pigouvianos para corregirla. Además, en su debate posterior con Keynes tras la *Teoría General* (1936), formuló el "efecto Pigou": una caída de precios en las crisis aumenta el valor real de los saldos monetarios y, con ello, el consumo, reactivando la economía sin necesidad de intervención pública.

- Teoría monetaria: → k de cambridge. Reformula la TCD desde un punto de vista microeconómico, preguntándose por las razones que llevan a los individuos a demandar dinero pese a no reportar intereses. El equilibrio en el mercado de dinero viene dado por $M_S = M_D = (1/V) \cdot P \cdot Y$
- Comercio internacional: Reformulará y explicará gráficamente la teoría de la demanda recíproca de Mill (Curva de Oferta Neta, Relación Real de Intercambio) y analizará las condiciones de mejora del saldo de la Balanza de Pagos tras una devaluación, reflejado en la condición Marshall-Lerner:$|\varepsilon_M| + |\varepsilon_X| > 1$.

Léon Walras (1834-1910)

Walras fue un economista francés que se basa de forma profusa en las matemáticas (él mismo reconoció en este hecho la influencia de Cournot, que había sido compañero de estudios de su padre, Auguste Walras, en la École Normale Supérieure de París, y quien animó a su hijo a aplicar el método matemático a la economía). Walras es el líder de la "Escuela de Lausana", a la que pertenece también Vilfredo Pareto, su sucesor en la cátedra.

Walras se cita a menudo como uno de los tres principales exponentes de la revolución marginalista porque:

- Coincide en el tiempo con Menger y Jevons en la publicación de sus trabajos.
- También se basa en el concepto de utilidad marginal, a la que él denomina "rareté".

No obstante, su teoría del equilibrio general le sitúa como uno de los padres de la economía neoclásica.

Su obra principal es *Elementos de economía política pura* (Walras, 1874). Para Walras, la economía estaba dividida en dos partes:

1. "Economía pura", que explica el funcionamiento del sistema económico por medio de un modelo competitivo hipotético.
2. Análisis de política económica, dividido en políticas que afectan a la producción y políticas que afectan a la distribución.

Su principal aportación se relaciona, como indica el título de la obra, con la economía pura, y es la teoría del equilibrio general.

Teoría del equilibrio general

La teoría del equilibrio general supera la perspectiva del equilibrio parcial de Marshall y analiza de forma simultánea los diferentes mercados de una economía y el proceso por el cual se alcanza el equilibrio general competitivo, es decir, el conjunto de precios y cantidades con el que cada agente maximiza su objetivo tomando los precios como dados y todos los mercados se vacían a la vez.

Ahora bien, la teoría del equilibrio general no es, en puridad, una aportación completamente pionera. De hecho, los fisiócratas ya habían aludido a la importancia de este concepto con trabajos como el *Tableau économique* de Quesnay[52]. Pero con Walras este análisis adquiere una nueva dimensión.

Se basa en los siguientes supuestos:

1. Contexto de competencia perfecta, que supone libertad de entrada, plena flexibilidad de precios, empresas con ingresos iguales a sus costes, etc.
2. Los agentes se clasifican en dos grupos, superando la noción de clase social clásica: consumidores y productores[53].
3. Las incógnitas son los precios y las cantidades de bienes y factores productivos[54].
4. En el modelo inicial, no existe producción, es decir, es un modelo de intercambio puro, de manera que los agentes, perfectamente informados, llegan a la subasta con una dotación de bienes y dinero que opera como restricción presupuestaria.

Subastador walrasiano y proceso de tâtonnement

A continuación, Walras introduce el concepto de subastador walrasiano. Se trata de un agente que gestiona el mercado, procesa todas las ofertas y demandas, determina por tanteo qué precios vacían el mercado y solo tras esto permite el comercio a los precios de equilibrio, es decir, no realiza transacciones a precios distintos del de vaciado.

[52] El *Tableau économique* de François Quesnay describe un modelo para la economía de las naciones según el cual la sociedad se divide en tres clases: propietarios de la tierra (aristócratas, alto clero), trabajadores estériles (artesanos, mercaderes), y trabajadores productivos (agricultores, ganaderos, pescadores). Este modelo prevé un flujo de mercancías con un esquema basado en la circulación sanguínea. Se trata de un equilibrio económico estacionario, cerrado, en el cual no existe distinción entre factores productivos y bienes producidos. Se considera uno de los antecedentes a los desarrollos del flujo circular de la renta.

[53] A partir de ahí se construirá un sistema de ecuaciones. Para construir un sistema de ecuaciones Walras distingue entre:

- En el mercado de bienes, los consumidores compran los bienes a las empresas que los ofertan
- En el mercado de ff.pp, los consumidores ofertan sus ff.pp. a las empresas que los demandan, obteniendo de ahí los ingresos.
- En el equilibrio se obtienen los precios relativos en términos del numerario y las asignaciones que cumplen: cada individuo maximiza su utilidad, el precio iguala la oferta a la demanda y no existe arbitraje.

[54] Si hay "m" productos y "n" ff.pp., "m" precios de los productos y "n" precios de los ff.pp., el número de incógnitas será 2m + 2n. Un bien debe hacer de numerario, por lo que la solución vendrá dada en precios relativos (los precios de los demás bienes se expresan en cantidades del bien numerario).

 © 2026 Manuel Cabado García

Este subastador modifica los precios a través de un proceso de tâtonnement, por el cual la economía evoluciona hacia el equilibrio general.

Dicho proceso consiste en que el subastador aumentará los precios de los bienes cuando haya exceso de demanda, y los reducirá cuando haya exceso de oferta. Así, el proceso conducirá a una eventual anulación de todos los excesos de oferta y de demanda.

Como se ve, para Walras, la variable de ajuste hacia el equilibrio es el precio, mientras que para Marshall es la cantidad.

Resultados

Los principales resultados derivados de su investigación son los siguientes:

1. Ley de Walras: la suma del valor monetario de los excesos de demanda en todos los mercados será igual a cero para cualquier vector de precios.

$$\sum_{j=1}^{n} p_j \cdot Z_j\left(P\right) = 0 \tag{1.6}$$

2. El equilibrio walrasiano es un vector de precios, P^*, que hace compatibles los planes de todos los agentes y vacía simultáneamente todos los mercados, es decir, que anula el exceso de demanda, Z_j, de cada bien:

$$Z_j\left(P^*\right) = 0 \quad \forall j = 1, \ldots, n \tag{1.7}$$

3. Características del equilibrio:

   - Existencia. Walras creyó asegurar la existencia del equilibrio contando ecuaciones e incógnitas: al tomar un bien como numerario desaparece un precio desconocido y, por la ley de Walras, una de las ecuaciones de mercado es redundante, de modo que ambos números coinciden. Sin embargo, que coincidan no garantiza una solución con precios y cantidades no negativos. Abraham Wald (1933) dio en los años treinta las primeras demostraciones rigurosas bajo supuestos restrictivos y, en la década de 1950, Kenneth Arrow y Gerard Debreu (1954) probaron la existencia con carácter general mediante un teorema de punto fijo.
   - Unicidad. Según Walras, pueden existir varios equilibrios, es decir, equilibrios de alta y baja productividad. Existen distintos métodos para la demostración como el Teorema de la Preferencia Revelada y de los sustitutivos brutos.
   - Estabilidad. Walras afirma que el proceso es estable gracias al proceso de tâtonnement. Es decir, el agente que gestiona el mercado actúa como "mano invisible" que permite alcanzar el equilibrio general.

Otras aportaciones

Walras también realiza otras aportaciones: Elimina el beneficio como remuneración a un cuarto factor productivo. No se puede tratar al empresario como un factor por separado, sino como una función que puede ser desempeñada por cualquier agente. Desaparece por tanto la idea de beneficio como remuneración a una clase socioeconómica. Walras afirma que existirá beneficio extraordinario, como un exceso de los ingresos sobre los costes, una ganancia accidental y no contractual que obtiene el empresario.

En el largo plazo los beneficios normales son iguales a cero. Walras afirma que un beneficio positivo en el largo plazo solo puede ser el resultado de una continua perturbación del estado estacionario a causa del progreso técnico.

Recepción y valoración posterior

El equilibrio general walrasiano ha recibido algunas críticas: se le considera demasiado matemático, demasiado abstracto y alejado de la realidad. No obstante, ha sido importante para el estudio de las interdependencias de los sectores. Por ejemplo, los estudios de Leontief en las Tablas Input-Output.

No se valoró su obra correctamente en un principio, tanto por el método matemático como por su carácter socialista: apoyaba la competencia y la libre empresa, pero defendía la expropiación de toda la tierra, de hecho, con el alquiler de las tierras no sería necesario recaudar impuestos.

Desarrollos posteriores

Hacia finales del siglo XIX, el edificio neoclásico estaba ya estructurado, pero faltaban una serie de piezas para completar sus teorías. Los desarrollos posteriores más importantes fueron:

- Teoría monetaria:
  - Irving Fisher en *The purchasing power of money* (1911), desarrollaría la famosa teoría neoclásica de la Teoría Cuantitativa del Dinero, que propone que el dinero es neutral, y que variaciones en la masa monetaria de una nación, se traducirán en variaciones en la misma magnitud del Nivel General de Precios.
  - Knut Wicksell y la teoría monetaria moderna. Para Wicksell un aumento de la oferta monetaria también lleva a un aumento de los precios. No obstante, el mecanismo de transmisión no serán los saldos reales como en Fisher, sino los tipos de interés.
- Teoría de la demanda: Edgeworth y otros autores (Pareto, Hicks, Allen) concluyen que la utilidad no es medible cardinalmente, por lo que se pasa a una teoría ordinal de

 © 2026 Manuel Cabado García

la utilidad, de manera que la utilidad se representa mediante curvas de indiferencia, cuya pendiente es la relación marginal de sustitución entre los bienes.

- Teoría del Bienestar:
  - Pigou introdujo la idea de "fallo de mercado", que hacía necesario introducir elementos correctores ya que la asignación competitiva era ineficiente desde el punto de vista social.
  - Pareto instauró un concepto de eficiencia, el "óptimo de Pareto", como aquella situación en la que no es posible mejorar el bienestar de un agente sin empeorar el de algún otro. Se apoya para ello en las curvas de indiferencia de Edgeworth y en la "caja de Edgeworth", el diagrama que representa conjuntamente las dotaciones y preferencias de dos agentes: el lugar geométrico de las tangencias entre sus curvas de indiferencia, la "curva de contrato", recoge precisamente el conjunto de asignaciones eficientes en el sentido de Pareto. Al no permitir comparaciones interpersonales de utilidad, Pareto no busca un único óptimo social, sino que reconoce un número infinito de óptimos —cada punto de la curva de contrato—, sin que sea posible comparar el bienestar agregado que generan entre sí.
- Comercio internacional: El modelo neoclásico paradigmático en el ámbito del comercio internacional es el de Heckscher-Ohlin-Samuelson, que señala que la razón última del comercio es la distinta dotación relativa de factores.
- Crecimiento económico: El modelo neoclásico paradigmático en el ámbito del crecimiento económico es el de Solow, en el que la tasa de crecimiento de una economía se mantiene en el tiempo.

1.6. Conclusiones ⏳ 30"

Si la escuela clásica dio origen a la ciencia económica, la escuela neoclásica la consolida como tal. Los principales motivos del triunfo del pensamiento neoclásico son:

1. Permite una unificación en el tratamiento dado a los problemas del valor y de la distribución.
2. Aporta nuevos instrumentos muy útiles como el análisis marginal.
3. Sustituye el análisis clásico por clases sociales por el individualismo metodológico.

La principal implicación del triunfo del pensamiento neoclásico fue la consolidación de la fundamentación microeconómica, que es el enfoque que predomina hoy en día.

Será el estudio de Keynes de la economía, y la crisis económica de los años treinta, los que incluyan el análisis agregado en el estudio de la economía, con su preocupación en

controlar los agregados macroeconómicos. Sin embargo, en la actualidad esta distinción entre Microeconomía y Macroeconomía es más difusa, por cuanto los determinantes de la inflación, el desempleo o el déficit por cuenta corriente se explican a través de comportamientos individuales frente a la falta de información, la rigidez de precios o la ausencia de movilidad de los factores.

 © 2026 Manuel Cabado García

Apéndice

1.A.1. Cómo influyó la escuela clásica en los desarrollos posteriores

La escuela clásica influirá en el desarrollo de la teoría económica a lo largo de los dos siglos siguientes:

- La obra de Smith dará lugar a la ciencia económica como tal.
- La ley de rendimientos decrecientes y la teoría de la distribución de Ricardo serán el punto de partida para el análisis de la producción basado en el margen.
- El análisis de Malthus influirá en Keynes.
- La obra de Mill será un punto de partida para las teorías del bienestar.
- La influencia de Marx ha sido enorme en los regímenes socialistas.

En definitiva, si bien sus obras carecen de un análisis consistente, existen múltiples paradojas en sus conclusiones, y en algunos casos sus teorías están inconclusas, es innegable que la escuela clásica influye de forma determinante en el desarrollo posterior de la economía.

No obstante, esta escuela se encuentra muy centrada en la oferta, de manera que no puede resolver, por ejemplo, la paradoja del agua y los diamantes. La solución a estas cuestiones llegará con los economistas marginalistas.

1.A.2. Karl Marx

Marx plantea una crítica frontal a los clásicos, se asemeja a los mismos en la adopción de una teoría del valor trabajo, en la formulación de leyes para explicar la economía y en la adopción de modelos de economía cerrada sin Estado y basados en la competencia. No obstante, muestra importantes diferencias con los clásicos tradicionales:

- Marx no divide la sociedad en trabajadores, capitalistas, y terratenientes, sino en "dominantes" y "dominados" a través de una relación de producción, que se identifican con las clases de la burguesía y el proletariado.
- Marx no se considera un economista puro. Cree que la economía no puede subsistir como ciencia separada de la sociología, la historia, la antropología, etc., razón por la cual desarrolla su "teoría del materialismo histórico" con la que intenta construir una ciencia social unificada. Esta teoría señala que lo que mueve la historia es la lucha de clases, que supone un enfrentamiento por el reparto del excedente.
- Mientras que los economistas clásicos consideran que la economía se rige por leyes universales que no se pueden dominar, Marx argumenta que las leyes dependen del momento histórico.

Si bien sus aportaciones se extienden a los ámbitos de la política, la filosofía y la economía, Marx ante todo fue un filósofo que consideraba que su labor no debía consistir en meramente analizar la sociedad, sino también promover los cambios que consideraba deseables. Para ello, abogó por una revolución fundamental de la sociedad y la economía, en vez de la introducción de pequeños cambios. De hecho, acuñó los nombres y los términos con los que se definen en la actualidad al socialismo y al comunismo.

El método de Marx es el materialismo dialéctico, que es un tanto ecléctico. Combinará el deductivismo al emplear axiomas y el uso de la lógica, con el inductivismo, al partir del análisis de la realidad y buscar generalizaciones. Por último, tendrá un marcado carácter histórico, al afirmar que las leyes no son eternas ni universales, sino específicas al momento histórico.

Las cuatro aportaciones más importantes de Marx son los modos de producción, su teoría del valor, las leyes del movimiento capitalista, y sus recomendaciones de política económica.

Modos de producción

El concepto de "modo de producción" es un concepto abstracto que contiene tres elementos:

1. Relaciones de producción. Se trata de las relaciones contractuales con las que se reparte la producción entre dominantes y dominados, y están caracterizadas por un conflicto permanente. Además, Marx habla del "fetichismo de la mercancía", en el sentido de que dichas relaciones contractuales son falsas en la medida en que son controladas totalmente por el capitalista como propietario de los medios de producción y es el que realmente determina las estructuras de propiedad.
2. Estructura económica. Es el ecosistema que surge como resultado de todas las relaciones de producción en la economía.
3. Superestructura jurídica y política. Es el esquema que se configura sobre la estructura económica, y que tiene como fin perpetuar las relaciones de producción. Está representada por el Estado y por la ideología que lo legitima.

Teoría del valor

Producto vs. mercancía. Marx distingue entre "producto" y "mercancía". Todas las economías producen "productos", es decir, aquellos bienes que tienen "valor de uso", pero en el capitalismo los "productos" adquieren la forma de "mercancías" (que tienen valor de uso y de cambio).

Esta doble naturaleza de la mercancía procede, a su vez, del carácter dual de todo trabajo que la produce: el mismo trabajo es a la vez "trabajo concreto", es decir, la actividad útil específica (tejer, sastrear, etc.) que crea el valor de uso, y "trabajo abstracto",

 © 2026 Manuel Cabado García

el gasto indiferenciado de fuerza de trabajo humana en general, que es lo que crea el valor de cambio. No son dos tipos de trabajo separados, sino dos aspectos de un mismo acto productivo: todo trabajo concreto es simultáneamente trabajo abstracto en la medida en que se homogeneiza como parte del trabajo social total.

Para Marx, el valor de los bienes está determinado por el coste de producción, y el único coste de producción es el del trabajo productivo. Marx rechaza la idea de que el valor de los bienes venga determinado por las fuerzas de mercado (incluso a corto plazo).

Para determinar el valor del trabajo (y, por lo tanto, el de los bienes), Marx parte de una fuerza de trabajo que considera homogénea. Conviene matizar que el trabajo relevante no es el efectivamente empleado por cada productor concreto, sino el "tiempo de trabajo socialmente necesario", es decir, el que se requiere en las condiciones normales de producción vigentes y con el grado medio de destreza e intensidad del trabajo: un productor menos eficiente que el resto no incorpora más valor a su mercancía por tardar más tiempo en producirla. Sobre esta base, Marx distingue entre:

- "Trabajo muerto" o "capital constante" (c): es el trabajo asociado a una máquina, que no produce ninguna plusvalía. En efecto, Marx niega que la maquinaria sea productiva por sí misma, ya que considera que sólo es productiva por la cantidad de trabajo que se utiliza en ella.
    - "Trabajo vivo", que representa el valor de cambio, y está formado a su vez por:
    - "Capital variable" (v): la parte del valor generado por el trabajo vivo que se paga al trabajador como salario, equivalente al valor socialmente necesario para reproducir su fuerza de trabajo.
- "Plusvalor" (pl): el trabajo se remunera por debajo del valor que realmente produce, dando lugar a la plusvalía, que procede únicamente del trabajo vivo y no del intercambio. Es la diferencia entre el valor total generado por el trabajo y lo pagado como capital variable. Plusvalor = valor generado por el trabajo – capital variable.[55]

La idea de Marx es que como el capitalista domina las relaciones de producción, puede aprovecharse de esto para obtener una ganancia, es decir, remunerar al trabajo por un valor menor al que realmente produce, dando lugar a una plusvalía que procede del trabajo exclusivamente y surge en la producción, y no del proceso de intercambio.

Marx supuso que las fuerzas de un mercado competitivo ocasionarían que la tasa de beneficios fuese la misma para todos los sectores y empresas, pues una ganancia mayor en un sector ocasionaría una desviación de los recursos que resultaría a largo plazo una

[55] De estas variables se derivan tres ratios que modelizan el comportamiento dinámico del sistema: la tasa de explotación $e = pl/v$, la composición orgánica del capital $k = c/v$ y la tasa de beneficio $r = pl/(c+v)$. Combinándolas se obtiene $r = e/(1+k)$: la tasa de beneficio depende positivamente de la explotación y negativamente de la mecanización —la relación que explica, según Marx, la caída tendencial del beneficio.

tasa de beneficio uniforme. Esta es la "Teoría de la competencia del capital" (aunque ello no implica que la composición orgánica del capital sea la misma).

Leyes del movimiento capitalista

Marx describe cinco leyes o tendencias generales inherentes a la dinámica capitalista. Cada una de ellas brota de la naturaleza dinámica de la economía y tiene su raíz en el conflicto entre "fuerzas productivas" dinámicas y las "relaciones de producción" estáticas:

1. Ley de la acumulación y la tasa decreciente de ganancia. Los capitalistas tratan de obtener las máximas plusvalías posibles y, como éstas proceden del trabajo, tratarán de explotarlo lo máximo posible. Los capitalistas deberán buscar métodos de producción intensivos en trabajo, a fin de maximizar sus beneficios. Sin embargo, los capitalistas continúan acumulando capital. Así se da una excesiva acumulación de capital: si un capitalista individual instala un mayor nivel de mecanización, su plusvalía aumenta porque produce a un menor coste (sustituye trabajo por capital, ↓valor en cambio), pero puede seguir vendiendo al mismo precio que el resto de empresas, menos mecanizadas. Pero si otros capitalistas le copian y este proceso se acaba extendiendo, entonces aumentará la competencia y el precio se reducirá[56].
2. Ley de la concentración creciente y de la centralización de la industria. La búsqueda del beneficio lleva a una mayor sustitución del trabajo por capital donde los capitalistas tenderán a crear empresas de gran escala, con una división del trabajo más marcada y una capacidad de producción mucho mayor. Esto puede dar lugar a sobreproducción, reduciendo los precios hasta un punto en el que sólo los productores más eficientes sobrevivirán. Así, la industria se irá centralizando progresivamente y el poder económico se concentrará en un número de manos cada vez menor.
3. Ley del creciente ejército industrial de reserva. El capital sustituye al trabajo, lo que genera paro y da lugar a un creciente "ejército industrial de reserva", es decir, una masa de desempleados cuya competencia por los puestos de trabajo mantiene los salarios cerca del nivel de subsistencia. Un desempleo que es de dos tipos: el paro tecnológico, causado por la sustitución de trabajo por maquinaria, y el desempleo cíclico, causado por el exceso de producción, que a su vez es resultado de la creciente concentración y centralización.
4. Ley de la miseria creciente del proletariado. Conforme crece el "ejército de reserva", crece la miseria del proletariado. Debido a la creciente competencia en los mercados, los capitalistas buscarán mantener los beneficios a través de una reducción de los salarios, la precarización del empleo (incorporación de mujeres y niños a la fuerza

[56] Crítica: Marx no tiene en cuenta que, al aumentar el capital, la productividad del trabajo aumenta, lo cual puede aumentar la plusvalía. Por lo tanto, no es tan claro que un aumento del capital vaya a reducir el beneficio.

 © 2026 Manuel Cabado García

de trabajo) y un aumento de las horas laborales exigidas (jornadas más largas para dar un mayor uso a las máquinas).

5. Ley de las crisis y depresiones. La miseria creciente del proletariado tiende a la revolución social, esto es, a que el proletariado acabe apoderándose de los medios de producción. Marx vincula los ciclos económicos al gasto en inversión. Cuando el ejército de desempleados aumenta y los salarios disminuyen, los capitalistas tenderán a contratar más trabajo y a reducir capital. Cuando los salarios suben, los capitalistas sustituyen trabajo por capital. Las crisis son una parte integral del proceso capitalista. Éstas se harán cada vez más recurrentes debido a la creciente precariedad y competencia, hasta que desencadenen la revolución social (pues el ejército industrial de reserva aumentaría a medida que las crisis fueran siendo más graves).

Implicación principal de las leyes del movimiento capitalista: desaparición del capitalismo. En efecto, aunque los capitalistas aprovechan inicialmente su carácter de clase dominante para apropiarse de la plusvalía de los trabajadores, el proceso de empobrecimiento de éstos acabará desembocando en una crisis que dé lugar a un nuevo "modo de producción": el socialismo, donde los medios de producción están en manos de los trabajadores.

Política económica

En el *Manifiesto comunista* (1848), Marx (junto con Engels) expone sus recomendaciones de política económica:

1. Expropiación de la tierra.
2. Sistema fiscal progresivo.
3. Abolición del derecho de herencia.
4. Confiscación del patrimonio de los emigrados
5. Centralización del crédito en manos del Estado.
6. Multiplicación de empresas públicas.
7. Planificación de la producción.
8. Obligación a trabajar.
9. Educación gratuita y pública para los niños
10. Sistema productivo alejado del individualismo y la iniciativa privada.
11. Estructura productiva resultado de la combinación entre agricultura e industria.

El problema de la transformación

Un problema teórico central en la obra de Marx, conocido como el "problema de la transformación", consiste en derivar los precios de producción a partir de los valores-trabajo.

Si las mercancías se vendieran por su valor, con una tasa de explotación común, la tasa de beneficio $r = e/(1+k)$ sería menor en los sectores con mayor composición orgánica del capital. Como la competencia iguala la tasa de beneficio entre industrias y la composición orgánica difiere entre ellas, los precios tienen que apartarse de los valores, y la cuestión es cómo hacerlo sin romper el vínculo entre trabajo y precio.[57]

El primer ataque técnico sistemático llegó en 1907 de la mano de Bortkiewicz (1907), que formalizó matemáticamente el procedimiento de transformación de Marx y mostró que este solo cuadra —los precios totales igualan a los valores totales y la plusvalía total iguala al beneficio total— si se transforman los precios de los outputs, pero no los de los inputs (materias primas y medios de producción), que Marx seguía valorando en términos de trabajo. A partir de esta crítica se abrieron dos grandes líneas de respuesta, ninguna de ellas hegemónica:

- Sraffa (1960) reconstruyó la teoría de los precios de producción sin recurrir en absoluto a la teoría del valor-trabajo, mostrando que el sistema de precios y la tasa de beneficio general pueden determinarse directamente a partir de las condiciones técnicas de producción.
- La llamada "Nueva Interpretación" de Foley (1982) y Gérard Duménil defendió, en cambio, que redefiniendo la relación entre valor y precio en términos monetarios las dos igualdades de Marx sí pueden preservarse.

Más de un siglo después, el debate sigue sin una resolución generalmente aceptada: el problema de la transformación no invalida por sí solo el diagnóstico histórico o sociológico de Marx sobre el capitalismo, pero sí debilita su pretensión de fundamentarlo en una teoría cuantitativa rigurosa y cerrada del valor.

De las cinco leyes del movimiento capitalista, algunas tendencias parciales sí se han observado históricamente: la limitación de la jornada laboral, el aumento de la composición orgánica del capital, la concentración empresarial en ciertos sectores y la recurrencia de crisis cíclicas. Sin embargo, la implicación central de las cinco —la miseria creciente del proletariado y el colapso final del capitalismo— no se ha materializado: los salarios reales han crecido junto con la productividad, en buena medida porque, como ya se apuntó en la crítica a la primera ley, Marx infravaloró la capacidad del progreso técnico para compensar la caída de la plusvalía relativa.

1.A.3. La escuela austriaca

La escuela austriaca es una escuela de pensamiento económico heterodoxo que está basada en el individualismo metodológico — el concepto de que los fenómenos sociales resultan

[57] La intuición es que una siderúrgica, muy mecanizada, emplea pocos trabajadores por cada euro invertido y genera, por tanto, poca plusvalía por euro. Si vendiera su acero por su valor, obtendría una rentabilidad inferior a la de un taller textil intensivo en mano de obra y nadie invertiría en ella.

 © 2026 Manuel Cabado García

de las motivaciones y acciones de los individuos. La escuela austriaca se ha caracterizado por su fuerte crítica hacia las teorías económicas neoclásicas, marxistas, keynesianas y monetaristas. Si bien existen diferentes puntos de vista sobre política económica que pueden tener los austriacos, la escuela austriaca tiende a autodefinirse como "la ciencia económica del libre mercado".

La escuela austriaca se originó en la Viena de finales del siglo XIX y principios del XX, con el trabajo de Carl Menger, Eugen Böhm von Bawerk, Friedrich von Wieser y otros. Menger no solo formuló el principio marginal de forma independiente a Jevons y Walras: dio lugar a una escuela de pleno derecho que, a diferencia de las otras dos tradiciones marginalistas, acabaría separándose del núcleo neoclásico para constituir una corriente heterodoxa propia.

Eugen von Böhm-Bawerk (1851-1914)

Discípulo de Menger, es el gran teórico del capital y del interés de la escuela (Böhm-Bawerk, 1884). A diferencia de las teorías anteriores sobre los factores productivos —para los fisiócratas, la tierra, y para los clásicos, el trabajo—, Böhm-Bawerk sostiene que tierra y trabajo son factores "originarios", mientras que el capital es un factor *producido* mediante el concurso de aquellos: se utiliza porque permite métodos de producción indirectos, más productivos —pescar con una red fabricada, en vez de con las manos—, aunque a costa de alargar el proceso.

Para medir la intensidad capitalista, acuña el concepto de período medio de producción: el promedio ponderado de todos los intervalos de tiempo durante los que quedan inmovilizadas las unidades de trabajo empleadas —directa o indirectamente— en la obtención de un bien. Cuanto más se alarga este período, mayor es el producto obtenido, aunque a una tasa decreciente.

Respecto al interés, Böhm-Bawerk lo entiende como el precio pagado por disfrutar de bienes presentes en lugar de esperar a los futuros, y distingue tres motivos por los que los bienes presentes son más valiosos:

1. Motivo psicológico (expectativa): los agentes esperan disponer de mayor riqueza en el futuro, y como la utilidad marginal decrece con la renta, prefieren consumir hoy.
2. Motivo psicológico (miopía): los agentes tienden a subestimar sistemáticamente sus necesidades futuras.
3. Motivo tecnológico: los bienes presentes gozan de superioridad técnica sobre los futuros, pues pueden invertirse ya en un proceso productivo indirecto.

Friedrich von Wieser (1851-1926)

El otro gran discípulo de Menger fue el primer autor en utilizar explícitamente el término "utilidad marginal" (Wieser, 1884) y desarrolló la teoría de la imputación: el valor de

los factores de producción no se determina de forma sucesiva partiendo de precios ya dados de los bienes de consumo (como pretendía Menger), sino que, en una economía industrial donde los productos se elaboran con los mismos productos, debe resolverse mediante un sistema de ecuaciones simultáneas —anticipo, de nuevo, del enfoque walrasiano.

El método austríaco y el Methodenstreit

La escuela austriaca comparte tres rasgos que la diferencian del resto del marginalismo: el subjetivismo radical de la teoría del valor, el individualismo metodológico llevado al extremo, y el rechazo del método matemático en favor de la deducción lógica pura (praxeología). Esta postura metodológica desembocó en la célebre polémica del método (*Methodenstreit*) (Menger, 1883) que enfrentó a Menger con Gustav von Schmoller, líder de la Escuela Histórica alemana, en la primera mitad de la década de 1880: mientras Schmoller defendía que la teoría económica debía construirse a partir de la observación empírica y la investigación histórica, Menger sostenía la primacía del razonamiento deductivo. El debate, agrio y prolongado, terminó con la victoria académica de los austríacos, aunque el historicismo influyó decisivamente en el posterior institucionalismo americano de Veblen, Commons y Mitchell.

Legado del siglo XX

La influencia de la escuela austriaca trascendió el período estrictamente neoclásico gracias a dos figuras: Joseph Schumpeter, cuya teoría del desarrollo económico —innovación, empresario innovador, "destrucción creativa" y ciclos económicos superpuestos (Kitchin, Juglar, Kondratiev)— constituye el germen de toda la moderna teoría del ciclo, y Friedrich Hayek, que combinó la teoría del capital de Böhm-Bawerk con la teoría monetaria de Wicksell para construir su propia teoría austríaca del ciclo económico, y protagonizó frente a Oskar Lange el célebre "debate del cálculo socialista" sobre la imposibilidad de una asignación racional de recursos sin precios de mercado.

Desde mediados del siglo XX, los economistas tradicionales han criticado la escuela austriaca moderna y consideran que su metodología, al rechazar elaborar teoría económica desde los modelos matemáticos, la econometría y el análisis macroeconómico, está fuera de la economía convencional, o es "heterodoxa". Aunque ha sido considerada heterodoxa desde fines de la década de 1930, atrajo un renovado interés en la década de 1970 después de que Hayek compartiera el Premio Nobel de Ciencias Económicas de 1974, y de nuevo tras la crisis financiera mundial de 2008.

 © 2026 Manuel Cabado García

1.A.4. La controversia del capital: el debate entre las dos Cambridge

Este fue un importante debate[58] económico de principios de la segunda mitad del siglo XX, protagonizado por Joan Robinson, Piero Sraffa, Pasinetti, Kaldor, postkeynesianos del Cambridge británico, que se enfrentaron a Samuelson y Solow, de la síntesis neoclásica del Cambridge de Estados Unidos.

Este debate comienza con una crítica de Robinson a la función de producción agregada neoclásica.

1. Su crítica es que, dado que el capital está constituido por un gran número de bienes heterogéneos, es imposible especificarlo en términos físicos, por lo que habrá que especificarlo en términos de valor.
2. El problema es que los precios de los bienes de capital dependen de los costes de producción y de la tasa "normal" de beneficio. Por tanto, un mismo stock de capital puede tener diferentes valores en función de cuál sea la tasa de beneficio.
3. Esto implica que para obtener el valor de capital haya que obtener antes la tasa de beneficio, por lo que la tasa de beneficio no puede obtenerse a partir de la función de producción agregada. Se llega así a un sistema de razonamiento según el cual no es posible la existencia de una función de producción agregada sin especificar a priori un sistema de precios y de distribución.

Como escribió la propia Robinson (Robinson, 1953) en el artículo que dio origen al debate:

> *"La función de producción ha sido un poderoso instrumento de mala educación. Al estudiante de teoría económica se le enseña a escribir $Q = f(L, K)$, donde L es una cantidad de trabajo,K una cantidad de capital y Q una tasa de producción de mercancías. Se le indica que suponga que todos los trabajadores son iguales, y que mida L en horas-hombre de trabajo; se le dice algo sobre el*

[58] Idea general de la Controversia del Capital: De acuerdo a Robinson, el capital fijo es heterogéneo y no se puede agregar como el capital financiero. Este último se mide en unidades monetarias y fácilmente se puede añadir.

Adicionalmente Sraffa sugirió que esta es una relación inversa a la causalidad asumida por teoría neoclásica de la producción entre la tasa de ganancia y la cantidad de capital. De acuerdo con esa teoría, específicamente, de acuerdo a la ley de los rendimientos decrecientes, un aumento en la cantidad de capital utilizado debe reducir la tasa de ganancia. Sraffa mostró matemáticamente que, por el contrario, un cambio en la tasa de ganancia provoca un cambio no lineal del valor de las inversiones.

Este análisis sugiere que un uso más intensivo de un factor de producción — no solo el capital— puede llevar a situaciones que incrementarían el precio de ese factor. Según sus proponentes, este análisis levanta serias dudas sobre la visión neoclásica de los precios como indicador de escasez en general y, en particular, sobre la sugerencia del principio de sustitución, que se deriva de la teoría de oferta y demanda, concretamente, en relación al funcionamiento de los bienes de capital.

Sin embargo, los economistas neoclásicos no ven ningún gran problema. Según ellos es suficiente agregar el valor en dinero de todos los bienes con el fin de obtener el monto total de capital. Esto es, obviamente, una simplificación pero una simplificación permisible a niveles introductorios y como mecanismo para estudiar relaciones generales, no una representación fiel de la realidad.

> *problema del número índice al elegir una unidad de producto; y entonces se le apresura a la siguiente cuestión, con la esperanza de que olvide preguntar en qué unidades se mide K. Antes de que llegue a preguntarlo, ya se ha convertido en profesor, y así los malos hábitos de pensamiento se transmiten de una generación a la siguiente."*

El problema se agrava con los llamados "efectos Wicksell": ante un cambio en el tipo de interés, los precios relativos de los bienes de capital varían, de modo que el valor de un mismo stock físico de capital cambia aunque no cambie la técnica. Más grave aún es el fenómeno del "reswitching", o retorno de técnicas, que Sraffa (1960) dejó apuntado: una misma técnica puede ser la más rentable tanto a tipos de interés muy altos como muy bajos, y ser desplazada por otra en el rango intermedio, lo que contradice la relación monótona y decreciente entre tipo de interés e intensidad de capital que postula la teoría neoclásica.[59]

A mediados de los años cincuenta (1956) Solow y Swan intentan resolver este problema de circularidad construyendo modelos en los que la producción se realice con un solo bien de capital. Si bien esto resuelve el problema, es una solución poco realista.

En 1962, Samuelson introducirá el concepto de "función de producción subrogada":

- Su supuesto esencial es que, para cada técnica de producción, la intensidad del capital (la relación K/L) es la misma en el sector de bienes de consumo que en el de bienes de capital.
- Así, conforme aumenta el capital, cae su productividad y disminuye el beneficio.
- Esta idea será criticada por Garegnani, que muestra que, si la relación K/L difiere entre sectores, la relación entre el beneficio y K es ambigua. El propio Samuelson acabó admitiendo en 1966 la validez del reswitching, tras el simposio del *Quarterly Journal of Economics* en el que Pasinetti refutó el teorema de Levhari.

Este debate perderá vigor a partir de ese momento.

Hoy, este problema se considera un fenómeno matemático más que una crítica fundamental. Los economistas modernos entienden que estas no linealidades pueden ocurrir, pero las tratan con herramientas más avanzadas. En lugar de trabajar con modelos agregados simples, utilizan modelos dinámicos y microfundados que pueden manejar la heterogeneidad del capital y la complejidad de las decisiones de producción sin tropezar con estas aparentes paradojas.

Por ejemplo, el "pricing kernel" (también llamado factor de descuento estocástico), un concepto de valoración de activos utilizado en Macroeconomía y finanzas, permite valorar el capital de manera coherente en modelos dinámicos sin caer en las inconsistencias lógicas

[59] La intuición es que, si una técnica vuelve a ser la mejor cuando el tipo de interés sigue bajando, ya no puede decirse que un capital más "barato" lleve siempre a usar técnicas más intensivas en capital, que es la base de la función de producción agregada neoclásica.

 © 2026 Manuel Cabado García

que preocupaban a los economistas de Cambridge. Formalmente, es la tasa marginal de sustitución intertemporal de un agente representativo: mide, en términos de utilidad, cuánto vale hoy un flujo de caja futuro e incierto. De la ecuación de Euler se deriva que el precio de cualquier activo —una máquina, un bono, una acción— es, simplemente, el valor esperado de sus pagos futuros descontados por este factor. Es, por tanto, una herramienta posterior al propio debate: se formaliza con Rubinstein (1976),Lucas (1978) y Breeden (1979), más de una década después del artículo de Robinson que dio origen a la controversia, y se consolida después como el objeto unificador de la valoración de activos con Hansen y Jagannathan (1991). De esta forma:

- Evita la agregación: en lugar de intentar sumar "capital" como una masa homogénea, los modelos modernos que utilizan un pricing kernel se centran en valorar los activos individuales de la economía. Cada máquina, cada edificio, se valora por separado en función de sus flujos de caja esperados y el precio del riesgo. La defensa moderna de la teoría neoclásica ya no pasa, de hecho, por sostener que exista una función de producción agregada bien definida: en un modelo de equilibrio general dinámico, todos los mercados se vacían simultáneamente y los precios relativos se determinan por el principio del margen aplicado a cada activo, sin necesidad de agregar el capital en ningún momento. Esto elude por completo la necesidad de crear una medida agregada incoherente.
- Se basa en principios microfundados: El pricing kernel se deriva directamente de las preferencias y el comportamiento de los agentes económicos (como las funciones de utilidad de los hogares). De esta forma, el valor del capital no se basa en una definición circular que depende de la tasa de ganancia, sino en algo más fundamental: cómo las personas valoran el consumo presente frente al consumo futuro y cómo perciben el riesgo. La tasa de ganancia es el resultado de este proceso de valoración, no la base para él.
- Valora el capital de manera dinámica: Los modelos modernos son dinámicos, lo que significa que tienen en cuenta el tiempo y la incertidumbre. Esto permite valorar el capital de forma coherente a lo largo del tiempo, algo que los modelos estáticos de la teoría neoclásica tradicional no podían hacer de forma robusta y que estaba en el corazón de las críticas de la Vieja Cambridge.

En definitiva, la economía ha pasado de intentar medir el "capital" como una sustancia abstracta a usar una herramienta sofisticada que valora activos específicos. Este cambio de enfoque ha hecho que las preocupaciones lógicas del debate de las Dos Cambridge sean prácticamente irrelevantes para los economistas que trabajan con modelos modernos y dinámicos.

1.A.5. Las cuatro leyes fundamentales de los clásicos

Los clásicos adaptan la idea de orden natural de los fisiócratas subrayando que dicho orden natural viene descrito por cuatro leyes fundamentales, defendidas en mayor o menor medida por todos ellos.

1. Ley de los rendimientos marginales decrecientes de Turgot: La tecnología presenta rendimientos decrecientes a partir de un determinado punto: Dados un factor fijo (tierra) y otro variable, las sucesivas aplicaciones del input variable harán que el producto crezca, primero en proporción creciente y después en proporción decreciente hasta llegar a un máximo (esta es la conocida como "Ley de las Proporciones Variables"). Al desarrollar este principio, Turgot habla de las aplicaciones adicionales de factor variable, con lo que también se anticipa el análisis marginalista.
2. Ley de la población de Malthus: Defiende el continuo crecimiento de la población cuando las condiciones económicas son favorables.
3. Mecanismo de flujo de especie (o ley de los metales preciosos): Ante un superávit comercial entra oro en el país (suponiendo un patrón oro puro en el que las transacciones comerciales se saldan en oro), así que aumentan la masa monetaria y consecuentemente los precios. Esto hará que la competitividad vía precio sea cada vez menor y el superávit se reduzca (ajuste automático del equilibrio exterior a largo plazo.)
4. Ley de Say: Toda oferta creará su propia demanda por lo que siempre vaciará el mercado. Si existe sobreproducción de una determinada mercancía se deberá, según Say, bien a una producción mal dirigida o bien a una disminución en la producción de otras mercancías con las que comprar aquella en la que existe sobreproducción.

1.A.6. Jean Baptiste Say (1767 – 1832)

Su principal aportación es la "Ley de Say", según la cual toda producción genera las rentas necesarias para comprarla, de modo que no puede haber un exceso de oferta generalizado y la economía tiende al pleno empleo de los recursos.[60] Por tanto, Say afirma que no puede existir un exceso de capacidad generalizado. Como mucho podía existir en la provisión de determinados productos.

Razonamiento:

- Antes de que los compradores puedan ser demandantes tienen que haber sido ellos mismos una fuente de oferta, obteniendo una remuneración por su trabajo. Es decir, el valor del producto es distribuido íntegramente entre los factores productivos. A

[60] La fórmula "la oferta crea su propia demanda" no es de Say, que hablaba de que "los productos se pagan con productos": su primera aparición publicada es de Keynes, en 1936, que la utilizó para criticar la ley.

 © 2026 Manuel Cabado García

nivel micro, una fábrica no demanda su propia producción, pero a nivel agregado la economía se encuentra integrada (no hay "otros"). En consecuencia, todo lo comprado debe ser igual a todo lo producido.

- ¿Cómo se explican las situaciones de penuria y de demanda insuficiente? La explicación se encuentra en un desajuste en los mercados, adelantándose a la Ley de Walras, este autor señalaba que ante una mala asignación de la producción, pueden darse casos en donde en ciertos mercados exista un exceso de oferta, pero solo a condición de que en otros exista un exceso de demanda.
- De acuerdo con Say, las personas solo trabajarán hasta conseguir acumular un valor igual al del consumo que desean realizar. Incluso si deseasen ahorrar, éste se canalizaría en inversión, que se convertiría en compras de capital.

Pero, ¿funciona la Ley de Say cuando se tiene en cuenta el dinero? No: su uso como depósito de valor rompe la cadena de argumentos de Say. Una escasez de dinero podría generar un exceso de oferta de todo lo demás. Estas implicaciones más complejas de los efectos de la oferta monetaria sobre la economía serán consideradas más adelante, cuando se pase a adoptar el dinero fiat.

Otras críticas que se pueden hacer a la Ley de Say son la no consideración de círculos perversos y fallos de coordinación (los trabajadores no tienen trabajo porque las empresas no tienen clientes, y viceversa).

En suma, los mecanismos que garantizan que la demanda retire toda la oferta son los precios y tipos de interés totalmente flexibles, más el supuesto de que los individuos no atesoran dinero.

Teoría subjetiva del valor Esta fue otra de sus grandes aportaciones. Rechaza las ideas de Smith y Ricardo y postula la "teoría subjetiva del valor", según la cual el valor de un bien depende de: i) La utilidad que reporta, ii) Su escasez. Su teoría tendrá mucha influencia sobre los marginalistas.

Teoría de la distribución Rechaza también a Ricardo y Smith, señalando que cada factor productivo es remunerado proporcionalmente de acuerdo a su participación en el proceso productivo. Por tanto, es el precio del producto el que determina la remuneración de los factores, y no al revés.

1.A.7. Precedentes hispánicos: la Escuela de Salamanca

Aunque el consenso académico sitúa el origen de la ciencia económica en 1776, conviene no ignorar que, dos siglos antes, un grupo de teólogos y juristas españoles del siglo XVI —la Escuela de Salamanca, con figuras como Luis de Molina, Domingo de Soto y Martín de

Azpilcueta— anticiparon ya algunas de las ideas centrales que los clásicos "redescubrirían" más tarde.

Molina y Azpilcueta formularon una temprana teoría subjetiva del valor, sosteniendo que el valor de los bienes no es intrínseco, sino que depende de su utilidad para los individuos y de las circunstancias del mercado —un precursor directo de la revolución marginalista, casi tres siglos anterior—. Azpilcueta observó además que la llegada masiva de oro y plata americanos a España provocaba subidas de precios, sentando las bases de la teoría cuantitativa del dinero antes que Jean Bodin (1568) o el propio David Hume. El escaso reconocimiento internacional de esta escuela se explica, sobre todo, porque sus ideas quedaron encuadradas dentro de tratados de teología moral y derecho —no como una disciplina autónoma reconocible como "economía"— y no tuvieron continuidad directa con la ciencia económica que se consolidó en Gran Bretaña a partir del siglo XVIII. A esto se suma, según algunos autores, el efecto de la llamada leyenda negra británica sobre la producción intelectual española. Su contribución quedó así relegada a un segundo plano hasta que Schumpeter y los autores austríacos la reivindicaron en el siglo XX.

1.A.8. Knut Wicksell y el proceso acumulativo

Knut Wicksell (Estocolmo, 1851 – Estocolmo, 1926) inició su formación en matemáticas y no se orientó hacia la economía hasta 1887, cuando viajó a Inglaterra, Francia, Alemania y Austria para estudiarla. Su acceso a la cátedra de economía política y hacienda pública de la Universidad de Lund se retrasó por su heterodoxia social y religiosa: en 1900-1901, el obispo de Lund se opuso a su nombramiento pese a que el tribunal lo situaba primero, alegando que sus ideas —neomalthusianismo, defensa del amor libre— lo hacían inadecuado para el cargo. Finalmente fue nombrado profesor asociado en 1901 y catedrático pleno pocos años después. Años más tarde, en 1908, protagonizó un episodio distinto: pronunció un discurso satirizando el dogma de la Inmaculada Concepción como acto de desobediencia civil en apoyo de un preso condenado por blasfemia, lo que le costó una condena que cumplió en 1909, con 57 años —tiempo que aprovechó para traducir *La riqueza de las naciones* al sueco. Sus obras principales son *Sobre el valor, el capital y la renta* (Wicksell, 1893), *Interés y precios* (Wicksell, 1898) y las *Conferencias sobre economía política* (Wicksell, 1901), (Wicksell, 1906).

Wicksell parte del pensamiento clásico de Ricardo y Say, corregido por sus maestros Böhm-Bawerk y Walras, pero termina rompiendo con el enfoque neoclásico de su tiempo en el terreno monetario. Consideró ilógico superponer, sin conectarlas, una explicación subjetiva de los precios particulares (la utilidad marginal) y una explicación objetiva de los movimientos generales de los precios (la teoría cuantitativa del dinero). Además, juzgó incompatible la Ley de Say con la propia teoría cuantitativa, ya que si la oferta global iguala siempre a la demanda global, no podría producirse ningún movimiento de precios

 © 2026 Manuel Cabado García

que explicar. Por ello, Wicksell rechazó la Ley de Say en una economía monetaria: al mediar el dinero como depósito de valor, la política bancaria y crediticia puede generar desajustes entre oferta y demanda agregadas sin que varíe la cantidad física de bienes.

Para explicar estos movimientos de precios, Wicksell distingue entre el tipo de interés natural y el monetario (o efectivo), distinción que apoya en la teoría del capital de Böhm-Bawerk y que convierte en el centro de su teoría[61]:

- Tipo de interés natural: el rendimiento esperado por los empresarios de sus inversiones, determinado por la productividad marginal del capital.
- Tipo de interés monetario (o efectivo): el que se fija en los mercados de crédito, con cierta autonomía respecto al natural por la acción de la política bancaria.

Cuando ambos tipos coinciden, el ahorro y la inversión se igualan y el nivel de precios permanece estable —lo que Wicksell llama "equilibrio monetario" (Wicksell, 1898). Si el tipo monetario cae por debajo del natural, la inversión supera al ahorro, suben los precios y mejoran las expectativas de beneficio, lo que a su vez alimenta nuevas inversiones: se pone en marcha un "proceso acumulativo" al alza que no se corrige espontáneamente, sino solo mediante una política crediticia restrictiva. De forma simétrica, un tipo monetario por encima del natural desencadena un proceso acumulativo a la baja. Esta idea de que los desequilibrios entre ahorro e inversión pueden prolongarse y agravarse espontáneamente —sin el mecanismo autocorrector que suponía la Ley de Say— convierte a Wicksell en un precursor directo de la macroeconomía y de la revolución keynesiana, y sienta las bases de la Escuela Sueca (Ohlin, Myrdal) que desarrollará después estas ideas.

Wicksell hizo además una última aportación, de consecuencias muy posteriores: mostró que el valor del stock de capital de una economía cambia al variar el tipo de interés, a través de dos canales distintos —un efecto precio (la revalorización de los bienes de capital existentes) y un efecto real (el cambio en las cantidades físicas de capital elegidas con la nueva técnica de producción)—, de forma que la relación entre el tipo de interés y la cantidad de capital no tiene por qué ser monótona cuando existe más de un bien de capital. Este "efecto Wicksell" es precisamente el mecanismo técnico que Robinson y Sraffa explotarán medio siglo después en la controversia del capital entre las dos Cambridge (1.A.4), al mostrar que el valor del capital no puede definirse de forma independiente de la tasa de beneficio que se pretende explicar con él.

[61] La idea tiene un antecedente en Henry Thornton, que en 1802 ya advirtió que un tipo del Banco de Inglaterra inferior a la rentabilidad del capital estimulaba el crédito y los precios.

Bibliografía

Arrow, K. J. y G. Debreu (1954). Existence of an Equilibrium for a Competitive Economy. *Econometrica 22*(3), 265–290.

Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy 82*(6), 1095–1117.

Bartels, C., F. Kersting y N. Wolf (2023). Testing Marx: Capital Accumulation, Income Inequality, and Socialism in Late Nineteenth-Century Germany. *The Review of Economics and Statistics 107*(4), 935.

Böhm-Bawerk, E. v. (1884). *Kapital und Kapitalzins*.

Bortkiewicz, L. v. (1907). Wertrechnung und Preisrechnung im Marxschen System. *Archiv für Sozialwissenschaft und Sozialpolitik 25*, 10–51.

Breeden, D. T. (1979). An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities. *Journal of Financial Economics 7*(3), 265–296.

Carlyle, T. (1849). Occasional Discourse on the Negro Question. En *Fraser's Magazine for Town and Country*. London.

Clark, J. B. (1899). *The Distribution of Wealth: A Theory of Wages, Interest and Profits*. Macmillan.

Cournot, A. A. (1838). *Recherches sur les principes mathématiques de la théorie des richesses*. Paris: Hachette.

Dupuit, J. (1844). De la mesure de l'utilité des travaux publics. *Annales des Ponts et Chaussées 8*, 332–375.

Ehrlich, P. R. (1968). *The Population Bomb*.

Fisher, I. (1907). *The Rate of Interest: Its Nature, Determination and Relation to Economic Phenomena*. New York: Macmillan.

Foley, D. K. (1982). The Value of Money, the Value of Labor Power, and the Marxian Transformation Problem. *Review of Radical Political Economics 14*(2), 37–47.

Gossen, H. H. (1854). *Entwicklung der Gesetze des menschlichen Verkehrs und der daraus fließenden Regeln für menschliches Handeln*. Braunschweig: Friedrich Vieweg.

Hansen, L. P. y R. Jagannathan (1991). Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy 99*(2), 225–262.

 © 2026 Manuel Cabado García

Hume, D. (1752). Of the Balance of Trade. En *Political Discourses*.

Jevons, W. S. (1871). *The Theory of Political Economy*. London: Macmillan.

Juhász, R. (2018). Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade. *American Economic Review 108*(11), 3339–3376.

Keynes, J. M. (1924). Alfred Marshall, 1842-1924. *The Economic Journal 34*(135), 311–372.

Lloyd, W. F. (1834). *A Lecture on the Notion of Value as Distinguishable Not Only from Utility, but also from Value in Exchange*.

Lucas, R. E. (1978). Asset Prices in an Exchange Economy. *Econometrica 46*(6), 1429–1445.

Malthus, T. R. (1798). *An Essay on the Principle of Population*.

Marshall, A. (1890). *Principles of Economics*. London: MacMillan and Co.

Martín Martín, V. (2010). William Stanley Jevons, economista y pionero. *Cuadernos de Información Económica* (219), 159–169.

Marx, K. (1852). *Der 18te Brumaire des Louis Napoleon*.

Marx, K. (1867). *Das Kapital*.

Meadows, D. H., D. L. Meadows, J. Randers y W. W. Behrens (1972). The Limits to Growth. Informe técnico, Club of Rome.

Menger, C. (1871). *Grundsätze der Volkswirthschaftslehre*. Wien: Wilhelm Braumüller.

Menger, C. (1883). *Untersuchungen über die Methode der Socialwissenschaften und der Politischen Oekonomie insbesondere*.

Mill, J. S. (1836). On the Definition of Political Economy; and on the Method of Investigation Proper to It.

Mill, J. S. (1848). *Principles of Political Economy*.

Mises, L. v. (1949). *Human Action: A Treatise on Economics*. New Haven: Yale University Press.

Ricardo, D. (1817). *On the Principles of Political Economy and Taxation*.

Robinson, J. (1953). The Production Function and the Theory of Capital. *The Review of Economic Studies 21*(2), 81–106.

Rubinstein, M. (1976). The Valuation of Uncertain Income Streams and the Pricing of Options. *Bell Journal of Economics and Management Science 7*, 407–425.

Schumpeter, J. A. (1911). *Theorie der wirtschaftlichen Entwicklung*. Leipzig: Duncker & Humblot.

Smith, A. (1759). *The Theory of Moral Sentiments*. London: A. Millar.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*.

Sraffa, P. (1960). *Production of Commodities by Means of Commodities*.

Turgot, A. R. J. (1767). Observations sur le mémoire de M. de Saint-Péravy en faveur de l'impôt indirect.

Veblen, T. (1900). Preconceptions of Economic Science. *Quarterly Journal of Economics*.

von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Hamburg: Perthes.

Wald, A. (1933). Über die eindeutige positive Lösbarkeit der neuen Produktionsgleichungen. *Ergebnisse eines mathematischen Kolloquiums*.

Walras, L. (1874). *Éléments d'Économie Politique Pure: Ou Théorie de la Richesse Sociale*.

Wicksell, K. (1893). *Über Wert, Kapital und Rente nach den neueren nationalökonomischen Theorien*. Jena: Gustav Fischer.

Wicksell, K. (1898). *Geldzins und Güterpreise*. Jena: Gustav Fischer.

Wicksell, K. (1900). Om gränsproduktiviteten såsom grundval för den nationalekonomiska fördelningen. *Ekonomisk Tidskrift 2*, 305–337.

Wicksell, K. (1901). *Föreläsningar i nationalekonomi, Häft I*. Stockholm/Lund: Fritzes/Berlingska.

Wicksell, K. (1906). *Föreläsningar i nationalekonomi, Häft II: Om penningar och kredit*. Stockholm/Lund: Fritzes/Berlingska.

Wicksteed, P. H. (1894). *An Essay on the Co-ordination of the Laws of Distribution*.

Wieser, F. v. (1884). *Über den Ursprung und die Hauptgesetze des wirtschaftlichen Werthes*.

 © 2026 Manuel Cabado García

Tema 2

Pensamiento económico (II): de Keynes a la actualidad. Referencia a la Síntesis Neoclásica, al Monetarismo, a la Nueva Macroeconomía Clásica y a los modelos Neokeynesianos de primera y segunda generación.

2.1. Introducción ⏳ 1'

"A largo plazo todos estaremos muertos" y "Las personas que se creen libres de influencia intelectual suelen ser esclavas de un economista muerto"[1]. Con estas dos frases, el economista John Maynard Keynes no solo desafió la Economía Clásica incapaz de responder a la Gran Depresión, sino que fundó la Macroeconomía moderna con su *Teoría General* (1936), defendiendo la intervención estatal con fines estabilizadores.

No obstante, la Economía no es una disciplina estática: cada escuela de pensamiento nace como crítica a la anterior, enfrentando los retos de su tiempo. Así, la Síntesis Neoclásica refinó el legado keynesiano, hasta que el Monetarismo lo cuestionó por su manejo

[1]Las frases son tan icónicas porque capturan perfectamente el espíritu de la revolución que Keynes estaba proponiendo:

- "A largo plazo todos estaremos muertos" desafía a los economistas clásicos que argumentaban que la economía se corregiría sola con el tiempo, ignorando el sufrimiento de la gente en el presente.
- La frase sobre el "economista muerto" critica a quienes se aferraban ciegamente a ideas obsoletas sin darse cuenta.

de la inflación.

Más adelante, la Nueva Economía Clásica introdujo las expectativas racionales, y la Nueva Economía Keynesiana respondió incorporando imperfecciones del mercado. Este vaivén teórico revela cómo cada corriente genera nuevas herramientas y políticas para superar las limitaciones de las anteriores y, de esta forma, adaptarse a un mundo en constante cambio.

Por consiguiente, el objetivo del tema es el análisis de las doctrinas económicas a través del siguiente esquema:

1. Keynes.
2. Síntesis Neoclásica.
3. Monetaristas.
4. Nueva Macroeconomía Clásica.
5. Nueva Economía Keynesiana.
6. Otros desarrollos.

2.2. John Maynard Keynes ⏳ 4'30"

La publicación de la *Teoría General de la Ocupación, el Interés y el Dinero* en 1936 supone un cambio de paradigma respecto al pensamiento ortodoxo clásico. Se inscribe en una lenta recuperación de la Gran Depresión de los años 30 en la que las fuerzas de mercado no son capaces de devolver a la economía a su nivel de pleno empleo.

Críticas al planteamiento neoclásico Keynes plantea una crítica al planteamiento neoclásico[2], ya que no son capaces de explicar las causas de la crisis ni encontrar soluciones:

1. Enfoque temporal(corto vs largo plazo): Keynes se distancia de los clásicos en su enfoque temporal. Mientras los neoclásicos se centran en el largo plazo, confiando en que los mercados se ajustan con el tiempo, Keynes pone el foco en el corto plazo, donde los desequilibrios como el desempleo requieren intervención inmediata. Su célebre frase "en el largo plazo todos estaremos muertos" resume su rechazo a esperar que el tiempo resuelva los problemas económicos.
2. Mecanismos de ajuste (salarios nominales rígidos a la baja): respecto a la efectividad de los mecanismos de ajuste, los clásicos creen que los precios y salarios son flexibles, lo que permite que los mercados se autorregulen tras un shock. Keynes, en cambio,

[2]John Maynard Keynes denominó "modelo clásico", en la nota del Capítulo 1 de su *Teoría General* (1936), al modelo macroeconómico implícito en los trabajos de los seguidores de Ricardo, entre los que incluía explícitamente a economistas marginalistas como Marshall, Edgeworth y, sobre todo, Pigou. No obstante, es más adecuada la denominación de modelo neoclásico, ya que estos autores pertenecen realmente a la escuela marginalista posterior, y no al núcleo clásico original de Smith, Ricardo o Mill.

 © 2026 Manuel Cabado García

parte de salarios nominales rígidos a la baja y, sobre todo, sostiene que aunque fuesen flexibles su reducción no garantizaría el pleno empleo, porque deprime la demanda efectiva. Por eso el sistema no se corrige por sí solo y aparece el desempleo involuntario, que no puede explicarse desde la teoría clásica.[3]

3. Ajuste por cantidades: en cuanto al mecanismo de ajuste, los neoclásicos sostienen que los precios son el canal principal para restaurar el equilibrio. Keynes discrepa: para él, los ajustes se producen a través de las cantidades, como el nivel de producción y empleo. Esto implica que la economía puede estabilizarse en un punto de baja actividad sin alcanzar el pleno empleo, algo que los clásicos no contemplan.
4. Ilusión monetaria: la visión del comportamiento de los agentes también marca una diferencia clave. Los clásicos suponen que los agentes solo atienden a magnitudes reales. Keynes, en cambio, admite que los trabajadores sufren "ilusión monetaria", es decir, que reaccionan a variables nominales, como el salario nominal, sin descontar la evolución de los precios, lo que distorsiona sus decisiones económicas.[4]
5. Rechazo de la ley de Say: finalmente, Keynes rechaza la Ley de Say, que sostiene que toda oferta genera su propia demanda. Los clásicos la aceptan, confiando en que el tipo de interés y la no acumulación de dinero garantizan el equilibrio. Keynes invierte esta lógica: es la demanda efectiva la que determina la producción. Si los agentes deciden atesorar dinero o si la inversión cae, puede haber insuficiencia de demanda y desempleo persistente e involuntario.
6. Demanda de Dinero por Especulación: Keynes desafió la Teoría Cuantitativa del Dinero al introducir un nuevo motivo para la demanda de dinero: la especulación. La gente atesora dinero no solo para transacciones, sino también como una forma de riqueza, lo que rompe con la dicotomía clásica[5] y con el supuesto clásico de la neutralidad del dinero, explicando por qué la política monetaria puede tener efectos reales en la economía.

[3] La rigidez de los precios, además de la de los salarios, es un supuesto que introducirá después la Síntesis Neoclásica al formalizar el modelo keynesiano en el esquema IS-LM.

[4] La intuición es que un trabajador rechaza una rebaja del 2 % en su nómina, pero acepta sin protestar una subida del 1 % cuando los precios crecen un 3 %, aunque en ambos casos pierde poder adquisitivo. El término "racionalidad limitada", con el que a veces se asocia esta idea, procede de Herbert Simon y no forma parte del vocabulario de Keynes.

[5] La dicotomía clásica es la idea central de la Macroeconomía clásica que establece una separación total entre las variables reales y nominales.

Las variables reales (producción, empleo, PIB real) se analizan de forma independiente de las variables nominales (oferta de dinero, nivel de precios). Según este concepto, los cambios en la cantidad de dinero solo afectan a los precios, dejando las variables reales sin cambios. A esto se le llama neutralidad del dinero.

Para los clásicos, el dinero es un simple "velo" que facilita el comercio, pero no influye en la capacidad productiva de la economía. El PIB real y el empleo están determinados únicamente por factores reales como la tecnología, la mano de obra y el capital. La dicotomía clásica fue uno de los principales blancos de la crítica keynesiana, que demostró cómo los cambios monetarios sí tienen un impacto directo y significativo en la producción y el empleo.

A continuación se analizan sus implicaciones en el mercado de trabajo, bienes y de dinero para, finalmente, introducir sus recomendaciones de política económica.

Mercado de trabajo

Keynes interpreta el equilibrio neoclásico del mercado de trabajo como un "caso especial que no es típico en la sociedad en la que vivimos". Es un caso especial que corresponde a una situación en la que la demanda es suficientemente elevada como para absorber toda la producción.

Así, la insuficiencia de demanda durante el periodo de la Gran Depresión llevaba a un nivel de desempleo involuntario masivo que no era el nivel de empleo compatible con el pleno empleo neoclásico.

Demanda de trabajo Keynes mantiene los supuestos macroeconómicos de la función de producción y acepta la demanda de trabajo tal y como la postulan los clásicos. Se demanda trabajo hasta el punto en el que el salario nominal se iguala al valor de la productividad marginal del trabajo, es decir, que depende negativamente del salario real:

$$L_D = f\left(\frac{W}{P}\right) \tag{2.1}$$

Oferta de trabajo

- La demanda agregada determina la producción y, a través de la función de producción se obtiene el nivel de empleo
- El salario nominal se determina vía negociación.
- El trabajador de Keynes se guía por el salario nominal, que es sobre el que puede influir la negociación colectiva, por lo que tiene ilusión monetaria.
- De este modo, si caen los precios, caerá la cantidad vendida y por tanto la demanda de trabajo. Con ilusión monetaria no se aceptará una reducción del salario nominal, ya que Keynes argumenta que existen salarios rígidos a la baja.
- Esto dará lugar a desempleo involuntario, ya que el mercado de trabajo no podrá vaciarse.
- Por lo tanto, la oferta de trabajo se mantendrá constante en el salario mínimo nominal mientras no se supere un nivel de empleo L_m, y a partir del mismo crecerá con el nivel de empleo:

$$L_S = W_m + f(L), \quad f' > 0 \text{ si } L > L_m,\ f = 0 \text{ si } L \leq L_m \tag{2.2}$$

Solución Según Keynes, en una situación con recursos ociosos (que, según él, será lo más frecuente), una política económica expansiva que estimule la demanda agregada podrá

 © 2026 Manuel Cabado García

aumentar los niveles de empleo y de producción. La ilusión monetaria de los trabajadores permitiría a las empresas aumentar los precios en mayor proporción que los salarios nominales (por lo que los salarios reales disminuirían), y esto permitiría aumentar el empleo y la producción.

Frente al remedio neoclásico de bajar los salarios nominales para reducir el paro, Keynes plantea que el efecto sobre el empleo es incierto, porque depende de sus repercusiones sobre los determinantes de la Demanda Efectiva. Por un lado, una caída de salarios reduce costes y precios, abarata el tipo de interés y estimula la inversión. Por otro, redistribuye renta de los trabajadores –con alta propensión a consumir– hacia los rentistas, encarece el valor real de las deudas y puede generar expectativas de deflación futura. En una economía deprimida, Keynes concluye que estos efectos negativos tienden a dominar, por lo que rechaza la flexibilidad salarial como vía de salida de la crisis.

Este razonamiento constituye el núcleo del Principio de la Demanda Efectiva: el nivel de empleo y de producción no lo fija el mercado de trabajo por sí solo, sino el punto en el que la Demanda Agregada esperada por las empresas iguala a los ingresos mínimos que estas exigen para contratar ese volumen de empleo. Por eso el desempleo resultante es involuntario: hay trabajadores dispuestos a trabajar al salario real vigente, o incluso a uno algo menor, que no encuentran empleo, y no pueden hacer nada para remediarlo porque la insuficiencia procede de la Demanda Agregada, que escapa a su control.[6]

Mercado de bienes

Mientras que en el modelo neoclásico la variable que se determinaba en el sector real era el tipo de interés[7], en el enfoque keynesiano las que se determinan son la renta y el nivel general de precios.

Demanda agregada

La demanda agregada depende de tres componentes principales

$$DA = C + I + G \tag{2.3}$$

que se analizarán a continuación:

1. Consumo[8], C, considerado por Keynes el fin último de la actividad económica, donde

[6] La intuición es que el parado no puede "ofrecerse más barato" para conseguir empleo: si las ventas de las empresas no se recuperan, rebajar su salario no crea el puesto que falta.

[7] De acuerdo al enfoque neoclásico, en el sector real se determina el tipo de interés real. La renta no se determina porque ésta ya viene dada (pues se considera exógena al nivel de pleno empleo).

[8] Keynes es el primer economista que elabora una función de demanda de consumo agregada, y lo hace en su *Teoría General* de 1936, donde considera el consumo como el fin último de la actividad económica, en contraposición a los neoclásicos, que consideraban el consumo como residual una vez que los tipos de interés habían determinado el ahorro. La demanda de consumo no procede del comportamiento optimizador del agente individual, siendo desarrollado en base a la introspección y en la observación de su entorno ("ley psicológica").

el principal determinante del consumo y el ahorro es la renta absoluta disponible del período, con un componente autónomo (consumo mínimo o de subsistencia) y otro en función de la renta disponible:

$$C = C_0 + c \cdot Y^d \tag{2.4}$$

muy poco influido por los tipos de interés, ya que depende de la renta actual, y donde la propensión media al consumo, PMe_C, es decreciente con la renta (ahorro como bien de lujo), y la propensión marginal a consumir, PMg_C, es decir, la parte de cada euro adicional de renta disponible que se destina a consumo (el parámetro c), se sitúa entre 0 y 1. Así, al existir consumo autónomo, la propensión media será superior a la marginal, $PMe_C > PMg_C$, y el ahorro será un residuo del consumo, al contrario de lo que afirmaban los clásicos (para ellos el consumo era residuo del ahorro).

2. Inversión, I, para Keynes la inversión es una variable fundamental, determinada por la Eficiencia Marginal del Capital (EMK), que recoge los rendimientos que esperan obtener los empresarios, y, sobre todo, por su estado de ánimo, que Keynes llamó "animal spirits", es decir, el impulso espontáneo a actuar en lugar de esperar en un contexto de incertidumbre en el que no es posible calcular probabilidades.[9] Estos son un componente autónomo y volátil que da a la inversión un carácter impredecible, algo a lo que Keynes concedía mucha más importancia que al tipo de interés. Así, la inversión efectiva y la deseada no siempre coinciden, generando desequilibrios. Por lo que la inversión dependerá de un componente autónomo y de un componente endógeno:

$$I = I_0 - b \cdot r \tag{2.5}$$

3. Gasto Público, G, se considera una variable exógena, al igual que los clásicos[10].

$$G = G_0 \tag{2.6}$$

Teoría del multiplicador Agrupando los términos y suponiendo impuestos proporcionales, se obtiene el multiplicador keynesiano del gasto público, α_G, que mide cuánto aumenta la renta por cada unidad de aumento exógeno de un componente de la demanda

[9] La intuición es que un empresario no decide abrir una nueva tienda tras calcular con precisión su rentabilidad, que es imposible de conocer, sino porque "ve" que el momento es bueno, y ese optimismo puede evaporarse de un día para otro. Keynes los describe como "un impulso espontáneo a la acción en lugar de la inacción" (*Teoría General*, cap. 12).

[10] Keynes consideraba que los burócratas/políticos/funcionarios tenían mayor conocimiento, un horizonte temporal más largo o eran incorruptibles. Aunque, Keynes le daba una gran importancia para dinamizar la economía. Mejor incrementar el gasto que reducir impuestos con el fin de estimular la demanda efectiva para acabar con el paro involuntario. Una vez superada la crisis, se debería volver al equilibrio presupuestario.

 © 2026 Manuel Cabado García

agregada. Su valor es mayor que 1, dado que la propensión marginal a consumir, c, y el tipo impositivo, τ, están entre 0 y 1:[11]

$$Y = \underbrace{\frac{1}{1 - c\,(1 - \tau)}}_{\alpha_G} \cdot (A_0 - b \cdot r) \tag{2.7}$$

así, el efecto multiplicador se deriva de la idea de que el gasto de una persona se convierte en el ingreso de otra, lo que genera una reacción en cadena[12].

De este marco se deriva una implicación contraintuitiva, la paradoja del ahorro: si todos los hogares intentan ahorrar más a la vez, el ahorro agregado no aumenta, porque la caída del consumo reduce la demanda agregada y la renta hasta que el ahorro vuelve a igualar a la inversión. Si la inversión depende a su vez de la renta, el ahorro total puede incluso disminuir.[13] Domina así el lado de la demanda: es la renta –y no el tipo de interés, como en el modelo clásico– la verdadera variable de ajuste keynesiana.

Mercado de dinero

En el análisis de Keynes el dinero, además de ser medio de cambio, es un depósito de valor. Los individuos atesoran saldos monetarios como forma de ahorrar. La demanda de dinero estará influida por tanto por el tipo de interés, que también actúa sobre el sector real. En consecuencia, se rompe la dicotomía clásica y la neutralidad del dinero.

Keynes argumentaba que los individuos demandan dinero por tres motivos principales.

1. Motivo de transacción, donde se demanda dinero para realizar pagos cotidianos, y su demanda está en función de la renta.
2. Motivo de precaución, por el que se guarda dinero para hacer frente a imprevistos, también en función de la renta.
3. Motivo especulación, siendo este motivo la gran innovación de Keynes. Así, se relaciona con la función del dinero como depósito de valor. La demanda de dinero por este motivo está inversamente relacionada con el tipo de interés: si los tipos son bajos, los individuos anticipan que subirán, lo que devaluará el precio de los bonos. Por tanto, optan por atesorar dinero en efectivo para evitar pérdidas de capital. Esta demanda es muy inestable, ya que depende de las expectativas, muy volátiles, de los tenedores de riqueza sobre la evolución futura del tipo de interés.

[11] La intuición es que los 100 euros que el Estado paga a un constructor se convierten en su renta, que gasta en parte en el supermercado, cuyo dueño a su vez gasta una parte, y así sucesivamente.

[12] El economista Richard Kahn (1931) fue el primero en desarrollar esta idea, centrándose en el empleo. Más tarde, John Maynard Keynes generalizó el concepto para aplicarlo al ingreso y lo convirtió en una herramienta clave de la Macroeconomía.

[13] La intuición es que el ahorro de una familia que deja de cenar fuera es la venta que pierde el restaurante: lo que es prudente para un hogar se vuelve contraproducente cuando lo hacen todos a la vez.

En resumen, la demanda de dinero keynesiana dependerá positivamente de la renta y negativamente del tipo de interés:

$$M_D = f(Y, r) \tag{2.8}$$

Esta visión rompió la dicotomía clásica y le dio a la política monetaria un papel activo en la gestión de la economía.

Rechazo de la ley de Say Si el dinero tiene una función de depósito de valor, atesorar dinero no es irracional, existiendo la posibilidad de que los agentes no gasten íntegramente su renta en consumo e inversión llevando a que la demanda sea menor que la producción lanzada al mercado y no cumpliéndose así la Ley de Say[14].

Implicaciones de política económica

Las políticas de demanda sí son efectivas para alterar la renta:

- Política fiscal: un aumento del gasto público estimula directamente la demanda agregada. Además, va a ser muy efectiva por el efecto de los multiplicadores.
- Política monetaria: una política monetaria expansiva que reduzca el tipo de interés estimula indirectamente la demanda agregada al aumentar la inversión.
    - De hecho, esta dinámica de la política monetaria es lo que da nombre a su obra, *The General Theory of Employment, Interest, and Money*: la oferta de dinero afecta al tipo de interés, lo cual afecta a la inversión y ésta, al empleo. El término "General" se debe a que Keynes entiende que lo anterior solo funciona si no hay pleno empleo (y, para él, esta situación de recursos ociosos será la situación general, mientras que el pleno empleo es, dirá, sólo un caso particular).
    - No obstante, Keynes señala que la política monetaria va a ser menos efectiva que la fiscal, ya que se da lo que él denomina los "dos eslabones débiles":
        - Demanda de inversión muy poco sensible al tipo de interés, ya que las expectativas son muy volátiles (componente autónomo, animal spirits).

[14] Importante, la identidad contable, $S = I$, se mantiene en ambas escuelas:

- Para los clásicos, la igualdad entre ahorro e inversión es un equilibrio de mercado que se alcanza gracias al ajuste del tipo de interés. Es decir, el tipo de interés flexible garantiza que todo lo que se ahorra se invierte.
- Para Keynes, en cambio, esta igualdad es una identidad contable, pero no un equilibrio automático. Él sostiene que el tipo de interés no equilibra el mercado debido a la introducción del motivo especulación, que hace que las personas puedan preferir atesorar dinero líquido en lugar de invertirlo. En su modelo, el ajuste se produce a través de la producción y el ingreso. Si el ahorro no se invierte, la demanda cae, lo que reduce la producción y, por ende, el ingreso. Esta reducción del ingreso continúa hasta que el ahorro se iguala a la inversión.

 © 2026 Manuel Cabado García

- Demanda de dinero muy sensible al tipo de interés (motivo especulación), de modo que un aumento de la oferta monetaria apenas reduce el tipo de interés[15].

No obstante, dado el carácter rupturista del modelo keynesiano se va a ver sujeto a una continua revisión para superar sus limitaciones y tratar de aunarlo con el pensamiento clásico, surgiendo así la Síntesis Neoclásica.

2.3. Síntesis Neoclásica ⏳ 3'

La SNC se refiere a la doctrina económica que aparece en EE.UU a principios de los años 40 y que dominaría el pensamiento económico hasta los años 70, cuando tomó el relevo el monetarismo. Fue Paul Samuelson con su libro *Economía*, 1955, quien dio a esta doctrina el sentido de síntesis, ya que se puede considerar:

- En el corto plazo existirían rigideces de precios y salarios de modo que el modelo keynesiano resulta adecuado para analizar las fluctuaciones a corto plazo.
- En el largo plazo los precios y salarios serían flexibles y el modelo clásico y neoclásico permitiría estudiar las cuestiones económicas de largo plazo.

A continuación, se analizan los principales aportes de la SNC.

Modelo IS-LM

La Síntesis Neoclásica utiliza como elemento central el modelo IS-LM[16], formulado por Hicks (1937) en su artículo "Mr. Keynes and the Classics" como respuesta a la *Teoría General*, y ampliado después por Modigliani (1944) y por Hansen (1953), quien rebautizó el modelo IS-LL original de Hicks como IS-LM, denominación que ha perdurado hasta hoy.

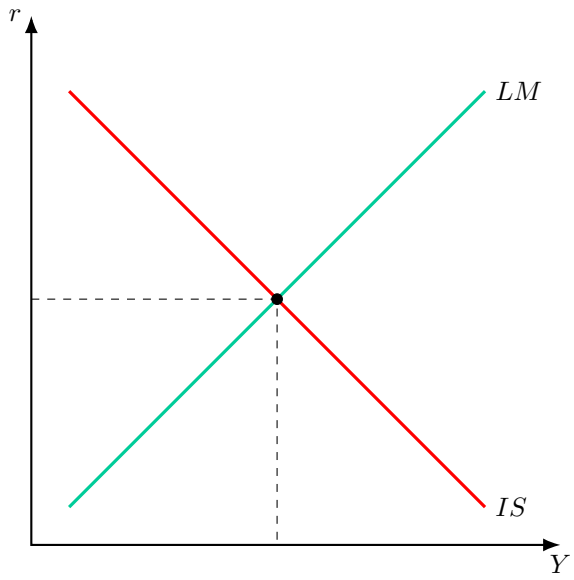
El modelo IS-LM analiza el equilibrio en una economía cerrada con precios rígidos a partir de dos curvas como se observa en la Figura 2.1:

- Curva IS: que representa las combinaciones de tipo de interés y renta tales que el mercado de bienes está en equilibrio.
- Curva LM: que representa las combinaciones de tipo de interés y renta tales que el mercado de dinero está en equilibrio.

[15] En el caso extremo, la trampa de liquidez, el tipo de interés es tan bajo que todos esperan que suba, la demanda de dinero se vuelve infinitamente sensible al tipo de interés y cualquier aumento de la oferta monetaria se atesora sin reducirlo, por lo que la política monetaria pierde su eficacia.

[16] El nombre del modelo procede de las iniciales inglesas de sus dos curvas: IS ("Investment-Saving", inversión-ahorro), que recoge el equilibrio del mercado de bienes, y LM ("Liquidity preference-Money supply", preferencia por la liquidez-oferta monetaria), que recoge el equilibrio del mercado de dinero.

Figura 2.1: Modelo IS-LM



La pendiente de la curva IS dependerá de la sensibilidad de la inversión al tipo de interés y la de la LM de la sensibilidad de la demanda de dinero al tipo de interés. En función de estas sensibilidades la política fiscal y la política monetaria tendrán una mayor o menor capacidad de afectar al output.

Implicaciones de política económica Los principales rasgos generales de la Síntesis Neoclásica, respecto a la política económica, son:

- La inestabilidad del sector privado y la existencia de perturbaciones justifican la intervención del estado con políticas de estabilización.
- El crecimiento y la suavización de las fluctuaciones cíclicas son los objetivos de la política económica.
- El uso mixto de la política fiscal y de la política monetaria, mediante fine tuning, permitiría alcanzar el pleno empleo y estabilizar las fluctuaciones cíclicas.

Perfeccionamiento de los elementos keynesianos

La Síntesis Neoclásica trata, además, de perfeccionar los elementos principales del análisis keynesiano:

- Demanda de consumo, ante la evidencia empírica de Kuznets[17]:

[17]En datos de corte transversal/sección cruzada y en series temporales cortas, el consumo seguía la función keynesiana. Sin embargo, en series temporales largas la función de consumo ya no dependía del consumo autónomo, por lo que PMe_C es igual a la PMg_C, dando lugar al "enigma del consumo", donde diferentes autores propusieron teorías para darle respuesta, tanto por el lado keynesiano, como por el monetarista con Friedman y su teoría de la renta permanente.

 © 2026 Manuel Cabado García

- Teoría de la renta relativa: la teoría de Duesenberry postula que el consumo depende de la renta relativa, es decir, de la renta del individuo en comparación con la de su grupo social de referencia y con la renta más alta que haya alcanzado en el pasado. Este último elemento genera un efecto trinquete que explica por qué el consumo es rígido a la baja.[18] Sin embargo, esta teoría se diferencia de otras al no considerar la riqueza ni la renta vitalicia como variables explicativas.
- Teoría del ciclo vital: la Teoría del Ciclo de Vida de Modigliani postula que el consumo depende de la renta vitalicia esperada. Los individuos buscan una senda de consumo estable a lo largo de su vida, ahorrando en las etapas de altos ingresos para desahorrar en las de bajos ingresos (como la jubilación). El ahorro es clave para suavizar el consumo intertemporal.

- Demanda de inversión, destacando:
  - Teorías del acelerador: la teoría del acelerador, cuyo principio se remonta a Aftalion (1909) y J. M. Clark (1917) y cuyas versiones flexibles se desarrollan entre los años cuarenta y sesenta, postula que la inversión de las empresas responde directamente a los cambios en la demanda. Si la demanda aumenta, las empresas invierten para ampliar su capacidad productiva y satisfacerla: $I = v \cdot \Delta Y$. Así, la inversión es proporcional a la variación de la demanda, ΔY, y el factor de proporcionalidad, v, es el coeficiente acelerador, que coincide con la relación capital-producto deseada y suele ser mayor que 1.[19] Con estas teorías también se explicarán los ciclos económicos.
  - La q de Tobin: fundamenta la inversión keynesiana al relacionar el valor de mercado de una empresa, VME, con el coste de reposición de su capital, CRK, es decir: $q = \frac{VME}{CRK}$. Si $q > 1$, la empresa invertirá porque su capital vale más en el mercado que lo que cuesta reponerlo. Esto impulsa la expansión, vinculando las expectativas del mercado, en las que influyen los "animal spirits", con las decisiones de inversión.

- Demanda de dinero: profundiza en el modelo de demanda de dinero de Keynes tratando de mejorar sus fundamentos, y explicar la creciente evidencia empírica disponible sobre la demanda de dinero:
  - Motivo transacción: modelo de inventarios, el modelo Baumol (1952) y Tobin (1956) trata de explicar la demanda de dinero por motivo de transacción me-

[18] La intuición es que una familia que ha llegado a ganar 4.000 euros al mes y pasa a ganar 3.000 no renuncia de golpe a su nivel de vida, sino que reduce su ahorro para mantenerlo.

[19] La intuición es que, si una fábrica necesita 3 euros de maquinaria por cada euro de producción anual, un aumento de la demanda de 100 euros exige invertir 300. Por eso pequeñas variaciones de la demanda generan grandes oscilaciones de la inversión.

diante modelos de gestión de inventarios, asumiendo que los agentes minimizan el coste de mantener los saldos monetarios necesarios para realizar transacciones.

- Motivo precaución: modelo de costes de ruptura, El modelo de Whalen (1966) parte del enfoque de Baumol y Tobin, pero introduce la existencia de incertidumbre sobre los gastos e ingresos, lo que genera unos costes de ruptura de no tener suficientes saldos monetarios, y permite, así, analizar el motivo de demanda de dinero por precaución.
- Motivo especulación: modelo de cartera, Tobin (1958) trata de fundamentar la demanda de dinero por motivo de especulación, considerando que es resultado de las decisiones de optimización de los agentes sobre su cartera en un entorno de incertidumbre suponiendo que los agentes eligen entre dos activos el dinero que es libre de riesgo y ofrece rendimiento nulo y los bonos que son activos con riesgo y rendimiento positivo. El resultado es que la demanda de dinero por motivo de especulación depende negativamente del tipo de interés.

Curva de Phillips

Alban William Phillips publicó en 1958 un trabajo donde mostraba la relación entre el desempleo y la variación de los salarios monetarios en Reino Unido entre 1861 y 1957 (publicado en la revista *Economica*). La relación derivada del análisis empírico era negativa, no lineal (y con rigideces salariales a la baja) con muy buen ajuste a los datos observados.

Lipsey es el primero en ofrecer una explicación teórica de la relación empírica en los 60s. Para ello se basa en dos supuestos fundamentales:

1. Una relación lineal positiva entre el crecimiento de los salarios y el exceso de demanda de trabajo
2. Una relación no lineal negativa entre el exceso de demanda de trabajo y el desempleo, similar a la Curva Beveridge.

Combinando ambos supuestos se podía expresar el incremento en los salarios como una función del nivel de desempleo

$$\pi^w = f(u) \tag{2.9}$$

Posteriormente, Samuelson y Solow (1960) estiman la Curva de Phillips para EE.UU y proponen la existencia de una relación negativa no lineal (trade-off) entre inflación y desempleo, que será la Curva de Phillips que se empieza a plantear desde ese momento

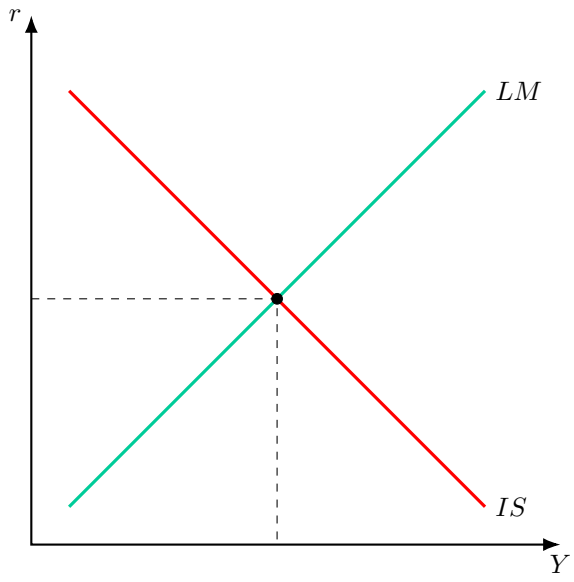
$$\pi = f(u) \tag{2.10}$$

Samuelson y Solow sugieren que esta relación puede explotarse como un "menú" de

 © 2026 Manuel Cabado García

política económica, eligiendo la combinación de inflación y desempleo preferida para cada periodo[20]. De esta forma, sería posible alcanzar menores tasas de desempleo de forma permanente mediante impulsos reiterados de la demanda agregada, a cambio de aceptar también de forma permanente una mayor inflación y pasar, como se observa en la Figura 2.2, del punto A al punto B.

Figura 2.2: Curva de Phillips



Esta confianza en la posibilidad de explotar de forma estable el trade-off entre inflación y desempleo no tardaría, sin embargo, en ser puesta en entredicho por una nueva generación de economistas —los monetaristas— que cuestionarían tanto sus fundamentos teóricos como su capacidad para explicar la estanflación de los años 70.

2.4. Monetaristas ⏳ 3'

El monetarismo[21] es una escuela que surge en los 50 y 60 encabezada por Milton Friedman, uno de los miembros más destacados de la Escuela de Chicago y Premio Nobel en 1976.

Antes de entrar en sus ideas sustantivas, conviene señalar la impronta metodológica de Friedman: en su ensayo "The Methodology of Positive Economics" (1953), defendió

[20] Los propios autores advirtieron de que la relación podría desplazarse a largo plazo, una cautela que la interpretación posterior de su trabajo tendió a olvidar.

[21] El término 'monetarismo' fue acuñado por Karl Brunner (economista suizo), y lo definió como "aquel movimiento intelectual que se apoya en la creencia de que el principal determinante de la demanda global en términos nominales es el dinero, y, de manera concreta, sus tasas de variación". Se define el monetarismo como aquella doctrina económica que estudia los efectos de las variaciones de la oferta monetaria sobre las variables económicas relevantes (como precios, empleo y producción).

El monetarismo plantea que mientras que las autoridades monetarias (banco central u otros) tienen el control sobre la oferta nominal, las personas basan sus decisiones con respecto a la cantidad de dinero real que esperan obtener.

la primacía de la economía positiva sobre la normativa y la tesis de la irrelevancia de los supuestos –una teoría debe juzgarse por el poder predictivo de sus conclusiones, no por el realismo de sus premisas–, en contraposición a autores como Gunnar Myrdal, para quien los conceptos económicos están inevitablemente cargados de juicios de valor. Este instrumentalismo metodológico, de raíz popperiana, es el que sustenta el enfoque empírico con el que Friedman construirá su crítica del edificio keynesiano.

Principales ideas del monetarismo en contraposición con la SNC:

1. Flexibilidad de precios y salarios como una aproximación aceptable de la realidad.
2. La oferta agregada es la principal determinante de los resultados económicos, además de introducir el concepto de la tasa natural de paro (TNP)
3. La inflación es un fenómeno monetario, causado por un aumento de la oferta monetaria superior al crecimiento de la producción, de acuerdo con la ecuación cuantitativa $MV = PY$, donde la velocidad de circulación V es estable en el corto plazo.
4. La política monetaria solo tiene efecto sobre las variables reales a corto plazo. La demanda de dinero será relativamente estable, como se verá, al depender de muchas más variables.
5. Las fluctuaciones económicas han sido provocadas por política económica imprudente, mientras que para la SNC la economía está sujeta a shocks erráticos y es inestable.
6. Para los monetaristas la política monetaria es más eficaz que la política fiscal, para la Síntesis al revés.

Revisión de los componentes keynesianos

Así, los monetaristas van a realizar una importante revisión de los componentes keynesianos:

- Demanda de consumo: La crítica monetarista se materializará en la Teoría de la Renta Permanente (1957). Esta teoría parte de la hipótesis de que los agentes buscan maximizar intertemporalmente su utilidad. En consecuencia, buscarán suavizar su consumo a lo largo del tiempo y éste pasará a depender de la renta permanente, definida como "aquel flujo de renta que puede ser mantenido perfectamente en el tiempo y que puede ser consumido íntegramente sin afectar a la fuente de riqueza de la que deriva". Así, la renta permanente implica consumir los intereses de la riqueza[22]:

$$Y^P = r \cdot W \tag{2.11}$$

[22]La riqueza, W, es la suma de los saldos monetarios, M, el valor actualizado de los bonos poseídos, B/r, el capital o activos reales, K, las acciones, A, y el valor actual de los salarios presentes y futuros esperados, w, es decir, la riqueza humana: $W = M + \frac{B}{r} + K + A + w$

 © 2026 Manuel Cabado García

por lo tanto, el consumo, será proporcional a la renta permanente, reflejando el deseo del individuo de suavizar el consumo:

$$C^P = k \cdot Y^P$$

Además, la renta permanente no es observable. De ahí que, según Friedman, los individuos la tengan que estimar, y lo hacen sobre la base de la hipótesis de las expectativas adaptativas (HEA), según la cual los agentes corrigen su previsión en cada periodo en una fracción del error cometido en el periodo anterior, de modo que la previsión acaba siendo una media ponderada de los valores pasados. Aplicada al consumo, el individuo estima su renta permanente observando sus rentas pasadas.[23]

- Demanda de inversión: para los monetaristas, la inversión depende fundamentalmente del tipo de interés, ya que, al ser racionales, basan sus decisiones en la rentabilidad. La inversión es mucho más estable.
- Demanda de dinero: Friedman concibe el dinero como un bien más, de manera que su demanda no tiene que ser justificada por razones particulares (como hacían los keynesianos):

  - Así, la demanda de dinero dependerá, como la de cualquier otro activo, de tres factores principales:

    1. La riqueza del individuo, medida por la renta permanente, y que, según los monetaristas, será el factor determinante de la demanda de dinero. De ahí que la demanda de dinero monetarista sea muy estable (a diferencia de la keynesiana), ya que la renta permanente lo es.
    2. La rentabilidad del dinero en relación a otros activos financieros y reales.
    3. Las preferencias de los individuos.

  - Concretamente, la demanda de dinero dependerá positivamente de la renta permanente, del nivel general de precios y de la proporción de las rentas del trabajo –las rentas del trabajo son, generalmente, más inciertas que las provenientes de activos financieros, por lo que los agentes desearán mantener más dinero por motivo precaución–, y negativamente de la rentabilidad de otros activos y de las expectativas de inflación

$$M^D = f(\underset{+}{Y^P}, \underset{+}{\frac{W_L}{W}}, \underset{-}{r}, \underset{-}{\pi^e}) \cdot P \tag{2.12}$$

[23] La intuición es que quien ha visto subir los precios un 2 % durante años espera otro 2 %, y solo revisa poco a poco esa cifra si la inflación se dispara, por lo que se equivoca siempre en la misma dirección mientras dura el cambio.

De esta forma, se rechazarán los dos eslabones débiles de Keynes, por lo que la política monetaria será más efectiva que la fiscal aunque solo en el corto plazo. A largo plazo, las recomendaciones de política monetaria son de uso reglado y no discrecional, ya que el uso discrecional solo llevará a aumentos de la inflación, cumpliéndose así la tesis aceleracionista de la inflación.

Curva de Phillips aumentada con expectativas

La segunda fase del monetarismo de Friedman critica a la Síntesis Neoclásica por su enfoque en la curva de Phillips. Argumenta que la relación entre inflación y desempleo es solo temporal, ya que los agentes económicos negocian los salarios nominales basándose en sus expectativas de inflación, preocupándose por los salarios reales.

Además, Friedman introduce el concepto de tasa natural de paro (TNP), es decir, el nivel de desempleo que existe cuando el mercado de trabajo está en equilibrio y las expectativas se cumplen, determinado por las características estructurales de los mercados de trabajo y de bienes y no por la demanda agregada[24]. De esta forma, intentos por reducir el desempleo por debajo de esta tasa solo causan una inflación acelerada. Este modelo explica la estanflación de los años 70, un fenómeno que la curva original era incapaz de prever.

Así, la curva de Phillips aumentada con expectativas, en términos de precios, se puede expresar como la inflación esperada más un múltiplo de la diferencia entre la tasa natural de paro y la tasa de paro

$$\pi = \pi^e + \alpha \cdot (u_n - u) \tag{2.13}$$

donde α recoge la sensibilidad de la inflación a la brecha entre ambas tasas de paro. Por lo tanto, si la tasa de desempleo es inferior a la tasa natural de paro, la inflación será superior a la esperada.

Ahora, tal y como se observa en la Figura 2.3:

- A medida que los trabajadores perciban que sus salarios reales no han aumentado adaptarán sus expectativas de inflación, desplazando la curva de Phillips hacia arriba. En el punto C, la economía se encuentra en la tasa natural de paro, pero la tasa de inflación es superior a la que existía inicialmente.

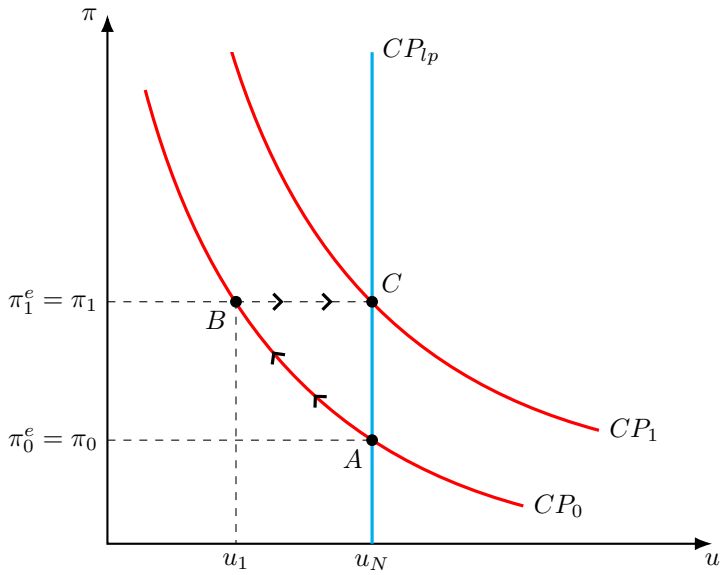
[24] En otras palabras, sería el nivel de desempleo resultante del equilibrio general Walrasiano que recoge las características estructurales del mercado de trabajo y de los bienes –fricciones e imperfecciones, costes de información, de movilidad, estructura impositiva– y en la que no hay desempleo involuntario.

A menudo, la tasa natural de paro (TNP) se identifica con el paro friccional, que hace referencia a aquel que se produce por la falta de emparejamiento instantáneo entre la oferta y la demanda de trabajo. Es decir, se debe a buscadores de empleo que aún no han aceptado una oferta de trabajo a la espera de encontrar una que les satisfaga, y/o a empresas que aún no han asignado una vacante a la espera de encontrar un candidato que les convenza. Edmund Phelps llegó de forma independiente a la misma idea en esos mismos años.

 © 2026 Manuel Cabado García

- A largo plazo, a medida que los trabajadores revisan sus expectativas e incorporan los errores cometidos, la curva de Phillips será vertical en la tasa natural de desempleo, que se obtiene uniendo los puntos A y C.

Figura 2.3: Curva de Phillips monetarista



El propio mecanismo de expectativas adaptativas que sostiene esta versión monetarista de la Curva de Phillips, sin embargo, pronto sería cuestionado por carecer de microfundamentos y por permitir errores sistemáticos y persistentes en los agentes: de esa crítica surgiría la Nueva Macroeconomía Clásica, que sustituiría las expectativas adaptativas por expectativas racionales.

2.5. Nueva Macroeconomía Clásica ⏳ 3'

La Nueva Macroeconomía Clásica (NMC) transformó la teoría económica al introducir expectativas racionales, microfundamentos rigurosos y formalización matemática. Sustituyó las relaciones keynesianas ad-hoc por un agente representativo que optimiza intertemporalmente.

Lucas, en su influyente crítica de 1976, mostró que los modelos econométricos tradicionales fallan al evaluar políticas, ya que los agentes, al anticiparlas racionalmente, modifican su comportamiento. Esto implica que los parámetros estimados no son constantes, lo que invalida su capacidad predictiva ante cambios en las reglas de política económica.

Hipótesis de las expectativas racionales

De esta forma, la Nueva Macroeconomía Clásica introduce la hipótesis de expectativas racionales (HER) y critica la HEA de Friedman porque permite al Gobierno "engañar"

sistemáticamente a los agentes económicos a corto plazo.

Bajo HER, los agentes ya no cometen errores sistemáticos porque utilizan eficientemente toda la información de la que disponen. Por lo tanto, el valor esperado de las variables económicas se halla mediante la esperanza de las variables condicionada a toda la información disponible en el momento de formular las expectativas, Ω_{t-1}, utilizando eficientemente dicha información

$$P_t^e = \mathbb{E}\left(P_t | \Omega_{t-1}\right) \tag{2.14}$$

Por lo tanto, el valor de la variable en el momento t será igual al valor esperado de la variable más un término de error o ruido blanco

$$P_t = P_t^e + \varepsilon \tag{2.15}$$

Así, Robert Lucas fue el primer economista en aplicar a la Macroeconomía la HER, formulada por John Muth en 1961, y concluyó que, al combinarla con la teoría de la tasa natural de paro, los errores de previsión solo pueden tener un carácter aleatorio, lo que dio origen a la NMC.

Curva de oferta de Lucas Bajo estos supuestos, se obtiene la curva de oferta de Lucas, según la cual la producción, y_t, solo se aparta de su nivel natural, y_N (asociado a la tasa natural de desempleo), cuando el nivel de precios difiere del esperado, es decir, cuando hay errores en las expectativas. El parámetro θ mide la intensidad de esa respuesta y depende de la variabilidad relativa de las perturbaciones: el productor no sabe si una subida de su precio refleja un cambio en la demanda de su mercado o inflación general, y cuanto más frecuentes sean las perturbaciones específicas frente a las generales, más atribuirá la subida a su mercado y más aumentará la producción:

$$y_t = y_{N_t} + \theta\left(P_t - P_t^e\right) \tag{2.16}$$

es decir, la producción agregada solo se desvía de su nivel natural por variaciones no anticipadas del nivel de precios. Por consiguiente, frente a cambios de precios anticipados, la curva es vertical incluso en el corto plazo, y solo las sorpresas de precios le dan pendiente positiva.[25]

[25] La intuición es que un agricultor que ve subir el precio de sus naranjas no sabe si es porque la gente quiere más naranjas o porque todo está subiendo. Si produce más y resulta que era inflación general, ha trabajado de más por un error de percepción.

 © 2026 Manuel Cabado García

Curva de Phillips Además, considerando la ley de Okun[26] según la cual existe una relación estable y negativa entre desempleo y PIB, la curva de oferta agregada de Lucas se puede interpretar como una representación alternativa de la curva de Phillips aumentada con expectativas, por lo que reorganizando la expresión:

$$u_t = u_{N_t} - \frac{1}{\alpha} \cdot (\pi_t - \pi_t^e), \quad \alpha > 0 \tag{2.17}$$

Implicaciones de política monetaria

Estos desarrollos tienen una serie de implicaciones de política económica:

- Proposición de Inefectividad de la Política, formulada por Sargent y Wallace (1975): bajo expectativas racionales y vaciado continuo de mercados, cualquier regla de política monetaria sistemática –anticipable por los agentes– es neutral incluso a corto plazo, afectando solo a los precios. Únicamente las sorpresas monetarias no anticipadas tienen efecto real sobre la producción.
- Inconsistencia dinámica de la política monetaria, introducida por Kydland y Prescott (1977) y aplicada específicamente al ámbito monetario por Barro y Gordon (1983): una política óptima ex ante puede dejar de serlo ex post, de modo que la autoridad tiene incentivos a desviarse de lo anunciado una vez que los agentes ya han actuado en consecuencia. Por ejemplo, tras anunciar una política antiinflacionista que modera las expectativas de precios, la autoridad se ve tentada a generar una sorpresa inflacionista para impulsar el empleo, aunque unos agentes racionales anticiparán ese incentivo y ajustarán sus expectativas de partida en consecuencia. Concretamente, si la política monetaria es discrecional, la autoridad monetaria tendrá incentivos para incrementar el nivel de producción de manera ex-post, elevando el nivel de inflación. No obstante, aunque la economía se encuentre dinámicamente en el equilibrio, esta experimentará la misma tasa natural de desempleo, pero con una inflación superior (sesgo inflacionista). Por consiguiente, el bienestar de dicha sociedad será inferior –dado que la inflación y el desempleo son males–, siendo clave para la NMC el establecimiento de reglas[27].
- Equivalencia Ricardiana, David Ricardo planteó el hipotético caso de que deuda e impuestos fuesen sustitutivos. Barro (1974) recuperó esta idea y defendió que el efecto expansivo de un aumento del gasto público será independiente de cómo se

[26] La Ley de Okun establece una relación empírica entre las desviaciones del desempleo de su tasa natural y las desviaciones de la producción real de su producción potencial. Es decir, cuando la producción real es inferior a la potencial, el desempleo es superior a su tasa natural, y viceversa. Esta ley cuantifica el coste en términos de producción de tener un desempleo por encima de su tasa natural.

[27] Otras soluciones al problema de la inconsistencia temporal son la reputación del banco central, que está dispuesto a aceptar más paro para crearse una reputación antiinflacionista, la delegación en un banco central independiente con fuertes preferencias antiinflacionistas y los contratos óptimos, en los que la remuneración del banco central depende de su éxito en reducir la inflación.

financie, vía impuestos o vía déficit, pues los agentes no perciben la deuda como riqueza neta (es decir, no hay ilusión de deuda), ya que conocen la restricción intertemporal del Gobierno y saben que, tarde o temprano, tendrá que satisfacerla, por lo que aumentará los impuestos en el futuro.

- Desagradable aritmética monetarista, desarrollada por Sargent y Wallace (1981), es una situación en la que la política monetaria no es capaz de controlar la inflación por culpa de una política fiscal sistemáticamente expansiva. Se llega a una situación en la que la financiación del déficit público a través de deuda resulta más inflacionista que su monetización. Así, en última instancia, y en el contexto particular de este modelo, la inflación sería más un fenómeno fiscal que un fenómeno monetario, poniéndose de manifiesto la interrelación entre la política monetaria y la fiscal. Para Sargent y Wallace, si la monetización del déficit en el futuro es inevitable, cuanto más tarde en realizarse peor, ya que mayor será el volumen de intereses y mayor, por tanto, la necesidad de monetización.

Por consiguiente, abogarán por políticas fiscales de oferta, no de demanda, y una política monetaria reglada.

Teoría de los ciclos de la NMC

Basándose en su teoría de la oferta, Lucas postula que los shocks monetarios inesperados confunden a los agentes con información imperfecta, quienes los interpretan como cambios en los precios relativos. Esto lleva a un aumento temporal de la oferta y el empleo (expansión). Al corregir su error, los productores reducen la producción, generando una recesión.

Sin embargo, esta teoría solo explica ciclos económicos de corta duración. Así, la teoría monetaria de los ciclos fue superada por la teoría de los ciclos reales.

Teoría de los ciclos reales La teoría de los ciclos reales de Kydland y Prescott (1982) explica las fluctuaciones económicas como la respuesta óptima de los agentes a perturbaciones reales, sobre todo de productividad, y transformó de manera revolucionaria la teoría de ciclos y el análisis macroeconómico al inaugurar los modelos DSGE[28]. Estos se basan en dos aspectos:

- Introducen una fuente de perturbación real (inicialmente, shocks de productividad y, posteriormente, otras fuentes reales tales como el gasto público).
- Asumen mecanismos de optimización intertemporal de consumo y ocio. De este modo, se acaba con la idea de que las recesiones son "malas" y las fases de expansión

[28] La intuición es que, si un año la tecnología rinde menos, a los trabajadores les compensa trabajar menos ese año y más cuando la productividad se recupere. La recesión sería así una respuesta racional, no un fallo del mercado.

 © 2026 Manuel Cabado García

"buenas": las recesiones no serán deseables para los agentes, pero son el resultado de su conducta maximizadora, por lo que los ciclos son un fenómeno de equilibrio.

Implicaciones de política económica El ciclo es, en todo momento, un fenómeno Pareto eficiente, ya que es consecuencia de las decisiones óptimas de los agentes, siendo todo el desempleo voluntario, por lo que no es recomendable ni deseable una política económica estabilizadora. Esta teoría tiene una gran fundamentación microeconómica a la vez que es la primera en identificar la causa de los ciclos como algo ajeno a la Política Económica.

Precisamente esa conclusión de eficiencia y esa ausencia de espacio para la política económica resultaban difíciles de compatibilizar con fenómenos tan persistentes como el desempleo involuntario prolongado: aceptando la Hipótesis de Expectativas Racionales, la Nueva Economía Keynesiana buscaría recuperar un papel activo para la política económica introduciendo competencia imperfecta y rigideces de precios y salarios con microfundamentos propios.

2.6. Nueva Economía Keynesiana ⏳ 3'

Sin embargo, la NMC fue pronto criticada, especialmente por su supuesto extremo de que los mercados están continuamente en equilibrio, por lo que no lograban explicar la existencia de desempleo involuntario. Así, surgió a principios de los años 80 una nueva generación de economistas (Akerlof, Bernanke, Blanchard, Summers, Fischer o Yellen) que desarrollaron la llamada Nueva Economía Keynesiana (NEK), una nueva corriente que, partiendo del mismo rigor microeconómico que la NMC, introdujo fricciones de distinto tipo que acercaban los resultados de los modelos más a la realidad.

Aunque no existía un programa de investigación único, los modelos neokeynesianos, según Mankiw y Romer, comparten dos características esenciales:

1. El dinero no es neutral, es decir, las fluctuaciones nominales tienen efectos reales debido a las rigideces.
2. Las imperfecciones de los mercados y la rigidez de precios o salarios son cruciales para explicar los ciclos económicos, ofreciendo amplias posibilidades para entender la falta de respuesta de los precios.

Se distinguen dos generaciones.

Primera generación

Los primeros modelos son los de primera generación, que se desarrollan desde finales de los 70 y durante los 80, y se centran en el análisis de las rigideces salariales y de precios y

en sus implicaciones de PE usando modelos estáticos de equilibrio parcial con mercados de competencia imperfecta.

Admiten que los modelos keynesianos previos adolecían de problemas como la falta de microfundamentación y fallos a la hora de justificar el supuesto de rigideces. Las principales aportaciones de los neo-keynesianos de primera generación son:

- Modelos de rigideces salariales:
  - Nominales: los economistas neokeynesianos explican las rigideces nominales con modelos de contratos multiperiodo (Fischer) y negociaciones descoordinadas (Taylor). Argumentan que estas rigideces, en un contexto de expectativas racionales, hacen que los shocks nominales tengan efectos reales, ya que los precios y salarios se ajustan lentamente.
  - Reales: la teoría neokeynesiana del mercado de trabajo micro desarrolla una serie de modelos en los que el salario real es rígido en el sentido de que no se mueve para vaciar el mercado, bien por la actuación optimizadora de los trabajadores (modelo de insiders-outsiders), bien por la actuación optimizadora de las empresas (modelo de salarios de eficiencia y modelo de contratos implícitos).
- Modelos de rigideces de precios:
  - Costes de menú, de Mankiw (1985), son todos aquellos costes en los que puede incurrir una empresa en el proceso de cambio de sus precios de venta, y pueden deberse a:
    - El tiempo de reunión y deliberación del nuevo precio.
    - La impresión de los nuevos precios.
    - Deterioro de la imagen (ante continuos cambios en el precio) y pérdida de clientes.
  - Cuasirracionalidad (o racionalidad incompleta), de Akerlof y Yellen (1985), un comportamiento cuasirracional es aquel en el que el agente no altera sus decisiones ante una perturbación si la pérdida prevista por no hacerlo es pequeña, es decir, de segundo orden: al partir de un óptimo, esa pérdida crece con el cuadrado de la perturbación y es despreciable si esta es pequeña, aunque sus efectos agregados sobre la producción y el empleo puedan ser importantes.[29] La cuasirracionalidad puede darse por diversos motivos: pereza, falta de información, falta de tiempo, existencia de costes –de hecho, los costes de menú serían un ejemplo de cuasirracionalidad–.

[29] La intuición es que a un restaurante no le compensa reimprimir la carta porque el precio óptimo de un plato haya subido diez céntimos, pero si todos los restaurantes razonan igual, el nivel de precios apenas se mueve y la perturbación se traslada a las cantidades.

 © 2026 Manuel Cabado García

- Aversión al riesgo del empresario, de Greenwald y Stiglitz (1988), la idea es que si los empresarios son aversos al riesgo la negativa a ajustar los precios puede ser una conducta óptima del empresario.
- Fallos de coordinación, Blanchard y Kiyotaki (1987) demostrarán que, si bien para cada empresa individual puede ser óptimo mantener su precio en respuesta a una caída de la demanda cuando existen costes de menú o comportamiento cuasi racional o aversión al riesgo del empresario, el hecho de que todas las empresas hagan esto puede dar lugar a importantes efectos a nivel macroeconómico: es decir, puede generar externalidades de demanda agregada.
- Escalonamiento de precios, la forma en que las empresas modifican sus precios también puede generar rigideces nominales: por ejemplo, cuando las empresas revisan sus precios no de forma continua, sino con una regularidad predeterminada (variación de los precios según periodicidad), o cuando lo hacen solo si el nuevo precio óptimo está fuera de una horquilla que se marcaron previamente como límite (modificación de precios en función de la evolución de los precios de alguna variable importante). El grado de rigidez del nivel general de precios depende de:
  - Frecuencia del ajuste: cuanto menor sea la frecuencia de ajuste, mayores fricciones nominales.
  - Momento de ajuste de cada empresa: cuanta menos sincronía exista, menor será la cuantía del ajuste.

- Rigideces en el mercado de crédito: las imperfecciones que se presentan en el mercado de crédito dan lugar a un tipo de interés real fijo que no vacía el mercado de crédito. Ello da lugar a equilibrios con racionamientos en el mercado financiero (credit crunch). En el modelo de Stiglitz y Weiss (1981), los prestamistas se enfrentan al riesgo de que los prestatarios no devuelvan el crédito que se les ha concedido, generándose un problema de información asimétrica.
- NAIRU, o tasa de paro no aceleradora de la inflación, es decir, la tasa de desempleo con la que la inflación se mantiene estable porque las aspiraciones salariales de los trabajadores son compatibles con los márgenes que desean fijar las empresas[30]. Los economistas de la NEK realizaron importantes aportaciones en el ámbito de las imperfecciones del mercado laboral, desarrollando el denominado modelo de la NAIRU (Layard y Nickell (1985) y Carlin y Soskice (1990)), que explica cómo el desempleo estructural persiste incluso en equilibrio, debido a factores como la

[30]La intuición es que, si el paro es muy bajo, los trabajadores piden subidas que las empresas trasladan a los precios, lo que provoca nuevas peticiones salariales y la inflación se acelera. Si el paro es muy alto, ocurre lo contrario. La NAIRU es el punto de equilibrio entre ambos casos. A diferencia de la tasa natural de Friedman, surge de mercados de competencia imperfecta y puede depender de la historia reciente del desempleo.

negociación salarial, la regulación y las expectativas.

Segunda generación

La primera generación de economistas neokeynesianos no pudo consolidarse, en parte, por usar metodologías obsoletas. Sin embargo, una segunda oleada transformó la Macroeconomía al aceptar las reglas metodológicas de la NMC, pero con cambios fundamentales en la teoría de los ciclos.

Estos nuevos neokeynesianos, como Clarida, Galí, Gertler, Woodford o Bernanke, integraron los principios de los modelos DSGE (Equilibrio General Dinámico Estocástico), con agentes racionales y decisiones microfundamentadas. No obstante, se diferenciaron en dos aspectos clave:

1. Rechazo de la competencia perfecta y adopción de competencia monopolística, con rigideces y ajustes escalonados de precios y salarios.
2. Reintroducción de los aspectos monetarios y la política monetaria en el centro del análisis, incluyendo una autoridad monetaria activa.

Así, estos modelos de segunda generación, basados en la Teoría de los Ciclos Reales, constituyeron la Nueva Síntesis, siendo el marco dominante hasta la Gran Recesión.

Modelo canónico de tres ecuaciones de la NEK

Las principales aportaciones de los economistas de la segunda generación se verán al desarrollar el modelo canónico de la NEK.

Así, estos modelos sintetizan las condiciones de optimización mediante tres elementos:

1. La curva IS ampliada con expectativas, se obtiene del problema de maximización de utilidad intertemporal de los hogares y representa el lado de la demanda del modelo

$$x_t = \mathbb{E}\left(x_{t+1}\right) - \frac{1}{\sigma} \cdot \left(i_t - \mathbb{E}\left[\pi_{t+1}\right] - r_t^n\right) + u_t \tag{2.18}$$

   establece que, por el lado de la demanda existe una relación de equilibrio entre el output gap, x_t, y el tipo de interés nominal, i_t, condicionada a las expectativas sobre la inflación futura, $\mathbb{E}\left[\pi_{t+1}\right]$, y sobre el output gap futuro, $\mathbb{E}\left[x_{t+1}\right]$, y al shock de demanda, u_t.

2. La Curva de Phillips ampliada con expectativas, se obtiene del problema de optimización de las empresas y de la fijación de precios (a la Calvo (1983), donde en cada periodo solo una fracción de empresas puede reajustar su precio), siendo una versión forward-looking de la curva de Phillips tradicional, dando lugar al lado de la oferta del modelo

$$\pi_t = \beta \cdot \mathbb{E}\left(\pi_{t+1}\right) + \kappa \cdot x_t + e_t \tag{2.19}$$

 © 2026 Manuel Cabado García

que relaciona la inflación actual, π_t, con las expectativas de inflación futura y con el output gap, x_t, donde κ mide la sensibilidad al ciclo económico.

3. Regla monetaria, que sigue el banco central, de tipo Taylor, para ajustarse a unos objetivos de inflación, π_T y output gap, donde los parámetros ϕ son el peso relativo que le asigna el banco al gap de inflación y de producción[31]

$$i_t = r_t^n + \pi_t + \phi_\pi \cdot (\pi_t - \pi_T) + \phi_x \cdot x_t + \varepsilon_t \tag{2.20}$$

así, si la inflación o el output gap aumentan por encima de sus valores objetivo, el banco central sube los tipos para enfriar la economía. Para que el equilibrio quede determinado es necesario que el tipo de interés nominal reaccione más que proporcionalmente ante la inflación, de modo que el tipo real suba cuando esta aumenta (Principio de Taylor). En la formulación anterior, en la que la respuesta total a la inflación es $1 + \phi_\pi$, basta con que $\phi_\pi > 0$.[32]

El resultado es un sistema de ecuaciones que podría considerarse una modelización del modelo IS-LM. Blanchard y Galí (2007) bautizaron como "divina coincidencia" al resultado de que, en ausencia de rigideces salariales, estabilizar la inflación equivale a estabilizar también el output gap, de modo que ambos objetivos del banco central dejan de estar en conflicto.

Implicaciones de política económica

Durante las décadas de 1990 y 2000, la Nueva Síntesis (combinando la NMC y la NEK) estableció un consenso en política económica, con tres pilares centrales que la Gran Recesión pondría en tela de juicio.

Consenso Pre-Crisis 2008 Este periodo coincidió con la denominada "Gran Moderación"[33]:

1. Preeminencia de la Política Monetaria: se consideraba la principal herramienta para gestionar la demanda agregada a corto plazo, con un banco central independiente como pilar fundamental para influir en las expectativas.
2. Poca Relevancia de los Mercados Financieros: se ignoraba su impacto macroeconómico, salvo por el canal de crédito, y la regulación financiera no se consideraba una herramienta clave.

[31] En la formulación inicial de Taylor el valor era de 0,5 cada uno y el objetivo de inflación y los tipos reales eran del 2 %

[32] La intuición es que, si la inflación sube un punto y el banco central solo sube el tipo nominal medio punto, el tipo real baja, la economía se calienta aún más y la inflación sigue subiendo.

[33] Este fue un periodo que va aproximadamente desde mediados de los años 80 hasta 2007, caracterizado por: baja volatilidad macroeconómica, confianza en la política monetaria y dominio de la Nueva Síntesis.

3. Poca Importancia de la Política Fiscal: se promovía su neutralidad para evitar el dominio fiscal, con escepticismo sobre su efectividad, incluso aceptando la Equivalencia Ricardiana.

Replanteando la Política Macroeconómica Post-Crisis La crisis subprime expuso las fallas de este consenso. Economistas neokeynesianos como Olivier Blanchard, Giovanni Dell'Ariccia y Paolo Mauro, en su influyente nota "Rethinking Macroeconomic Policy" (2010), lideraron una revisión de los planteamientos anteriores. Los nuevos elementos esenciales del debate son:

1. Regulación Macroprudencial: Reconocer su importancia para controlar los excesos de riesgo del sistema financiero.
2. Problemas de la política monetaria: Admisión de su fracaso al no haber considerado el sistema financiero, sugiriendo un nuevo diseño.
3. Nuevo Papel de la política fiscal: Se revaluó su poder estabilizador, especialmente en la "trampa de liquidez" del zero lower bound, donde su multiplicador se vuelve muy efectivo. No obstante, algunos autores, como Taylor, mantuvieron su escepticismo, mientras otros, como Blanchard, defendieron su uso para evitar una recesión más profunda.
4. Modelos HANK: los modelos HANK (Heterogeneous Agent New Keynesian) comenzaron a surgir y ganar prominencia a partir de la década de 2010, especialmente después de la Gran Recesión. Son una respuesta a las limitaciones de los modelos DSGE estándar de agente representativo (RANK - Representative Agent New Keynesian) para explicar fenómenos como la desigualdad, la persistencia de las recesiones y la transmisión de la política monetaria a través de canales distributivos.

Finalmente, en el marco de la crisis del coronavirus se retomaron las bondades del expansionismo fiscal combinado con políticas monetarias de tipo expansivo.

Junto a este recorrido central de la Macroeconomía, conviene repasar brevemente otros desarrollos del pensamiento económico contemporáneo que, sin formar parte del núcleo estricto del debate keynesiano, han marcado también la disciplina en las últimas décadas.

2.7. Otros desarrollos ⏳ 2'

Por último se analizarán brevemente las aportaciones en otros campos diferentes a la Macroeconomía.

 © 2026 Manuel Cabado García

Crecimiento económico

Estos modelos tratan de explicar el incremento espectacular de la renta en los últimos dos siglos:

- Modelo de Solow: en los años 50, Robert Solow (Nobel en 1987) formuló el modelo neoclásico, aún hoy una referencia clave, donde el progreso técnico es el motor del crecimiento. Sin embargo, la teoría del crecimiento económico se volvió cada vez más abstracta y menos conectada con la realidad empírica, por lo que el interés se desplazó hacia el estudio de los ciclos económicos y de las teorías del desarrollo.
- Modelos de crecimiento endógeno: no fue hasta mediados de los 80, con los trabajos de Romer y Lucas, cuando la teoría del crecimiento se revitalizó gracias a los modelos endógenos, donde el progreso técnico dejó de ser un factor externo con los modelos AK y de competencia imperfecta. En este último punto cabe destacar los últimos premios Nobel en este aspecto:
    - Paul Romer demostró que la inversión en conocimiento e innovación puede sostener el crecimiento a largo plazo, al tratar el saber como un bien sin rendimientos decrecientes.
    - William Nordhaus amplió esta visión al incorporar los costes ambientales de la tecnología, mostrando que el crecimiento económico no puede desligarse de sus consecuencias climáticas, lo que dio lugar a los Modelos de Evaluación Integrada (IAMs), como DICE, PAGE y FUND.
    - Daron Acemoglu, por su parte, reveló que las instituciones inclusivas son fundamentales para fomentar la innovación y garantizar un desarrollo sostenido.

De esta forma, estos economistas, laureados con el Nobel en 2018 (Romer y Nordhaus) y 2024 (Acemoglu), transformaron la teoría del crecimiento al explicar la prosperidad desde dinámicas internas —tecnología, medio ambiente e instituciones— sin depender de supuestos externos.

Comercio internacional

El comercio internacional siempre ha sido una materia de estudio por parte de la Economía. Desde la segunda mitad del siglo XX hasta hoy destacan las aportaciones de:

- Modelo de Heckscher-Ohlin-Samuelson: los autores neoclásicos expandirán las teorías del comercio internacional, mejorando el instrumental analítico y superando ciertos supuestos restrictivos de los clásicos. Así, la teoría de la ventaja comparativa de Ricardo no vendría dada solo por la tecnología, sino por la dotación factorial de los países como en el modelo de Heckscher-Ohlin-Samuelson, explicando de esta forma el comercio entre naciones.

- Nueva Teoría del Comercio Internacional: sin embargo, las teorías neoclásicas no son capaces de explicar por qué el 50 % de los intercambios comerciales sean de carácter intraindustrial. De esta forma, la Nueva Teoría del Comercio Internacional, con autores como Paul Krugman, explica este hecho mediante modelos de competencia imperfecta y rendimientos crecientes, por lo que el comercio provoca un efecto escala (aumento del tamaño del mercado) y un efecto selección (traspaso de la producción de empresas menos eficientes a más eficientes), además de un aumento de la variedad de bienes ofrecidos.
- Novísima Teoría del Comercio Internacional, desarrollada por autores como Melitz, se centra en por qué las empresas comercian, no solo los países. Incorpora la heterogeneidad empresarial, demostrando cómo la mayor innovación a nivel individual de empresa y el efecto selección generan nuevas ganancias del comercio.

Microeconomía

La Microeconomía ha experimentado una revolución en las últimas décadas:

- Teoría de la Organización industrial: esta rama estudia cómo la estructura del mercado (monopolios, oligopolios, competencia monopolística) y las estrategias de las empresas afectan la competencia, la innovación y el bienestar. Se enfoca en las imperfecciones del mercado y el comportamiento estratégico de las firmas, alejándose de los modelos de competencia perfecta.
- Economía del comportamiento: estudios como los de Daniel Kahneman y Richard Thaler demostraron que las decisiones humanas están llenas de sesgos. Esto llevó a la creación de modelos más realistas que explican fenómenos como por qué la gente no ahorra lo suficiente o por qué no toma la mejor decisión en un mercado. Esta rama ha influido directamente en el diseño de políticas públicas.
- Teoría de juegos: se aplica al análisis de interacciones estratégicas entre agentes económicos (empresas, gobiernos, reguladores), revelando cómo sus decisiones interdependientes afectan los resultados. Esto ha sido crucial para el Diseño de Mercados, permitiendo estructurar reglas y mecanismos que mejoran la eficiencia y la competencia en sectores complejos. Ejemplos incluyen el diseño de subastas para licencias de telecomunicaciones (optimizando la asignación de frecuencias para telefonía móvil e internet) y la asignación de recursos escasos en otros mercados regulados.
- Economía experimental: el uso de laboratorios controlados y experimentos de campo ha permitido a los economistas probar teorías en un entorno real. Este enfoque empírico ha proporcionado evidencia sólida sobre el comportamiento humano y la efectividad de ciertas políticas. Un ejemplo conciso y relevante es el trabajo de Esther Duflo y Abhijit Banerjee, ganadores del Nobel en 2019. Su uso de experimentos de

 © 2026 Manuel Cabado García

campo aleatorizados (RCTs) en países en desarrollo demostró empíricamente qué políticas son más efectivas para combatir la pobreza.

2.8. Conclusiones ⏳ 30"

A lo largo de este tema se ha visto cómo cada escuela nace como respuesta crítica a los límites de la anterior: la Síntesis Neoclásica depuró y formalizó el legado de Keynes, el Monetarismo le opuso la neutralidad del dinero a largo plazo, la Nueva Macroeconomía Clásica introdujo las expectativas racionales frente al componente adaptativo del Monetarismo, y la Nueva Economía Keynesiana recuperó, con microfundamentos propios, un papel activo para la política económica. Sobre este recorrido, cabe destacar además tres enfoques que convergen en una transformación profunda del pensamiento económico contemporáneo:

1. El giro hacia la economía empírica, evidenciado por los Nobel recientes, refleja una voluntad de conectar la teoría con la realidad observable. Concretamente, las principales tendencias son la aplicación de métodos de causalidad para la evaluación de políticas (Card, Angrist, Imbens), la investigación en economía del desarrollo con experimentos de campo (Banerjee, Duflo, Kremer) y el estudio del papel de los mercados financieros y los bancos en las crisis (Bernanke, Diamond, Dybvig).
2. Integración de la Macroeconomía y la Microeconomía, esta evolución exige una integración más estrecha entre ambas, reconociendo que las decisiones individuales y las políticas agregadas están entrelazadas en un sistema complejo.
3. El uso riguroso de datos se convierte en una herramienta indispensable: no solo para validar hipótesis, sino para diseñar políticas públicas más eficaces y adaptadas a los desafíos reales. Así, la economía actual se aleja de modelos abstractos y se orienta hacia una comprensión más precisa, interdisciplinar y pragmática del comportamiento humano y de los mecanismos del mercado.

No obstante, las dos frases de Keynes siguen plenamente vigentes: incluso después del final, los economistas siguen influyendo en el presente, moldeando decisiones que afectan a millones.

Apéndice

2.A.1. Lista de ganadores del Premio Nobel de Economía

A continuación, se presenta la lista de ganadores del Premio Nobel de Economía, por año, nombres, razón y principales contribuciones de cada uno.[34]

Cuadro 2.1: Ganadores del Premio Nobel de Economía

Año	Laureado	Motivo	Principales aportaciones
1969	Ragnar Frisch Jan Tinbergen	Por el desarrollo y aplicación de modelos dinámicos para el análisis de los procesos económicos	Teorema de Frisch–Waugh–Lovell, Variación conjetural Econometría, Instrumentos de política económica
1970	Paul Samuelson	Por el trabajo científico a través del cual ha desarrollado la teoría económica estática y dinámica y contribuido activamente a elevar el nivel del análisis en la ciencia económica.	Teorema de la envolvente, Estática comparativa, Preferencia revelada
1971	Simon Kuznets	Por sus empíricamente fundamentadas interpretaciones del crecimiento económico que ha conducido a una nueva y más profunda comprensión de la estructura económica y social y el proceso de desarrollo.	Curva de Kuznets, Contabilidad de la renta nacional, Ciclos de Kuznets

[34] Elaboración propia a partir de la fuente oficial nobelprize.org/prizes/economic-sciences/, la lista de laureados de en.wikipedia.org/wiki/List_of_Nobel_Memorial_Prize_laureates_in_Economic_Sciences y cdi.mecon.gob.ar/premios-nobel.

 © 2026 Manuel Cabado García

Año	Laureado	Motivo	Principales aportaciones
1972	John Hicks Kenneth Arrow	Por sus contribuciones pioneras a la teoría del equilibrio económico general y la teoría del bienestar.	Modelo IS-LM (Hicks), Teorema de imposibilidad de Arrow, Equilibrio general Arrow-Debreu
1973	Wassily Leontief	Por el desarrollo del método input-output y por su aplicación a importantes problemas económicos.	Análisis input-output, Paradoja de Leontief
1974	Gunnar Myrdal Friedrich Hayek	Por su trabajo pionero en la teoría del dinero y las fluctuaciones económicas y por sus penetrantes análisis de la interdependencia de los fenómenos económicos, sociales e institucionales.	Causación circular acumulativa (Myrdal), Orden espontáneo (Hayek), Teoría austriaca del ciclo económico
1975	Leonid Kantorovich Tjalling Koopmans	Por sus contribuciones a la teoría de la óptima localización de recursos.	Programación lineal, Análisis de actividades, Asignación óptima de recursos
1976	Milton Friedman	Por sus resultados en los campos del análisis del consumo, historia y teoría monetaria y por su demostración de la complejidad de la política de estabilización.	Hipótesis de la renta permanente, Monetarismo, Tasa natural de desempleo
1977	Bertil Ohlin James Meade	Por su rupturista contribución a la teoría del comercio internacional y los movimientos internacionales de capitales.	Modelo Heckscher-Ohlin, Teoría de uniones aduaneras

Año	**Laureado**	**Motivo**	**Principales aportaciones**
1978	Herbert A. Simon	Por su investigación pionera en el proceso de adopción de decisiones en las organizaciones económicas.	Racionalidad limitada, Satisficing
1979	Theodore Schultz Arthur Lewis	Por su investigación pionera en el desarrollo económico con atención particular a los problemas de los países en desarrollo.	Teoría del capital humano (Schultz), Modelo de Lewis (economía dual)
1980	Lawrence R. Klein	Por la creación de modelos econométricos y la aplicación al análisis de las fluctuaciones económicas y políticas económicas.	Modelos macroeconométricos, Modelo Klein-Goldberger
1981	James Tobin	Por su análisis de los mercados financieros y sus relaciones con las decisiones de gasto, empleo, producción y precios.	Tasa Tobin, Modelo Tobit, Q de Tobin, Modelo Baumol–Tobin
1982	George J. Stigler	Por sus estudios de las estructuras industriales, el funcionamiento de los mercados y las causas y efectos de la reglamentación pública.	Captura del regulador
1983	Gerard Debreu	Por haber incorporado nuevos métodos analíticos a la teoría económica y por su rigurosa reformulación de la teoría del equilibrio general.	Equilibrio general Arrow-Debreu, Teoría del valor

 © 2026 Manuel Cabado García

Año	Laureado	Motivo	Principales aportaciones
1984	Richard Stone	Por haber hecho contribuciones fundamentales al desarrollo de los sistemas de cuentas nacionales y haber por tanto mejorado substancialmente las bases del análisis económico empírico.	Sistema de Cuentas Nacionales
1985	Franco Modigliani	Por sus pioneros análisis del ahorro y de los mercados financieros.	Hipótesis del ciclo vital, Teorema de Modigliani-Miller
1986	James M. Buchanan	Por su desarrollo de las bases contractuales y constitucionales de la teoría de la adopción de decisiones económicas y políticas.	Elección pública (Public Choice), Economía constitucional
1987	Robert M. Solow	Por sus contribuciones a la teoría del crecimiento económico.	Modelo de crecimiento de Solow, Residuo de Solow (PTF)
1988	Maurice Allais	Por sus contribuciones pioneras a la teoría de los mercados y la eficiente utilización de los recursos.	Paradoja de Allais
1989	Trygve Haavelmo	Por su clarificación de los fundamentos de la teoría de la probabilidad para la econometría y su análisis de estructuras económicas simultáneas.	Enfoque probabilístico en econometría, Estructuras simultáneas

Año	**Laureado**	**Motivo**	**Principales aportaciones**
1990	Harry M. Markowitz Merton Howard Miller William F. Sharpe	Por su trabajo pionero en la teoría de la economía financiera.	Teoría de carteras (media-varianza), Teorema de Modigliani-Miller, CAPM
1991	Ronald H. Coase	Por su descubrimiento y clarificación del significado de los costes de transacción y los derechos de propiedad para la estructura institucional y el funcionamiento de la economía.	Teorema de Coase, Costes de transacción, Teoría de la firma
1992	Gary S. Becker	Por haber extendido el dominio del análisis microeconómico a un amplio campo del comportamiento y la interacción humanos, incluyendo comportamientos no mercantiles.	Capital humano, Economía de la discriminación, Economía de la familia
1993	Robert W. Fogel Douglass C. North	Por haber renovado la investigación de la historia económica aplicando la teoría económica y métodos cuantitativos para explicar el cambio económico e institucional.	Cliometría (Fogel), Instituciones y cambio económico (North)

 © 2026 Manuel Cabado García

Año	Laureado	Motivo	Principales aportaciones
1994	John C. Harsanyi John F. Nash Reinhard Selten	Por sus pioneros análisis del equilibrio en la teoría de los juegos no cooperativos.	Equilibrio de Nash, Información incompleta (Harsanyi), Equilibrio perfecto en subjuegos (Selten)
1995	Robert Lucas	Por haber desarrollado y aplicado la hipótesis de las expectativas racionales y haber por tanto transformado el análisis macroeconómico y profundizado la comprensión de la política económica.	Expectativas racionales, Crítica de Lucas
1996	James A. Mirrlees William Vickrey	Por sus fundamentales contribuciones a la teoría económica de los incentivos bajo información asimétrica.	Imposición óptima (Mirrlees), Subasta de Vickrey, Peaje de congestión (Vickrey)
1997	Robert C. Merton Myron S. Scholes	Por desarrollar un nuevo método para determinar el valor de los derivados.	Modelo Black-Scholes-Merton, Valoración de opciones
1998	Amartya Sen	Por sus contribuciones a la economía del bienestar.	Teoría del desarrollo humano, enfoque basado en las capacidades (Capability approach)
1999	Robert A. Mundell	Por sus análisis sobre las políticas fiscales y monetarias bajo diferentes sistemas monetarios y sus análisis de las áreas óptimas de divisas.	Zona monetaria óptima, Economía de la oferta,Modelo Mundell–Fleming, Efecto Mundell–Tobin

Año	**Laureado**	**Motivo**	**Principales aportaciones**
2000	James J. Heckman Daniel L. McFadden	Por su desarrollo de la teoría y los métodos de análisis de muestras selectivas Por su desarrollo de la teoría y los métodos de análisis de la elección discreta	Corrección de Heckman, Modelos de elección discreta (McFadden)
2001	George A. Akerlof Michael Spence Joseph E. Stiglitz	Por sus análisis de los mercados con información asimétrica.	Mercado de "limones"/selección adversa (Akerlof), Señalización (Spence), Screening/riesgo moral (Stiglitz)
2002	Daniel Kahneman Vernon L. Smith	Por haber integrado los conocimientos de la investigación psicológica en la ciencia económica, especialmente en lo relativo al juicio humano y la toma de decisiones en condiciones de incertidumbre. Por haber establecido los experimentos de laboratorio como herramienta en el análisis económico empírico, especialmente en el estudio de mecanismos de mercado alternativos.	Economía del comportamiento, teoría prospectiva, aversión a la pérdida, sesgos cognitivos. Economía experimental, subastas combinatorias.

 © 2026 Manuel Cabado García

Año	Laureado	Motivo	Principales aportaciones
2003	Robert F. Engle Clive W. J. Granger	Por haber desarrollado métodos de analizar las series temporales con volatilidad variante en el tiempo (ARCH) Por haber desarrollado métodos de análisis de series temporales con tendencias comunes (cointegración).	Modelo autorregresivo con heterocedasticidad condicional (autoregressive conditional heteroskedasticity, ARCH) Cointegración, Causalidad de Granger
2004	Finn E. Kydland Edward C. Prescott	Por sus contribuciones a la dinámica macroeconómica: "la consistencia del tiempo en la política macroeconómica y las fuerzas que regulan los ciclos económicos".	Teoría del Ciclo Económico Real (RBC theory), Inconsistencia dinámica en política monetaria Filtro de Hodrick-Prescott
2005	Robert J. Aumann Thomas C. Schelling	Por haber ampliado nuestra comprensión del conflicto y la cooperación mediante el análisis de la Teoría de Juegos.	Equilibrio correlacionado, Teorema del acuerdo de Aumann Punto de Schelling (teoría de juegos), Egonomics (autocontrol y el "yo dividido")
2006	Edmund S. Phelps	Por su análisis en política macroeconómica.	Regla de oro del ahorro, Tasa natural de desempleo, Discriminación estadística

Año	**Laureado**	**Motivo**	**Principales aportaciones**
2007	Leonid Hurwicz Eric S. Maskin Roger B. Myerson	Por haber establecido las bases de la teoría del diseño de mecanismos.	Diseño de mecanismos (teoría de juegos)
2008	Paul Krugman	Por sus aportes a la teoría de la Economía Internacional y la Geografía Económica.	Nueva teoría del comercio, Nueva Geografía Económica, Efecto mercado interior (Home market effect)
2009	Elinor Ostrom Oliver Williamson	Por sus aportes a la "Nueva economía institucional".	Análisis institucional y marco de desarrollo, Gobierno de los bienes comunes Nueva economía institucional
2010	Peter A. Diamond Dale T. Mortensen Christopher A. Pissarides	Por sus estudios sobre el desempleo.	Teorema de eficacia de Diamond-Mirrlees, Modelo del coco de Diamond. Teoría de búsqueda y emparejamiento (Matching theory)

 © 2026 Manuel Cabado García

Año	**Laureado**	**Motivo**	**Principales aportaciones**
2011	Thomas Sargent Christopher A. Sims	Por sus investigaciones empíricas sobre causa y efecto en la Macroeconomía.	Proposición de ineficacia de la política (Policy-ineffectiveness proposition, PIP) Modelo Vectorial Autorregresivo (VAR) en Macroeconomía, Teoría fiscal del nivel de precios, Desatención racional (Rational inattention)
2012	Alvin E. Roth Lloyd S. Shapley	Por sus trabajos en la teoría de las asignaciones estables y el diseño de mercado.	Problema del matrimonio estable (también problema de emparejamiento estable o SMP) ,Costes/mercados repugnantes Valor de Shapley, Juego estocástico, Juego potencial, Índice de potencia de Shapley-Shubik, Teorema de Bondareva-Shapley, Algoritmo de Gale-Shapley, Lema de Shapley-Folkman.

Año	**Laureado**	**Motivo**	**Principales aportaciones**
2013	Eugene Fama Lars Peter Hansen Robert Shiller	Por sus análisis empíricos sobre los precios de los activos financieros.	Modelo de tres factores de Fama-French, Hipótesis de mercado eficiente débil, semifuerte y fuerte. Método de los momentos generalizado (Generalized Method of Moments , GMM) Índice Case-Shiller, Índice CAPE
2014	Jean Tirole	Por sus análisis del poder de los mercados y su regulación.	Equilibrio perfecto de Markov, Mercado bilateral
2015	Angus Deaton	Por sus análisis sobre los sistemas de demanda, el consumo y los ingresos, la pobreza y el bienestar.	Sistema de demanda casi ideal (Almost Ideal Demand System, AIDS), Pseudo-paneles, Encuestas de hogares en países en desarrollo
2016	Oliver Hart Bengt Holmström	Por su trabajo en la teoría del contrato.	Moral Hazard, Contratos incompletos
2017	Richard H. Thaler	Por sus contribuciones a la economía conductual.	Nudge Theory, Mental Accounting, Choice Architecture

 © 2026 Manuel Cabado García

Año	Laureado	Motivo	Principales aportaciones
2018	William Nordhaus Paul Romer	Por haber integrado el cambio climático en el análisis macroeconómico a largo plazo. Por integrar las innovaciones tecnólogicas en el análisis macroeconómico a largo plazo.	Modelo DICE (Dynamic Integrated Climate-Economy model) Incorporación de la I+D a la teoría del crecimiento endógeno
2019	Michael Kremer Esther Duflo Abhijit Banerjee	Por su enfoque experimental para aliviar la pobreza global.	Uso de Randomized Control Trial (RCT) en la economía del desarrollo
2020	Paul Milgrom Robert Wilson	Por las mejoras en la teoría de las subastas y las invenciones de nuevos formatos de subastas.	Subastas simultáneas de rondas múltiples (SMRA), Teorema del no intercambio, Diseño del mercado, Efectos de reputación (teoría de juegos), Juegos supermodulares, Estática comparativa monótona, Principio de vinculación, Subasta de aceptación diferida

Año	**Laureado**	**Motivo**	**Principales aportaciones**
2021	Guido Imbens Joshua Angrist David Card	Por sus contribuciones empíricas a la economía del trabajo Por sus contribuciones metodológicas al análisis de las relaciones causales.	Experimentos naturales sobre economía laboral (incluyendo Diferencias en diferencias) Local average treatment effect (LATE), experimentos naturales para estimar los vínculos causales
2022	Ben Bernanke Douglas Diamond Philip H. Dybvig	Por su investigación sobre bancos y crisis financieras	Análisis de la Gran Depresión Modelo de Diamond–Dybvig
2023	Claudia Goldin	Por haber contribuido a mejorar nuestra comprensión de los resultados de las mujeres en el mercado laboral.	Análisis de la experiencia histórica de las mujeres en la economía
2024	Daron Acemoglu Simon Johnson James A. Robinson	Por los estudios sobre cómo se forman las instituciones y cómo estas afectan a la prosperidad.	Instituciones inclusivas vs. extractivas, Formación de instituciones y prosperidad

 © 2026 Manuel Cabado García

Año	Laureado	Motivo	Principales aportaciones
2025	Joel Mokyr Philippe Aghion Peter Howitt	Por haber explicado el crecimiento económico impulsado por la innovación: una mitad a Joel Mokyr, por identificar los prerrequisitos para el crecimiento sostenido a través del progreso tecnológico; la otra mitad, conjunta, a Philippe Aghion y Peter Howitt, por la teoría del crecimiento sostenido mediante destrucción creativa.	Destrucción creativa, Crecimiento endógeno, Historia económica de la tecnología
2026	Robert E. Hall Michael Woodford	Por los fundamentos microeconómicos de la estabilización macroeconómica.	Consumo como paseo aleatorio (Hall, 1978), Modelo neokeynesiano de tres ecuaciones (Woodford, 2003), Tipo de interés natural y output gap

2.A.2. La evolución metodológica de la investigación macroeconómica

En línea con el recorrido histórico de este tema —de la Síntesis Neoclásica a la Nueva Economía Keynesiana—, conviene cerrar con una perspectiva metaanalítica distinta: no qué ha dicho la Macroeconomía en cada etapa, sino cómo ha cambiado la propia forma de investigarla y producirla como disciplina académica.

Glandon, Kuttner, Mazumder y Stroup (2022) ofrecen el estudio más sistemático disponible sobre esta transformación. Los autores recopilaron a mano, artículo por artículo, información de 1.894 trabajos publicados entre 1980 y 2018 en las diez revistas más influyentes del campo —cinco especializadas en Macroeconomía (Journal of Monetary Economics, Journal of Economic Dynamics and Control, Journal of Money, Credit and Banking, American Economic Journal: Macroeconomics y Review of Economic Dynamics) y cinco de interés general (American Economic Review, Econometrica, Journal of Political

Economy, Quarterly Journal of Economics y Review of Economic Studies)—, clasificando cada uno según su objetivo epistemológico y su metodología cuantitativa.

Los resultados dibujan un panorama muy distinto del de los años 80. La teoría formal se ha vuelto casi omnipresente: los modelos formales, prácticamente todos microfundamentados, aparecen en el 70 % de los artículos de 2016-2018, y los trabajos sin sección teórica sustancial son ya una rareza —un giro que conecta directamente con la crítica de este mismo tema a Keynes por carecer de "microfundamentos": su función de consumo, presentada en 1936 como una "ley psicológica", sería hoy inadmisible sin ese andamiaje—. La disciplina ha abandonado, además, en gran medida el programa positivista de contrastar hipótesis al estilo popperiano: solo el 10 % de los artículos de 2016-2018 se proponen falsificar o corroborar explícitamente una teoría. Desde 1990, el enfoque dominante ha pasado a ser el "ajuste de modelos" (model fitting): construir un modelo teórico y calibrar sus parámetros para que reproduzca ciertos momentos estadísticos de los datos, en vez de comprobar si sus predicciones se cumplen.

Esta teoría no solo se ha vuelto más prevalente, sino también más compleja computacionalmente: los modelos estáticos de equilibrio parcial o general han cedido terreno casi por completo a los DSGE, y si hace una década predominaba el agente representativo, ya casi un tercio de los artículos DSGE de 2016-2018 incorporan agentes heterogéneos[35] —el germen de los modelos HANK que este tema menciona como respuesta a las limitaciones del enfoque RANK estándar—. Los mercados frictionless son cada vez más raros: casi la mitad de los artículos de 2016-2018 modelizan el sector financiero, un interés que ya venía creciendo desde antes de la Gran Recesión, tras un período de menor atención entre 1980 y 1990. Metodológicamente, las técnicas de microeconometría han desplazado a los métodos de series temporales en paralelo a un giro desde los datos agregados hacia los microdatos y de panel. La inmensa mayoría de los artículos incluyen ya códigos JEL de campos distintos —sobre todo Microeconomía y economía financiera—, evidencia de una fuerte fertilización cruzada con el resto de la disciplina. Por último, los patrones de difusión de las ideas son muy heterogéneos: no siempre una innovación se transmite linealmente desde las revistas generalistas a las especializadas, y los tiempos de adopción varían enormemente de un caso a otro.

Este salto en la complejidad computacional de los modelos no es un simple efecto colateral, sino que se ha convertido en un objeto de estudio y enseñanza por derecho propio. Ya en 1997, Kenneth Judd (1997) sostenía que la computación no sustituye a la teoría económica, sino que la complementa y la extiende, permitiendo explorar modelos analíticamente intratables. Casi tres décadas después, Fernández-Villaverde y Guerrón (2024), en sus lecciones de computación de alto rendimiento (HPC) para economistas, constatan que ese giro se ha acelerado: la resolución de modelos con programación dinámica en varios

[35] Véase el epígrafe de Nueva Economía Keynesiana, subsección "Implicaciones de política económica", donde se desarrollan los modelos HANK frente a RANK.

 © 2026 Manuel Cabado García

estados, los DSGE altamente no lineales con múltiples perturbaciones, las restricciones ocasionalmente vinculantes o la estimación estructural de modelos sin forma cerrada exigen ya, de forma rutinaria, capacidad de cómputo paralelo —múltiples núcleos, GPU o incluso TPU— que hace solo una generación habría resultado impensable para un investigador individual. Los propios autores señalan que esta tendencia seguirá acelerándose por el abaratamiento del hardware, la disponibilidad de grandes bases de datos y el auge del aprendizaje automático, hasta el punto de que una parte sustancial de la carrera de cualquier macroeconomista actual consistirá, de forma creciente, en programar, evaluar código ajeno y supervisar el uso de métodos computacionales —la contrapartida práctica y menos visible del giro hacia los modelos heterogéneos (HANK) y la microeconometría ya documentado por Glandon et al.

Este giro —de la falsación de hipótesis al ajuste de modelos— admite además una lectura más profunda a la luz de la propia epistemología de la ciencia que atraviesa la historia contada en este tema. Thomas Kuhn (1962) sostenía que la ciencia no avanza de forma lineal mediante refutaciones sucesivas, como pretendía Popper, sino alternando periodos de "ciencia normal" —en los que una comunidad comparte un paradigma y resuelve problemas dentro de él— con revoluciones puntuales que lo sustituyen cuando se acumulan demasiadas anomalías. La propia revolución keynesiana de 1936 encaja con ese esquema: respondió a una anomalía que el paradigma neoclásico no lograba explicar —la persistencia del desempleo masivo durante la Gran Depresión—, aunque, a diferencia del guion kuhniano estricto, el paradigma anterior no fue sustituido sino absorbido: neoclásicos y keynesianos terminaron coexistiendo bajo la Síntesis Neoclásica, con el largo plazo reservado a los primeros y el corto plazo a los segundos. Imre Lakatos (1970) refinó después este esquema con el concepto de "programa de investigación": un núcleo teórico que sus defensores tratan como irrefutable, rodeado de un cinturón protector de hipótesis auxiliares que sí pueden revisarse sin abandonar el núcleo. Bajo esta lente, la Nueva Macroeconomía Clásica de Lucas, Kydland y Prescott puede leerse como un programa de investigación cuyo núcleo —la Hipótesis de Expectativas Racionales y la crítica de Lucas— se ha mantenido intacto desde los años 70, mientras que su cinturón protector se ha ido renovando sucesivamente: del modelo de islas de Lucas a los ciclos reales, y de estos al propio armazón DSGE que hoy domina la disciplina. El hallazgo de Glandon et al. de que la investigación macroeconómica ya no busca falsificar teorías, sino ajustar modelos dentro de un marco DSGE compartido, es exactamente lo que este esquema lakatosiano predeciría de un programa de investigación maduro y todavía "progresivo": la comunidad macroeconómica no está ya discutiendo su núcleo, sino perfeccionando su cinturón protector.

Esta transformación metodológica no es exclusiva de la Macroeconomía, sino que se enmarca en una dinámica más amplia de toda la disciplina económica. Galiani, Gálvez y Nachman (2025), mediante procesamiento de lenguaje natural y análisis de citas a gran

escala, muestran que la economía se ha especializado progresivamente por campos, con un declive relativo de los artículos puramente teóricos desde los años 90 en favor de la investigación aplicada y de la "teoría aplicada". Su hallazgo más matizado es que esta especialización no es uniforme: mientras la teoría y los métodos econométricos se especializan cada vez más, la economía aplicada se vuelve, paradójicamente, más multidisciplinar, incorporando herramientas y preguntas de otras ciencias sociales.

La propia crisis financiera de 2008 actuó, además, como un punto de inflexión estructural para la disciplina en su conjunto —no solo para el contenido teórico de la Macroeconomía, como ya se ha visto en este tema con el tránsito del "Consenso Pre-Crisis" a "Replanteando la Política Macroeconómica Post-Crisis", sino para su organización social como comunidad científica—. Baccini, Barabesi y Debernardi (2026), mediante análisis de redes multicapa que combinan relaciones de citación y similitud temática entre revistas, documentan una reconfiguración significativa de las comunidades de investigación y de los patrones de colaboración entre economistas a raíz de la crisis: fragmentación de subcampos previamente unificados y reorientación temática hacia las cuestiones que la Gran Recesión puso de relieve —en términos kuhnianos, una acumulación de anomalías empíricas (la incapacidad de los modelos DSGE estándar para anticipar o explicar la crisis) que forzó una revisión, aunque no un abandono, del programa de investigación dominante—.

Cliometría

Este giro generalizado hacia la cuantificación y la evidencia empírica tiene, de hecho, un antecedente histórico bien anterior a cualquiera de las escuelas tratadas en este tema: la Cliometría, es decir, el estudio de la historia económica con teoría económica y métodos estadísticos, cuyos primeros practicantes —Alfred Conrad y John Meyer— publicaron en 1957 (1957) el trabajo fundacional que aplicaba por primera vez teoría económica y contraste estadístico riguroso a preguntas de historia económica. Su método más característico, popularizado después por Robert Fogel (Nobel de Economía en 1993, ver tabla anterior) en su célebre estudio sobre el impacto del ferrocarril en el crecimiento de Estados Unidos[36], es el análisis contrafactual: construir un modelo formal de lo que habría ocurrido en ausencia de un determinado factor histórico, para poder medir su efecto causal real en lugar de narrarlo cualitativamente.

La Cliometría del siglo XXI ha dado, sin embargo, un vuelco radical respecto a esos primeros debates narrativos de mediados del siglo pasado: su centro de gravedad se ha desplazado hacia el estudio de la persistencia histórica y el desarrollo comparado a muy largo plazo, apoyándose en bases de datos masivas, sistemas de información geográfica y

[36] Fogel, R. W. (1964), *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore, Johns Hopkins University Press. El uso de un modelo contrafactual explícito —estimar el PIB sin ferrocarril, sustituyéndolo por transporte fluvial y canales— generó una fuerte controversia entre los historiadores económicos "tradicionales" y los "nuevos historiadores económicos" (cliómetras), precisamente por el peso metodológico que Fogel otorgaba a la simulación frente a la narrativa histórica.

 © 2026 Manuel Cabado García

variables instrumentales cada vez más sofisticadas para demostrar empíricamente cómo eventos ocurridos hace siglos —a veces hace milenios— siguen determinando la riqueza o la pobreza de regiones enteras en la actualidad. El ejemplo más influyente de esta corriente es el trabajo de Acemoglu, Johnson y Robinson (2001), que instrumentaliza la calidad institucional de las antiguas colonias europeas a través de la tasa de mortalidad de los primeros colonos[37]: donde esa mortalidad era alta —el África subsahariana, buena parte del trópico—, las potencias coloniales se limitaron a construir instituciones extractivas orientadas al mero saqueo de recursos, mientras que donde era baja —Norteamérica, Australia—, se asentaron poblaciones estables que exigieron instituciones que protegiesen la propiedad privada. Los autores muestran que esa diferencia institucional temprana persiste hasta hoy y explica una parte sustancial de la brecha de renta entre países —el mismo Daron Acemoglu, no por casualidad, protagonista ya de este tema como premio Nobel de 2024 por la formación de las instituciones y su efecto sobre la prosperidad (ver "Otros desarrollos" más arriba)—.

Nathan Nunn (2008) aplica una lógica cliométrica equivalente al comercio de esclavos africano, cruzando digitalmente los registros históricos de embarque de los siglos XV al XIX con mapas étnicos del continente: las regiones que sufrieron una mayor extracción de esclavos presentan hoy, de forma estadísticamente significativa, peor desempeño económico y mayor fragmentación étnica. Oded Galor y David Weil (2000), por su parte, ofrecen el correlato teórico-cuantitativo de esta línea de trabajo al modelizar formalmente la transición histórica desde la trampa malthusiana —donde cualquier mejora tecnológica se traducía en más población, no en más renta per cápita— hacia el crecimiento moderno sostenido, sentando las bases de lo que después se conocería como Teoría del Crecimiento Unificado. Y, en un ejemplo más reciente de la creciente sofisticación técnica de la disciplina, Dalgaard, Kaarsen, Olsson y Selaya (2022) recurren a la teledetección satelital para reconstruir la red de calzadas del Imperio Romano del siglo II d.C., mostrando que su densidad histórica predice con notable precisión la infraestructura viaria y el desarrollo económico local de la Europa actual —dos milenios después—. Autores como Stelios Michalopoulos han seguido una senda metodológica similar, combinando mapas satelitales y fronteras étnicas precoloniales para estudiar la persistencia de conflictos y la geografía económica.

En España, uno de los máximos referentes de este programa cliométrico es Leandro Prados de la Escosura (Universidad Carlos III de Madrid), cuyas series históricas de PIB, productividad y capital para España desde 1850 —adoptadas como referencia oficial por el Banco de España— sitúan el crecimiento económico español en perspectiva comparada internacional (2017).

[37] Las series de mortalidad de colonos europeos usadas como instrumento han sido objeto de crítica metodológica posterior. Véase Albouy, D. Y. (2012), "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment", *American Economic Review*, 102(6), pp. 3059-3076, que cuestiona la fiabilidad y homogeneidad de las series de mortalidad originales.

- Junto con Carlos Álvarez-Nogal, extiende esa reconstrucción cuantitativa a los cinco siglos anteriores (1270-1850) y compara directamente la trayectoria española con la de Inglaterra y el resto de Europa occidental, mostrando cómo España parte de una renta per cápita similar en el siglo XIV para acabar muy rezagada hacia 1850 (2013).
- Su trabajo más reciente retoma y actualiza esa comparación centrándose en los términos de intercambio entre España y Gran Bretaña durante la Revolución Industrial, como explicación adicional de la divergencia (2024).
- Más allá del PIB, con el Historical Index of Human Development extiende el mismo programa cliométrico —medir y contrastar en vez de narrar— al estudio del bienestar (salud, educación, nivel de vida) a muy largo plazo para 157 países desde 1870 (2015).

Ese mismo principio inaugurado por Conrad y Meyer en 1957 —medir y contrastar en vez de narrar— es, casi setenta años después, el que atraviesa todas las tendencias descritas en este apéndice: del giro DSGE al auge de los microdatos, de la especialización por campos a la reconfiguración post-crisis, la disciplina económica en su conjunto ha ido convergiendo, con matices e idas y venidas, hacia el programa metodológico que la Cliometría inauguró —y que su propia deriva reciente, hacia la persistencia histórica de muy largo plazo, demuestra que sigue plenamente viva—.

 © 2026 Manuel Cabado García

Bibliografía

Acemoglu, D., S. Johnson y J. A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review 91*(5), 1369–1401.

Akerlof, G. A. y J. L. Yellen (1985). A Near-Rational Model of the Business Cycle, with Wage and Price Inertia. *The Quarterly Journal of Economics 100*(Supplement), 823–838.

Baccini, A., L. Barabesi y C. Debernardi (2026). Exploring the Shape of Economics: A Multilayer Network Analysis of Social Communities and Intellectual Similarity Among Journals Before and After the 2008 Financial Crisis. arXiv:2508.09079v2.

Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy 82*(6), 1095–1117.

Barro, R. J. y D. B. Gordon (1983). Rules, Discretion and Reputation in a Model of Monetary Policy. *Journal of Monetary Economics 12*(1), 101–121.

Baumol, W. J. (1952). The Transactions Demand for Cash: An Inventory Theoretic Approach. *The Quarterly Journal of Economics 66*(4), 545–556.

Blanchard, O., G. Dell'Ariccia y P. Mauro (2010, feb). Rethinking Macroeconomic Policy. IMF Staff Position Note SPN/10/03, International Monetary Fund.

Blanchard, O. y J. Galí (2007). Real Wage Rigidities and the New Keynesian Model. *Journal of Money, Credit and Banking 39*(Supplement 1), 35–65.

Blanchard, O. J. y N. Kiyotaki (1987). Monopolistic Competition and the Effects of Aggregate Demand. *The American Economic Review 77*(4), 647–666.

Calvo, G. A. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics 12*(3), 383–398.

Carlin, W. y D. Soskice (1990). *Macroeconomics and the Wage Bargain: A Modern Approach to Employment, Inflation and the Exchange Rate*. Oxford: Oxford University Press.

Conrad, A. H. y J. R. Meyer (1957). Economic Theory, Statistical Inference, and Economic History. *The Journal of Economic History 17*(4), 524–544.

Dalgaard, C.-J., N. Kaarsen, O. Olsson y P. Selaya (2022). Roman Roads to Prosperity: Persistence and Non-Persistence of Public Infrastructure. *Journal of Comparative Economics 50*(4), 896–916.

de la Escosura, L. P. (2015). World Human Development: 1870–2007. *Review of Income and Wealth 61*(2), 220–247.

de la Escosura, L. P. (2017). *Spanish Economic Growth, 1850–2015*. Cham: Palgrave Macmillan.

de la Escosura, L. P. (2024). The Terms of Trade Between Spain and Britain and the Industrial Revolution. En *A Millennial View of Spain's Development: Essays in Economic History*, pp. 277–321. Cham: Springer / Fundación Rafael del Pino.

Fernández-Villaverde, J. y P. Guerrón (2024). Why High-performance Computing? (Lectures on High-performance Computing for Economists I). Material docente, University of Pennsylvania / Boston College, 27 de agosto de 2024.

Friedman, M. (1953). The Methodology of Positive Economics. En *Essays in Positive Economics*, pp. 3–43. Chicago: University of Chicago Press.

Galiani, S., R. H. Gálvez y I. Nachman (2025). Specialization Trends in Economics Research: A Large-Scale Study Using Natural Language Processing and Citation Analysis. *Economic Inquiry 63*(1), 289–329.

Galor, O. y D. N. Weil (2000). Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. *The American Economic Review 90*(4), 806–828.

Glandon, P. J., K. Kuttner, S. Mazumder y C. Stroup (2022, jan). Macroeconomic Research, Present and Past. NBER Working Paper 29628, National Bureau of Economic Research.

Greenwald, B. C. y J. E. Stiglitz (1988). Examining Alternative Macroeconomic Theories. *Brookings Papers on Economic Activity 19*(1), 207–260.

Hansen, A. H. (1953). *A Guide to Keynes*. New York: McGraw-Hill.

Hicks, J. R. (1937). Mr. Keynes and the Çlassics"; A Suggested Interpretation. *Econometrica 5*(2), 147–159.

Judd, K. L. (1997). Computational Economics and Economic Theory: Substitutes or Complements? *Journal of Economic Dynamics and Control 21*(6), 907–942.

Kahn, R. F. (1931). The Relation of Home Investment to Unemployment. *The Economic Journal 41*(162), 173–198.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. London: Macmillan.

 © 2026 Manuel Cabado García

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kydland, F. E. y E. C. Prescott (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy 85*(3), 473–491.

Kydland, F. E. y E. C. Prescott (1982). Time to Build and Aggregate Fluctuations. *Econometrica 50*(6), 1345–1370.

Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. En I. Lakatos y A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, pp. 91–196. Cambridge: Cambridge University Press.

Layard, R. y S. Nickell (1985). The Causes of British Unemployment. *National Institute Economic Review 111*(1), 62–85.

Mankiw, N. G. (1985). Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly. *The Quarterly Journal of Economics 100*(2), 529–538.

Modigliani, F. (1944). Liquidity Preference and the Theory of Interest and Money. *Econometrica 12*(1), 45–88.

Nunn, N. (2008). The Long-Term Effects of Africa's Slave Trades. *The Quarterly Journal of Economics 123*(1), 139–176.

Samuelson, P. A. y R. M. Solow (1960). Analytical Aspects of Anti-Inflation Policy. *The American Economic Review 50*(2), 177–194.

Sargent, T. J. y N. Wallace (1975). Rational.$^{\text{Ex}}$pectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule. *Journal of Political Economy 83*(2), 241–254.

Sargent, T. J. y N. Wallace (1981). Some Unpleasant Monetarist Arithmetic. *Federal Reserve Bank of Minneapolis Quarterly Review 5*(3), 1–17.

Stiglitz, J. E. y A. Weiss (1981). Credit Rationing in Markets with Imperfect Information. *The American Economic Review 71*(3), 393–410.

Tobin, J. (1956). The Interest-Elasticity of Transactions Demand for Cash. *The Review of Economics and Statistics 38*(3), 241–247.

Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies 25*(2), 65–86.

Whalen, E. L. (1966). A Rationalization of the Precautionary Demand for Cash. *The Quarterly Journal of Economics 80*(2), 314–324.

Álvarez Nogal, C. y L. P. de la Escosura (2013). The Rise and Fall of Spain (1270–1850). *The Economic History Review 66*(1), 1–37.

 © 2026 Manuel Cabado García

Tema 3

Teoría de la demanda del consumidor. Axiomas sobre las preferencias, función de utilidad y función de demanda marshalliana. Otros desarrollos.

3.1. Introducción ⏳ 1'

Los consumidores eligen qué comprar cada día, casi sin pensarlo. Pero, ¿por qué compran lo que compran? En este sentido, la Teoría de la Demanda del consumidor pretende explicar, a través de la función de demanda, el comportamiento del consumidor racional en el mercado, en un contexto competitivo y con información perfecta.

Su objetivo es responder a dos preguntas principales:

1. ¿Qué bienes y qué cantidades adquieren los consumidores?
2. ¿Cómo se determinan sus preferencias y qué factores determinan las variaciones en las decisiones de los mismos cuando cambian los parámetros del problema?

Esta teoría se encuadra dentro de la Microeconomía, que estudia el comportamiento de los distintos agentes (consumidores y empresas) cuya interacción en el mercado determina la asignación de los recursos escasos.

Históricamente, la revolución marginalista y su continuación, la Economía Neoclásica, son las primeras escuelas de pensamiento que tratan de encontrar formas de demanda que resuman el concepto del consumidor aplicando el concepto de utilidad[1] –cuyo primer

[1] La utilidad se define como la capacidad de los bienes para satisfacer necesidades y se implanta en la teoría económica a partir de la primera Revolución Marginalista.

antecedente riguroso se remonta a Bernoulli[2]–, acuñado posteriormente por autores como Bentham y Gossen[3].

Las primeras aportaciones proceden del enfoque cardinal de Marshall que, por apoyarse en supuestos excesivamente restrictivos, dará paso al enfoque ordinal, donde la utilidad es medida ordinalmente mediante la introducción del concepto de preferencias a partir de las aportaciones de Pareto, Hicks, Slutsky y Fisher.

Por consiguiente, el objetivo de esta exposición será el análisis de la teoría neoclásica de la demanda del consumidor[4] basada en el enfoque ordinal a través del siguiente esquema:

1. Enfoque cardinal de la teoría neoclásica.
2. Enfoque ordinal de la teoría neoclásica.
3. Estática comparativa, para analizar cómo varía la demanda al variar precios y renta
4. Condiciones de un sistema de demanda completo.
5. Demanda de mercado.

3.2. Enfoque cardinal de la teoría neoclásica ⏳ 1'

Antes de que se impusiera el enfoque ordinal, Marshall desarrolló un enfoque cardinal de la demanda, asentado en una función de utilidad medible, aditiva, separable y cuasilineal respecto al dinero:

$$U(x_1, x_2, M) = u(x_1) + u(x_2) + u(M) \tag{3.1}$$

[2]En 1738, Daniel Bernoulli recurrió por primera vez de forma rigurosa a la utilidad marginal decreciente para resolver la paradoja de San Petersburgo, un juego con valor esperado monetario infinito por el que, sin embargo, ningún individuo estaría dispuesto a pagar una cantidad elevada. Bernoulli mostró que si la utilidad del dinero es cóncava –marginalmente decreciente–, el valor esperado de la utilidad del juego es finito, anticipando en más de un siglo el concepto de utilidad marginal que formalizaría la Revolución Marginalista.

[3]En 1854, Gossen formuló sus tres leyes, que han tenido una gran relevancia en el desarrollo de la Teoría de Demanda:

1. Ley de saturación de necesidades: "la utilidad de un bien disminuye conforme aumenta la cantidad disponible de este bien, hasta llegar a la saciedad". Es decir, Umg decreciente.
2. Ley de igualdad de las UMg ponderadas: una persona distribuye su renta entre los bienes de modo que la última unidad monetaria gastada en cada uno le reporte la misma utilidad marginal. Es decir, en el óptimo, las UMg ponderadas por su precio se igualan.
3. Ley de la escasez como condición previa al valor económico, lo que implica la existencia de un problema económico y, por tanto, la posibilidad de optimización de los recursos.

[4]El análisis de la Teoría NC de la Demanda se basa en:

- El individualismo metodológico, es decir, en el examen del comportamiento del individuo (consumidor).
- Un concepto subjetivo de utilidad.
- Racionalidad: dados precios y renta, el individuo planifica su gasto de manera que haga máxima su satisfacción (autores como Kahneman y Tversky o Richard Thaler considerarán irracionalidad de las decisiones). - Información completa: contexto de certidumbre.

 © 2026 Manuel Cabado García

donde la utilidad total del consumidor es simplemente la suma de la utilidad que reporta cada bien y la del dinero no gastado, sin que exista interacción alguna entre ellas –de ahí que se diga aditiva y separable–. Bajo el supuesto de que la utilidad marginal del dinero es constante, resolviendo el problema de maximización se obtiene que el consumidor iguala la utilidad marginal de cada bien a su precio,$u'(x_i) = p_i$, de modo que la demanda de cada bien depende únicamente de su propio precio.

Este enfoque fue superado por resultar excesivamente restrictivo:

1. Considera los bienes independientes entre sí, impidiendo ver su grado de complementariedad o sustituibilidad.
2. La utilidad marginal del dinero constante es poco realista[5] y, además, elimina el efecto renta y el efecto sustitución cruzado.
3. Exige un conocimiento excesivo por parte de los consumidores.
4. Sobre todo, este enfoque utilitarista añade a la cardinalidad –que solo permite comparar diferencias de utilidad para un mismo individuo– el supuesto, distinto y adicional, de comparabilidad interpersonal de utilidades: no basta con que cada individuo pueda ordenar sus propias cestas, sino que haría falta poder afirmar que un mismo euro reporta la misma satisfacción a dos personas distintas, algo que no es observable ni verificable empíricamente y que introduce de matute juicios de valor en lo que debería ser un análisis positivo.

El enfoque ordinal prescindirá por completo de esa exigencia: bastará con que cada individuo sea capaz de ordenar sus propias preferencias, sin necesidad de medir ni comparar la satisfacción entre distintos consumidores –es precisamente ese punto de partida, mucho menos restrictivo, el que se desarrolla a continuación–.

3.3. Enfoque ordinal de la teoría neoclásica ⏳ 9'

El problema ordinal de teoría de demanda del consumidor se divide en tres partes principalmente:

1. Determinar el conjunto de consumo.
2. Determinación de un criterio para evaluar los planes de consumo: preferencias, para una posterior construcción de una Función de Utilidad que se analizará a partir de curvas de indiferencia.
3. Determinar los planes de consumo a los que puede aspirar un consumidor dada su restricción presupuestaria, completando así el problema de demanda del consumidor.

[5] Si la utilidad marginal del dinero es constante, la última unidad monetaria reporta la misma utilidad independientemente de la cantidad que se posea. Por ejemplo, no es lo mismo recibir una unidad monetaria con una renta de 10 que recibir dicha unidad con una renta de 10000.

Dado un conjunto de cestas factibles y unos recursos, el consumidor elegirá aquella cesta que mejor se acomode a sus necesidades.

Esta teoría parte de los siguientes supuestos:

1. Racionalidad del consumidor ya que el consumidor maximiza su utilidad.
2. Utilidad ordinal, el consumidor puede ordenar las cestas en función del beneficio que le reportan.
3. Solo ordena las preferencias, nunca mide la utilidad.
4. Su utilidad depende de la cantidad consumida de cada mercancía.
5. Además de información perfecta, contexto estático, la decisión de consumo se toma de manera aislada de otras decisiones como ocio o trabajo, se consideran bienes de consumo final.

Conjunto de elección

El enfoque ordinal de la utilidad basa esa comparación en el concepto de preferencias. Solo se exige que los individuos puedan ordenar las distintas cestas de bienes de un conjunto de elección, X, que se supone:

- No vacío: Tiene que consumir algo
- Acotado inferiormente: por la condición de no negatividad del consumo.
- Cerrado inferiormente: los límites pertenecen al conjunto (que el conjunto sea cerrado implica que los límites se incluyen en el mismo. El límite inferior se incluye en el conjunto, el superior no se considera hasta que se introduce la restricción presupuestaria).
- Convexo[6]. Cualquier combinación lineal de cestas que se encuentren en el conjunto también pertenecerá al conjunto.

Además, se supone que la utilidad total depende de las cantidades consumidas:

$$u = f(x_1, \ldots, x_n) \tag{3.2}$$

Preferencias

Para poder establecer una adecuación dentro del conjunto de elección, se introduce la relación binaria "es al menos tan bueno como", $\succsim$. Para limitar la discrecionalidad de

[6]Un conjunto se define como convexo si, para cada par de sus puntos, cada punto sobre el segmento que los une está contenido en el conjunto. Un conjunto es convexo si se puede ir de un punto a otro en línea recta sin salir del mismo. Si dos planes de consumo son posibles, también lo será cualquier plan formado a partir del promedio ponderado. Este supuesto implica perfecta divisibilidad de los bienes.

Así, un conjunto es convexo cuando cualquier combinación lineal de dos puntos pertenecientes al mismo está también incluida en el conjunto.

 © 2026 Manuel Cabado García

las preferencias[7] y construir una función de utilidad se introducen una serie de axiomas respecto al consumidor:

- **Axioma 1** Completitud, el consumidor es capaz de comparar cualquier par de cestas:[8] para todo $x, y \in X$ se tiene que $x \succsim y$ o $y \succsim x$ o ambas.
- **Axioma 2** Transitividad, para todo $x, y, z \in X$, si $x \succsim y$ e $y \succsim z$ entonces $x \succsim z$, es decir, las preferencias son consistentes
- **Axioma 3** Reflexividad, una cesta es al menos tan buena como ella misma, $x \succsim x$ para todo $x \in X$. Este axioma permite que ninguna cesta quede fuera del orden del consumidor

Estos tres axiomas conjuntamente aseguran que el comportamiento del consumidor es racional –son los axiomas de racionalidad– y garantizan un preorden completo, es decir, las diferentes cestas admiten relaciones de preferencia o indiferencia, aunque no garantizan la existencia de una función de utilidad. Esto es, son condición necesaria pero no suficiente para la representación mediante una función de utilidad como, por ejemplo, el orden lexicográfico[9].

Por ello será necesario incluir los siguientes axiomas de regularidad:

- **Axioma 4** Continuidad, no existen saltos en las preferencias. Evita la posibilidad de comportamientos extremistas: garantiza, junto con el pre-orden completo, que las preferencias admitan alguna función de utilidad continua (Teorema de Debreu)[10]. Matemáticamente, si se prefiere x a y sobre una curva de preferencias, las cantidades similares a x serán también preferidas sobre y, lo que se expresa formalmente como sigue:$\forall x, y, z \in X$ si $x \succsim y \succsim z$ entonces $\exists \lambda \in [0, 1]$ tal que $y \sim \lambda \cdot x + (1 - \lambda)\, z$.

Además, los siguientes axiomas de deseabilidad limitarán la forma funcional (para darle una forma concreta, deseable a la función)

[7] Con limitar la discrecionalidad de las preferencias se hace referencia a no escoger una muestra demasiado amplia sino limitada para que sea operativa y de buen comportamiento

[8] La intuición es que el consumidor nunca se queda sin saber qué prefiere: ante dos cestas cualesquiera, siempre puede decir cuál le gusta más o que le son indiferentes. Si se cumplen las dos relaciones a la vez, es indiferente entre ambas.

[9] El orden lexicográfico compara los componentes de un conjunto uno a uno y ordena los elementos de acuerdo al primer elemento diferente que se encuentra. No es posible dibujar curvas de indiferencia. Proporciona un preorden completo, en el que además no hay dos cestas distintas indiferentes entre sí. Cumple con completitud, reflexividad, transitividad, y monotonía, pero no se puede construir ninguna función de utilidad, porque es necesaria la propiedad de continuidad.

Para saber más, acudir al anexo.

[10] La continuidad descarta preferencias como las lexicográficas (Apéndice, 3.A.3), que cumplen completitud, reflexividad y transitividad pero no continuidad, y por ello no admiten ninguna función de utilidad. No debe confundirse con la diferenciabilidad (Axioma 7): la continuidad garantiza que exista una función de utilidad, no que esta sea derivable en todo punto –una función tipo Leontief, por ejemplo, es continua pero presenta un vértice donde la RMS no está definida.

- **Axioma 5** Convexidad estricta, hace referencia a la preferencia por la variedad, por lo que los consumidores no se van a los extremos. A medida que disminuye el consumo de un bien va siendo más difícil sustituirlo.
- **Axioma 6** Monotonía, los individuos prefieren más a menos ("cuanto más mejor") en cualquier dirección y, como los individuos no se sacian, la función de utilidad será creciente[11].
- **Axioma 7** Diferenciabilidad, la pendiente de la curva de indiferencia es única en todos sus puntos. Las curvas de indiferencia no tienen puntos angulares o "picos" donde no se pueda derivar. Es decir, admite derivadas de primer orden en cualquier dirección, aportando así la suavidad que la mera continuidad no garantiza[12].

Estas preferencias son de buen comportamiento[13] y por lo tanto la función de utilidad también lo será.

Teorema de Debreu Si las preferencias de un individuo cumplen los axiomas de completitud, transitividad, reflexividad y continuidad, existe una función de utilidad continua que representa las preferencias. La monotonía no es necesaria para esta existencia –es una propiedad adicional que garantiza que dicha función sea creciente–, pero sí lo es la continuidad, que descarta representaciones como la de las preferencias lexicográficas.

Además, para evitar en el futuro soluciones de esquina, se añade el supuesto de que la función (preferencias) sea asintótica a los ejes.

La función de utilidad

Por consiguiente, los siete axiomas anteriores garantizan la representación de las preferencias a través de una función de utilidad, $u(x)$, que relaciona cestas de bienes con un índice de utilidad

$$u : X \to \mathbb{R} / x \succsim y \Leftrightarrow u(x) \geq u(y) \tag{3.3}$$

[14]

[11] La monotonía es la versión fuerte de un supuesto más general de deseabilidad, cuya versión más débil es la no saciedad local (véase la subsección "Curvas de indiferencia"): esta solo exige que, en cualquier punto, exista alguna cesta cercana estrictamente mejor –no que "más de todo" sea siempre mejor–. La monotonía implica no saciedad local, pero no al revés.

[12] A diferencia de la continuidad, que solo exige ausencia de saltos, la diferenciabilidad exige además ausencia de vértices: una función continua puede tener puntos angulares –como la tipo Leontief– donde la pendiente cambia bruscamente, mientras que una función diferenciable tiene pendiente única y bien definida en todo punto, lo que permite aplicar el cálculo diferencial (utilidades marginales, RMS, condiciones de primer orden) sin excepciones.

[13] En matemáticas, el "buen comportamiento" quiere decir también que no se está violando ninguna premisa necesaria para la aplicación satisfactoria de cualquiera que sea el análisis que esté en discusión.

[14] u es una función del conjunto de bienes, X, para los reales (números reales $\mathbb{R}$) tal que, /, el bien x es al menos tan bueno como el bien y, $x \succsim y$, si y solo si, $\Leftrightarrow$, esto tiene una traslación en términos de utilidad, $u(x) \geq u(y)$.

 © 2026 Manuel Cabado García

Propiedades A su vez, esta función verifica las siguientes propiedades:

1. Ordenación[15], de los vectores de consumo (Axiomas de pre-orden completo 1,2, y 3), mediante una representación numérica de la ordenación de preferencias individual de forma que las que se prefieran tengan un número más alto.
2. Continua, por serlo las preferencias Axioma 4.
3. Dos veces diferenciable –un supuesto adicional al Axioma 7, que por sí solo solo garantiza diferenciabilidad de primer orden–, lo que facilita la obtención de la cesta óptima de los agentes y permite aplicar las condiciones de segundo orden, realizar la estática comparativa de la demanda mediante el teorema de la función implícita (de donde se deriva la ecuación de Slutsky), comprobar formalmente la RMS decreciente y calcular elasticidades de sustitución.
4. Monótona y creciente, por la monotonía de las preferencias (Axioma 6). La derivada de la función respecto a los bienes es positiva siempre, lo cual implica que no hay males.
5. Estrictamente cuasicóncava[16], por la convexidad de las preferencias Axioma 5. Garantiza el cumplimiento de las condiciones de segundo orden en el problema de maximización de utilidad.

Función de utilidad en un plano de tres dimensiones Gráficamente, en un plano de utilidad, u, bien x y bien y como el de la Figura 3.1 se representa la función de utilidad para las combinaciones de consumo de ambos bienes lo que da lugar a la región de dicha función de utilidad.

Curvas de indiferencia

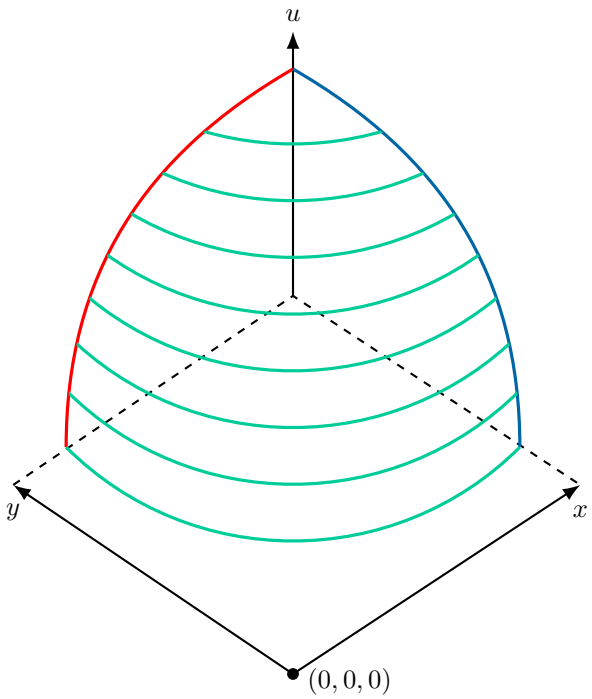
Sin embargo, para simplificar la representación gráfica de la función de utilidad en el caso de cestas con dos bienes, se puede realizar en el plano cortes horizontales a la función de utilidad para obtener las curvas de nivel, como se observa en la Figura 3.2, en un plano

[15] Aclaración entre ordenación e ITM+, concepto que suele aparecer:

- La ordenación es la idea principal: la función de utilidad solo indica si una cesta de bienes es mejor, peor o igual que otra. Por ejemplo, "me gusta más la pizza que la hamburguesa".
- La propiedad de independencia de transformaciones monótonas positivas es la confirmación matemática de esto. Significa que los números que se asignan a esas preferencias no importan, solo su orden. Si le asigno un valor de 10 a la pizza y 5 a la hamburguesa, el hecho de que la pizza me guste más no cambia si en su lugar les asigno valores de 100 y 50. Lo que importa es que 100 es mayor que 50.

[16] Es decir, que dada una cesta que reporta mayor utilidad que otra, cualquier combinación lineal de ambas reportará más utilidad que la que menos utilidad aporte. Esta propiedad permite establecer que la maximización de U(x) es equivalente a elegir la cesta que reporta más utilidad. Es cuasicóncava y no cóncava porque las transformaciones monótonas positivas preservan la ordenación de preferencias pero alteran la forma.

Figura 3.1: Función de utilidad en 3 dimensiones



de x e y. Estas curvas de nivel son las denominadas curvas de indiferencia, introducidas por Edgeworth (1881) y desarrolladas por Pareto (1906) y Hicks .

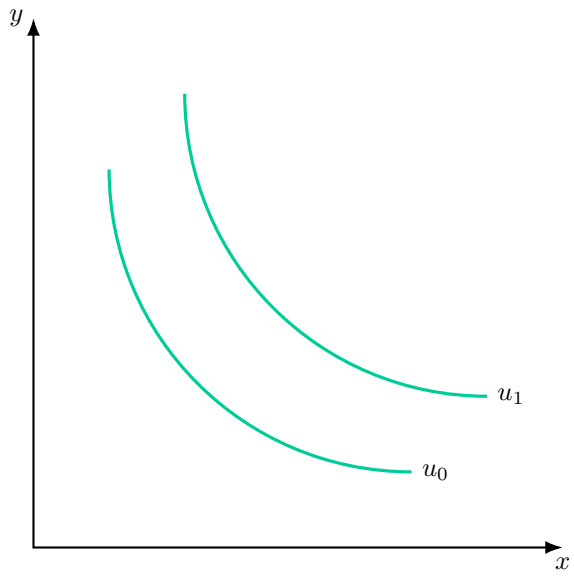
Curvas de indiferencia Las curvas de indiferencia se definen como el lugar geométrico de combinaciones de bienes –cestas– que proporcionan el mismo nivel de utilidad –son indiferentes entre sí–. Cada curva de indiferencia tiene un nivel de utilidad asociado: $u(x, y) = \bar{u}$.

Propiedades A su vez, las curvas de indiferencia tienen las siguientes propiedades:

1. Ubicuidad: existen infinitas curvas de indiferencia, lo cual implica que por cualquier espacio del plano pasa una curva de indiferencia.
2. Las curvas más alejadas del origen significan una mayor utilidad
3. No intersección: Las curvas de un mismo individuo nunca se cruzan, por la transitividad de las preferencias. Si lo hicieran, una misma cesta tendría dos niveles de utilidad, lo cual sería incoherente.
4. Continuidad, las curvas de indiferencia son continuas por la continuidad de la función de utilidad, reflejando que pequeñas variaciones en bienes no alteran bruscamente las preferencias.
5. No existe saciedad local –versión más débil de la monotonía del Axioma 6–, es decir, alrededor de una combinación de bienes siempre hay otra combinación de bienes que

 © 2026 Manuel Cabado García

Figura 3.2: Mapa de curvas de indiferencia



reporta más satisfacción.

6. Convexidad, son convexas respecto al origen[17]. La pendiente de la curva de indiferencia disminuye a medida que se desciende a lo largo de la curva. La forma viene determinada por la convexidad de las preferencias, y provoca que la función de utilidad sea cuasi-cóncava.
   Implica el Equilibrio en el consumo y que las combinaciones lineales son preferidas. Los agentes preferirán una cesta de mercado equilibrada a otra que contenga mucho de un bien y muy poco del otro.
7. Pendiente negativa (por la propiedad de monotonía[18]) y decreciente viene representada por la relación marginal de sustitución, $RMS_{x,y}$. Así, para que un consumidor se desprenda de una cantidad de un bien debe ser compensado con más cantidad del otro.

Relación Marginal de Sustitución, RMS indica la tasa a la que el consumidor está dispuesto a sustituir un bien por otro manteniendo constante su nivel de utilidad[19]

$$RMS_{x,y} = -\left.\frac{\partial y}{\partial x}\right|_{\bar{u}} = \frac{UMg_x}{UMg_y} \tag{3.4}$$

[17]Las combinaciones que son preferidas o indiferentes forman parte de un conjunto convexo

[18]Por la monotonía se evita que haya males y que el individuo llegue a saturarse de algún bien.

[19]La intuición es que la RMS mide cuánto de un bien está dispuesto a entregar el consumidor a cambio de una unidad más del otro sin quedar ni mejor ni peor. Es una valoración subjetiva, a diferencia de la relación de precios, que es la tasa a la que el mercado permite intercambiar los bienes. Por ejemplo, si el consumidor está dispuesto a renunciar a dos cafés por un té más, su RMS entre té y café es dos.

siendo el grado de convexidad de la curva de indiferencia un indicador de la sustituibilidad entre los bienes, que se mide con la elasticidad de sustitución, σ, es decir, la variación porcentual de la proporción en que se consumen los dos bienes ante una variación porcentual de la RMS[20]

$$\sigma = \frac{\partial (y/x)}{\partial RMS_{x,y}} \cdot \frac{RMS_{x,y}}{y/x} \tag{3.5}$$

obteniendo diferentes curvas de indiferencia[21] como se observa en la Figura 3.3, de esta forma:

- Para bienes perfectamente sustitutivos la RMS es constante. Matemáticamente, $\sigma = \infty$[22] y la curva de indiferencia sería lineal. Ejemplos de este tipo de bienes son pepsi y coca-cola, café y té o azúcar y sacarina.
- Para bienes complementarios perfectos –aquellos en los que el aumento en la cantidad consumida de un bien solo aumenta la utilidad si se acompaña del aumento en la cantidad consumida del otro– la RMS no está definida[23], dando lugar a una función de utilidad tipo Leontief en forma de L con $\sigma = 0$. Ejemplos de bienes son zapatos del pie izquierdo y zapatos del pie derecho, brocas y taladros, raquetas y pelotas de tenis o tornillos y tacos.
- Entre estos dos casos extremos se encontrarían las curvas de indiferencia con RMS decreciente, es decir, bienes con sustituibilidad limitada. Un ejemplo sería la función de utilidad tipo Cobb-Douglas, con $\sigma = 1$.

[20] La intuición es que σ mide lo fácil que resulta sustituir un bien por otro. Si es alta, un pequeño cambio en la valoración relativa de los bienes lleva a cambiar mucho la proporción en que se consumen, y en el caso extremo de los sustitutivos perfectos σ es infinita. Si es nula, la proporción no cambia en absoluto, como ocurre con los complementarios perfectos.

[21] Otros casos (levantamiento de supuestos)

- Bienes neutrales: Su consumo le es indiferente al individuo. Por ello las curvas de indiferencia son rectas verticales u horizontales según qué bien sea neutro.
- Males: Las curvas de indiferencia tendrán pendiente positiva y serán crecientes. La utilidad crecerá hacia el eje de abscisas u ordenadas dependiendo de qué bien sea el mal.
- Curvas de indiferencia cóncavas: La RMS es creciente al aumentar X (dados los bienes X e Y). Estas curvas de indiferencia no implican que los bienes sean males, únicamente rompen el supuesto de convexidad, es decir la preferencia por la variedad. Para saber analíticamente si un bien es un mal, se calcula la UMg (primera derivada) del bien y si es negativa es un mal.

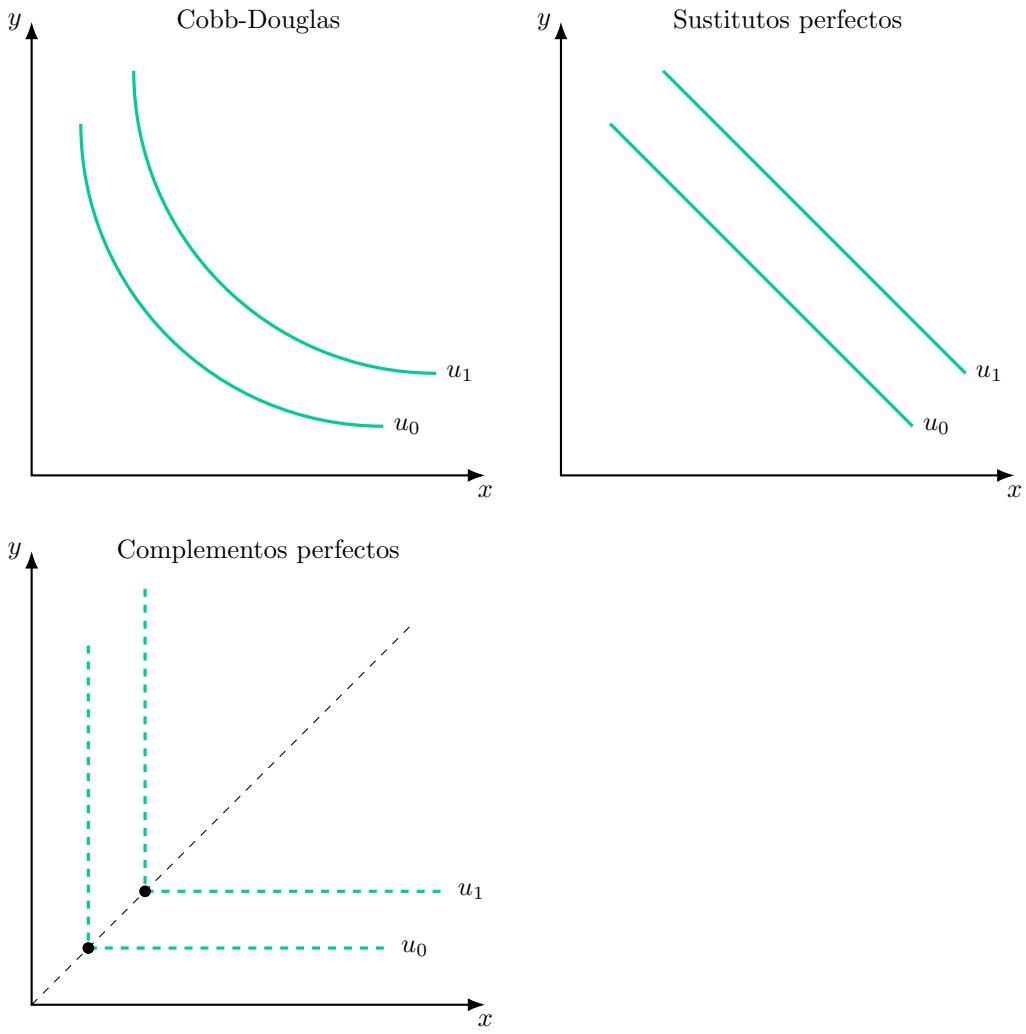
[22] Esto se debe a que al ser la RMS constante el valor de $\partial RMS_{x,y}$ en la ecuación 3.5 es 0, por lo tanto el valor de σ es ∞.

[23] En el segmento horizontal, la RMS es cero, ya que no hay disposición a sustituir un bien por el otro. En el segmento vertical, la RMS es infinita. Finalmente, en el vértice —el punto clave de consumo óptimo— la RMS es indefinida, dado el quiebre de la curva. Esto subraya la imposibilidad de sustitución entre los bienes, ya que solo aumentan la utilidad en una proporción fija.

Así, en el caso de una función de utilidad tipo Leontief, el consumo solo tiene sentido en su vértice.

 © 2026 Manuel Cabado García

Figura 3.3: Tipos de funciones de utilidad



Restricción presupuestaria

Por otro lado, el consumidor se enfrenta al problema de la escasez que se puede abordar desde dos puntos de vista:

1. Restricción física del gasto de elección.
2. Restricción económica de la renta.

Se analizará la restricción de renta suponiendo mercados completos –los precios de los bienes son positivos y conocidos por todos los consumidores – y agentes precio aceptantes, por lo que la restricción presupuestaria para dos bienes a una renta, m, dada será

$$p_x \cdot x + p_y \cdot y \leq m \tag{3.6}$$

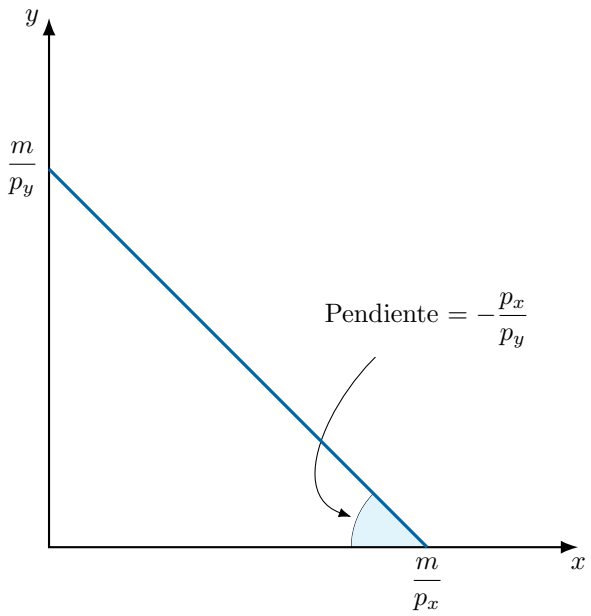
donde p_x y p_y son los precios de los bienes x e y respectivamente.

Propiedades A su vez, el conjunto presupuestario posee las siguientes propiedades:

1. Acotado, por arriba por la restricción presupuestaria y por abajo por la no negatividad del consumo de los bienes.
2. Convexo, cualquier combinación convexa de dos puntos pertenecientes al conjunto pertenece también al mismo.
3. Cerrado, cualquier combinación de bienes sobre la recta o sobre los ejes es asequible, incluye extremos.
4. No vacío, con una renta positiva y al menos el precio de un bien finito, el consumidor podrá consumir una cantidad positiva de al menos un bien.

Recta de balance Gráficamente, en un plano bien x y bien y como el de la Figura 3.4 se representa la recta de balance cuya pendiente viene determinada por la ratio de precios, $-\frac{p_x}{p_y}$, y representa el coste de oportunidad de un bien en términos del otro[24], la cual refleja la tasa de cambio entre los dos bienes en el mercado, es decir, tasa objetiva[25].

Figura 3.4: Restricción presupuestaria



[24]En los extremos la cantidad consumida del otro bien es cero. En el caso de $\frac{m}{p_x}$, viene de que el consumo de y es 0, por lo que $m = p_x \cdot x$ entonces $x = \frac{m}{p_x}$ y lo mismo aplica en el extremo de solo consumir el bien y.

[25]Coste de oportunidad al reducir la cantidad de y que puede obtenerse al reducir la cantidad de x en una unidad y suponiendo que los precios están dados. Este coste de oportunidad será constante.

 © 2026 Manuel Cabado García

Equilibrio del consumidor

Por consiguiente, el problema de maximización al que se enfrenta el consumidor es el siguiente

$$\text{máx} \quad u(x,y) \tag{3.7}$$

sujeto a

$$p_x \cdot x + p_y \cdot y \leq m \tag{3.8}$$

El problema se puede resolver a través del método de Lagrange, obteniéndose la siguiente condición de primer orden, CPO[26][27]:

$$\frac{UMg_{x_i}}{p_i} = \lambda, \quad \forall i \tag{3.9}$$

o lo que es lo mismo

$$RMS_{x,y} = \frac{UMg_x}{UMg_y} = \frac{p_x}{p_y} \tag{3.10}$$

la cual implica que la relación marginal de sustitución se iguale a la ratio de precios, es decir, la tasa subjetiva a la que el consumidor está dispuesto a sustituir un bien por otro es igual al coste de oportunidad de un bien en términos de otro (tasa objetiva del mercado)[28][29].

[26] Debido a la no saciedad local, el símbolo de desigualdad, $\leq$, de la restricción presupuestaria pasa a ser una igualdad, $=$, a la hora de resolver el Lagrangiano.

Así, la desigualdad de la restricción ($\leq$) se convierte en una igualdad ($=$) porque, por el principio de no saciedad, el consumidor siempre gasta todo su ingreso para obtener la máxima satisfacción. La solución a este problema, usando el método de Lagrange, proporciona la condición clave de equilibrio: la Relación Marginal de Sustitución (RMS) debe ser igual a la relación de precios. Esto significa que la valoración subjetiva del consumidor se alinea perfectamente con la valoración objetiva del mercado.

[27] La condición que proporciona la solución (elección del consumidor) viene dada por la igualación entre las UMg ponderadas por el precio de cada bien y la igualación de todas ellas al multiplicador de Lagrange. El consumidor distribuye toda su renta entre los diferentes bienes hasta que las UMg ponderadas por el precio se igualan (Segunda Ley de Gossen).

Siendo λ el multiplicador de Lagrange (precio sombra de la restricción) que, en este caso, representa la UMg de la renta, es decir, lo que aumenta la utilidad máxima al aumentar la renta en una unidad. A diferencia del enfoque cardinal, λ no es constante: depende de los precios y de la renta.

[28] Curiosidad:

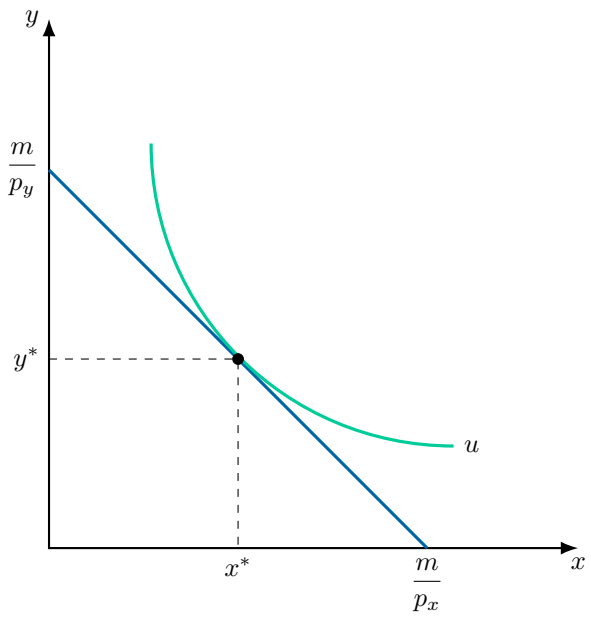
- Si $RMS > P_1/P_2$, gráficamente la pendiente de la curva de indiferencia es superior a la de la recta de balance,la valoración subjetiva del consumidor del bien 1 (RMS) será superior a la valoración objetiva (precio relativo) de dicho bien. Por tanto, el consumidor aumentará su utilidad sustituyendo unidades del bien 2 por unidades del bien 1.
- Si $RMS < P_1/P_2$, ocurrirá lo contrario, donde la pendiente de la curva de indiferencia es inferior a la de la recta de balance

El consumidor continuará reajustando su consumo hasta que su RMS sea exactamente igual al precio relativo, $RMS = P_1/P_2$, momento en el que alcanza su punto de consumo óptimo y no tiene incentivo para realizar más sustituciones. Gráficamente, la pendiente de la curva de indiferencia y la pendiente de la recta de balance serán iguales.

[29] El coste de oportunidad y el precio relativo (valoración objetiva) pueden divergir cuando no se cumplen las condiciones ideales del mercado. Esto sucede, por ejemplo:

Equilibrio Gráficamente, en un plano de bien x y bien y como el de la Figura 3.5, se determina así un punto de equilibrio que viene dado por el punto de tangencia entre la curva de indiferencia y la restricción presupuestaria, es decir, donde se igualan sus pendientes.

Figura 3.5: Equilibrio del consumidor



Demanda ordinaria o marshalliana A partir de la condición de equilibrio es posible obtener las funciones de demanda del consumidor para cada uno de los bienes que dependen de los precios y de la renta: $x = x(p, m)$. Estas funciones de demanda, obtenidas del problema de maximización de la utilidad, se denominan ordinarias o marshallianas, $x(p, m)$ y tienen las siguientes propiedades:

1. Es continua para todo vector de precios y renta positivos.
2. Habitualmente decreciente respecto al precio propio y creciente respecto a la renta –para bienes normales sin efecto Giffen–, aunque esta propiedad no es universal: el apartado de Estática comparativa desarrolla las excepciones (bienes inferiores, bienes Giffen).

- En mercados imperfectos: Si hay monopolios u oligopolios, los precios no reflejan el coste de producción, alterando el verdadero coste de oportunidad.
- En mercados con bienes no transables: Si un bien no se puede comprar o vender, su coste de oportunidad es subjetivo y no está dictado por un precio de mercado.
- Con información asimétrica: Si los precios no reflejan toda la información disponible, el coste de oportunidad percibido por un consumidor puede diferir del coste real.

En la teoría neoclásica, sin embargo, se asume un mercado ideal donde el precio relativo es el coste de oportunidad, lo que simplifica el modelo para entender la elección racional del consumidor.

 © 2026 Manuel Cabado García

3. Homogénea de grado cero en precios y renta (el consumidor se guía por los precios relativos. Implica ausencia de ilusión monetaria, lo que permite prescindir de la referencia del dinero)
4. Cumple con la ley de Walras: gasta toda su renta.
5. Diferenciable, con carácter general, lo que permite aplicar el análisis de estática comparativa que se desarrolla a continuación.
6. Única si las curvas de indiferencia son estrictamente convexas.

A partir de la agregación horizontal de las demandas individuales –sumando, a cada precio, la cantidad demandada por cada consumidor dada su renta– siempre es posible obtener la curva de demanda de mercado, sin más condición que la de que los agentes sean precio aceptantes y sus demandas no estén interrelacionadas. Ahora bien, esa demanda agregada dependerá en general de cómo esté repartida la renta entre los consumidores. Solo si, además, la demanda depende de la renta total y no de su distribución (condición de Gorman[30]) podrá expresarse la demanda de mercado como función únicamente de la renta agregada, con independencia del reparto.

3.4. Estática comparativa ⏳ 6'

Una vez identificada la demanda marshalliana, cabe preguntarse cómo se ve afectada ésta por cambios en las variables que la determinan: renta, precio propio y precios de otros bienes. Es decir, una vez respondida la pregunta de ¿qué bienes y qué cantidades consumirá el consumidor?, se trata de responder a la pregunta de ¿cómo varía la decisión adoptada por el consumidor cuando varían los parámetros del problema?[31]

Variaciones en los precios

Comenzando con variaciones en los precios, gráficamente, en un plano de bien x y bien y como el de la Figura 3.6

- Se representa la restricción presupuestaria y la curva de indiferencia tangente a la misma dando lugar al equilibrio en A.
- Se supone una disminución del precio del bien x, por lo que la recta presupuestaria pivota sobre el eje y expandiendo el conjunto presupuestario y permitiendo al con-

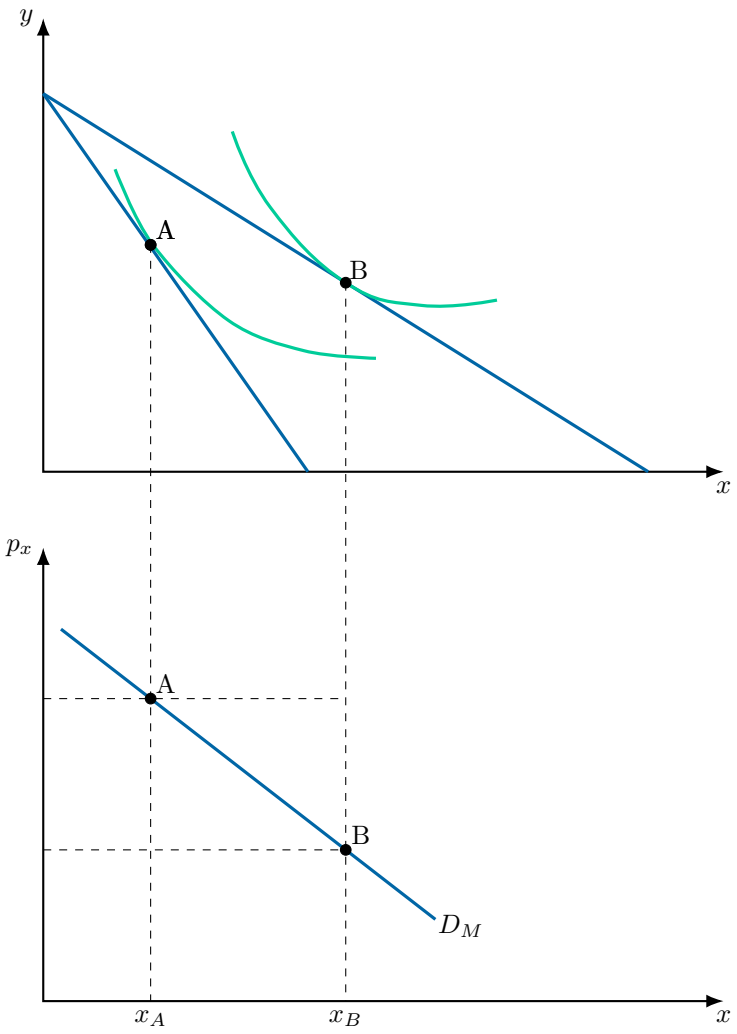
[30] Gorman, W. M. (1953), "Community Preference Fields",*Econometrica*, 21(1), pp. 63-80. También conocida como forma polar de Gorman (*Gorman polar form*), desarrollada en Gorman (1961),*Metroeconomica*.

[31] El objetivo será determinar si la teoría neoclásica de la demanda permite identificar regularidades robustas (que se cumplan bajo cualquier circunstancia: "leyes de demanda") y poder así derivar la curva de demanda. Desgraciadamente, se verá que la teoría neoclásica de la demanda no ofrece muchas regularidades robustas, pero sí algunas, que se comentan a continuación.

sumidor alcanzar una curva de indiferencia más alejada del origen que le reportará más utilidad.

- En el nuevo equilibrio B el consumidor adquiere una mayor cantidad de ambos bienes. Esto implicaría que, en principio, parece que se cumple la condición de negatividad (que se verá más adelante), y en un plano adyacente de precio de bien x, p_x, y cantidad de x, la demanda tiene pendiente negativa, es decir, al disminuir el precio de un bien, aumenta su consumo.

Figura 3.6: Obtención de la curva de demanda



No obstante, es necesario analizar con detalle la variación del precio, ya que la variación de la cantidad demandada ante un cambio en los precios es el resultado de la interacción de dos efectos:

- Efecto sustitución: es el cambio en la cantidad demandada que se produce porque el bien se ha vuelto más barato o más caro en comparación con los demás, dejando

 © 2026 Manuel Cabado García

al margen que el consumidor sea ahora más rico o más pobre, es decir, manteniendo constante su renta real.[32]

- Efecto renta: es el cambio en la cantidad demandada que se produce porque, con la misma renta monetaria, el consumidor puede comprar más si el precio baja, o menos si sube, es decir, porque ha variado su renta real.[33]

ES y ER: visiones de Slutsky y Hicks A su vez, para analizar el efecto sustitución es necesario definir el concepto de renta real constante, existiendo dos criterios de compensación que reparten el mismo efecto total en un punto intermedio distinto:

- Slutsky, ante una variación de precios, la renta real constante es aquella que permite adquirir la cesta anterior a nuevos precios, es decir, el consumidor se sitúa sobre la misma restricción presupuestaria.
- Hicks, ante una variación de precios, la renta real constante es aquella que permite alcanzar la misma utilidad anterior a nuevos precios, es decir, el consumidor se sitúa sobre la misma curva de indiferencia.

Gráficamente, como se observa en la Figura 3.7, ante una reducción del precio de x el efecto total es el paso de A a B en ambos casos, pero el punto intermedio difiere:

- Slutsky traslada paralelamente la nueva restricción presupuestaria hasta que roce la cesta inicial A, dando el punto S ($\text{ES}_\text{S} = A \to S$,$\text{ER}_\text{S} = S \to B$). Trasladando este punto –igual que se hizo antes con A y B para obtener la curva de demanda ordinaria– a un plano de precio y cantidad de x, se obtiene la demanda compensada de Slutsky,D_S.
- Hicks traslada la nueva restricción hasta que sea tangente a la curva de indiferencia inicial, dando el punto H ($\text{ES}_\text{H} = A \to H$,$\text{ER}_\text{H} = H \to B$). De la misma forma, trasladando H se obtiene la demanda compensada o hicksiana,D_H.

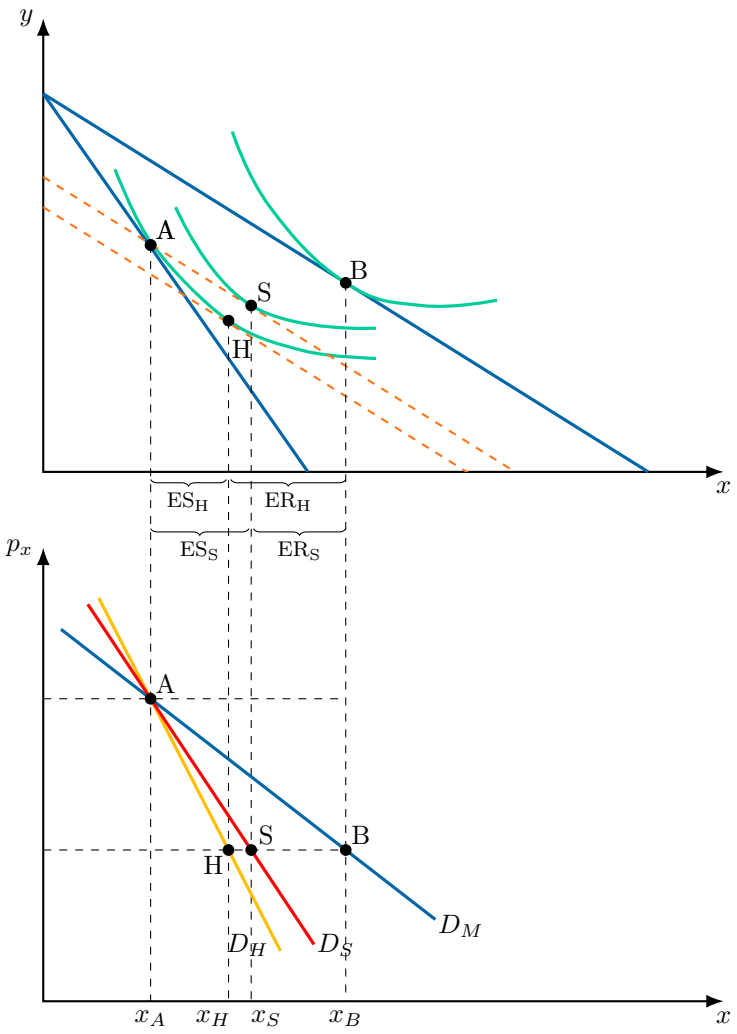
Para un cambio de precio discreto, S y H son en general distintos. Solo coinciden en el límite de una variación infinitesimal, el caso que recoge la ecuación de Slutsky.

Con el método de Hicks se descomponen los dos efectos: la demanda hicksiana recoge exclusivamente el efecto sustitución –no depende de la renta–, y el efecto renta se obtiene por diferencia respecto al efecto total.

[32] La intuición es que, cuando un bien se abarata, conviene sustituir otros bienes por él, y cuando se encarece, conviene sustituirlo por otros. Por eso este efecto siempre va en sentido contrario a la variación del precio. Por ejemplo, si baja el precio del café, el café sale más barato en comparación con el té, y el consumidor tiende a tomar más café y menos té.

[33] La intuición es que una bajada de precios hace al consumidor más rico sin que gane más dinero, y una subida lo hace más pobre. Cómo reacciona a ese cambio depende del bien: si es normal, comprará más al enriquecerse, y si es inferior, comprará menos. Siguiendo el ejemplo, al bajar el precio del café al consumidor le sobra dinero, y si el café es un bien normal tomará todavía más café.

Figura 3.7: Descomposición de efecto renta y sustitución



Ecuación de Slutsky Analíticamente, el efecto renta y el efecto sustitución se relacionan a través de la ecuación de Slutsky, que establece que la variación de la cantidad demandada al variar el precio, es decir el efecto total, es igual a la suma del efecto sustitución –dado por el término que representa la variación de la demanda al variar el precio manteniendo constante la renta real– más el efecto renta –dado por el término que representa la variación de la demanda al variar la renta–

$$\underbrace{\frac{\partial x\left(p,m\right)}{\partial p_x}}_{\text{Efecto total}} = \underbrace{\frac{\partial h_x\left(p,u\right)}{\partial p_x}}_{\text{Efecto sustitución}} \underbrace{-\frac{\partial x\left(p,m\right)}{\partial m}\cdot x\left(p,m\right)}_{\text{Efecto renta}} \tag{3.11}$$

Tipos de bienes Además, a partir del signo y la magnitud del efecto renta y del efecto sustitución es posible caracterizar los bienes y la pendiente de la función de demanda. Concretamente, ante un aumento del precio de un bien:

 © 2026 Manuel Cabado García

- El efecto sustitución propio es negativo, por la convexidad de las preferencias[34].
- El efecto renta puede ser:
  - Negativo: se trata de un bien normal, al reducirse la renta se reduce el consumo. En este caso, la demanda tendrá pendiente negativa, puesto que el efecto sustitución y el efecto renta son negativos.
  - Positivo: se trata de un bien inferior, al aumentar (reducirse) la renta, se reduce (aumenta) su consumo. Existen dos posibilidades, ante el cambio del precio:
    - Si $|ER| < |ES|$, el bien es inferior y al aumentar el precio se reduce su consumo.
    - Si $|ER| > |ES|$, el bien es inferior pero ahora al aumentar su precio se aumenta su consumo. Se trata de un bien Giffen, cuya función de demanda tiene pendiente positiva[35][36].

Elasticidad Por otro lado, una medida para establecer la sensibilidad de la demanda es la de elasticidad, distinguiendo entre

- Elasticidad precio de la demanda de un bien, que es la variación porcentual de la cantidad demandada de un bien ante una variación porcentual de su precio

$$\varepsilon_P = \frac{\partial x}{\partial p_x} \cdot \frac{p_x}{x} \tag{3.12}$$

  y dependiendo de su valor
  - Si $|\varepsilon_P| > 1$ la demanda es elástica, es decir, muy sensible al precio. Viajes aéreos, coches de lujo, comidas en restaurantes, y la mayoría de los productos de marca con sustitutos fáciles.

[34] En ese sentido, su signo es contrario al de la variación del precio

[35] Caso excepcional de los bienes Giffen: si un bien es inferior y absorbe una gran parte de la renta, un aumento de su precio puede hacer subir su demanda. Esto ocurre porque el encarecimiento empobrece tanto al consumidor que este se ve obligado a renunciar a otros bienes más caros, comprando aún más del bien barato para poder subsistir –el efecto renta, positivo por tratarse de un bien inferior, supera al efecto sustitución, siempre negativo–. El ejemplo clásico de manual –las patatas durante la hambruna irlandesa del siglo XIX, popularizado por Marshall– carece en realidad de verificación empírica sólida y ha sido cuestionado por historiadores económicos (Rosen, 1999). La evidencia empírica más sólida de comportamiento Giffen real procede de Jensen y Miller (2008), que documentaron el fenómeno subsidiando el precio de alimentos básicos entre hogares muy pobres de China: evidencia fuerte para el arroz en Hunan y más débil para el trigo en Gansu. No existen "bienes Giffen" en sentido estricto, sino situaciones puntuales, verificadas empíricamente, en las que se confirma esta paradoja.

[36] Bien Veblen: si sube el precio, sube la demanda –al contrario que un bien Giffen, aquí no hay ningún motivo de subsistencia, sino que el precio elevado forma parte del propio atractivo del bien (consumo conspicuo u ostentoso, en el sentido dado por Veblen (1899)). Se trata típicamente de bienes de lujo –relojes de alta gama, bolsos de marcas exclusivas, ciertos vinos de gama alta–, donde un precio más bajo podría incluso reducir su atractivo social. A diferencia de los bienes Giffen, que son inferiores, los bienes Veblen son bienes de lujo. Ambos comparten una pendiente de demanda positiva, pero por razones opuestas.

  - Si $|\varepsilon_P| = 1$ elasticidad unitaria, una variación porcentual del precio provoca una variación porcentual igual, y de signo contrario, en la cantidad demandada, de modo que el gasto en el bien no cambia.
  - Si $|\varepsilon_P| < 1$ la demanda es inelástica, es decir, poco sensible al precio. Ejemplos: Medicamentos vitales (como la insulina), la gasolina a corto plazo, la sal, y los servicios de agua y electricidad.

- Elasticidad precio cruzada, que es la variación porcentual de la demanda de un bien ante cambios en el precio de otro bien

$$\varepsilon_C = \frac{\partial y}{\partial p_x} \cdot \frac{p_x}{y} \tag{3.13}$$

  y dependiendo de su valor

  - Si $\varepsilon_C > 0$ los bienes son sustitutivos brutos. Ejemplos: Coca-Cola y Pepsi, mantequilla y margarina, el té y el café.
  - Si $\varepsilon_C = 0$, los bienes son independientes. Ejemplos: El precio de una bicicleta no afecta la demanda de leche. Un ordenador y una caja de galletas.
  - Si $\varepsilon_C < 0$ los bienes son complementarios. Ejemplos: Automóviles y gasolina, consolas de videojuegos y videojuegos, palomitas de maíz y entradas de cine.

Variaciones de la renta monetaria

Por otro lado, ante un aumento en la renta, gráficamente en un plano de bien x e y como el de la Figura 3.8 la recta de balance se desplazará paralelamente, pues se mantiene constante la relación entre los precios. Las combinaciones óptimas de bienes al variar la renta manteniendo los precios constantes representan la curva de renta-consumo, CRC.

Curva de Engel Derivada de la curva de renta-consumo se obtiene la curva de Engel, que relaciona el nivel de renta con la cantidad comprada de un bien. Además, esta relación permite distinguir entre un bien normal y un bien inferior:

$$\text{Si } \frac{\partial x\,(p,m)}{\partial m} \geq 0 \text{ es un bien normal} \tag{3.14}$$

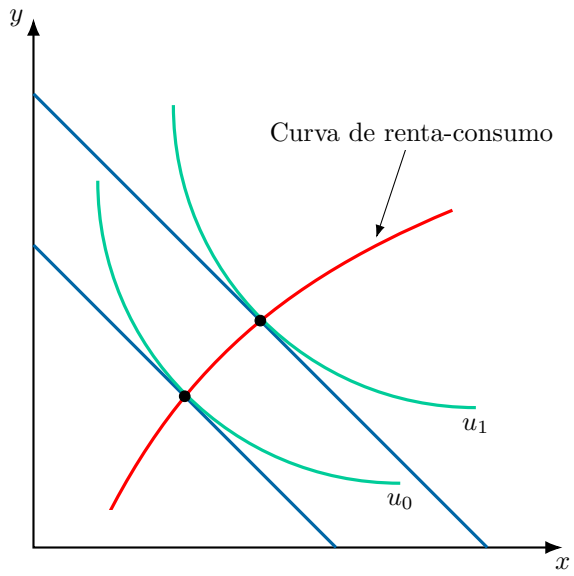
$$\text{Si } \frac{\partial x\,(p,m)}{\partial m} < 0 \text{ es un bien inferior} \tag{3.15}$$

Dentro de los bienes normales se encuentran:

- Los de primera necesidad, cuyo consumo aumenta en menor proporción que la renta, y

 © 2026 Manuel Cabado García

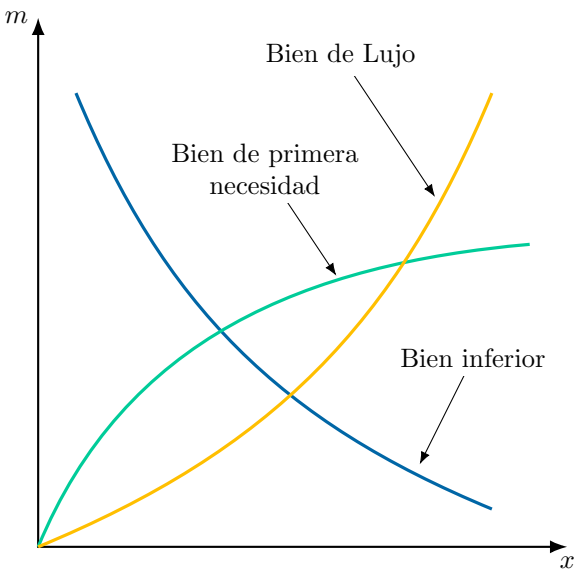
Figura 3.8: Efecto de las variaciones en la renta



- Los de lujo, cuyo consumo aumenta en mayor proporción que las variaciones de renta.

Bajo el supuesto de monotonicidad, no todos los bienes de una economía pueden ser inferiores. Además, un bien puede ser inferior a partir de un determinado nivel de renta. Gráficamente, en un plano de renta, m, y cantidad, x, como el de la Figura 3.9, se representan las curvas de Engel de un bien de primera necesidad, de uno de lujo y de uno inferior.

Figura 3.9: Curvas de Engel



En términos porcentuales se puede expresar la relación entre un bien y la renta me-

diante el concepto de elasticidad-renta:

$$\varepsilon_{x,M} = \frac{\partial x\,(p,m)}{\partial m} \cdot \frac{m}{x} = \begin{cases} \text{Bien normal} & \text{si} \quad \varepsilon_{x,M} \geq 0 \\ \text{Bien de primera necesidad} & \text{si} \quad 0 < \varepsilon_{x,M} < 1 \\ \text{Bien de lujo} & \text{si} \quad \varepsilon_{x,M} > 1 \\ \text{Bien inferior} & \text{si} \quad \varepsilon_{x,M} < 0 \end{cases} \tag{3.16}$$

Preferencias homotéticas Las preferencias son homotéticas cuando la RMS depende solo de la proporción en que se consumen los bienes y no de la cantidad total consumida.[37] En ese caso, todos los bienes tienen elasticidad renta igual a uno y la curva de renta-consumo es una recta que pasa por el origen.

3.5. Condiciones de un sistema de demanda completo ⏳ 1'30"

Con todo lo visto hasta el momento, un sistema completo de demanda debe cumplir:

1. Condición de agregación de Engel, establece que la suma de las elasticidades renta de todos los bienes, ponderadas por su participación en el gasto total, debe ser igual a uno:

$$\sum_{i=1}^{n} s_i \cdot \varepsilon_{i,m} = 1 \tag{3.17}$$

   Esto implica que no todos los bienes pueden ser ni inferiores ni neutros, es decir, el consumidor tiene que consumir algún bien normal. En consecuencia, si se incrementa la renta, el exceso de la misma deberá ser gastado de acuerdo con sus elasticidades renta, no pudiendo darse el caso en el que el consumidor comprase menos cantidad de todos los bienes, ya que no se cumpliría el axioma de monotonía.

2. Condición de agregación de Cournot, que muestra cómo un cambio en el precio de un bien afecta a la demanda de todos los bienes. La suma de las elasticidades de todos los bienes respecto al precio de uno de ellos, incluida la del propio bien y ponderadas por su participación en el gasto, es igual a la participación de ese bien en el gasto cambiada de signo:

$$\sum_{i=1}^{n} s_i \cdot \varepsilon_{i,j} = -s_j \tag{3.18}$$

[37] Gráficamente, las curvas de indiferencia tienen la misma pendiente a lo largo de cualquier recta trazada desde el origen. La intuición es que el consumidor reparte su renta en las mismas proporciones sea cual sea su nivel: si su renta se duplica, compra el doble de cada bien. Un ejemplo es la función Cobb-Douglas.

 © 2026 Manuel Cabado García

La principal implicación es que las respuestas de todas las demandas a un mismo precio no son independientes entre sí. Si el precio de un bien disminuye y el consumidor lo consume más, el consumo de otros bienes debe ajustarse para que el gasto total siga siendo igual a la renta. En particular, si la demanda de un bien es elástica, algún otro bien tiene que ser sustitutivo bruto suyo.

3. Condición de simetría: los efectos sustitución cruzados son simétricos, es decir, el efecto sustitución que una variación del precio del bien j tiene sobre la demanda del bien i es igual al que una variación del precio de i tiene sobre la demanda de j:[38]

$$S_{i,j} = S_{j,i} \tag{3.19}$$

Esta condición implica que las preferencias del consumidor son consistentes.

4. Condición de homogeneidad, que se deriva de que la demanda marshalliana es homogénea de grado cero en precios y renta: la suma de todas las elasticidades precio de un bien, la propia y las cruzadas, es igual a su elasticidad renta cambiada de signo:

$$\sum_{j=1}^{n} \varepsilon_{i,j} = -\varepsilon_{i,m} \tag{3.20}$$

Esto implica que, ante variaciones de precios y renta en la misma proporción, la cantidad consumida no varía (no hay ilusión monetaria).

5. Condición de negatividad, que establece que el efecto sustitución de la demanda de un bien sobre su propio precio es negativo o cero.

$$S_{i,i} \leq 0 \tag{3.21}$$

Esto implica que cuando el precio de un bien aumenta, el efecto sustitución siempre hará que la demanda de ese bien disminuya o se mantenga igual. Esta es la base de la ley de la demanda: las curvas de demanda marshalliana de bienes ordinarios tienen pendiente negativa.

3.6. Demanda de mercado ⏳ 1'

Por último, la demanda de mercado para un bien es la suma horizontal de las demandas de los consumidores individuales. Es decir, la cantidad demandada en el mercado para cada precio es la suma de las demandas individuales de todos los consumidores.

Analíticamente, se suponen dos bienes, x e y, y dos consumidores con las siguientes

[38] La intuición es que, una vez aislado el efecto sustitución, es decir, con la utilidad constante, la relación entre dos bienes es recíproca: el efecto que una subida de un euro en el precio del café tiene sobre la demanda de té es igual al que una subida de un euro en el precio del té tiene sobre la demanda de café.

funciones de demanda sobre el bien x:

$$x_1 = x_1\left(p_x, p_y, m_1\right) \tag{3.22}$$

$$x_2 = x_2\left(p_x, p_y, m_2\right) \tag{3.23}$$

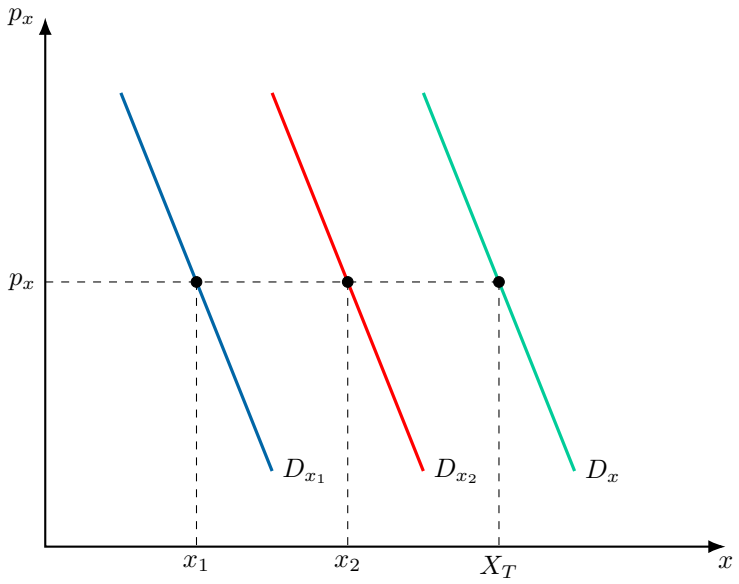
siendo x_1 y x_2 las demandas respectivas del consumidor 1 y 2 sobre el bien x.

Suponiendo que ambos se enfrentan a los mismos precios y son precio-aceptantes, entonces la demanda total de X es igual a la suma de las cantidades demandadas por los consumidores:

$$X_T = x_1 + x_2 \Rightarrow X_T = D_x\left(p_x, p_y, m_1, m_2\right) \tag{3.24}$$

Gráficamente, en un plano de precios de x y cantidad demandada de x como el de la Figura 3.10 se representa la curva de demanda de x del consumidor 1, D_{x_1}, y del consumidor 2, D_{x_2}, obteniendo la curva de demanda de mercado, D_x, mediante la agregación de las demandas individuales.

Figura 3.10: Obtención de la curva de demanda de mercado



3.7. Conclusiones ⏳ 30"

Como conclusión, la teoría neoclásica ha construido una teoría de la demanda que constituye un punto de referencia para análisis posteriores.

Sin embargo, este enfoque recibe algunas críticas, cada una de las cuales ha dado lugar a desarrollos posteriores que tratan de superarla:

1. No considera la existencia de incertidumbre, un problema abordado por el análisis del consumo en un contexto de incertidumbre.

 © 2026 Manuel Cabado García

2. Se basa en una función de utilidad que es subjetiva y no observable, limitación que superan tanto la teoría de la preferencia revelada como la teoría de la dualidad, ambas con base empírica.
3. Trata los bienes como bloques homogéneos, sin descomponerlos en sus características, lo que no explica fenómenos como la fidelidad de marca, el lanzamiento de nuevos productos o la publicidad. Esta limitación la aborda la demanda de características de Lancaster (1966).
4. Trata el consumo como un acto instantáneo, sin coste de oportunidad del tiempo empleado en consumir. Esta limitación la aborda la función de producción de los hogares de Becker (1965).

Apéndice

3.A.1. Enfoque cardinal de la teoría neoclásica

Los supuestos de este enfoque son que la utilidad es medible en unidades monetarias –y por tanto comparable entre individuos–, que es creciente, que la utilidad marginal de cada bien es decreciente (primera ley de Gossen) y que la utilidad marginal del dinero es constante, al tomarse como unidad de medida,$u'(M) = 1$. Resolviendo el lagrangiano

$$\mathcal{L} = u(x_1) + u(x_2) + u(M) - \lambda \cdot (p_1 x_1 + p_2 x_2 + M - Y)$$

se obtiene

$$\frac{u'(x_1)}{p_1} = \frac{u'(x_2)}{p_2} = \lambda = u'(M) = 1$$

de donde$u'(x_i) = p_i$: el consumidor demanda cada bien hasta que su utilidad marginal iguala el precio, con lo que la demanda depende únicamente de su propio precio,$x_i = x(p_i)$.

Este enfoque fue superado por desarrollos posteriores por varias razones. El supuesto de utilidad marginal del dinero constante es poco realista y, al implicarλ constante, elimina todo efecto renta y todo efecto sustitución cruzado –salvo que la función sea aditiva y separable, no toda función aditiva y separable carece de efectos cruzados, como puede comprobarse con el contraejemplo $U(x_1, x_2) = \ln(x_1) + x_2$, que sigue siendo aditiva y separable pero genera$x_1 = p_2/p_1$, dependiente del precio del otro bien–. Además, sobre la cardinalidad –que solo garantiza comparar diferencias de utilidad dentro de un mismo individuo– este enfoque superpone el supuesto, adicional e independiente, de comparabilidad interpersonal de utilidades, que resulta excesivamente exigente. Por todo ello, Pareto rompe con la fusión entre marginalismo y utilitarismo bentamita –acuñando el término "ofelimidad" para distinguir la utilidad estrictamente individual y subjetiva de cualquier agregado social– y se impone el enfoque ordinal desarrollado en el cuerpo del tema.

3.A.2. Tipos de función de utilidad

En cuanto a la forma de las curvas, esta estará en función de los diferentes tipos de utilidad que puedan darse, y dará información sobre los bienes:

Función Cobb–Douglas: aquella que cumple con todas las características mencionadas previamente y las más típicas

$$U(X_1, X_2) = A \cdot X_1^{\alpha} \cdot X_2^{\beta}, \quad \alpha, \beta > 0$$

donde α y β son constantes positivas y la RMS

$$|RMS| = \frac{\alpha \cdot X_2}{\beta \cdot X_1}$$

 © 2026 Manuel Cabado García

Sustitutivos perfectos: Una persona con estas preferencias estará dispuesta a renunciar a la misma cantidad de Y para conseguir una unidad más de X independientemente de cuánto esté consumiendo de X

$$U\left(X_1, X_2\right) = a \cdot X_1 + b \cdot X_2$$

donde la RMS es constante

$$|RMS| = \frac{a}{b}$$

Complementarios perfectos: Función de coeficientes fijos tipo Leontief:

$$U(X_1, X_2) = Min\left\{\frac{X_1}{\alpha}, \frac{X_2}{\beta}\right\}$$

donde la RMS no existe y la UMg es constante.

Funciones CES (Elasticidad de Sustitución Constante)

$$U = A\left(\alpha \cdot X_1^{\rho} + \beta X_2^{\rho}\right)^{\frac{\varepsilon}{\rho}}$$

donde $\rho \leq 1$ es el parámetro de sustitución y ε es el grado de homogeneidad y $\alpha y \beta > 0$ son los parámetros que representan las preferencias relativas que el individuo tiene por los bienes. $1/1 - \rho$ es la elasticidad de sustitución, aunque dicho concepto tiene más sentido verlo en la teoría de la producción

Respecto a la RMS:

$$RMS = \frac{\alpha}{\beta} \cdot \left(\frac{X_2}{X_1}\right)^{1-\rho}$$

3.A.3. Preferencias lexicográficas

El funcionamiento de las preferencias lexicográficas puede ilustrarse con la búsqueda de una palabra en un diccionario: se da prioridad absoluta a la primera letra, de modo que solo si dos palabras empiezan por "A" se atiende a la segunda letra, y solo si también coinciden se pasa a la tercera, y así sucesivamente. Ninguna palabra que empiece por "B" será nunca preferida a otra que empiece por "A", por muy larga o llamativa que sea la palabra con "B".

Las preferencias lexicográficas en la teoría del consumidor funcionan exactamente igual: Un consumidor con preferencias lexicográficas sobre dos bienes (por ejemplo, Bien 1 y Bien 2) siempre priorizará absolutamente la cantidad del Bien 1. Solo si dos cestas de bienes tienen la misma cantidad del Bien 1, entonces el consumidor se fijará en la cantidad del Bien 2 para decidir cuál prefiere.

Ejemplo intuitivo:

- Cesta A: (5 unidades de Bien 1, 10 unidades de Bien 2)

- Cesta B: (4 unidades de Bien 1, 1000 unidades de Bien 2)

Un consumidor con preferencias lexicográficas siempre preferirá la Cesta A a la Cesta B, sin importar que la Cesta B tenga muchísimo más del Bien 2. La cantidad de Bien 1 (5 vs 4) decide la preferencia de inmediato. La enorme cantidad de Bien 2 en la Cesta B es completamente irrelevante una vez que el Bien 1 ha sido evaluado y es diferente.

Únicamente si las cantidades del Bien 1 coincidieran, por ejemplo:

- Cesta C: (5 unidades de Bien 1, 1 unidad de Bien 2)
- Cesta D: (5 unidades de Bien 1, 2 unidades de Bien 2) Entonces el consumidor preferiría la Cesta D, porque las cantidades de Bien 1 son iguales, y la Cesta D tiene más del Bien 2.

Su papel en la teoría de la demanda del consumidor El papel de las preferencias lexicográficas en la teoría de la demanda es principalmente ilustrativo y de contraste, más que de uso común para modelar el comportamiento real:

- Rompen la "continuidad": La principal razón por la que son problemáticas es que no son continuas. Esto significa que no es posible introducir cambios infinitesimales en la cantidad de un bien esperando una respuesta "suave" en las preferencias. Siempre hay un salto abrupto en la preferencia al cambiar la cantidad del bien más prioritario.
- No se pueden representar con una función de utilidad continua: Debido a esa falta de continuidad, no es posible construir una función de utilidad matemática "tradicional" y continua que represente estas preferencias. Esto es un gran problema para la teoría, ya que la mayor parte de las herramientas habituales (derivadas, optimización) requieren funciones continuas.
- No hay curvas de indiferencia "suaves": En un gráfico, las curvas de indiferencia de preferencias lexicográficas no son las curvas suaves habituales en el resto de la teoría. Serían puntos o líneas muy extrañas que no permiten la convexidad habitual.
- No es posible la optimización estándar: La teoría de la demanda del consumidor busca un punto óptimo donde la Relación Marginal de Sustitución (RMS) sea igual a la relación de precios (el punto de tangencia de la curva de indiferencia con la recta presupuestaria). Con preferencias lexicográficas, no hay una RMS bien definida ni un punto de tangencia único y estable, lo que hace que los métodos de optimización estándar fallen.
- Excepción teórica: Aunque son una rareza matemática y no se usan para modelar el comportamiento de la mayoría de los bienes, se estudian para mostrar la importancia de los supuestos subyacentes a la teoría del consumidor, como la continuidad y la convexidad. Son un buen ejemplo de un tipo de preferencia "irracional" o "extrema" que no puede ser manejada por las herramientas estándar.

 © 2026 Manuel Cabado García

El interés de las preferencias lexicográficas para la teoría de la demanda no reside en su realismo descriptivo, sino en delimitar con precisión qué axioma resulta realmente indispensable para garantizar la existencia de una función de utilidad: satisfacen completitud, transitividad y reflexividad –los axiomas del pre-orden completo– y, sin embargo, no admiten ninguna función de utilidad que las represente, precisamente porque fallan en continuidad. Aunque en la vida real las preferencias lexicográficas puras son raras (porque casi siempre existe disposición a sacrificar algo de una cosa por mucha de otra), pueden identificarse situaciones que se aproximan a este comportamiento, especialmente cuando existe una prioridad absoluta por un factor concreto[39], y constituyen así el contraejemplo canónico que confirma, por la vía negativa, la necesidad del Axioma 4 dentro del Teorema de Debreu: sin continuidad, ni la racionalidad del consumidor ni la posibilidad de ordenar sus cestas bastan para construir el aparato analítico –derivadas, curvas de indiferencia, condiciones de tangencia– sobre el que se apoya el resto de la teoría de la demanda.

3.A.4. Teoría de la preferencia revelada

La teoría de la preferencia revelada fue desarrollada por Samuelson (1938). Si bien la teoría neoclásica de la demanda tiene un enfoque completamente deductivo –partiendo de unos supuestos se construye una teoría que permite realizar predicciones sobre el comportamiento del consumidor–, la teoría de la preferencia revelada tendrá un enfoque completamente inductivo –partiendo de la observación del comportamiento del consumidor, se deduce la naturaleza de las preferencias–.

El gran valor añadido que tiene esta teoría es que permite realizar las mismas predicciones acerca del comportamiento de los consumidores con unos supuestos mucho menos restrictivos.

Concretamente, la teoría de la preferencia revelada se basa en los siguientes supuestos:

1. Ante un vector de precios p y una dotación de renta m, el consumidor gasta toda su renta en la compra de bienes y servicios
2. Solo una cesta de bienes x es escogida para cada nivel de renta y precios, es decir, solo hay una cesta óptima en cada ocasión.
3. Solo existe un vector de precios p y una dotación monetaria m en donde cada cesta óptima es escogida, es decir, no puede haber una cesta óptima que sea escogida en distintas ocasiones.

[39] Algunos ejemplos:

- La seguridad en un viaje aéreo: se elige siempre la aerolínea con menor riesgo de accidente, sin importar lo cómoda o lujosa que sea la otra.
- Medicamentos y efectos secundarios: se elige siempre el medicamento con mayor probabilidad de curación, incluso si eso implica soportar efectos secundarios más desagradables.
- Alimentos y alergias graves: se elige siempre el alimento del que se tiene total certeza de que no contiene el alérgeno, sin importar lo apetitoso que sea otro plato que pueda contenerlo.

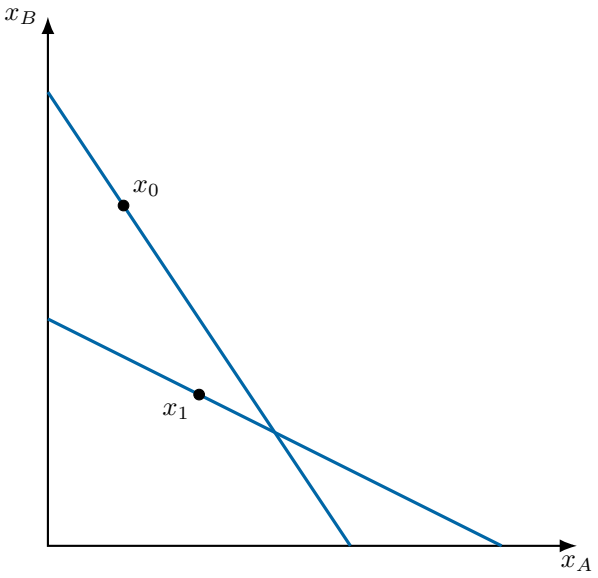
4. Las preferencias del consumidor son consistentes, quedando reflejado en los dos axiomas de la preferencia revelada.

Axioma débil de preferencia revelada

El axioma débil de preferencia revelada establece que si al nivel de precios p_0 un consumidor elige la cesta x_0 estando también disponible la cesta x_1, entonces cuando el consumidor elige la cesta x_1 es porque se ha producido un cambio de precios y a este nuevo nivel, p_1, x_0 ya no está disponible. Es decir, si x_1 era asequible a los precios p_0 y aun así se eligió x_0, entonces x_0 no puede ser asequible a los precios p_1 cuando se elige x_1

$$p_0 \cdot x_1 \leq p_0 \cdot x_0 \Rightarrow p_1 \cdot x_0 > p_1 \cdot x_1 \tag{3.25}$$

Figura 3.11: Axioma débil de preferencia revelada



Axioma fuerte de preferencia revelada

El axioma fuerte de preferencia revelada impone la transitividad de las preferencias. Por lo tanto, cambios consecutivos en los precios permiten establecer un orden en la deseabilidad de las cestas en función de las cestas que estuvieran disponibles a cada nivel de precios y la cesta óptima finalmente elegida en cada uno de estos.

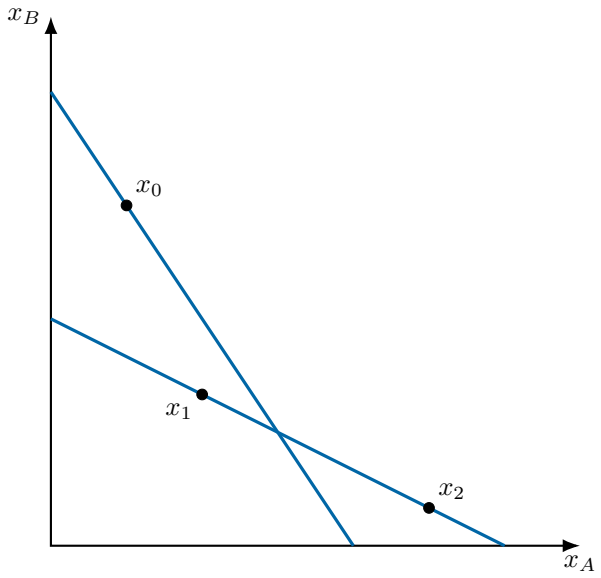
Dadas tres cestas de bienes, x_0, x_1 y x_2, no comparables entre sí a un mismo nivel de precios, se puede establecer un orden de preferencia mediante la relación indirecta de preferencia revelada entre las distintas cestas.

Si x_0 se revela preferida a x_1, x_1 se revela preferida a x_2, entonces el orden de preferencias es $x_0 > x_1 > x_2$.

 © 2026 Manuel Cabado García

Este no es más que una versión más fuerte del axioma previo, que refuerza la consistencia y la transitividad de las preferencias reveladas.

Figura 3.12: Axioma fuerte de preferencia revelada



Aplicaciones: efecto renta y efecto sustitución

Por otro lado, la teoría de la preferencia revelada puede ser utilizada para descomponer los efectos que produce una variación de precios sobre la demanda del consumidor.

La teoría de la preferencia revelada no se podrá apoyar sobre las curvas de indiferencia para calcular la variación compensatoria y así hallar el efecto sustitución en términos de Hicks. En su lugar se servirá de la definición de Slutsky de renta real: la compensación de renta se calcula desplazando la recta presupuestaria, a los nuevos precios, hasta que la cesta inicial vuelva a estar disponible.

A través de este cálculo se hallará la cesta inicial, x_0, la cesta final, x_1, y la cesta intermedia x^*.

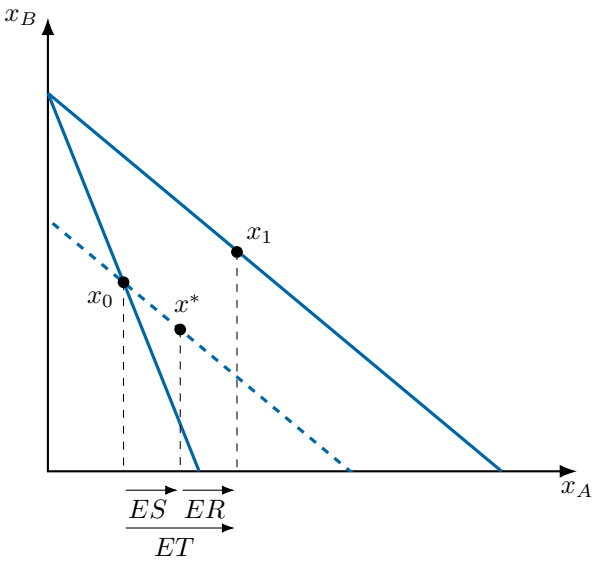
Esta cesta intermedia podría ubicarse a la izquierda de la cesta inicial x_0, a la derecha o ser exactamente igual que x_0.

No obstante, el axioma débil de preferencia revelada establece que a la izquierda no puede encontrarse ya que esa cesta ya estaba disponible al nivel de precios inicial y, sin embargo, no era elegida.

Por su parte, x_0 tampoco puede ser la cesta intermedia, ya que una misma cesta no puede ser óptima para dos niveles de precios distintos. Por lo tanto, x^* se encuentra necesariamente a la derecha de x_0.

Así, el efecto sustitución, al igual que en el modelo neoclásico, tiene un efecto opuesto a la dirección a la que se mueve el precio. Como el precio cae, se demanda más del bien

Figura 3.13: Preferencia revelada: efecto renta y efecto sustitución



que se ha abaratado.

De esta forma, $x^* - x_0$ representa el efecto sustitución, $x_1 - x^*$ representa el efecto renta y $x_1 - x_0$ representa el efecto total.

Conclusión

En conclusión, la teoría de la preferencia revelada de Samuelson es completamente equivalente a la teoría neoclásica de la demanda de consumo, ya que todas las predicciones realizadas por la segunda también son realizadas por la primera.

Por lo tanto, un consumidor que satisfaga los supuestos relativamente poco restrictivos de la teoría de la preferencia revelada se comportará como si tuviera unas preferencias que satisfagan la teoría neoclásica del consumidor.

En definitiva, esto otorga una mayor robustez a la teoría neoclásica, ya que sigue siendo válida con unos supuestos mucho más laxos.

3.A.5. Métodos de los precios hedónicos

Los precios hedónicos son un concepto y un método económico que se utiliza para estimar el valor de las diferentes características o atributos que componen un bien o servicio, especialmente aquellos que son heterogéneos y cuyas características no se venden por separado en el mercado.

La idea fundamental es que el precio de un bien no es un todo monolítico, sino más bien la suma de los valores implícitos de sus características individuales. Se asume que los consumidores valoran las características de un bien más que el bien en sí mismo.

 © 2026 Manuel Cabado García

Para entenderlo mejor, piénsese en una vivienda: el precio de una casa no se basa solo en "ser una casa". Influyen multitud de atributos:

- Características estructurales: número de habitaciones, baños, metros cuadrados construidos, antigüedad, tipo de construcción, calidad de los materiales.
- Características de ubicación: cercanía a servicios (escuelas, hospitales, transporte público), seguridad del barrio, calidad del vecindario.
- Características ambientales o de calidad de vida: cercanía a parques, vistas (al mar, a la montaña), nivel de ruido, calidad del aire.

El método de precios hedónicos busca descomponer el precio total del bien para asignar un "precio implícito" a cada una de estas características. Para calcularlo se utilizan modelos econométricos, generalmente regresiones multivariables. La idea es establecer una relación matemática entre el precio del bien (variable dependiente) y sus múltiples características (variables independientes).

De esta forma, los precios hedónicos tienen diversas aplicaciones, entre las más importantes:

1. Valoración de bienes y servicios no comercializables directamente en el mercado: Es especialmente útil para estimar el valor de bienes ambientales (calidad del aire, acceso a espacios verdes, paisajes) o servicios públicos, que no tienen un precio de mercado explícito, pero influyen en el precio de otros bienes (como la vivienda).
2. Ajuste de calidad en índices de precios: Permiten ajustar los índices de precios (como el Índice de Precios al Consumo o el Índice de Precios de Vivienda) por cambios en la calidad de los bienes. Si un producto sube de precio porque incorpora mejores características, el método hedónico puede diferenciar si la subida es por calidad (no es inflación "pura") o por una subida de precio real.
3. Análisis del mercado inmobiliario: Para entender cómo diferentes atributos de una propiedad (tamaño, ubicación, número de habitaciones, etc.) afectan su precio y cómo valoran los compradores esas características.
4. Evaluación de políticas públicas: Permiten evaluar el impacto económico de ciertas políticas (ej. mejoras en la calidad ambiental de un área) al observar cómo estas influyen en el valor de las propiedades cercanas.
5. Marketing y diseño de productos: Ayudan a las empresas a entender qué características son más valoradas por los consumidores para optimizar el diseño de sus productos y estrategias de precios.

El método de precios hedónicos conecta directamente con el equilibrio del consumidor desarrollado en el cuerpo del tema: los precios implícitos estimados para cada característica no son más que una aproximación empírica a las utilidades marginales relativas que,

en el óptimo, determinan la relación marginal de sustitución entre atributos –la misma condición de tangencia que arroja la maximización de la utilidad sujeta a la restricción presupuestaria (Apéndice, 3.A.7)–. De ahí que la regresión hedónica pueda leerse como el correlato econométrico de la teoría ordinal de la demanda: mientras esta explica en qué condiciones se alcanza el equilibrio, el método hedónico permite recuperar empíricamente los precios sombra que ese equilibrio revela cuando el bien en cuestión no se transa por separado en el mercado.

3.A.6. La función de producción de los hogares de Becker

Gary Becker (1965) propone un enfoque distinto del consumidor: los bienes de mercado no reportan utilidad por sí mismos, sino que son insumos que el hogar combina con tiempo para producir los verdaderos bienes que generan satisfacción –de forma análoga a como una empresa combina factores para producir un output–.

Formalmente, el consumidor obtiene utilidad de una serie de bienes producidos en el hogar, Z_i, cada uno obtenido mediante una función de producción doméstica que combina un bien de mercado,x_i, y el tiempo dedicado a él,t_i:

$$Z_i = f_i\left(x_i, t_i\right), \qquad U = U\left(Z_1, \ldots, Z_n\right)$$

El consumidor maximiza esta utilidad sujeto a dos restricciones, no solo una: la restricción monetaria habitual y una restricción de tiempo, dado que el tiempo total disponible,T, se reparte entre tiempo de trabajo,t_w, y tiempo dedicado a la producción doméstica:

$$\sum_i p_i x_i = w \cdot t_w + V, \qquad \sum_i t_i + t_w = T$$

Combinando ambas restricciones –sustituyendo t_w– se obtiene una única restricción de "renta completa":

$$\sum_i \left(p_i x_i + w \cdot t_i\right) = w \cdot T + V$$

donde $w \cdot T + V$ es la renta completa: el dinero disponible más el valor del tiempo total, valorado al salario.

De esta restricción combinada se deduce que el precio efectivo –o precio sombra– de cada bien producido en el hogar no es solo su precio de mercado, sino que incluye también el coste de oportunidad del tiempo empleado en producirlo, ligado al salario w. Esto explica, por ejemplo, por qué los individuos con salarios más altos sustituyen tiempo por bienes de mercado –comida precocinada en vez de cocinar, tintorería en vez de lavar en casa–: cuanto mayor es w, más caro resulta, en términos de renta completa, el tiempo empleado en cualquier actividad.

Este enfoque, en la misma línea crítica que la demanda de características de Lancaster,

 © 2026 Manuel Cabado García

supera la limitación neoclásica de tratar el consumo como un acto instantáneo sin coste de oportunidad temporal, y ha sido la base de desarrollos posteriores en la economía de la familia, la oferta de trabajo y la economía del tiempo.

A escala agregada, este enfoque tiene implicaciones macroeconómicas relevantes: al encarecerse el tiempo con el salario, un aumento salarial no solo desplaza la oferta de trabajo –vía el efecto sustitución habitual entre ocio y trabajo–, sino que también reasigna el tiempo dentro del propio hogar hacia actividades que requieren menos tiempo relativo, lo que ayuda a explicar la caída histórica del tiempo dedicado a la producción doméstica y el auge paralelo de sectores de mercado que sustituyen esa producción –restauración, guarderías, servicios domésticos–. Esta reasignación del tiempo del hogar hacia el mercado es, de hecho, una de las explicaciones citadas para el aumento sostenido de la participación laboral femenina en las economías desarrolladas durante la segunda mitad del siglo XX.

Esto conecta directamente con un problema de medición: el PIB solo contabiliza la producción que pasa por el mercado, de modo que toda la producción doméstica –cuidado de hijos, limpieza, cocina– queda completamente fuera de las cuentas nacionales, pese a generar un valor económico real.Nordhaus y Tobin (1972) fueron pioneros al proponer una Medida del Bienestar Económico (MEW) que ajustaba el PIB incorporando el valor del ocio y del trabajo no remunerado, aplicando precisamente la lógica de Becker. Trabajos más recientes, como el de Landefeld et al. (2009), han construido cuentas satélite oficiales de la producción doméstica en Estados Unidos: incorporarla elevaría el PIB medido convencionalmente en torno a una cuarta parte –con una tendencia decreciente desde cerca del 40 % en 1965, a medida que más mujeres se incorporaban al mercado laboral–. Estas estimaciones tienen implicaciones de política evidentes: decisiones como la fiscalidad de los servicios domésticos, la provisión pública de cuidados o las comparaciones internacionales de bienestar dependen críticamente de si se tiene en cuenta, o no, esta producción invisible para el PIB.

3.A.7. El problema de maximización de la utilidad mediante el método de Lagrange

El problema de maximización de la utilidad presentado en el epígrafe “Equilibrio del consumidor” admite dos vías de resolución equivalentes: despejar la restricción presupuestaria para una de las variables y sustituirla directamente en la función de utilidad, o recurrir al método de los multiplicadores de Lagrange, más general y aplicable incluso cuando la restricción no admite un despeje explícito para ninguna de las variables. A continuación se desarrolla esta segunda vía paso a paso, siguiendo la misma lógica –plantear primero el problema, formalizar después el multiplicador y solo al final interpretarlo– con la que

suele presentarse este método en los manuales de optimización estática[40].

Por la monotonía de las preferencias (Axioma 6), que implica la no saciedad local, el consumidor agota siempre su renta, de modo que la restricción presupuestaria de la desigualdad $p_x \cdot x + p_y \cdot y \leq m$ se convierte, en el óptimo, en una igualdad. El problema puede escribirse entonces como

$$\max_{x,y} \quad u(x,y) \quad \text{sujeto a} \quad p_x \cdot x + p_y \cdot y = m.$$

El primer paso del método de Lagrange consiste en escribir la restricción con un cero en un lado de la igualdad y construir el Lagrangiano sumando al objetivo el multiplicador, λ, multiplicado por dicha expresión

$$\mathcal{L}(x,y,\lambda) = u(x,y) - \lambda\left(p_x \cdot x + p_y \cdot y - m\right).$$

El segundo paso consiste en derivar el Lagrangiano respecto a cada una de las variables de elección y respecto al propio multiplicador, e igualar cada derivada a cero

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= UMg_x - \lambda \cdot p_x = 0, \\
\frac{\partial \mathcal{L}}{\partial y} &= UMg_y - \lambda \cdot p_y = 0, \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= -\left(p_x \cdot x + p_y \cdot y - m\right) = 0.
\end{aligned}$$

La última condición no es más que la restricción presupuestaria original. De las dos primeras se despeja directamente

$$\lambda = \frac{UMg_x}{p_x} = \frac{UMg_y}{p_y},$$

la misma condición ya avanzada en el cuerpo del tema, y dividiendo una entre otra se obtiene

$$\frac{UMg_x}{UMg_y} = \frac{p_x}{p_y} \Leftrightarrow RMS_{x,y} = \frac{p_x}{p_y},$$

la condición de tangencia entre la curva de indiferencia y la recta de balance: el consumidor deja de reajustar su cesta cuando la tasa a la que está dispuesto a sustituir un bien por otro coincide con la tasa a la que el mercado le permite hacerlo.[41]

[40] Aunque es un manual de macroeconomía, la exposición del método de Lagrange en Doepke, Lehnert y Sellgren (1999) –planteando primero el problema, formalizando después el Lagrangiano y solo al final interpretando el multiplicador– resulta especialmente clara para ganar intuición antes de formalizar, por lo que se ha tomado como modelo expositivo en este apartado.

[41] Esta derivación presupone una solución interior, es decir, que el consumidor adquiere una cantidad estrictamente positiva de ambos bienes. El supuesto de que las preferencias son asintóticas a los ejes –introducido junto con el Teorema de Debreu– garantiza precisamente esa solución interior. Cuando ese supuesto no se sostiene y puede haber soluciones de esquina, las condiciones de Lagrange se generalizan

 © 2026 Manuel Cabado García

Queda por precisar qué representa exactamente el multiplicador λ. Supóngase que el consumidor se encuentra en la cesta óptima y que su renta aumenta en una cantidad infinitesimal, dm. Si esa renta adicional se dedicase íntegramente al bien x, permitiría comprar dm/p_x unidades adicionales de dicho bien, lo que elevaría la utilidad en $UMg_x \cdot dm/p_x$. Por la primera condición de primer orden, $UMg_x = \lambda \cdot p_x$, de modo que ese incremento de utilidad es exactamente $\lambda \cdot dm$ –y el mismo resultado se obtendría si la renta adicional se dedicase al bien y, precisamente porque en el óptimo las utilidades marginales ponderadas por el precio están igualadas entre todos los bienes–. Cada unidad adicional de renta reporta, por tanto, λ unidades de utilidad, con independencia de en qué bien se gaste: λ es el precio sombra de la restricción presupuestaria, la tasa a la que se relaja el máximo de utilidad alcanzable cuando se relaja la restricción de renta, es decir, la utilidad marginal de la renta ya avanzada en el cuerpo del tema.

Por último, las condiciones de primer orden anteriores son necesarias, pero no bastarían por sí solas para garantizar que la cesta hallada sea un máximo si no fuera por los axiomas de deseabilidad (Axiomas 5 a 7), que garantizan que la función de utilidad es estrictamente cuasicóncava y dos veces diferenciable. En un problema de maximización de una función estrictamente cuasicóncava sujeta a una restricción lineal, dichas condiciones de primer orden identifican un único máximo global, sin necesidad de comprobar punto por punto el hessiano orlado (Mas-Colell et al., 1995). De este modo, la derivación completa por el método de Lagrange confirma, paso a paso, la condición de equilibrio ya avanzada en el cuerpo del tema.

mediante las condiciones de Kuhn-Tucker, que permiten que alguna de las restricciones de no negatividad sea activa en el óptimo (Jehle y Reny, 2011).

Bibliografía

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal 75*(299), 493–517.

Doepke, M., A. Lehnert y A. W. Sellgren (1999). Macroeconomics. Apuntes de doctorado inéditos, concebidos como complemento al manual de Robert J. Barro, *Macroeconomics*.

Edgeworth, F. Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London: C. Kegan Paul and Co.

Jehle, G. A. y P. J. Reny (2011). *Advanced Microeconomic Theory* (3rd ed.). Harlow: Prentice Hall.

Jensen, R. T. y N. H. Miller (2008). Giffen Behavior and Subsistence Consumption. *American Economic Review 98*(4), 1553–1577.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy 74*(2), 132–157.

Landefeld, J. S., B. M. Fraumeni y C. M. Vojtech (2009). Accounting for Household Production: A Prototype Satellite Account Using the American Time Use Survey. *Review of Income and Wealth 55*(2), 205–225.

Mas-Colell, A., M. D. Whinston y J. R. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Nicholson, W. y C. Snyder (2017). *Microeconomic Theory: Basic Principles and Extensions* (12th ed.). Boston: Cengage Learning.

Nordhaus, W. D. y J. Tobin (1972). Is Growth Obsolete?

Pareto, V. (1906). *Manuale di Economia Politica, con una Introduzione alla Scienza Sociale*. núm. 13 de Piccola Biblioteca Scientifica. Milano: Società Editrice Libraria.

Rosen, S. (1999). Potato Paradoxes. *Journal of Political Economy 107*(S6), S294–S313.

Samuelson, P. A. (1938). A Note on the Pure Theory of Consumer's Behaviour. *Economica 5*(17), 61–71.

Veblen, T. (1899). *The Theory of the Leisure Class: An Economic Study in the Evolution of Institutions*. New York: The Macmillan Company.

 © 2026 Manuel Cabado García

Tema 4

Desarrollos de la teoría de la demanda: dualidad. Referencia a la integrabilidad de las preferencias y medidas de cambio en el bienestar. Elección en situaciones de riesgo. Otros desarrollos recientes.

4.1. Introducción ⏳ 1'

El consumidor, uno de los principales agentes económicos, tiene como objetivo maximizar la satisfacción que reporta el consumo de los bienes y servicios que se ofrecen en el mercado. Del estudio de su comportamiento se encarga la teoría de la demanda del consumidor, que trata de responder a dos preguntas:

1. Dados unos determinados precios y nivel de renta, ¿qué decisión de consumo adoptan los agentes?
2. ¿Cómo varían los agentes la decisión al modificar los parámetros de partida?

El enfoque ordinal de la utilidad permite derivar la función de utilidad y, a partir de esta, la demanda. Este enfoque se basa en:

- El supuesto de racionalidad y la existencia de la función de utilidad, aunque la utilidad no es directamente observable, y por lo tanto las demandas serán difíciles de determinar.
- Existencia de certidumbre, donde se supone que se conoce toda la información relevante, pero en la práctica existe incertidumbre al no conocerse con precisión todas

las variables bajo las cuales los agentes toman sus decisiones.

Los desarrollos de la propia teoría de la demanda neoclásica superan estas dos limitaciones, respectivamente, a través de:

- Teoría de la dualidad, un enfoque alternativo que permite obtener las funciones de demanda a partir de funciones observables o estimables, como la de gasto, sin necesidad de especificar la función de utilidad.
- Teoría de la elección en situaciones de incertidumbre, cuando la información es imperfecta y, aunque simétrica, se diferencia según Knight (1921) en:
  - Elección en situaciones de riesgo, cuando el agente puede asignar probabilidad a los distintos resultados de su decisión.
  - Elección en situaciones de incertidumbre, cuando no se pueden asignar probabilidades o estas son subjetivas.

Por consiguiente, el objetivo fundamental de la exposición es el desarrollo de las dos líneas de investigación anteriores a través del siguiente esquema:

1. Teoría de la dualidad, que se desarrollará matemáticamente y se expondrá su aplicabilidad.
2. Elección en situaciones de riesgo, donde se realizará un recorrido histórico, se mostrará su aplicabilidad y sus críticas[1].

4.2. Teoría de la dualidad ⏳ 10'30"

Comenzando con la teoría de la dualidad, el desarrollo de determinadas técnicas matemáticas en la primera mitad del siglo XX dio lugar al concepto de dualidad en los programas de optimización, que surge para hacer operativo el enfoque ordinal mencionado anteriormente.

La teoría de la dualidad estudia la decisión del consumidor mediante dos problemas equivalentes, maximizar la utilidad con una renta dada (primal) y minimizar el gasto necesario para alcanzar una utilidad dada (dual), cuyas soluciones están relacionadas entre sí[2]. La dualidad permitirá relacionar funciones de utilidad y demanda mediante el estudio del problema primal y del problema dual, siendo aplicable en el estudio de la teoría del consumidor, la economía de bienestar y muchas otras áreas de la Economía.

[1] La elección en situaciones de incertidumbre, al no admitir probabilidades objetivas, se apuntará conceptualmente en este bloque y se desarrollará en el Apéndice.

[2] La intuición es que da igual preguntar "¿cuál es la mayor satisfacción que puedo obtener con 100 euros?" que "¿cuánto dinero necesito como mínimo para alcanzar esa satisfacción?": ambas preguntas llevan a la misma cesta.

 © 2026 Manuel Cabado García

Elementos comunes Debido al carácter análogo de ambos problemas, los dos partirán del mismo conjunto de hipótesis y elementos. Específicamente compartirán la misma especificación de:

- Conjunto de elección: conjunto de consumo es $X = \mathbb{R}_+^K$[3]. A este conjunto factible se le imponen una serie de propiedades: no vacío, cerrado, acotado inferiormente y convexo. Sobre este conjunto se definen las preferencias del agente.
- Preferencias: se definen como una relación binaria "es al menos tan bueno como", $\succsim$, cumplen los axiomas de racionalidad[4] y los axiomas de regularidad[5], que garantizan que la relación binaria pueda representarse a través de una función de utilidad ordinal, $u(x)$.
- Conjunto presupuestario: Se parte del supuesto de que el consumidor opera en un mercado de competencia perfecta. Para adquirir los bienes disponibles el agente cuenta con una renta monetaria, $m > 0$. Este conjunto será no vacío, cerrado, acotado y convexo. El consumidor elige entre un conjunto dados unos precios, $p \gg 0$[6], y un nivel de renta, $m > 0$.

Problema primal

En este enfoque, el agente económico, con una restricción presupuestaria fija, busca la combinación de bienes que le permita maximizar la utilidad total.

El problema de maximización de utilidad se plantea como:

$$\text{máx} \quad u(x) \tag{4.1}$$

sujeto a

$$p \cdot x \leq m \tag{4.2}$$

La función de Lagrange correspondiente al problema de maximización:

$$\mathcal{L}(x, \lambda) = u(x) - \lambda \cdot (p \cdot x - m) \tag{4.3}$$

cuyas CPO, correspondientes a dos bienes, $K = 2$, son las siguientes:

$$\frac{\partial \mathcal{L}}{\partial x_1} : \frac{UMg_{x_1}}{p_1} = \lambda \tag{4.4}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} : \frac{UMg_{x_2}}{p_2} = \lambda \tag{4.5}$$

[3]El conjunto pertenece a los números reales, de dimensión K bienes y es estrictamente positivo.

[4]Se imponen los axiomas de completitud, reflexividad y transitividad, que garantiza relaciones de preferencia de pre-orden completo débil o relación de orden débil.

[5]Se imponen los axiomas de continuidad, monotonía estricta y convexidad estricta.

[6]El símbolo $\gg$ significa "mucho mayor que", en este caso los precios son mucho mayores que 0.

donde el multiplicador de Lagrange, λ, es la utilidad marginal de la renta gastada[7].

Combinando las ecuaciones (4.4) y (4.5) se obtiene la Relación Marginal de Sustitución:

$$RMS_{x_1,x_2} = \frac{UMg_{x_1}}{UMg_{x_2}} = \frac{p_1}{p_2} \tag{4.6}$$

que es la condición de tangencia entre la curva de indiferencia más alta y la restricción presupuestaria.

Sistema Completo de Funciones de Demanda Ordinarias o Marshallianas A partir de la solución y valorando en el óptimo (introduciendo la CPO en la RP) se obtiene el sistema de funciones de demanda ordinarias o marshallianas, $x^* = x(p, m)$, que (al no poner subíndice) expresa que la demanda de un bien dependerá de un vector de precios. Cumple las siguientes propiedades:

1. El sistema de Funciones de Demanda Ordinarias existe (teorema de Weierstrass) y es único (teorema convexidad), si las curvas de indiferencia son estrictamente convexas y por tanto la función de utilidad es estrictamente cuasi-cóncava.
2. A lo largo de la misma: i) Permanecen constantes la renta nominal y las preferencias de los agentes, ii) Varían el precio del bien que se trate, la renta real y el nivel de utilidad.
3. Observables y normales (es decir, se tiene en cuenta la variación del consumo debido a la variación de la renta real)
4. Continuas para todo el vector de precios y renta positivos.
5. Habitualmente decrecientes con respecto al precio propio y crecientes con respecto a la renta, aunque esta propiedad no es universal, quedando patente en excepciones como los bienes inferiores (demanda decreciente en la renta) y los bienes Giffen (demanda creciente en su propio precio).
6. Homogéneas de grado cero en renta y precios, lo que implica inexistencia de ilusión monetaria.
7. Se cumple la ley de Walras: gasta la totalidad de la renta.

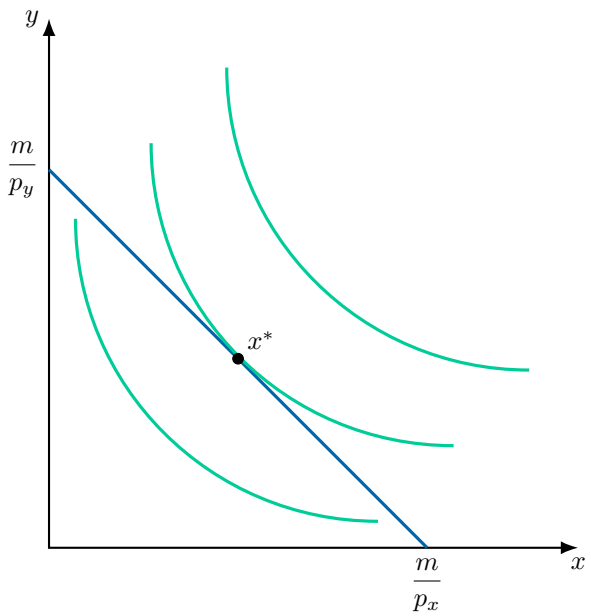
Análisis gráfico Gráficamente, en un plano de bienes x_1 y x_2 como el de la Figura 4.1 se representa el equilibrio derivado de la resolución de problema primal. Se dibuja en primer lugar la restricción del problema, es decir, la recta presupuestaria y, a continuación, se traza el punto de tangencia con la curva de indiferencia correspondiente, siendo este punto

[7] La interpretación económica es que las CPO establecen que la renta se satura y que la UMg del bien k se iguala al coste marginal de adquirirlo (su precio), multiplicado por el multiplicador de Lagrange, λ, que representa la utilidad marginal de la renta. Dicho de otra forma, se iguala la UMg de los bienes entre sí ponderadas por su precio. Es decir, lo que aumenta la utilidad máxima cuando la renta aumenta en una unidad, esto es, la UMg de la renta.

 © 2026 Manuel Cabado García

el vector de equilibrio x^*[8]. Por consiguiente, este vector de bienes x^* maximiza la utilidad del consumidor, dada la restricción presupuestaria, los precios y la renta[9].

Figura 4.1: Problema de Maximización de Utilidad



Aplicación Una vez resuelto el problema, sustituyendo las demandas óptimas, x^*, en la función de utilidad se obtiene la función indirecta de utilidad:

$$u\left(x\left(p,m\right)\right)=v\left(p,m\right) \tag{4.7}$$

Función Indirecta de Utilidad La función $v\left(p,m\right)$ es la llamada función indirecta[10] de utilidad, que representa la utilidad máxima alcanzable dados unos precios y nivel de renta, sin necesidad de solucionar el primal constantemente[11]. Posee las siguientes características:

1. Continua en p y m, por serlo la función de utilidad y las funciones de demanda marshallianas.

[8] Se denomina x-estrella

[9] Cabe señalar que $x^*=x\left(p,m\right)$ determina unas funciones que son homogéneas de grado 0 en precios y renta monetaria. Esto indica que las variables relevantes son las relativas, y que un cambio en la unidad de cuenta no modifica el equilibrio (ausencia de ilusión monetaria), al no alterar la restricción presupuestaria.

[10] Indirecta ya que depende de precios y renta y no de x.

[11] La FIU permite conocer los niveles de utilidad óptimos para cada conjunto dado de parámetros sin necesidad de repetir la resolución del problema primal (función de utilidad indirecta porque dependerá de precios y renta vía proceso de maximización, en vez de depender directamente de x como la función de utilidad $U(x)$)

2. Homogénea de grado 0 en p y m, por serlo también las funciones de demanda ordinarias[12].
3. No creciente en precios, p, y estrictamente creciente en renta, m, dado que una mejora del presupuesto siempre incrementa el bienestar y viceversa (Teorema de la Envolvente).
4. Cuasiconvexa en p y m, dado que la utilidad máxima que puede obtenerse de conjuntos presupuestarios promedios es inferior a la que se alcanza en conjuntos presupuestarios extremos.

Una vez caracterizado el problema primal y su solución, el problema dual permite abordar la decisión del consumidor desde una perspectiva alternativa, invirtiendo los papeles de la función objetivo y la restricción.

Problema dual

Ahora se estudiará el problema dual o problema de minimización de gasto para $p \gg 0$ y un nivel de utilidad, $\bar{u} > u(0)$, donde el gasto pasa a ser la función objetivo y el nivel de utilidad dado $\bar{u}$ la restricción del problema:[13]

$$\text{mín} \quad p \cdot x \tag{4.8}$$

sujeto a

$$u(x) \geq \bar{u} \tag{4.9}$$

conocido como el problema dual de la teoría de la demanda, y cuya función auxiliar de Lagrange es:

$$\mathcal{L}(x, \mu) = p \cdot x - \mu \cdot (u(x) - \bar{u}) \tag{4.10}$$

siendo las CPO:

$$\frac{\partial \mathcal{L}}{\partial x_1} : UMg_{x1} = \frac{p_1}{\mu} \tag{4.11}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} : UMg_{x_2} = \frac{p_2}{\mu} \tag{4.12}$$

donde el multiplicador de Lagrange, μ, tiene una interpretación inversa al primal ya que ahora representa el gasto marginal de la utilidad, es decir, cuánto debería aumentar el gasto si u aumentase marginalmente.

[12] La homogeneidad de la función de demanda ordinaria implica que ante incrementos proporcionales de los p y de la m la utilidad alcanzable no aumenta. Por consiguiente, las variables relevantes son las relativas

[13] La utilidad $u(0)$ es la utilidad proveniente del conjunto de consumo $x = (0, 0, \ldots, 0)$. La restricción $u > u(0)$ excluye aquellas situaciones no interesantes.

 © 2026 Manuel Cabado García

Combinando las ecuaciones (4.11) y (4.12) se obtiene la Relación Marginal de Sustitución:

$$RMS_{x_1,x_2} = \frac{UMg_{x_1}}{UMg_{x_2}} = \frac{p_1}{p_2} \tag{4.13}$$

Despejando μ de cualquiera de las dos CPO se obtiene, además, la relación de dualidad entre ambos multiplicadores:

$$\mu = \frac{p_1}{UMg_{x_1}} = \frac{p_2}{UMg_{x_2}} \quad \Rightarrow \quad \frac{1}{\mu} = \lambda \tag{4.14}$$

Es decir, el multiplicador del dual es el inverso del multiplicador del primal: mientras que λ mide cuánto aumenta la utilidad ante una unidad adicional de renta, μ mide cuánto debe aumentar el gasto para conseguir una unidad adicional de utilidad –son la misma tasa de intercambio entre renta y utilidad, vista desde direcciones opuestas–.

Sistema de funciones de demanda compensadas o hicksianas La solución del programa anterior define el sistema de funciones de demanda compensadas o hicksianas, $x^* = h(p, u)$.
El calificativo de compensadas proviene del hecho de que el presupuesto o gasto ya no es fijo sino variable, de modo que ante cualquier variación en los precios, la utilidad se mantendrá constante y dicho gasto se reajustará para acomodar o compensar dicha variación. Es decir, que indican las cantidades que comprarían los consumidores si se les compensara totalmente por los efectos renta de las variaciones del precio. Las demandas hicksianas solo tienen en cuenta el ES, no el ER.
Propiedades:

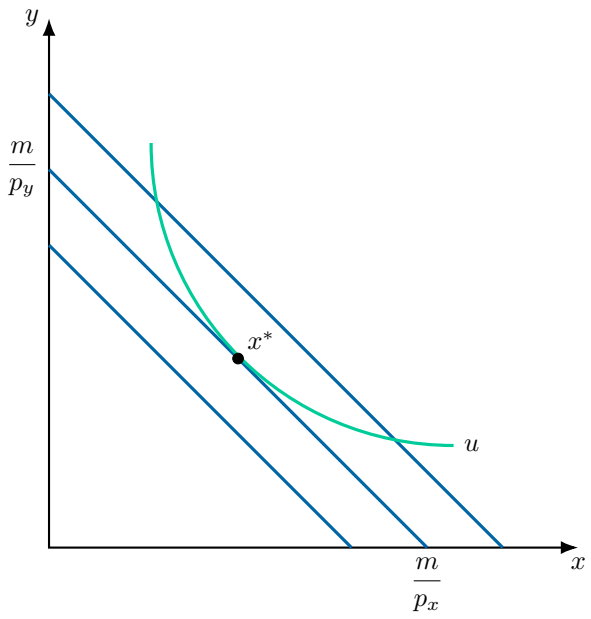
1. El sistema de Funciones de Demanda Compensada existe y es único (Teorema de las Funciones Implícitas).
2. Las funciones de demanda compensada son continuas en $(p, \bar{u})$ (Teorema del Máximo).
3. No es observable (porque U no lo es). No pueden estimarse empíricamente de manera directa. Será gracias a las relaciones de dualidad que sí podrán deducirse a partir de funciones marshallianas.
4. A lo largo de la misma: i) permanecen constantes el nivel de utilidad, es decir, la renta real en el sentido de Hicks, y las preferencias del agente, ii) varían los precios y la renta nominal, que se ajusta para compensar cada cambio de precio.
5. Es homogénea de grado 0 en P (sólo cuentan los precios relativos).
6. Se cumple el "no exceso de utilidad" (no excess utility, Mas-Colell, Whinston y Green, 1995), esto es, en el óptimo la restricción de utilidad siempre se cumple

con igualdad, de modo que el consumidor alcanza exactamente el nivel de utilidad requerido y no más.

7. Solo mide los ES. Las funciones de demanda compensadas cumplen la "ley de demanda compensada", es decir, precio y cantidad nunca se moverán en la misma dirección (no hay posibilidad de bienes Giffen en las demandas hicksianas por su carácter compensado). En el caso de un bien normal, la curva de demanda compensada es siempre más inclinada que la marshalliana.

Análisis gráfico Gráficamente, en un plano de bienes x_1 y x_2 como el de la Figura 4.2 se representa –al contrario que lo que se hizo en la Figura 4.1 con el problema primal– en primer lugar la curva de indiferencia que corresponde al nivel de utilidad u, que pasa a ser la restricción del problema. En el problema dual, el equilibrio, $x^* = h(p, u)$, es la consecuencia de confrontar esta curva con la recta de balance tangente a la misma, teniendo en cuenta que la pendiente de la recta seguirá siendo el cociente de los precios de los dos bienes considerados.

Figura 4.2: Problema de Minimización de Costes



Aplicación De forma análoga a la función indirecta de utilidad, sustituyendo la demanda hicksiana, x^*, en la función objetivo, $p \cdot x$, en el óptimo se obtiene la función de valor (óptimo) del problema, que es la función de gasto[14]:

$$p \cdot h(p, \bar{u}) = g(p, \bar{u}) \tag{4.15}$$

[14]Nótese que la función de gasto es la inversa de la función indirecta de utilidad porque contienen la misma información y ambas están determinadas por la misma función de utilidad.

 © 2026 Manuel Cabado García

Función de gasto La función $g(p, \bar{u})$ es la denominada función de gasto, que determina el gasto mínimo en el que el consumidor tiene que incurrir a unos precios, p, para alcanzar un nivel determinado de utilidad, sin necesidad de estar resolviendo el dual constantemente. Posee las siguientes características:

1. Estrictamente positiva.
2. Continua en precios, p, y utilidad, u, por serlo las funciones de demanda compensadas.
3. Homogénea de grado 1 en precios, por serlo de grado 0 las funciones de demanda compensadas. Esto implica que ante variaciones proporcionales de los precios, el gasto variará en esa misma proporción.
4. Estrictamente creciente en u y no decreciente en p, un aumento de la utilidad requerida aumenta siempre el gasto mínimo para lograrlo y viceversa.
5. Cóncava en precios. Ante un aumento de ciertos precios, el consumidor podría comportarse de manera pasiva y no modificar su cesta. Sin embargo, su comportamiento optimizador le lleva a sustituir bienes, por lo que el incremento del gasto será menor que el derivado del incremento de precios.

Recapitulando, se han obtenido todos los elementos necesarios (funciones de demanda y funciones valor (gasto), con sus propiedades) para pasar a continuación a establecer las relaciones existentes entre los diferentes componentes del problema primal y el dual.

Teorema básico de la dualidad

Una vez estudiados el problema primal y el dual se podrá establecer el denominado Teorema Básico de la Dualidad.

Teorema Básico de la Dualidad determina que, si una cesta x^* soluciona el primal con renta m y se plantea el dual exigiendo precisamente la utilidad máxima alcanzada, $u(x^*)$, entonces x^* también soluciona el dual y el gasto mínimo coincide con m.

De esta forma, ambos problemas son idénticos siempre que se realice el debido intercambio de información entre problemas:

1. Existencia: Existe una solución al problema primal si existe una solución al problema dual.
2. Identidad entre ambas soluciones: Cuando la renta, la cual actúa como restricción en el problema primal es igual al gasto mínimo en el problema dual, y la utilidad que actúa como restricción en el problema dual es igual a la utilidad máxima en el primal entonces en el óptimo ambas soluciones serán idénticas. Es decir, siendo el vector de precios el mismo en ambos problemas, en el óptimo las cantidades demandadas

según la función de demanda ordinaria (marshalliana) y compensada (hicksiana) coinciden.

Además, es importante señalar que la solución del primal y el dual solo coinciden en un punto, esto es en la cesta óptima, porque ante variaciones en los precios relativos la demanda hicksiana mantiene constante la utilidad, mientras que la demanda ordinaria incorpora además el efecto de esas variaciones sobre la capacidad adquisitiva, esto es, el ER.

No obstante, existe un par de relaciones de equivalencia/dualidad entre las funciones resultado del problema primal y del problema dual que se derivan de este "Teorema Básico de Dualidad". Son relaciones referidas tanto a las funciones de valor (FIU y FG) como a las funciones de demanda (marshallianas y hicksianas).

Dualidad en las funciones de valor y de demanda Concretamente, estas relaciones se manifiestan de las siguientes dos formas:

1. Dualidad entre las funciones de valor: La función indirecta de utilidad es la inversa de la función de gasto, pudiéndose obtener mediante la sustitución y la reorganización de términos y su interpretación es que todo nivel de utilidad óptimo –máximo– tiene asociado un nivel de gasto óptimo –mínimo–, y viceversa. Por ejemplo, para una función de utilidad de sustitutos perfectos, $u = x_1 + x_2$:

$$v(p, \bar{m}) = \frac{m}{p_1 + p_2} \longrightarrow g(p, \bar{u}) = (p_1 + p_2) \cdot u \tag{4.16}$$

[15]

2. Dualidad entre funciones de demanda: las funciones de demanda marshalliana y hicksiana pueden obtenerse mediante el problema inverso, a partir de la sustitución de argumentos –la demanda marshalliana se puede obtener a partir de la hicksiana sustituyendo por la función indirecta de utilidad–, siendo así una extensión del teorema básico de la dualidad:

$$x_i(p, m) = h_i(p, v(p, m)) \tag{4.17}$$

$$h_i(p, \bar{u}) = x_i(p, g(p, \bar{u})) \tag{4.18}$$

[15] La forma concreta de esta dualidad depende de la función de utilidad: para una Cobb-Douglas, $u = x_1^a x_2^{1-a}$, la demanda marshalliana es $x_1 = am/p_1$, $x_2 = (1-a)\, m/p_2$, de donde:

$$v(p, m) = m \cdot a^a (1-a)^{1-a} \cdot p_1^{-a} p_2^{-(1-a)} \longrightarrow g(p, \bar{u}) = \bar{u} \cdot \frac{p_1^a p_2^{1-a}}{a^a (1-a)^{1-a}}$$

es decir, potencias de los precios con exponentes iguales a las cuotas de gasto, en vez de la fracción simple del ejemplo de sustitutos perfectos.

 © 2026 Manuel Cabado García

Cuadro 4.1: Resumen de la teoría de la dualidad

Problema	**Funciones de demanda**	**Denominación y propiedades**
Primal $\begin{array}{ll} \text{máx} & u(x) \\ \text{s.a} & p \cdot x = m \end{array}$	$x = x(p, m)$	ordinales o marshallianas observables homogéneas de grado 0 en p y m
Dual $\begin{array}{ll} \text{mín} & p \cdot x \\ \text{s.a} & u(x) = \bar{u} \end{array}$	$x = h(p, \bar{u})$	compensadas o hicksianas no observables homogéneas de grado 0 en p (relativos)
	Solución	**Denominación y propiedades**
	$u(x(p, m)) = v(p, m)$	función indirecta de utilidad no observables homogéneas de grado 0 en p y m
	$p \cdot h(p, \bar{u}) = g(p, \bar{u})$	función de gasto observables homogéneas de grado 1 en p

Aplicaciones de la teoría de la dualidad

Por otro lado, del análisis de la dualidad se deducen una serie de resultados complementarios de enorme importancia a la hora de realizar estimaciones econométricas, análisis de bienestar y clasificar los bienes en base a su complementariedad o sustituibilidad.

Obtención demanda ordinaria y compensada sin conocer la función de utilidad

Una de esas aplicaciones es la obtención de forma simple de las funciones de demanda ordinaria y compensada sin necesidad de conocer la función de utilidad a través de la Identidad de Roy y el Lema de Shephard.

La Identidad de Roy (1947), pone en relación la función indirecta de utilidad con las funciones de demanda marshallianas. Por consiguiente, dada la función indirecta de utilidad se obtiene la demanda marshalliana[16]:

$$x_k(p, m) = -\frac{\partial v(p, m) / \partial p_k}{\partial v(p, m) / \partial m}, \qquad \forall k = 1, \ldots, K \tag{4.19}$$

Por su parte, el Lema de Shephard[17], muestra la relación entre la función de gasto y las funciones de demanda hicksianas. Por lo tanto, dada la función de gasto se pueden

[16] No usar el término entero, en vez de $x_k(p, m)$, usar x_k^m, o $x(\cdot)$, y lo mismo para v y g, ya que se sabe lo que son

[17] No confundir con el Lema de Hotelling, que relaciona la función de beneficios con las funciones de oferta y demanda de factores de la empresa.

conocer las demandas compensadas/hicksianas:

$$\frac{\partial g\left(p,u\right)}{\partial p_k} = h_k\left(p,u\right), \qquad \forall k = 1,\ldots,K \tag{4.20}$$

Obtención ecuación de Slutsky

En segundo lugar, la ecuación de Slutsky[18] relaciona las derivadas de las funciones hicksianas y marshallianas, permitiendo descomponer los efectos de la variación de un precio en efecto renta y efecto sustitución[19].

Analíticamente, se parte de una situación donde las demandas hicksianas y marshallianas coinciden[20] para la asignación óptima:

$$h\left(p,u\right) = x\left(p,m\right) \tag{4.21}$$

donde se puede hacer que la demanda marshalliana dependa de u sustituyendo m por la función de gasto, $g\left(p,\bar{u}\right)$:

$$h\left(p,u\right) = x\left(p,g\left(p,\bar{u}\right)\right) \tag{4.22}$$

por lo tanto es análogo derivar la demanda hicksiana del bien x_i con respecto al precio del bien j, p_j, que derivar la demanda marshalliana del bien x_i siguiendo la regla de la cadena[21]:

$$\frac{\partial h_i\left(p,u\right)}{\partial p_j} = \frac{\partial x_i\left(p,m\right)}{\partial p_j} + \frac{\partial x_i\left(p,m\right)}{\partial m}\cdot\frac{\partial g\left(p,u\right)}{\partial p_j} \tag{4.23}$$

Aplicando el lema de Shephard, se podrá sustituir el último término que está multiplicando por la demanda hicksiana, que será equivalente a la marshalliana:

$$\frac{\partial h_i\left(p,u\right)}{\partial p_j} = \frac{\partial x_i\left(p,m\right)}{\partial p_j} + \frac{\partial x_i\left(p,m\right)}{\partial m}\cdot x_j\left(p,m\right) \tag{4.24}$$

[18] Introducida por Slutsky (1915) y desarrollada más adelante por Hicks y Allen (1934)

[19] El Tema 3.A ya utiliza esta ecuación, pero solo en su versión particular para un bien frente a su propio precio ($\partial x/\partial p_x$), como herramienta descriptiva para clasificar bienes en normales, inferiores o Giffen, dando por supuesta la demanda hicksiana sin derivarla. La clasificación de sustitutivos/complementarios mediante elasticidad cruzada se trata allí aparte, sin pasar por esta ecuación. Aquí, en cambio, se deriva formalmente su versión general, para cualesquiera bienes i y j, partiendo de que en el óptimo la demanda hicksiana y la marshalliana coinciden y aplicando el análisis de dualidad: es tanto la justificación matemática del caso particular que aquel tema ya empleaba, como la base necesaria para clasificar bienes complementarios/sustitutivos netos y para la simetría que exige el Teorema de la Integrabilidad, que sí requieren $i \neq j$.

[20] Que las demandas coincidan en un punto no quiere decir que sean idénticas siempre

[21] Truco para presentación rápida y limpia: poner todo término dentro de los paréntesis de forma $(\cdot)$, por ejemplo $h_i\left(\cdot\right)$. Se irá más rápido y en el último paso que hay que cambiar los términos solo habrá que borrar e intercambiar dos términos, cambiar el signo $+$ por menos $-$ y ya se tendrá lista la ecuación de Slutsky

 © 2026 Manuel Cabado García

y reorganizando los términos y cambiando el signo + por un −[22] en (4.24) se obtendrá la ecuación de Slutsky, donde el efecto total, ET, es la suma del efecto sustitución, ES, y del efecto renta, ER:

$$\underbrace{\frac{\partial x_i(p,m)}{\partial p_j}}_{\text{Efecto total}} = \underbrace{\frac{\partial h_i(p,u)}{\partial p_j}}_{\text{Efecto sustitución}} \underbrace{-\frac{\partial x_i(p,m)}{\partial m} \cdot x_j(p,m)}_{\text{Efecto renta}} \tag{4.25}$$

Además, la ecuación de Slutsky se puede expresar en términos de elasticidades donde la elasticidad precio de la demanda marshalliana de x_i respecto a p_j, $\varepsilon_{i,j}$, será igual a la elasticidad precio de la demanda hicksiana de x_i respecto de p_j, $\varepsilon^C_{i,j}$, menos la elasticidad con respecto a la renta, m, de x_i, ε_m, multiplicada por la participación en el gasto del bien x_j, s_j[23]:

$$\varepsilon_{i,j} = \varepsilon^C_{i,j} - s_j \cdot \varepsilon_m \tag{4.26}$$

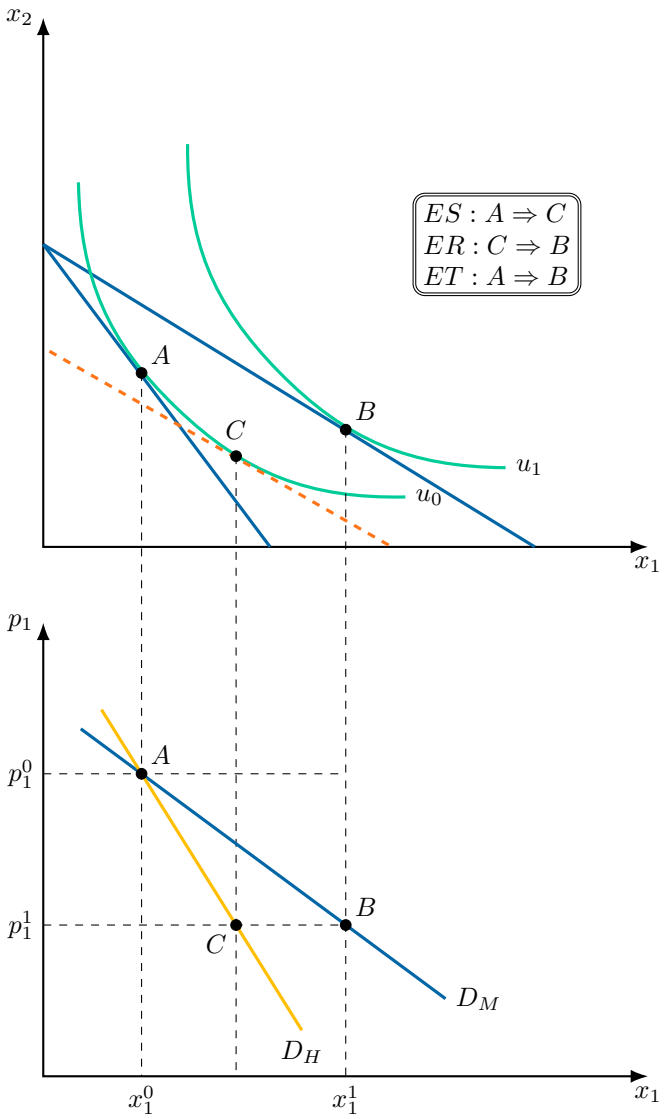
Análisis gráfico Gráficamente, en un plano de los bienes x_1 y x_2 como el de la Figura 4.3:

- Se parte de una situación inicial con una recta presupuestaria y una curva de indiferencia, u_0, tangente dando lugar al punto de equilibrio A.
- A través de un ejercicio de estática comparativa con una disminución en el precio del bien x_1, $\downarrow p_1$, se podrá descomponer el efecto renta y sustitución.
- Ante la disminución del precio, la recta presupuestaria pivotará sobre el eje del bien x_2, expandiendo el conjunto presupuestario y permitiendo al consumidor alcanzar una nueva curva de indiferencia, u_1, que le reportará más utilidad en el punto B. Para obtener el efecto sustitución, se desplazará la nueva recta presupuestaria hasta que sea tangente con la antigua curva de indiferencia u_0 en el punto C.
- Por lo tanto, el efecto sustitución sería el paso de A a C, el efecto renta el paso de C a B y el efecto total el paso de A a B.
- Para obtener la demanda marshalliana y hicksiana se trasladarán todos estos puntos a un plano adyacente de precios del bien x_1, p_1, y cantidades del bien x_1.
- El punto A se obtendrá para el precio inicial, p^0_1, y la cantidad inicial, x^0_1, mientras que el punto B se dará para el nuevo precio, p^1_1, y la nueva cantidad, x^1_1.
- Uniendo ambos puntos se obtendrá la demanda marshalliana, D_M.
- El punto C se correspondería con el punto B solo que comprando una menor cantidad y uniendo el punto A con el punto C se obtendrá la demanda hicksiana, D_H.

[22] El signo negativo representa la relación inversa entre cambios en precios y cambios en la capacidad de compra, en el caso analizado, la reducción del precio supone un incremento de la capacidad de compra.

[23] Donde $s_j = \dfrac{p_j \cdot x_j}{m}$

Figura 4.3: Descomposición efecto renta y sustitución



En suma, de la ecuación de Slutsky se obtiene que:

- El efecto sustitución es negativo, por el carácter semidefinido negativo de la matriz de efecto sustitución[24].
- El signo del efecto renta depende del bien:
  - Si la elasticidad renta es positiva entonces el bien es normal.
  - Si la elasticidad renta es negativa entonces el bien es inferior.
- Si el bien es normal, el efecto renta refuerza al efecto sustitución y la demanda tiene pendiente negativa. Si el bien es inferior, el efecto renta se opone al efecto sustitución y el resultado depende de cuál de los dos domine:

[24] En ese sentido, su signo es contrario al de la variación del precio

 © 2026 Manuel Cabado García

  - Si $|ES| > |ER|$, la demanda sigue teniendo pendiente negativa.
  - Si $|ES| < |ER|$, la demanda tiene pendiente positiva y se está ante un bien Giffen.

- De lo anterior se derivan las diferencias en la pendiente de la demanda hicksiana y marshalliana:
  - Para bienes normales, la demanda hicksiana tendrá una mayor pendiente que la marshalliana debido al efecto renta negativo presente únicamente en la demanda marshalliana[25].
  - Para bienes inferiores ocurre lo contrario: la demanda marshalliana es más inclinada que la hicksiana, porque el efecto renta compensa en parte el efecto sustitución. En el caso de bienes frontera[26], con efecto renta nulo, $ER = 0$, ambas tienen la misma pendiente.
- El signo del efecto sustitución cruzado, $\frac{\partial h_i}{\partial p_j}$, puede variar pero las relaciones sí son simétricas, $\frac{\partial h_i}{\partial p_j} = \frac{\partial h_j}{\partial p_i}$:
  - Si el efecto sustitución cruzado, es decir, la variación de la demanda hicksiana del bien i ante un cambio del precio del bien j, es positivo, los bienes i y j son sustitutivos netos[27].
  - Si es negativo, son complementarios netos.
  - Si es nulo, son independientes.

Problema de la integrabilidad de preferencias

En tercer lugar, una de las principales preguntas para validar la teoría neoclásica de la demanda es qué características deben cumplir las demandas estimadas para garantizar que provienen de unas preferencias de buen comportamiento. El pionero en resolver el problema de la integrabilidad de preferencias fue el ingeniero y economista italiano Antonelli (1886).

La respuesta a esta pregunta se conoce hoy en día como el Teorema de la Integrabilidad[28], que establece que un sistema de demandas marshallianas está generado por

[25] Las curvas de demanda hicksiana solo recogen el efecto sustitución, ya que a lo largo de ellas la utilidad se mantiene constante: la renta monetaria se ajusta tras cada cambio de precio para compensar exactamente la variación del poder adquisitivo, de modo que el efecto renta queda anulado. Por eso, para un bien normal, la demanda hicksiana es más inclinada que la marshalliana

[26] El término bienes frontera designa el caso límite entre bien normal e inferior, en el que el efecto renta es exactamente nulo ($ER = 0$) y las demandas marshalliana e hicksiana coinciden en pendiente –es el caso, por ejemplo, de una función de utilidad cuasilineal–.

[27] La intuición es que, con la utilidad constante, si se encarece el café y el consumidor compra más té, el té y el café son sustitutivos netos. Si compra menos azúcar, el café y el azúcar son complementarios netos.

[28] Resuelto formalmente por Hurwicz y Uzawa (1971), a partir del trabajo pionero de Antonelli.

una función de utilidad creciente y cuasicóncava si y solo si dicho sistema satisface las siguientes propiedades:

1. Ley de Walras, el sistema deberá agotar el total del gasto.

$$\Sigma p_i \cdot x_i = m \tag{4.27}$$

2. Matriz de efectos sustitución simétrica, es decir, el efecto sustitución que una variación del precio del bien j tiene sobre la demanda del bien i es igual al que una variación del precio de i tiene sobre la demanda de j. Por el Teorema de Frobenius, esta simetría garantiza que el sistema de demandas pueda integrarse en una función de gasto[29]. Esta es la condición de simetría de los efectos sustitución cruzados[30].
3. Matriz de efectos sustitución semidefinida negativa, lo que equivale a la concavidad de la función de gasto en precios.

Por consiguiente, a partir del sistema de demandas marshallianas se podrá integrar la función de gasto y por el Teorema Básico de la Dualidad de la función de gasto se podrá derivar la función indirecta de utilidad, de la que se extrae la función de utilidad.

Desde un punto de vista práctico la integrabilidad es relevante para el análisis de bienestar, pues se podrá recuperar la función de gasto que es clave para realizar dicho análisis.

Sistemas de demanda empíricos

Esta relevancia práctica de la integrabilidad culminó en el desarrollo de los sistemas de demanda flexibles: funciones de gasto que se ajustan directamente a los datos observados y en las que las restricciones teóricas de homogeneidad y simetría se traducen en condiciones sencillas sobre los parámetros, que pueden imponerse o contrastarse, mientras que la agregación se cumple por construcción. El más influyente es el Sistema Casi Ideal de Demanda (AIDS) de Deaton y Muellbauer (1980), que sigue siendo hoy la referencia estándar en econometría aplicada de la demanda. Su extensión cuadrática, el QUAIDS de Banks et al. (1997), es de uso habitual en organismos como el Banco Mundial, el FMI o la OCDE-FAO para analizar reformas de IVA y subsidios alimentarios (véase Apéndice 4.A.7).

[29] La importancia de estas características reside en el problema de integrabilidad:

- Matriz semidefinida negativa: garantiza que la matriz de sustitución sea consistente con una función de gasto cóncava.
- Matriz simétrica (Teorema de Frobenius): garantiza que ese sistema de derivadas sea integrable, es decir, que exista una única función de gasto –y, por tanto, una función de utilidad– de la que puedan derivarse las demandas observadas.

[30] Es decir, el orden de diferenciación de la función de gasto es irrelevante, $\frac{\partial^2 g}{\partial p_i \partial p_j} = \frac{\partial^2 g}{\partial p_j \partial p_i}$, de modo que ambos efectos sustitución cruzados coinciden en signo y en magnitud.

 © 2026 Manuel Cabado García

Cálculo monetario de alteraciones en el bienestar

A su vez, la teoría de la dualidad proporciona instrumentos que cuantifican monetariamente la variación de la utilidad. La función de gasto determina el gasto mínimo para alcanzar determinado nivel de utilidad.

Existen dos medidas alternativas basadas en la función de gasto:

- Analizando la demanda de Hicks mediante variaciones compensadas o equivalentes.
- Analizando la demanda marshalliana mediante el excedente del consumidor.

En función de qué precios de referencia se tomen se computará la variación equivalente o la variación compensatoria.

Variación equivalente La variación equivalente, VE, al mantener como referencia los precios iniciales se utiliza para cuantificar monetariamente el cambio de bienestar al producirse un cambio hipotético de precios. Por lo tanto, es la cantidad de renta que, a los precios iniciales, habría que dar al consumidor, o quitarle, para que alcanzase el mismo nivel de utilidad que obtiene tras el cambio de precios. Si es positiva, el cambio mejora su bienestar y, si es negativa, lo empeora[31].

Variación compensatoria La variación compensatoria, VC, al tomar como referente los precios finales, asume que el cambio en precios va a tener lugar y cuantifica el cambio de bienestar que corresponde a la compensación. Por lo tanto, es la cantidad de renta que, a los precios finales, habría que quitar al consumidor, o darle, para que volviese al nivel de utilidad que tenía antes del cambio de precios[32]. Si es positiva, el cambio mejora su bienestar y, si es negativa, lo empeora[33].

Análisis gráfico Gráficamente, en un plano de bienes x_1 y x_2 como el de la Figura 4.4, se continuará con el análisis de la Figura 4.3,

- Donde se puede observar que la diferencia entre nueva recta presupuestaria y la que se trasladó hasta la antigua utilidad u_0 es la variación compensatoria.

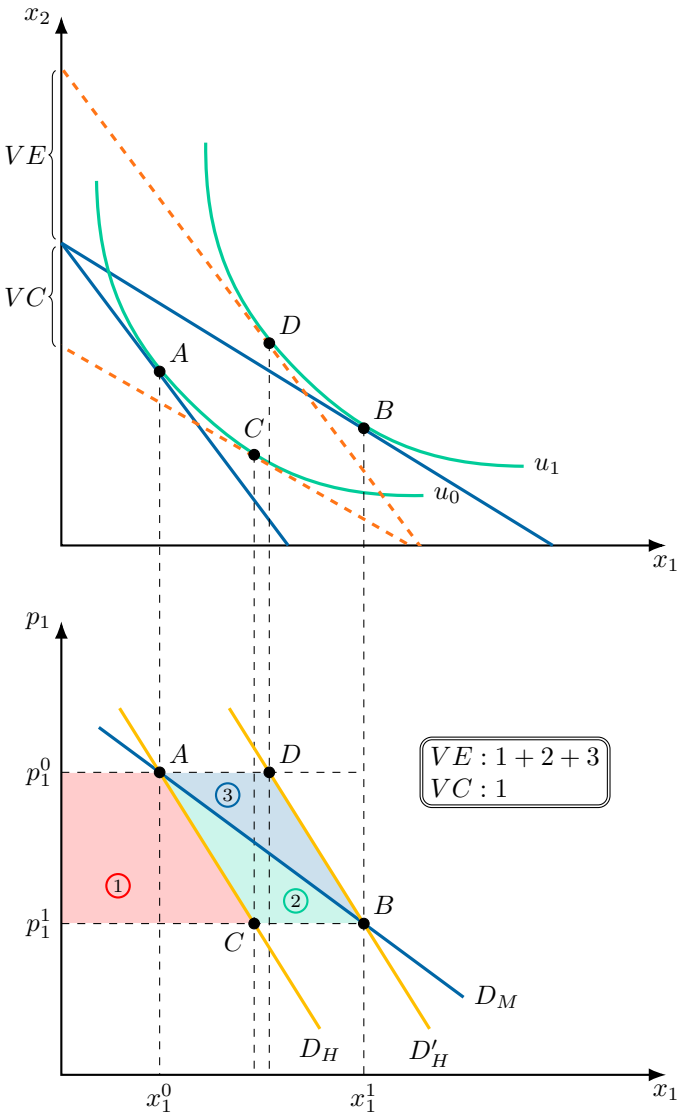
[31] En el primer caso el agente pedirá más dinero a cambio de que se mantengan los precios originales mientras que en el segundo el agente estará dispuesto a pagar dinero a cambio de mantener los precios previos.

[32] La intuición es que la VE pregunta cuánto dinero equivale al cambio sin que llegue a producirse y la VC, cuánto dinero lo compensa una vez producido. Por ejemplo, ante una bajada del precio del café, la VE es lo mínimo que el consumidor aceptaría por renunciar a la bajada y la VC, lo máximo que pagaría por disfrutarla.

[33] En el primer caso el agente estará dispuesto a pagar dinero a cambio de mantener los precios finales mientras que en el segundo el agente pedirá más dinero a cambio de mantener los precios finales y continuar percibiendo la misma utilidad inicial

- Por su parte, para obtener la variación equivalente se trasladará la antigua restricción hacia la nueva utilidad, u_1, hasta el punto de tangencia D, donde la diferencia entre esta y la nueva restricción es la variación equivalente, VE.
- Trasladando todos estos puntos a un plano adyacente de p_1 y x_1, donde, como consecuencia del cambio en el nivel de utilidad, la demanda hicksiana se expande a D'_H, surgiendo de unir los puntos D y B.
- Se pueden observar tres áreas: la 1, delimitada entre los dos precios, el punto A y el C, la 2, delimitada entre los puntos A, B y C y la 3, delimitada por los puntos A, D y B.
    - La Variación compensatoria comprendería el área 1, mientras que
    - La variación equivalente surgiría de la suma de las tres áreas.

Figura 4.4: Variación compensatoria y equivalente



 © 2026 Manuel Cabado García

Estudio analítico Analíticamente, la variación compensatoria sería la integral desde los nuevos precios, p_1, a los antiguos, p_0, de la demanda hicksiana inicial, $h(p, u_0)$:

$$VC = \int_{p_1}^{p_0} h(p, u_0) \cdot dp_1 \tag{4.28}$$

mientras que la variación equivalente sería la integral desde los nuevos precios, p_1, a los antiguos, p_0, de la demanda hicksiana final, $h(p, u_1)$:

$$VE = \int_{p_1}^{p_0} h(p, u_1) \cdot dp_1 \tag{4.29}$$

Entre los aspectos positivos de estos instrumentos está que permite jerarquizar un espectro de utilidades y con ello permitir cuantificar monetariamente cambios en el bienestar y permite estimar la variación de bienestar cuando varían uno o varios precios, algo que no tendrán otras medidas[34].

No obstante, entre los aspectos negativos destaca que estas medidas comparan con unos precios no observables –ausencia de contrafáctico–, lo que supone un problema y sobre todo que será necesario disponer de una función de gasto –equivalente a disponer de información completa–, cuando lo normal es carecer de ella.

Excedente del consumidor En caso de no disponer de información suficiente se podrán evaluar dichos impactos mediante la variación del excedente del consumidor marshalliano, VEC. El excedente del consumidor es la diferencia entre cuánto valora un bien y el pago que realiza por el mismo, por lo que su variación es el cambio en el bienestar derivado de la modificación del precio del bien:

$$VEC = \int_{p_1}^{p_0} D_M \cdot dp_1 \tag{4.30}$$

aunque lo realiza mediante la demanda marshalliana, que está sesgada[35].

Por ello destaca:

1. Ser una herramienta menos potente que las anteriores, dado que solo admite la variación de un precio y no un espectro.
2. No estará sesgada cuando el efecto renta sea nulo, como ocurre con preferencias cuasilineales.

[34] Además, permite realizar un análisis de bienestar dentro de modelos macroeconómicos como:

- Análisis coste beneficio, permitiendo evaluar el impacto de diferentes impuestos indirectos sobre bienes, permitiendo estimar la distorsión que estos producen y cuantificar la VC y la VE.
- Realizar un análisis de bienestar en el modelo canónico de la Nueva Economía Keynesiana, permitiendo evaluar cuál es la política monetaria óptima, o estudiar cómo responder ante un shock de inflación y desempleo simultáneos.

[35] Las medidas exactas del cambio de bienestar, la variación compensatoria y la equivalente, se obtienen con la demanda hicksiana, que mantiene constante la utilidad. El excedente, al usar la marshalliana, mezcla el efecto sustitución con el efecto renta.

3. En presencia de un efecto renta no nulo, tendrá un valor intermedio entre la variación equivalente y la compensatoria. Para un bien normal, la VEC subestima la VE y sobreestima la VC y su valor en la Figura 4.4 corresponde con la suma de las áreas 1 y 2.

Índices de precios Finalmente, si se desconoce la función marshalliana se podrá realizar un análisis de bienestar aún más reducido si se conocen las cestas de consumo iniciales, finales o ambas. La variación equivalente y la variación compensatoria se podrán aproximar mediante los índices de precios, que buscarán ajustar la renta del agente para que alcance la cesta de consumo, final o inicial, a los precios iniciales o finales respectivamente. Destacan:

- El Índice de Laspeyres, IL, que toma como referencia la cesta de consumo inicial, mide el cambio relativo en el coste necesario para mantener la cesta de consumo inicial (es decir, IPC o IPCA, aunque realmente son índices encadenados, poseen base móvil)
$$IL = \frac{p^1 \cdot x^0}{p^0 \cdot x^0} \tag{4.31}$$
si se considera un aumento de precios, el IL sobreestima el aumento del coste de mantener el mismo nivel de bienestar individual cuando se incrementen los precios, puesto que no se consideran las relaciones de sustituibilidad entre los bienes.
- El Índice de Paasche, IP, que toma como referencia la cesta de consumo final, mide el cambio relativo en el coste necesario para adquirir la cesta de consumo final a los precios nuevos (es decir, Deflactor del PIB)[36]
$$IP = \frac{p^1 \cdot x^1}{p^0 \cdot x^1} \tag{4.32}$$
subestima el aumento del coste de vida cuando suben los precios.

Estas aproximaciones buscan que el agente alcance una cesta de consumo determinada, no un nivel de utilidad constante, como hacen la VE y la VC. Por ello, presentan un sesgo predecible, ya que no consideran la sustitución entre bienes:

- El Índice de Paasche (IP) sirve para aproximar la VE: la sobrestima cuando los precios bajan y la subestima cuando suben.
- El Índice de Laspeyres (IL) sirve para aproximar la VC: la subestima cuando los precios bajan y la sobrestima cuando suben.

[36] Los índices de Laspeyres y Paasche son índices de precios de ponderaciones fijas, frente a los índices de ponderaciones encadenadas.

 © 2026 Manuel Cabado García

Concretamente, ante el impacto sobre el bienestar de la disminución del precio[37] que se ha analizado, el orden de magnitud de los instrumentos estudiados quedaría en el siguiente orden: $IL < VC < VEC < VE < IP$[38]. Este orden refleja el grado de aproximación al verdadero cambio en bienestar, desde el índice más conservador (Laspeyres) hasta el más generoso (Paasche), pasando por las medidas basadas en utilidad y demanda[39].

Cuadro 4.2: Orden de magnitud de medidas de bienestar

Tipo de bien	**Cambio de precios**	**Orden de magnitud (de menor a mayor)**	**Comentario clave**
Bien normal	↓ Bajada	$IL < VC < VEC < VE < IP$	Efecto renta positivo. Todas las medidas captan mejora
	↑ Subida	$IP < VE < VEC < VC < IL$	Efecto renta refuerza la pérdida de bienestar

[37] Si subiera el orden sería inverso

[38] No obstante, cabe destacar que estas herramientas solo serán válidas para evaluar el bienestar de un solo individuo. Para realizar un análisis colectivo mediante la agregación de preferencias individuales, estas herramientas solo serán válidas:

- Cuando todos los agentes tengan las mismas preferencias.
- Cuando las preferencias sean cuasi-lineales.
- Cuando solo sean relevantes las preferencias de un solo agente, concretamente el representativo que se utiliza en los modelos macroeconómicos con fundamentación microeconómica.

Para el resto de casos, el análisis de bienestar social tratará de rellenar estas lagunas buscando criterios desde el punto de vista colectivo o agregado.

[39] Comentarios:

- IL (Laspeyres): Usa la cesta inicial, ignora sustitución → subestima VC cuando bajan precios, sobreestima cuando suben.
- IP (Paasche): Usa la cesta final, ignora sustitución → sobreestima VE cuando bajan precios, subestima cuando suben.
- VC y VE: Basadas en demanda hicksiana, más precisas porque mantienen constante la utilidad.
- VEC: Basada en demanda marshalliana, intermedia entre VC y VE.

Tipo de bien	Cambio de precios	Orden de magnitud (de menor a mayor)	Comentario clave
Bien inferior	↓ Bajada	$IL < VC < VEC < VE < IP$	Efecto renta negativo. Puede contrarrestar el efecto sustitución
	↑ Subida	$IP < VE < VEC < VC < IL$	Riesgo de bien Giffen si ER > ES
Bien Giffen	↓ Bajada	$IL < VC < VEC < VE < IP$	Demanda disminuye pese a la bajada del precio (paradoja de Giffen). VC puede ser negativa
	↑ Subida	$IP < VE < VEC < VC < IL$	Demanda aumenta con el precio. VC puede ser positiva
Bien neutral	↓ o ↑	$IL = VC = VEC = VE = IP$	No hay efecto sustitución ni renta. Todas las medidas coinciden

4.3. Elección en situaciones de riesgo ⏳ 8'

Hasta ahora se ha analizado la decisión del consumidor bajo certeza, donde precios y renta son conocidos con seguridad. Sin embargo, en la realidad las elecciones de un agente

 © 2026 Manuel Cabado García

no dependen solo de sus propias decisiones, sino también de elementos desconocidos en el momento de decidir. Si el agente es capaz de asignar probabilidades a los distintos resultados de una decisión, entonces se encuentra en una situación de riesgo.

Estados de la naturaleza En dicha situación, los estados de la naturaleza son las distintas situaciones que pueden darse en el mundo, sobre las que el agente no tiene control, y que, combinadas con las acciones o estrategias que elige, determinan un resultado o pagos contingentes $(x_1, \ldots, x_n)$[40].

Lotería Para ello utiliza la lotería que es un instrumento que representa alternativas de riesgo, una lista $L = (p_1, \ldots, p_n)$ con $p_n \geq 0$ para todo n y $\sum_n p_n = 1$, donde p_n es la probabilidad de que el suceso n tenga lugar[41].Por lo tanto, una lotería es una distribución de probabilidad donde p_n es la probabilidad de que ocurra el suceso n asociado a un resultado x_n.

Criterio de decisión Con todo ello el agente podrá adoptar un criterio de decisión, que es un proceso mediante el cual toma decisiones reduciendo a términos ciertos cada estado de la naturaleza ponderado por su probabilidad.

Maximización de la renta esperada

Históricamente, el criterio de decisión consistía en maximizar la renta esperada de cada estrategia en un contexto de juego actuarialmente justo, que es aquel donde el coste de jugar es igual a la ganancia esperada, por lo que el valor neto del juego es 0:

$$\mathbb{E}(x) = \sum_{n=1}^{N} x_n \cdot p_n = 0 \tag{4.33}$$

Aunque es un método muy operativo, también tiene sus inconvenientes ya que:

1. Precisa de un elevado número de observaciones para fundamentar la decisión sobre la esperanza matemática, en caso contrario la varianza sería muy elevada y se perdería valor estadístico.
2. Conduce a la paradoja de San Petersburgo: un juego en el que se lanza una moneda hasta que sale cara, con un premio que se duplica en cada lanzamiento, tiene un valor esperado infinito, de modo que según este criterio un individuo debería estar dispuesto a pagar cualquier cantidad por jugarlo, cuando en la práctica nadie pagaría

[40] La intuición es que "llueve" o "no llueve" son estados de la naturaleza, mientras que mojarse o no depende además de haber cogido el paraguas.

[41] A su vez, pueden existir loterías compuestas cuyos resultados sean otras loterías, pero se prescindirá de su análisis dado que una lotería compuesta puede reducirse a una lotería simple.

más que unos pocos euros[42]. Por lo tanto, el criterio de maximización de ingreso esperado conduce a una contradicción, dado que ningún individuo racional aceptaría esto.

Para solucionar esta paradoja, Daniel Bernoulli propuso la maximización de la utilidad esperada

Maximización de la utilidad esperada

Bernoulli (1954) argumentó que los individuos no se preocupan directamente de la renta esperada sino de la utilidad que se deriva de la misma, es decir, maximizan su "esperanza moral" en vez de la renta esperada. Si se asume que la utilidad marginal de la riqueza es decreciente entonces el juego en el caso de San Petersburgo converge a una utilidad esperada finita pese a que la renta esperada sea infinita. Por lo tanto, los agentes solo estarán dispuestos a pagar una cantidad finita para jugar.

Matemáticamente, Bernoulli se basó en una función de utilidad logarítmica, donde x_n representa el premio en el n-ésimo intento:

$$u\left(x_n\right) = \ln\left(x_n\right) \tag{4.34}$$

que exhibe una utilidad marginal decreciente (es decir, $u^{'} > 0$ pero $u^{''} < 0$) y el valor de la utilidad esperada, UE, converge a un número finito

$$UE = \sum_{n=1}^{\infty} p_n \cdot u\left(x_n\right) = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot ln\left(2^n\right) \tag{4.35}$$

Sin embargo, esta solución no resuelve totalmente el problema debido a que la función de utilidad no está acotada, lo que da lugar a una reedición de la paradoja bajo otro tipo

[42] Concretamente, en la paradoja de San Petersburgo se propone el siguiente juego de doble o nada: se lanza una moneda hasta que aparezca cara. Si cara aparece en el intento n-ésimo, n, entonces el jugador recibe un premio de 2^n. Si x_n representa el premio cuando aparece cara en el intento n-ésimo entonces:

$$x_1 = 2, x_2 = 4, x_3 = 8, \ldots, x_n = 2^n$$

La probabilidad de conseguir cara por primera vez en el intento n-ésimo es $\left(\frac{1}{2}\right)^n$, que es la probabilidad de obtener $(n-1)$ cruces y entonces una cara. Por lo tanto, las probabilidades del premio dada la anterior ecuación son:

$$p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{8}, \ldots, p_n = \frac{1}{2^n}$$

Por consiguiente, el valor esperado del juego es infinito:

$$\begin{aligned}\mathbb{E}\left(x\right) &= \sum_{n=1}^{\infty} p_n \cdot x_n = \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot 2^n \\ &= 1 + 1 + 1 + \ldots + 1 = \infty\end{aligned}$$

 © 2026 Manuel Cabado García

de premios[43].

Teoría de la utilidad esperada

Hubo que esperar hasta los años 40 para que Von Neumann y Morgenstern (1944) desarrollasen los fundamentos matemáticos de la solución de Bernoulli a la paradoja de San Petersburgo. La idea central es que el agente posee una función de utilidad definida sobre el conjunto de resultados posibles y que cuando se enfrenta a loterías sobre dichos resultados elige la que maximiza el valor esperado de $u(x)$.

Por tanto, el agente decide sobre distribuciones de probabilidades de los resultados, disponiendo de una estructura de preferencias sobre dichas distribuciones, representadas por medio de una función que maximiza.

Analíticamente, los agentes poseen unas preferencias representables a través de una relación de preferencia binaria "es al menos tan bueno como", $\succsim$, sobre la lotería. Esta relación binaria cumplirá los siguientes axiomas:

- **Axioma 1** Completitud, cualquier par de loterías puede ser comparado.
- **Axioma 2** Transitividad, garantiza una relación de preferencias coherente.
- **Axioma 3** Continuidad, cambios marginales en la probabilidad no cambian la naturaleza de ordenación de dos loterías.
- **Axioma 4** Independencia, si se combinan dos loterías con una misma tercera lotería y en la misma proporción, la ordenación de preferencia entre las dos primeras no cambia[44].

El cumplimiento de los axiomas de racionalidad junto al de continuidad implica la existencia de una función de utilidad que preserva la ordenación de preferencias:[45]

$$u : L \to \mathbb{R} / L \succsim L' \Leftrightarrow u(L) \geq u(L') \tag{4.36}$$

Si a estos axiomas se añade el axioma de independencia, el teorema de la utilidad esperada garantiza que la relación binaria de preferencias podrá ser representada mediante

[43] Por ejemplo, Nicholson y Snyder (2017) en su ejemplo 7.1 proponen la función logarítmica donde los premios se puedan establecer como $x_n = e^{2^n}$, por lo que en ese caso

$$u(x_n) = ln\left(e^{2^n}\right) = 2^n$$

por lo que la utilidad esperada será infinita, reeditando la paradoja de San Petersburgo.

[44] La intuición es que, si se prefiere un viaje a Roma antes que uno a París, se debe seguir prefiriendo un sorteo con un 50 % de Roma y un 50 % de quedarse en casa antes que uno con un 50 % de París y un 50 % de quedarse en casa.

[45] u es una función del conjunto de loterías, L, para los reales (números reales $\mathbb{R}$) tal que, /, la lotería L es al menos tan buena como la lotería L', $L \succsim L'$, si y solo si, $\Leftrightarrow$, esto tiene una traslación en términos de utilidad, $u(L) \geq u(L')$.

una función de utilidad esperada. De esta forma, la utilidad de una lotería será la suma ponderada de las utilidades de Bernoulli por las probabilidades de los distintos sucesos[46]:

$$u\left(L\right) = \sum_{n=1}^{N} u_n \cdot p_n \tag{4.37}$$

garantizando este teorema que los agentes se comportarán como si asignasen utilidades a cada uno de los resultados. Así, se cumple el concepto de utilidad esperada de Bernoulli. Los jugadores no maximizan la esperanza matemática del juego, sino su utilidad esperada. La UE de una alternativa cualquiera (lotería) es la esperanza matemática de la utilidad de los resultados posibles.

Propiedades de la función de utilidad VNM

1. Es única salvo transformaciones afines positivas, es decir, un cambio de escala positivo más una traslación del origen.
2. Continuidad y monotonía: $u\left(x\right)$ es continua y creciente en los pagos, x.
3. Ordinal en loterías, cardinal en resultados: ordena loterías por utilidad esperada.
4. Separabilidad aditiva y linealidad en probabilidades: la utilidad de cada resultado es independiente del resto.
5. Normalización: puede acotarse entre 0 (peor resultado) y 1 (mejor resultado).
6. Diferenciabilidad: deseable que sea dos veces diferenciable para analizar actitudes frente al riesgo.

Por otro lado, en el caso de que el conjunto de resultados sea un intervalo continuo y la lotería pueda ser expresada como una función de densidad, entonces la función de utilidad se expresará de la siguiente forma:

$$u\left(L\right) = \int_{a}^{b} u\left(x\right) f\left(x\right) \cdot dx \tag{4.38}$$

Ventajas función de utilidad esperada Entre las ventajas de la función de utilidad esperada destacan:

- La idea de que se valore la esperanza de la utilidad y no al revés permite resolver la paradoja de San Petersburgo y explicar los comportamientos observados.
- La función de utilidad posee unas propiedades técnicas convenientes desde el punto de vista analítico, ya que será sensible a momentos de primer orden –la media– y superiores –como la varianza–.

[46] La idea de la UE ya fue introducida por Bernoulli, si bien VNM demostrarán que si las preferencias cumplen los axiomas señalados existe una función de utilidad esperada.

 © 2026 Manuel Cabado García

- La utilidad esperada pasa a tener un valor normativo, siendo una gran guía de decisión.

Finalmente, se observa que la función de utilidad esperada consta de dos componentes: uno subjetivo, la utilidad de Bernoulli, y otro objetivo, la distribución de probabilidad.

Actitudes ante el riesgo

La utilidad de Bernoulli, $u(x)$, representa las preferencias individuales frente a distribuciones de probabilidad definidas sobre cantidades monetarias ciertas. En este marco, se asume que $u(x)$ es creciente, reflejando que los individuos prefieren más riqueza a menos.

La forma concreta que adopte la función de utilidad permite identificar la actitud frente al riesgo del agente.

Equivalente cierto Para su adecuado análisis, conviene introducir antes el equivalente cierto, EC, que es la cantidad de dinero segura que proporciona la misma utilidad que la utilidad esperada de la lotería, $u(EC) = \mathbb{E}[u(x)]$[47].

Perfiles de riesgo Por lo tanto, un agente será:

- Averso al riesgo, cuando prefiere recibir con certeza el valor esperado de la lotería antes que jugarla, $u(\bar{x}) > \mathbb{E}[u(x)]$, por lo que la función de utilidad será cóncava[48].
- Neutral al riesgo, cuando le es indiferente jugar la lotería o recibir con certeza su valor esperado, $u(\bar{x}) = \mathbb{E}[u(x)]$, por lo que la función de utilidad será lineal.
- Amante del riesgo, cuando prefiere jugar la lotería antes que recibir con certeza su valor esperado, $u(\bar{x}) < \mathbb{E}[u(x)]$, por lo que la función de utilidad será convexa.

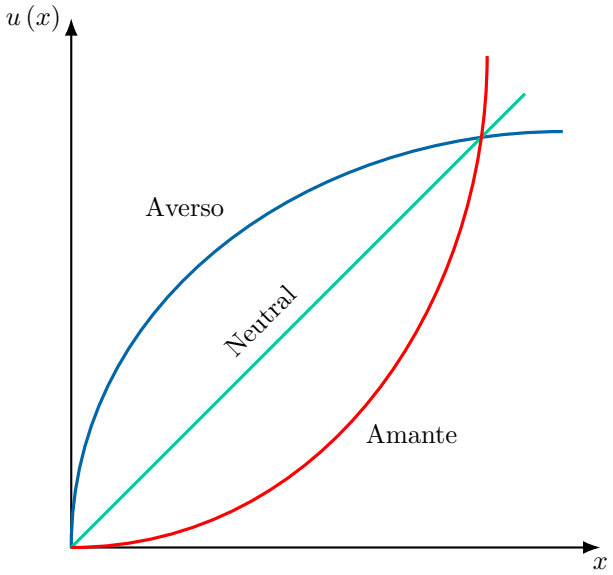
Gráficamente, en un plano de utilidad, $u(x)$, y cantidad de dinero, x, como el de la Figura 4.5, se representan las funciones de utilidad de un agente averso, cóncava, neutral, lineal, y amante del riesgo, convexa.

La intuición económica que subyace es el efecto asimétrico de la renta ganada frente a la renta perdida: partiendo de un nivel de renta dado, y ante variaciones equivalentes al alza y a la baja, la desutilidad marginal de la renta perdida es mayor que la utilidad marginal de la renta ganada –de ahí que la utilidad marginal de la riqueza sea decreciente para un agente averso–.

[47] La intuición es que es el precio mínimo al que el agente vendería su billete de lotería. Si le da igual quedarse el billete o recibir 40 euros seguros, su equivalente cierto es 40 euros.

[48] La intuición es que quien es averso al riesgo prefiere 50 euros seguros a lanzar una moneda para ganar 100 o nada, aunque ambas opciones valgan lo mismo en promedio. Formalmente, esta desigualdad es una aplicación directa de la desigualdad de Jensen: si u es cóncava, entonces $\mathbb{E}[u(x)] \leq u(\mathbb{E}[x]) = u(\bar{x})$ para cualquier variable aleatoria x, con igualdad solo si x es constante (sin riesgo) o u es lineal. Por eso la concavidad de u es exactamente lo que garantiza que la utilidad esperada de la lotería nunca supere a la utilidad del valor esperado, lo que da lugar a la aversión al riesgo.

Figura 4.5: Actitudes frente al riesgo



Prima de riesgo Gráficamente, en un plano de utilidad, $u(x)$ y de cantidad de dinero, x, como el de la Figura 4.6 se representa una función de utilidad, $u(x)$, de un agente averso al riesgo el cual pasará a jugar una lotería donde x_1 y x_2 son el menor y el mayor premio respectivamente, con sus respectivas correspondencias de utilidad, $u(x_1)$ y $u(x_2)$. Ambos puntos estarán unidos por la línea de utilidad esperada, que permitirá calcular a partir de la cantidad esperada del juego, $\bar{x}$, la utilidad esperada de la lotería, $u(EC)$ y su cantidad asociada, es decir, el equivalente cierto, EC. Se puede observar cómo la utilidad del valor esperado de la lotería, $u(\bar{x})$, es superior a su utilidad esperada, $u(EC)$. Por ello, el equivalente cierto es inferior al valor esperado, y la diferencia entre ambas cantidades de dinero es la prima de riesgo, PR:

$$PR = \bar{x} - EC \tag{4.39}$$

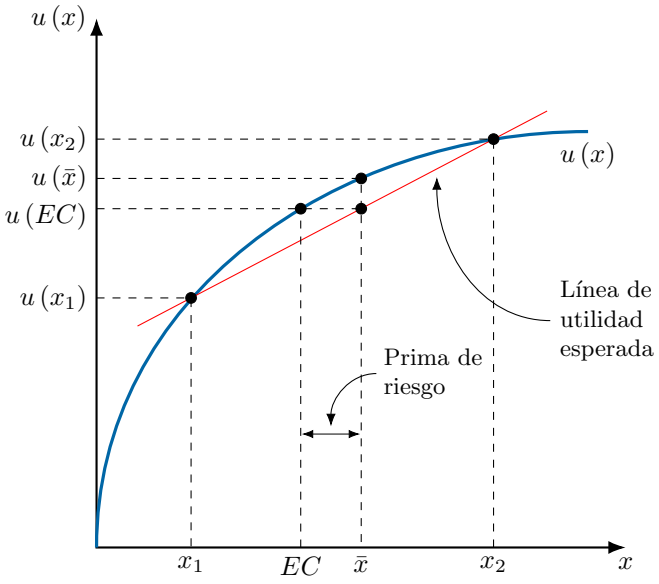
Concretamente, esta prima representa la cantidad de dinero que el agente estaría dispuesto a pagar para evitar el riesgo de la lotería y recibir una cantidad segura equivalente en utilidad.

Prima máxima y justa Además, se pueden definir dos magnitudes relacionadas con la prima de riesgo:

- La prima máxima, la cantidad máxima que el agente estaría dispuesto a pagar por renunciar a la lotería y eliminar el riesgo: la diferencia entre el mejor resultado posible y el equivalente cierto.
- La prima justa, la pérdida esperada del juego: la diferencia entre el mejor resultado

 © 2026 Manuel Cabado García

Figura 4.6: Equivalente cierto y prima de riesgo



posible y el valor esperado de la lotería.

Ambas magnitudes se relacionan con la prima de riesgo, ya que la diferencia entre el equivalente cierto y el valor esperado de la lotería –lo que el agente averso está dispuesto a pagar de más, por encima de la pérdida esperada, con tal de no jugar– es exactamente[49]:

$$PR = \bar{X} - EC = \text{Prima Máxima} \; - \; \text{Prima Justa} \tag{4.40}$$

Relación entre prima de riesgo y compañías de seguros De esta forma, la existencia de una prima de riesgo positiva para los agentes aversos abre un margen de negocio para las compañías de seguros: cualquier prima que cobren situada entre la prima justa (su coste esperado, el mínimo que necesitan cobrar para no perder dinero) y la prima máxima (lo máximo que el asegurado está dispuesto a pagar) resulta rentable para la aseguradora y, a la vez, beneficiosa para el asegurado, que preferirá pagarla antes que asumir el riesgo. Es precisamente ese margen, y no la mera existencia de aversión al riesgo, lo que explica la viabilidad de un mercado de seguros.

[49] Ejemplo numérico: un agente con riqueza de 10.000€ se enfrenta a un riesgo de accidente: con probabilidad 0,5 no ocurre nada (riqueza 10.000€) y con probabilidad 0,5 sufre una pérdida de 3.600€ (riqueza 6.400€). Con $u(w) = \sqrt{w}$: $EU = 0,5 \cdot 100 + 0,5 \cdot 80 = 90$, de donde $EC = 90^2 = 8.100$€. El valor esperado de la riqueza es $\bar{X} = 0,5 \cdot 10.000 + 0,5 \cdot 6.400 = 8.200$€. Así, la prima máxima es $10.000 - 8.100 = 1.900$€, la prima justa es $10.000 - 8.200 = 1.800$€, y $PR = \bar{X} - EC = 100$€, que coincide con $1.900 - 1.800$. Cualquier prima entre 1.800€ y 1.900€ –por ejemplo, 1.850€– es rentable para la aseguradora y preferible al riesgo para el asegurado.

Índices de aversión al riesgo

Si bien existen tres posibilidades teóricas respecto a la actitud ante el riesgo, desde un punto de vista empírico se observa que los agentes tienden a sentir[50] aversión al riesgo. El problema que surge es cuantificar esta aversión al riesgo.

De esta forma, para cuantificar el grado de aversión al riesgo se utilizarán dos coeficientes:

- Coeficiente de aversión al riesgo absoluta de Arrow-Pratt, RA, definido como el cociente, cambiado de signo, entre la segunda y la primera derivada de la función de Bernoulli[51]

$$RA = -\frac{u''(x)}{u'(x)} \tag{4.41}$$

  que recoge el grado de aversión mediante el grado de curvatura –cuanta más curvatura, más averso– y permite aproximar la prima de riesgo a través de la aproximación de Taylor –cuanto más alto sea RA más alta será la prima de riesgo–.

- Coeficiente de aversión relativa al riesgo, RR, que mide la aversión ante riesgos expresados como proporción de la riqueza, es decir, ante variaciones porcentuales de esta[52]:

$$RR = -\frac{u''(x)}{u'(x)} \cdot x \tag{4.42}$$

  dado que la comparación de actitudes frente al riesgo para un mismo agente puede depender de su nivel de riqueza. Además, RR puede ser también interpretado como la elasticidad de la utilidad marginal, UMg, con respecto a x, cambiada de signo: $-\varepsilon_{UMg,x}$.

En suma, cuanto mayores sean RA y RR, más averso al riesgo será el individuo.

Soluciones del mercado al problema del riesgo

Para quienes son aversos al riesgo, el mercado ofrece soluciones para reducir la incertidumbre a través de diversos mecanismos:

- Diversificación: consiste en combinar en una cartera activos cuyos rendimientos no están perfectamente correlacionados, lo que reduce su varianza sin sacrificar la rentabilidad esperada, tanto más cuanto menor sea la correlación[53].

[50] Además, la aversión al riesgo es conveniente desde el punto de vista analítico: la función de utilidad es cuasicóncava, las curvas de indiferencia convexas garantizan que haya un máximo.

[51] Como $u'(x)$ es siempre positivo, dependerá de la segunda derivada para ver el perfil de riesgo:

- Si $RA > 0$, el individuo es averso al riesgo.
- Si $RA < 0$, el individuo es amante del riesgo.
- Si $R(x) = 0$, el individuo es neutral ante el riesgo.

[52] La intuición es que el RA mide la aversión a jugarse una cantidad fija, por ejemplo 100 euros, y el RR la aversión a jugarse un porcentaje fijo del patrimonio, por ejemplo el 10 %.

[53] La intuición es que las pérdidas de unos activos tienden a compensarse con las ganancias de otros.

 © 2026 Manuel Cabado García

- Difusión del riesgo (Risk Spreading): El coste de un riesgo se reparte entre una gran cantidad de individuos. Ejemplos son las acciones de una empresa o los créditos sindicados. Al distribuir el riesgo, el colectivo actúa como si fuera neutral al riesgo, aunque sus miembros individualmente sean aversos.
- Puesta en común del riesgo (Risk Pooling): Los individuos agrupan sus riesgos y rentas, acordando recibir la media. Esto reduce la varianza de la renta de cada persona. Matemáticamente, la varianza de la renta total disminuye a medida que más individuos participan. Esto se observa en los fondos de inversión o las sociedades de garantía recíproca.
- Mercados contingentes (opciones y futuros): requieren contrapartes dispuestas a asumir el riesgo, ya sean especuladores o agentes con una exposición de signo contrario. Permiten redistribuir riesgos sin eliminarlos.
- Mercados de seguros: Redistribuyen riesgos entre agentes con diferentes actitudes. Basados en contratos contingentes y primas.

Métodos de reducción del riesgo: el mercado de seguros

A continuación se desarrolla una aplicación práctica de la elección bajo riesgo. La existencia de una prima de riesgo da cabida en la realidad a la existencia de compañías de seguros. El punto de partida de la decisión de aseguramiento es que el agente parte de una exposición al riesgo que desea reducir[54].

El problema del aseguramiento se plantea para:

- Un individuo con una riqueza inicial, W, y un activo valorado en R que corre el riesgo de ser robado con una probabilidad p.
- El agente, que es averso al riesgo, puede comprar un seguro que le compensa una cantidad K a cambio de una prima β, donde $0 < \beta < 1$.
- Los estados del mundo son:

  1. Que le roben con probabilidad p.
  2. Que no le roben con probabilidad $(1 - p)$

Los posibles resultados en términos de la riqueza del individuo serán respectivamente:

Por ejemplo, quien invierte a la vez en una empresa de paraguas y en otra de cremas solares gana con una lo que pierde con la otra, llueva o haga sol. El ejemplo, con una empresa de gafas de sol y otra de chubasqueros, procede de Varian (2014), cap. 12.

[54] Para analizar los contratos de seguros se considera la decisión de asegurarse en el marco del enfoque de Arrow-Debreu. Este enfoque implica que las preferencias, a diferencia del modelo de Von Neumann-Morgenstern se van a definir no sobre las loterías sino sobre los planes de consumo contingentes. De esta forma la resolución del problema del consumidor en un ambiente de riesgo va a ser similar a la resolución en un ambiente de certeza.

- Riqueza si le roban

$$w_1 = W - R + K - \beta \cdot K \tag{4.43}$$

- Riqueza si no le roban

$$w_2 = W - \beta \cdot K \tag{4.44}$$

La restricción presupuestaria para cada valor de K es:

$$w_2 = W - \frac{\beta \cdot (w_1 - W + R)}{1 - \beta}$$

Cuya pendiente viene dada por:

$$\frac{dw_2}{dw_1} = -\frac{\beta}{1 - \beta} < 0 \tag{4.45}$$

Esta es la tasa a la que el agente intercambia riqueza entre los estados. Es decir, el precio a pagar por pasar de un estado a otro.

Gráficamente, se acota la restricción presupuestaria en función de lo que se asegure el individuo:

$$\text{Si } K = R \longrightarrow w_1 = W - \beta \cdot K = w_2 \tag{4.46}$$

$$\text{Si } K = 0 \longrightarrow w_1 = W - R; \quad w_2 = W \tag{4.47}$$

El objetivo del individuo será por lo tanto elegir la cantidad de K que querrá asegurar. Se pasa a analizar esta cuestión y a ver cómo varía la solución en caso de considerar agentes con diferentes actitudes con respecto al riesgo y diferentes primas de riesgo.

El consumidor resuelve el siguiente problema de maximización condicionada, donde la función de utilidad esperada es del tipo Von Neumann y Morgenstern:

$$\max_{w_1, w_2} UE = \sum_{j=1}^{2} p_j \cdot u(w_j) = p \cdot u(w_1) + (1 - p) \cdot u(w_2)$$

sujeto a la restricción presupuestaria[55]

$$w_2 = W - \frac{\beta \cdot (w_1 - W + R)}{1 - \beta} \tag{4.48}$$

Derivando con respecto a las variables control (w_1, w_2) se obtiene la condición que da la cantidad óptima de seguro, igualando la pendiente de la restricción presupuestaria con

[55] Restricciones adicionales:

- La riqueza inicial del individuo nunca será inferior al valor del objeto a asegurar, $0 \leq R \leq W$.
- Nunca se podrá asegurar dicho objeto en un valor superior al valor del mismo, $0 \leq K \leq R$.
- La prima será positiva, es decir, que todo seguro supone un gasto para el individuo que lo contrata, $\beta > 0$.

 © 2026 Manuel Cabado García

la de la curva de indiferencia, esta última conocida como Relación Marginal de Sustitución Estocástica (RMSE)[56]:

$$-\frac{\beta}{1-\beta} = -\frac{p \cdot u'(w_1)}{(1-p) \cdot u'(w_2)} \tag{4.49}$$

La solución dependerá del perfil de riesgo del agente y del coste del seguro[57]. A continuación se analiza gráficamente.

La Decisión Óptima La decisión óptima es el punto donde la curva de indiferencia más alta toca la restricción presupuestaria como se observa en la Figura 4.7:

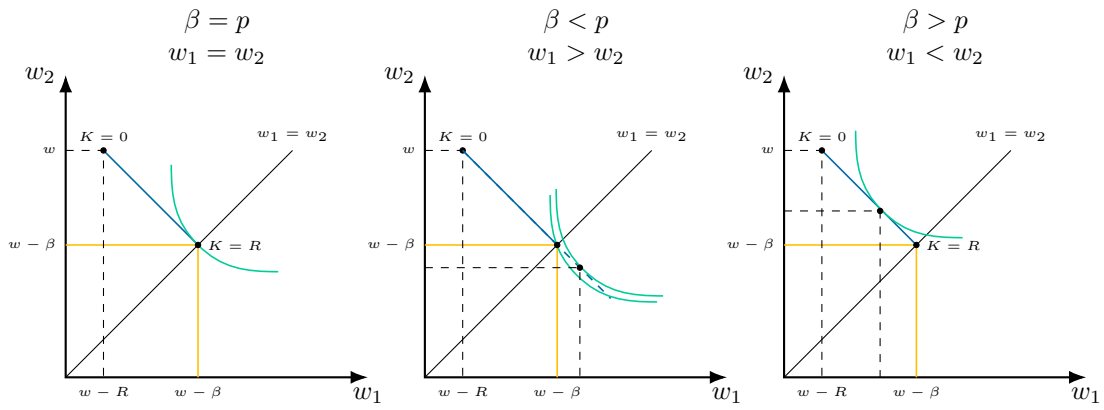
- Prima justa ($\beta = p$): El agente puede alcanzar la línea de certeza. Para maximizar su utilidad, el individuo se situará en el punto de tangencia con la línea de 45 grados (línea de certeza), lo que significa que optará por el seguro completo ($K = R$), ya que puede eliminar el riesgo por un precio justo.
- Prima demasiado baja ($\beta < p$): la restricción presupuestaria es más plana que la curva de indiferencia en su punto de corte con la línea de certeza, cuya pendiente es $-p/(1-p)$. En este caso, el seguro es tan ventajoso que el agente querrá sobreasegurarse ($K > R$), es decir, adquirir una cobertura superior al valor del bien, aunque en la práctica las aseguradoras no lo permiten y el agente se queda en el seguro completo. Este es el punto de tangencia óptimo, que se encuentra a la derecha de la línea de 45 grados, ya que el agente prefiere una situación con más riqueza si ocurre el evento negativo.
- Prima injusta ($\beta > p$): la restricción presupuestaria es más pronunciada que la curva de indiferencia sobre la línea de certeza, y la curva de indiferencia más alta tangente a ella se encuentra a la izquierda de la línea de certeza. Esto significa que el agente no puede eliminar completamente el riesgo de manera óptima, por lo que opta por un seguro parcial ($K < R$). Asume algo de riesgo, ya que el coste de eliminarlo por completo es demasiado alto.

En definitiva, la cantidad de seguro óptima dependerá de varios factores: a mayor grado de aversión al riesgo del agente, mayor será su demanda de seguro. A mayor probabilidad de siniestro, p, mayor demanda, porque aumenta la valoración relativa de la riqueza en el estado de siniestro (la RMSE crece) mientras que la pendiente de la restricción, que solo depende de β, no cambia. A mayor magnitud de la pérdida potencial, R, mayor demanda. Por último, con signo negativo, a mayor prima del seguro, β, menor demanda.

[56] La RMSE mide la tasa a la que el agente está dispuesto a sacrificar riqueza en un estado de la naturaleza a cambio de una unidad adicional de riqueza en el otro, y no es sino la relación marginal de sustitución habitual bajo certeza, ponderada por el cociente de probabilidades de ambos estados.

[57] En los mercados justos el individuo se sitúa donde la UMg de la riqueza es igual en todos los estados de la naturaleza («Teorema Fundamental de la Asunción de Riesgos», (Arrow, 1964)).

Figura 4.7: Decisión de aseguramiento



Críticas a la teoría de la utilidad esperada

La Teoría de la Utilidad Esperada, pese a su capacidad explicativa demostrada en las aplicaciones anteriores, no está exenta de críticas. Se exponen a continuación las más relevantes, que han llevado al desarrollo de teorías alternativas para explicar el comportamiento humano ante el riesgo.

Las primeras críticas se centran en la forma de la función de utilidad:

- Friedman y Savage (1948): Propusieron que la actitud ante el riesgo no es constante. En concreto, plantearon una función de utilidad cóncava para rentas bajas y altas y convexa en un tramo intermedio, lo que explica que un mismo individuo se asegure contra grandes pérdidas y, a la vez, compre lotería con la esperanza de un gran premio que le haga cambiar de nivel de renta. La crítica es que su modelo es solo descriptivo y no establece una forma funcional concreta.
- Markowitz (1952b): Afirmó que la utilidad no depende de la riqueza final, sino del cambio (ganancia o pérdida) respecto a un punto de referencia. Los individuos son aversos a pérdidas pequeñas y ganancias grandes, pero amantes de pérdidas grandes y ganancias pequeñas.

Las críticas más significativas surgen de experimentos que violan los axiomas de la Teoría de la Utilidad Esperada:

- Paradoja de Ellsberg (Ellsberg, 1961): Demostró la aversión a la ambigüedad. Los individuos prefieren loterías con probabilidades conocidas (riesgo) a loterías con probabilidades desconocidas (incertidumbre), violando en sentido estricto el axioma del elemento seguro (sure-thing axiom) de la probabilidad subjetiva, más que el axioma de independencia de la utilidad esperada en sentido literal[58].

[58] El axioma de independencia de Von Neumann-Morgenstern opera sobre probabilidades objetivas, mientras que la paradoja de Ellsberg surge precisamente de la ausencia de estas, por lo que su violación se formula con más precisión sobre el axioma del elemento seguro de Savage (1954).

 © 2026 Manuel Cabado García

- Paradoja de Allais (Allais, 1953): Mostró que los individuos violan el axioma de independencia. La gente prefiere un resultado seguro a una lotería con un valor esperado mayor, pero cuando se comparan dos loterías con resultados inciertos, prefieren la que ofrece el premio más alto, incluso si su probabilidad de ganar es menor[59].

Teoría de las perspectivas Para resolver estas inconsistencias, y recogiendo la idea de Markowitz sobre la dependencia del punto de referencia, Kahneman y Tversky (1979) desarrollaron la Teoría de las Perspectivas (Prospect Theory), que la amplía con dos elementos que Markowitz no incorporaba: los pesos decisorios y la asimetría explícita entre la pendiente de pérdidas y ganancias. Esta teoría se basa, por tanto, en dos conceptos principales:

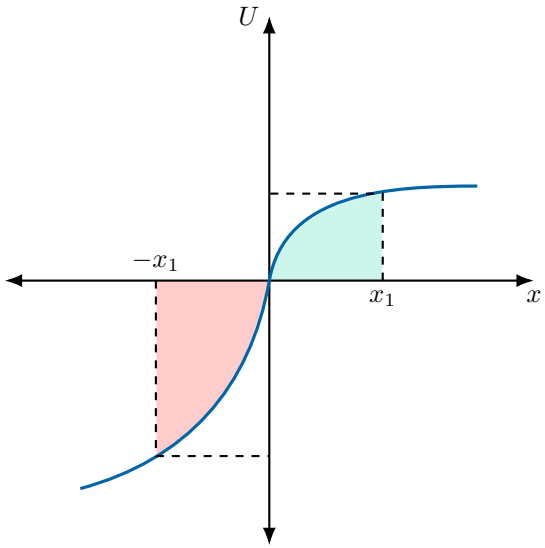
- Función de valor: sustituye a la función de utilidad de Bernoulli y no se define sobre el nivel de riqueza final, sino sobre las ganancias y pérdidas respecto a un punto de referencia (habitualmente el statu quo). Adopta forma de S, tal y como se observa en la Figura 4.8: cóncava en el tramo de ganancias (aversión al riesgo) y convexa en el de pérdidas (propensión al riesgo), con mayor pendiente en las pérdidas que en las ganancias, reflejo de que el dolor de perder es más intenso que el placer de ganar la misma cantidad[60].
- Pesos decisorios: los agentes no ponderan los resultados según la probabilidad objetiva, sino según un peso decisorio que la distorsiona de forma sistemática y no lineal, con forma de "S invertida": se sobreponderan las probabilidades bajas –de ahí el atractivo de la lotería o el sobreseguro frente a riesgos remotos, pese a su escaso valor esperado– y se infraponderan las probabilidades altas –por ejemplo, ante la elección entre un 80 % de probabilidad de ganar 4.000€ y la certeza de recibir 3.000€, la mayoría prefiere la opción segura, pese a que el valor esperado de la primera (3.200€) es mayor–. Esta distorsión es especialmente intensa cerca de los extremos: pasar de una probabilidad del 100 % (certeza) al 99 % se percibe como una caída de valor mucho mayor de lo que sugeriría esa única unidad porcentual –el efecto certeza–, precisamente el mecanismo que explica la paradoja de Allais.

En resumen, la Teoría de la Utilidad Esperada, aunque fundamental, fue criticada por no reflejar el comportamiento real. De esta forma, la Teoría de las Perspectivas ofreció una

[59] Concretamente, Allais (1953) mostró que la mayoría de los individuos prefiere la lotería A (100 % de probabilidad de 1 millón de euros) a la B (89 % de 1 millón, 10 % de 5 millones, 1 % de nada), pero al mismo tiempo prefiere la lotería D (10 % de 5 millones, 90 % de nada) a la C (11 % de 1 millón, 89 % de nada) –pese a que C y D se obtienen restando la misma probabilidad (0,89) a A y B respectivamente, lo que debería mantener el orden de preferencia si se cumpliera el axioma de independencia–.

[60] Una manifestación directa de esta asimetría es el efecto dotación (endowment effect): los individuos exigen, a cambio de desprenderse de algo que ya poseen, una cantidad de dinero mayor de la que estarían dispuestos a pagar por adquirirlo, pese a tratarse del mismo bien.

Figura 4.8: Función de valor de la Teoría de las Perspectivas



alternativa que, aunque no ha encontrado una aplicación universal, es la base de la economía conductual, ya que explica de forma empírica cómo las personas toman decisiones bajo riesgo e incertidumbre.[61]

4.4. Conclusiones ⏳ 30"

Como conclusión, las dos teorías analizadas anteriormente surgen para hacer frente a las limitaciones de la teoría neoclásica, abriendo líneas de investigación sobre el comportamiento del consumidor, tanto para hacer operativa la teoría tradicional como para modelizar situaciones reales en las que no se cumplen los supuestos de la misma.

Así, en el campo de la dualidad se encuentra:

- La teoría de la preferencia revelada, que permite recuperar las preferencias del consumidor directamente de su comportamiento observado, sin necesidad de partir de una función de utilidad –en la misma línea que el problema de integrabilidad ya visto–.
- La demanda de características, donde el consumidor no valora los bienes en sí mismos, sino los atributos o características que estos proporcionan.

Por otro lado, en el análisis de la incertidumbre se encuentran:

- Las reglas de decisión propias de la elección bajo incertidumbre, vistas en el Apéndice.

[61] El apartado 4.A.8 del Apéndice desarrolla la economía del comportamiento y su principal aplicación a la política pública, el *nudge* de Richard Thaler.

 © 2026 Manuel Cabado García

- La teoría de juegos, que extiende la elección bajo riesgo a contextos estratégicos, donde el resultado depende también de las decisiones de otros agentes.
- La teoría de selección adversa y riesgo moral, que analiza los problemas de información asimétrica –previos (selección adversa) y posteriores (riesgo moral) a la contratación– que pueden surgir precisamente en mercados de seguros como el ya analizado.

Apéndice

4.A.1. Elección en condiciones de incertidumbre

Como se ha señalado en el cuerpo, cuando el agente no puede asignar probabilidades objetivas a los distintos estados de la naturaleza se enfrenta a una situación de incertidumbre en sentido estricto (Knight, 1921). Existen dos vías principales para abordar la decisión en este contexto: suponer que el agente actúa como si estableciera probabilidades subjetivas, o recurrir a criterios de decisión que prescinden por completo de probabilidades.

Teoría de la probabilidad subjetiva El primer desarrollo riguroso de este enfoque corresponde a Savage (1954): el agente se comporta como si estableciera unas creencias o probabilidades subjetivas sobre los distintos estados de la naturaleza, lo que permite aplicar exactamente el mismo aparato formal de la utilidad esperada, unificando así el tratamiento del riesgo y la incertidumbre.

Imponiendo sobre las loterías estado-dependientes los axiomas de completitud, transitividad, continuidad, independencia extendida y preferencias estado-uniformes –la valoración de un mismo pago monetario no depende del estado de la naturaleza que finalmente se materialice–, existen unas probabilidades subjetivas $(\pi_1, \ldots, \pi_S)$, determinadas de forma única, y una función de utilidad de Bernoulli $u(\cdot)$ tales que:

$$(x_1, \ldots, x_S) \succsim (x'_1, \ldots, x'_S) \Leftrightarrow \sum_s \pi_s \cdot u(x_s) \geq \sum_s \pi_s \cdot u(x'_s)$$

El riesgo pasa a ser así un caso particular de esta utilidad esperada subjetiva, aquel en que las probabilidades subjetivas coinciden con las objetivas. Su ventaja es la capacidad unificadora. Su inconveniente es que hereda las críticas de la Teoría de la Utilidad Esperada, a las que se suma un problema propio: la paradoja de Ellsberg muestra que las probabilidades subjetivas inferidas del comportamiento observado no siempre son coherentes con ninguna asignación consistente, por la aversión a la ambigüedad.

Criterios de decisión clásicos Cuando ni siquiera resulta posible asignar probabilidades subjetivas mínimamente fiables, se recurre a criterios basados solo en la matriz de resultados posibles para cada combinación de estrategia y estado de la naturaleza. Considérese, a modo de ejemplo, la siguiente matriz de pagos con tres estrategias (A_1, A_2, A_3) y tres estados de la naturaleza (S_1, S_2, S_3):

	S_1	S_2	S_3
A_1	1	2	-3
A_2	4	-5	6
A_3	-7	8	9

 © 2026 Manuel Cabado García

1. Criterio de Wald o maximin (pesimista): se identifica el peor resultado de cada estrategia y se elige la de peor resultado menos malo. En el ejemplo, los peores resultados son $(-3, -5, -7)$, por lo que se elige A_1.
2. Criterio maximax (optimista): se identifica el mejor resultado de cada estrategia y se elige la de mejor resultado más alto. En el ejemplo, los mejores resultados son $(2, 6, 9)$, por lo que se elige A_3.
3. Criterio de Laplace o de razón insuficiente: ante la ausencia de información sobre las probabilidades, se asigna la misma probabilidad a cada estado (en el ejemplo, $1/3$) y se elige la estrategia de mayor valor esperado.
4. Criterio de Hurwicz: define un coeficiente de optimismo $\alpha \in [0, 1]$, propio de cada decisor, que pondera el mejor resultado de cada estrategia, M_i, mientras que $(1 - \alpha)$ pondera el peor, m_i:
$$\alpha \cdot M_i + (1 - \alpha) \cdot m_i$$
eligiendo la estrategia de mayor valor ponderado [62].
5. Criterio de Savage o del coste de oportunidad (minimax): se construye primero una matriz de arrepentimiento o coste de oportunidad, restando cada resultado del mejor resultado posible en ese mismo estado de la naturaleza. Se identifica después, para cada estrategia, el máximo coste de oportunidad –el peor de los arrepentimientos posibles–, y se elige la estrategia que minimiza ese máximo, de ahí el nombre minimax. En el ejemplo, los mejores resultados por estado son $(4, 8, 9)$, con lo que la matriz de arrepentimiento resulta:

	S_1	S_2	S_3
A_1	3	6	12
A_2	0	13	3
A_3	11	0	0

cuyos máximos por fila son $(12, 13, 11)$, por lo que se elige A_3.

4.A.2. Dominancia estocástica

Junto a la utilidad de Bernoulli, el segundo componente de la utilidad esperada es la distribución de probabilidad, $F(x)$, el elemento objetivo de la lotería, común a todos los agentes con independencia de su actitud ante el riesgo. Puede caracterizarse mediante sus momentos estadísticos: la esperanza (momento de primer orden: a mayor esperanza, mayor

[62] Esta familia de criterios tiene un reflejo directo en la economía del bienestar: el criterio maximin es precisamente la base de la función de bienestar social de Rawls, mientras que la agregación mediante utilidad esperada –análoga al criterio de Laplace, pero con ponderaciones propias de cada estado en lugar de equiprobabilidad– sustenta la función de bienestar social utilitarista de Harsanyi.

utilidad para cualquier agente que prefiera más riqueza a menos), la varianza (momento de segundo orden: refleja la dispersión o riesgo de la lotería y, para un agente averso, a mayor varianza, menor utilidad), la asimetría (momento de tercer orden: sesgo de la distribución hacia resultados mejores o peores) y la curtosis (momento de cuarto orden: concentración de los resultados en torno a la media).

Criterios de dominancia estocástica Cabe preguntarse si es posible comparar loterías de forma objetiva, con independencia de las preferencias concretas de cada agente. Dos criterios permiten hacerlo.

1. Dominancia estocástica de primer orden: una distribución $F(x)$ domina estocásticamente en primer orden a otra $G(x)$ si es preferida por cualquier agente que prefiera más riqueza a menos, sea cual sea su actitud frente al riesgo:

$$\int u(x)\, dF(x) \geq \int u(x)\, dG(x) \qquad \forall u(x) \text{ tal que } u'(x) > 0$$

2. Dominancia estocástica de segundo orden: restringiendo la comparación a loterías con idéntico valor esperado –para aislar el efecto puro del riesgo, ya que un agente puede aceptar un mayor riesgo si el pago esperado es suficientemente grande–, noción estrechamente relacionada con el concepto de aumento del riesgo (mean-preserving spread) de Rothschild y Stiglitz (1970), $F(x)$ domina estocásticamente en segundo orden a $G(x)$ si es preferida por cualquier agente averso al riesgo:

$$\int u(x)\, dF(x) \geq \int u(x)\, dG(x) \qquad \forall u(x) \text{ tal que } u'(x) > 0 \text{ y } u''(x) < 0$$

Ambos criterios, pese a su elegancia teórica, resultan poco operativos en la práctica: son extremadamente exigentes y solo permiten, en el mejor de los casos, ordenar parcialmente las loterías disponibles, sin cuantificar cuánto mejor es una respecto de otra. Por esta razón se recurre habitualmente al binomio media-varianza, que trata la media como un bien y la varianza como un mal. Aunque supone una pérdida de información respecto de la distribución completa –se prescinde de los momentos de orden superior, como la asimetría o la curtosis–, esta simplificación puede justificarse mediante el Teorema Central del Límite cuando el resultado final proviene de la suma de un número elevado de variables aleatorias, situación paradigmática en la teoría de selección de carteras desarrollada, entre otros, por Markowitz, Tobin y Sharpe [63].

[63] El binomio media-varianza es, además, exactamente óptimo –no una mera aproximación– bajo utilidad cuadrática, según se recoge en el Apéndice 4.A.5.

 © 2026 Manuel Cabado García

4.A.3. Decisión de cartera

De la misma forma que un agente averso al riesgo está dispuesto a renunciar a renta presente a cambio de mayor certidumbre –el modelo de aseguramiento ya desarrollado–, también está dispuesto a renunciar a la certidumbre perfecta a cambio de una compensación suficiente: es el planteamiento simétrico e inverso del modelo de aseguramiento.

Supuestos

- Un agente parte de una riqueza W invertida íntegramente en un activo sin riesgo (un bono) con rentabilidad cierta r.
- Puede invertir una proporción $\alpha \in [0,1]$ de su riqueza en un activo con riesgo (una acción), que ofrece un rendimiento e_1 con probabilidad $(1-p)$ y e_2 con probabilidad p, siendo $e_1 > r > e_2$ –de modo que la acción puede batir o no al bono, según el estado de la naturaleza que finalmente se materialice–.

Planteamiento y solución óptima Los posibles resultados en términos de la riqueza del individuo serán, respectivamente:

- Riqueza si la acción rinde e_1:

$$w_1 = (1-\alpha)\, W\,(1+r) + \alpha W\,(1+e_1)$$

- Riqueza si la acción rinde e_2:

$$w_2 = (1-\alpha)\, W\,(1+r) + \alpha W\,(1+e_2)$$

El consumidor resuelve el siguiente problema de maximización, donde la función de utilidad esperada es del tipo Von Neumann y Morgenstern:

$$\max_{\alpha} UE = (1-p)\cdot u\,(w_1) + p\cdot u\,(w_2)$$

La CPO establece que el agente asumirá riesgo hasta el punto en que la RMSE –la desutilidad marginal subjetiva de asumir riesgo– se iguale con el beneficio marginal objetivo que le ofrece la mayor rentabilidad esperada de la acción frente al bono:

$$\frac{u'\,(w_1)\cdot(1-p)}{u'\,(w_2)\cdot p} = \frac{r-e_2}{e_1-r}$$

Interpretación Si la cartera se invierte totalmente en el bono ($\alpha = 0$, sobre la línea de certeza), el agente no asume riesgo alguno. En la solución de tangencia, si el rendimiento esperado de la acción coincide exactamente con el del bono (rendimiento justo), el agente

averso encuentra óptimo no aceptar ningún riesgo –salta directamente a una solución de esquina–. Solo cuando la acción ofrece un rendimiento esperado superior al justo –una compensación extraordinaria que contrarresta la desutilidad de asumir riesgo– el agente averso decide diversificar su cartera y desplazarse hacia una solución interior, alejándose de la línea de certeza en proporción a α.

Este modelo, pese a su sencillez, permite explicar la demanda de activos de riesgo aun en presencia de aversión al riesgo por parte de los inversores, y constituye el fundamento microeconómico sobre el que se construyen desarrollos posteriores como la teoría de selección de carteras de Markowitz (1952a), el modelo CAPM de Sharpe (1964) y Lintner (1965), y la función de demanda de dinero –entendido como un activo– de Tobin (1958).

4.A.4. Preferencias tipo Epstein-Zin

En la utilidad esperada estándar de tipo CRRA, el coeficiente de aversión relativa al riesgo, γ, y la elasticidad de sustitución intertemporal (ESI), ψ, quedan atados al mismo parámetro: $\gamma = 1/\psi$. Esto obliga a un agente muy averso al riesgo a ser también, necesariamente, muy reacio a sustituir consumo entre periodos –pese a tratarse de dos actitudes conceptualmente distintas, una frente a la incertidumbre y otra frente al tiempo–.

Epstein y Zin (1989) resolvieron esta limitación mediante una utilidad recursiva que introduce γ y ψ como dos parámetros independientes:

$$U_t = \left[(1-\beta)\, c_t^{1-1/\psi} + \beta \left(\mathbb{E}_t \left[U_{t+1}^{1-\gamma} \right] \right)^{\frac{1-1/\psi}{1-\gamma}} \right]^{\frac{1}{1-1/\psi}}$$

[64]que recupera la utilidad esperada estándar como caso particular cuando $\gamma = 1/\psi$.

La motivación empírica de esta separación es la siguiente: si se calibra γ para que el modelo reproduzca la elevada prima de riesgo observada en los mercados de acciones –lo que exige un agente muy averso–, la ligadura $\gamma = 1/\psi$ obliga simultáneamente a una ESI muy baja, un agente que apenas estaría dispuesto a sustituir consumo entre periodos, en clara contradicción con el crecimiento suave y persistente del consumo agregado observado empíricamente. Esta tensión, conocida como el problema del tipo de interés libre de riesgo (risk-free rate puzzle), es la que las preferencias de Epstein-Zin permiten resolver al desvincular ambos parámetros.

Una implicación sustantiva de esta generalización, ausente en la utilidad esperada estándar, es que el agente pasa a tener preferencias sobre el momento en que se resuelve la incertidumbre, no solo sobre su nivel: si $\gamma > 1/\psi$, el agente prefiere que la incertidumbre se resuelva cuanto antes. Si $\gamma < 1/\psi$, prefiere que se resuelva tarde. Por último, si $\gamma = 1/\psi$, es indiferente al momento de resolución, recuperando de nuevo el caso estándar.

[64]La normalización con el factor $(1-\beta)$ en el término de consumo corriente procede de Epstein y Zin (1991), el trabajo empírico companion del artículo teórico de 1989.

 © 2026 Manuel Cabado García

Esta propiedad es precisamente la que explotan los modelos de riesgo a largo plazo (long-run risk) en macrofinanzas, como el de Bansal y Yaron (2004): si el consumo agregado contiene un componente persistente de crecimiento incierto, un agente con $\gamma > 1/\psi$ penaliza mucho más ese riesgo prolongado que un riesgo puramente transitorio de la misma varianza, lo que permite generar primas de riesgo elevadas –explicando así el equity premium puzzle– sin recurrir a niveles de aversión al riesgo poco realistas, a diferencia de la utilidad esperada estándar, donde ambos tipos de riesgo se tratan de forma equivalente.

4.A.5. Catálogo de funciones de utilidad y aversión al riesgo

Se recogen a continuación formas funcionales concretas de $u(x)$ y sus coeficientes de aversión al riesgo, propuestos originalmente por Arrow (1965) y Pratt (1964):

- CARA ($u(w) = -e^{-\alpha w}$, $r_A = \alpha$ constante, $r_R = \alpha w$ creciente): la más manejable en modelos financieros con rendimientos normales, ya que al ser constante la aversión absoluta, la cantidad en euros invertida en el activo arriesgado no depende de la riqueza del agente.
- CRRA ($u(w) = \dfrac{w^{1-\gamma} - 1}{1-\gamma}$, $r_A = \gamma/w$ decreciente, $r_R = \gamma$ constante): la más usada en Macroeconomía, ya que al ser constante la aversión relativa, la proporción de riqueza invertida en el activo arriesgado no cambia con el nivel de riqueza. Liga además la aversión al riesgo con la elasticidad de sustitución intertemporal.
- Logarítmica (caso límite de la CRRA con $\gamma = 1$): la función que usó el propio Bernoulli para resolver la paradoja de San Petersburgo. Es el puente natural entre las familias anteriores y la más simple de manejar en ejemplos[65].
- Cuadrática ($u(w) = aw - bw^2$, r_A creciente): es la única función de utilidad bajo la cual el criterio media-varianza de Markowitz es exactamente óptimo con cualquier distribución de los rendimientos, no una aproximación (si los rendimientos son normales, también lo es con cualquier otra función de utilidad). Sin embargo, tiene aversión absoluta creciente (poco realista) y no genera ahorro precautorio, al ser nula su tercera derivada.

4.A.6. Ampliación de la Teoría de las Perspectivas: aplicaciones

La Teoría de las Perspectivas de Kahneman y Tversky (1979) ha resultado especialmente fértil para explicar diversos enigmas financieros de difícil justificación desde la óptica de la utilidad esperada tradicional:

[65] Sustituyendo esta utilidad en la utilidad esperada sobre los dos estados contingentes se obtiene $EU = p \cdot \ln(w_1) + (1-p) \cdot \ln(w_2)$, que es formalmente una Cobb-Douglas con cuotas p y $1-p$ sobre las riquezas contingentes: de ahí que dé soluciones tan simples en los modelos de seguro y cartera, al repartir la riqueza entre estados manteniendo una cuota constante –sin que ello implique ausencia de efecto sustitución, que sigue presente, solo compensado exactamente por el efecto renta–.

1. Equity premium puzzle: la inversión en acciones ofrece, de forma sistemática y a largo plazo, una rentabilidad muy superior a la de la deuda pública –más de 6 puntos porcentuales–, un diferencial excesivo para explicarse solo por una aversión al riesgo razonable. La explicación conductual apela a la aversión miope a las pérdidas: cuanto más frecuentemente evalúan los inversores el valor de su cartera, más dolorosas resultan las fluctuaciones a corto plazo, lo que les lleva a exigir una prima de riesgo desproporcionadamente alta a la renta variable.
2. Disposition effect: la renuencia observada de los inversores a vender acciones que arrastran minusvalías (prefieren esperar a recuperarse) frente a su tendencia a materializar rápidamente las plusvalías, coherente con el valor marginal decreciente de ganancias y pérdidas de la función en forma de S.
3. La rigidez a la baja de los salarios nominales (una reducción salarial se percibe como pérdida, mientras que la falta de ajuste al alza por inflación se percibe como una mera renuncia a una ganancia, menos dolorosa) y de los precios de la vivienda en épocas de recesión (los propietarios se resisten a vender por debajo de su precio de referencia, generado en la época de bonanza previa).

Pese a su capacidad para recoger estos hechos estilizados, la Teoría de las Perspectivas no ha encontrado, durante décadas, una gran aplicación práctica generalizada –son relativamente pocos los manuales de finanzas que la incorporan de forma sistemática–. Con el desarrollo más reciente de la economía conductual, sin embargo, su influencia ha aumentado notablemente en campos como la Macroeconomía, la economía de la salud, la economía pública, e incluso el ámbito jurídico, para explicar sesgos característicos en las decisiones de los jurados.

4.A.7. Sistemas de demanda flexibles: el modelo AIDS

Antes de llegar al AIDS, la econometría de la demanda había ensayado ya dos estrategias para estimar sistemas de demanda completos y coherentes con la teoría. La primera fue el Sistema Lineal de Gasto (LES) de Stone (1954), construido sobre la función de utilidad de Stone-Geary: el consumidor cubre primero un mínimo de subsistencia γ_i de cada bien y reparte el resto de la renta –"supernumeraria"– en proporciones fijas β_i, dando lugar a demandas lineales en precios y renta. Su ventaja era la simplicidad y la interpretación económica directa. Su limitación era una estructura de sustitución excesivamente rígida –determinada enteramente por los β_i, sin margen para patrones de sustitución más ricos entre bienes–.

La segunda vía fue el modelo de Rotterdam, desarrollado independientemente por Theil (1965) y Barten (1964) en la Escuela de Rotterdam. A diferencia del LES, no parte de una función de utilidad concreta: expresa directamente, en forma diferencial, el cambio de cada cuota presupuestaria como función de los cambios de precios y renta, apoyándose en

 © 2026 Manuel Cabado García

la propia descomposición de Slutsky. Esto le daba mucha más flexibilidad de sustitución, pero a costa de no derivar de una función de gasto explícita, lo que dificultaba su uso para el análisis de bienestar (véase el apartado Cálculo monetario de alteraciones en el bienestar).

El AIDS de Deaton y Muellbauer (1980) nace precisamente para combinar lo mejor de ambos: la flexibilidad de sustitución del modelo de Rotterdam con el fundamento en una función de gasto explícita del LES –los propios autores presentan su modelo en estos términos–.

El problema de la integrabilidad, estudiado en el cuerpo del tema, admite una solución práctica y directamente estimable: partir de una forma funcional flexible para la función de gasto que, por construcción, satisfaga las restricciones teóricas de la demanda, en lugar de deducir dichas restricciones a posteriori de una función de utilidad concreta. El Sistema Casi Ideal de Demanda (AIDS) de Deaton y Muellbauer (1980) es el ejemplo canónico de esta estrategia.

El AIDS parte de una función de gasto de la familia PIGLOG (price-independent generalized logarithmic), que agrega de forma exacta sobre consumidores heterogéneos:

$$\ln c\left(u,p\right)=\left(1-u\right)\ln a\left(p\right)+u\ln b\left(p\right)$$

donde $\ln a\left(p\right)$ es una función translog de los precios y $\ln b\left(p\right)$ le añade un término de tipo Cobb-Douglas:

$$\ln a\left(p\right)=\alpha_0+\sum_k\alpha_k\ln p_k+\frac{1}{2}\sum_k\sum_j\gamma_{kj}\ln p_k\ln p_j$$

$$\ln b\left(p\right)=\ln a\left(p\right)+\beta_0\prod_k p_k^{\beta_k}$$

Aplicando el lema de Shephard a $\ln c\left(u,p\right)$ y despejando u de la propia función de gasto –igualándola a la renta m–, se obtienen las ecuaciones de participación presupuestaria del bien i:

$$w_i=\alpha_i+\sum_j\gamma_{ij}\ln p_j+\beta_i\ln\left(\frac{m}{P}\right)$$

donde $w_i=p_ix_i/m$ es la cuota presupuestaria del bien i y P es el índice de precios translog:

$$\ln P=\alpha_0+\sum_k\alpha_k\ln p_k+\frac{1}{2}\sum_k\sum_j\gamma_{kj}\ln p_k\ln p_j$$

Los parámetros tienen una lectura económica directa: α_i es la cuota presupuestaria de referencia del bien i –a precios y renta base–, β_i mide cómo evoluciona esa cuota con la renta real –positivo si el bien es de lujo, negativo si es de primera necesidad–, y γ_{ij} recoge el patrón de sustitución entre los bienes i y j ante variaciones de precios relativos.

Esta forma funcional tiene la ventaja de ser lineal en los parámetros una vez fijado P, lo que explica su enorme popularidad aplicada: en la práctica P suele aproximarse

linealmente por el índice de Stone, $\ln P^* = \sum_k w_k \ln p_k$, dando lugar a la variante conocida como AIDS lineal (LA/AIDS), estimable por mínimos cuadrados en dos etapas.

Las restricciones teóricas de la demanda, vistas en el problema de la integrabilidad, se traducen aquí en restricciones lineales sobre los parámetros, directamente contrastables:

- Agregación (adding-up): $\sum_i \alpha_i = 1$, $\sum_i \beta_i = 0$, $\sum_i \gamma_{ij} = 0$ para todo j.
- Homogeneidad de grado cero en precios y renta: $\sum_j \gamma_{ij} = 0$ para todo i.
- Simetría de Slutsky: $\gamma_{ij} = \gamma_{ji}$.

Imponer las dos primeras equivale a estimar el sistema completo bajo restricciones lineales. Contrastar la simetría, en cambio, es una prueba genuina de la teoría –y buena parte de la literatura aplicada ha encontrado evidencia mixta al respecto–.

El AIDS resultó determinante para la econometría de la demanda por combinar flexibilidad funcional –no impone de antemano formas restrictivas de sustitución, a diferencia de la Cobb-Douglas o la CES–, con estimación sencilla y verificación directa de las propiedades teóricas. Su extensión más influyente, el QUAIDS de Banks et al. (1997) –que añade un término cuadrático en $\ln(m/P)$ para capturar curvas de Engel no lineales–, es hoy la herramienta habitual de organismos como el Banco Mundial, el FMI o la OCDE-FAO al simular el efecto distributivo de reformas de IVA o de subsidios alimentarios sobre distintos deciles de renta.

A modo de ilustración de este uso, van Oordt (2018) estima un QUAIDS completo para los hogares sudafricanos con el fin de comparar dos instrumentos redistributivos: mantener el tipo cero del IVA sobre alimentos básicos o eliminarlo y destinar la recaudación adicional a ampliar las transferencias monetarias existentes. Gracias a que el QUAIDS permite simular la respuesta de sustitución de cada hogar ante el cambio de precios relativos –no solo el efecto mecánico de la subida de precio–, el estudio concluye que la transferencia monetaria domina al tipo cero si la recaudación puede destinarse íntegramente a ese fin, pero que en ausencia de esa garantía el tipo reducido sigue protegiendo mejor a los hogares más pobres.

Este tipo de ejercicio se ha convertido en la práctica habitual de las instituciones que asesoran reformas del IVA en países en desarrollo: el propio Banco Mundial, en su documento de referencia sobre el diseño de tipos de IVA (Thomas, 2024), se apoya en simulaciones basadas en el AIDS/QUAIDS –como la de Van Oordt para Sudáfrica (2018) o la de Thomas para Nueva Zelanda (2022)– para argumentar que sustituir tipos reducidos generalizados por una combinación de tipo único y transferencias focalizadas suele mejorar la eficiencia recaudatoria sin perjudicar a los hogares de renta baja, siempre que el sistema de transferencias sea suficientemente amplio y fiable.

 © 2026 Manuel Cabado García

4.A.8. Economía del comportamiento y el *nudge* de Thaler

La teoría de las perspectivas abrió el camino a la economía del comportamiento, que incorpora al análisis de la elección los sesgos sistemáticos observados en las decisiones reales. Richard Thaler recibió en 2017 el Premio Nobel de Economía "por sus contribuciones a la economía del comportamiento" (Royal Swedish Academy of Sciences, 2017), y su aportación más influyente en la política pública es el *nudge*, literalmente "empujón".

El *nudge* y el paternalismo libertario Thaler y Sunstein (2008) definieron el *nudge* como cualquier aspecto de la arquitectura de la elección que altera el comportamiento de las personas de forma predecible sin prohibir ninguna opción ni cambiar significativamente sus incentivos económicos. Para ser un simple *nudge*, la intervención debe poder evitarse de forma fácil y barata. Colocar la fruta a la altura de los ojos es un *nudge*, prohibir la comida basura no lo es.[66]

Su fundamento es el paternalismo libertario. Una política es paternalista si se elige con el objetivo de influir en las decisiones de los afectados de forma que mejore su situación, y es libertaria si no hay coacción y cada uno conserva la libertad de elegir otra cosa (Thaler y Sunstein, 2003, p. 175). El argumento central es que algún grado de paternalismo es inevitable, porque quien diseña una política tiene que decidir qué ocurre si el individuo no elige, es decir, fijar una opción por defecto (Thaler y Sunstein, 2003, p. 176). En palabras de los autores, ninguna ley natural establece que, si el trabajador no dice nada, el 0 % de su salario vaya a un plan de pensiones (Thaler y Sunstein, 2003, p. 177).[67]

La evidencia: el poder de la opción por defecto

- Afiliación automática. Madrian y Shea (2001) estudiaron una empresa estadounidense que pasó a afiliar automáticamente a sus nuevos empleados a su plan de pensiones, con libertad para darse de baja. La afiliación inicial saltó del 49 % al 86 % (Thaler y Sunstein, 2003, p. 177).
- *Save More Tomorrow.* Thaler y Benartzi (2004) diseñaron un plan en el que los trabajadores se comprometen por adelantado a destinar al ahorro una parte de sus futuras subidas salariales. El 78 % de quienes recibieron la oferta se adhirió, el 80 % de ellos permaneció en el plan hasta la cuarta subida y su tasa de ahorro media pasó del 3,5 % al 13,6 % en 40 meses (Thaler y Benartzi, 2004, p. S164).[68]

[66] La intuición es que el mismo menú puede presentarse de muchas formas y ninguna es neutral. Quien organiza ese contexto, el arquitecto de la elección, influye en lo que se elige aunque no prohíba nada (Thaler y Sunstein, 2008).

[67] La intuición es que un formulario siempre tiene que traer alguna casilla marcada de antemano, y marcar "no" también es una decisión que condiciona lo que harán quienes no lo lean con atención.

[68] El diseño combina varios sesgos. La decisión se toma con antelación, el aumento del ahorro coincide con la subida salarial, de modo que el trabajador nunca ve bajar su nómina y no lo percibe como una pérdida, y la escalada es automática, con lo que la inercia juega a favor del ahorro. El programa influyó

- Donación de órganos. Johnson y Goldstein (2003) compararon países europeos con consentimiento explícito, en los que hay que apuntarse para ser donante, con países de consentimiento presunto, en los que hay que borrarse. El consentimiento efectivo iba del 4,25 % de Dinamarca al 27,5 % de los Países Bajos en el primer grupo, frente a valores entre el 85,9 % de Suecia y el 99,98 % de Austria en el segundo, casi 60 puntos de diferencia (Johnson y Goldstein, 2003, pp. 1338-1339). España, líder mundial en donación de fallecidos, muestra a la vez el alcance y el límite de este resultado. Tiene consentimiento presunto desde 1979, pero esa ley ha quedado inactiva, porque en la práctica siempre se pregunta a la familia y su decisión es definitiva. Entre 1979 y 1989 ni la tasa de donación ni la de negativas cambiaron apenas respecto a otros países europeos. El despegue llegó con el sistema nacional de donación organizado a partir de 1989, que redujo las negativas de en torno al 30-40 % a cerca del 15 % (Fabre et al., 2010, p. 922).[69]
- Ahorro para la jubilación en Dinamarca. Con registros daneses, Chetty et al. (2014) estimaron que cada dólar de subvención fiscal al ahorro para la jubilación solo aumentaba el ahorro total en un céntimo, mientras que las aportaciones automáticas lo elevaban de forma sustancial. La razón es que solo un 15 % de los individuos son ahorradores activos, que responden a la subvención trasladando ahorro de unas cuentas a otras, y el 85 % restante son pasivos, insensibles a la subvención pero muy influidos por lo que se aporta automáticamente en su nombre (Chetty et al., 2014, p. 1141).

Las unidades de *nudge* La idea pasó pronto a la administración:

- El Reino Unido creó en julio de 2010 el *Behavioural Insights Team*, un equipo de ocho personas en la Oficina del Gabinete, con vínculos estrechos con Thaler (Behavioural Insights Team, 2011, pp. 3-4).
- En Estados Unidos, la Orden Ejecutiva 13707, de 15 de septiembre de 2015, animó a las agencias federales a identificar políticas y programas en los que aplicar las ciencias del comportamiento pudiera mejorar sustancialmente el bienestar público y la eficacia del gasto, y formalizó el *Social and Behavioral Sciences Team* para asesorarlas (Executive Office of the President, 2015, pp. 56365-56366).
- La OCDE dedicó en 2017 un informe a 112 casos de 60 organismos públicos de 23 países y concluyó que las ciencias del comportamiento ya no podían verse como una moda pasajera (OECD, 2017, pp. 3 y 13). En 2023 contaba más de 300 instituciones

en la *Pension Protection Act* aprobada por el Congreso de Estados Unidos en 2006 (Committee for the Prize in Economic Sciences in Memory of Alfred Nobel, 2017, p. 15).

[69] La Ley 30/1979 permite la extracción de órganos de fallecidos salvo que hubieran dejado constancia expresa de su oposición (Jefatura del Estado, 1979, art. 5.2). La norma vigente obliga a consultar su voluntad a través de la familia (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2012, art. 9.1). En 2025 España registró 51,9 donantes fallecidos por millón de habitantes (Ministerio de Sanidad, 2026).

 © 2026 Manuel Cabado García

que las aplicaban a la política pública en 63 países, más de 200 de ellas dentro de los gobiernos de 56 países (Hubble y Varazzani, 2023).

- En España no existe una unidad estatal. La CNMC constató en 2025 que la economía conductual se aplica "aún de forma limitada" en el ámbito público, con experiencias como la de la Agencia Tributaria desde su Plan Estratégico 2020-2023, y recomendó crear unidades de economía conductual, empezando por una en la Administración General del Estado (CNMC, 2025, pp. 44 y 109).

Límites y críticas

- Efectos menores a gran escala. DellaVigna y Linos (2022) reunieron 126 ensayos aleatorios con 23 millones de personas, incluidos todos los de dos de las mayores unidades de *nudge* de Estados Unidos. En los artículos académicos el efecto medio era de 8,7 puntos, un 33,4 % sobre el grupo de control, mientras que en las unidades de *nudge* era de 1,4 puntos, un 8,0 %. La publicación selectiva de los resultados favorables explicaba en torno al 70 % de la diferencia (DellaVigna y Linos, 2022, p. 81).
- Sesgo de publicación. Mertens et al. (2022) obtuvieron, con más de 200 estudios, un efecto de tamaño pequeño a mediano, con una d de Cohen, el tamaño del efecto medido en desviaciones típicas, de 0,45. Maier et al. (2022) mostraron que, al corregir el sesgo de publicación, ese efecto caía a 0,04 y no quedaba evidencia de eficacia.[70]
- Crítica conceptual. Glaeser (2006) advirtió de que la racionalidad limitada afecta también a quien regula, y que los consumidores tienen más incentivos que los funcionarios para corregir sus propios errores, por lo que reconocer los límites de la cognición puede reforzar el argumento a favor de un Estado limitado. Sugden y otros autores criticaron, además, que el paternalismo libertario se aparta de la soberanía del consumidor (Committee for the Prize in Economic Sciences in Memory of Alfred Nobel, 2017, p. 16).

En suma, el *nudge* no sustituye a los instrumentos clásicos de la política económica, como los impuestos, las subvenciones o la regulación, sino que los complementa allí donde la inercia y la atención limitada pesan más que los precios, como ilustra el caso danés del ahorro para la jubilación.

[70] La intuición es que, si solo se publican los experimentos que salen bien, la media de lo publicado sobrestima el efecto real, igual que una media de notas calculada solo con los aprobados.

Bibliografía

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica 21*(4), 503–546.

Antonelli, G. B. (1886). *Sulla teoria matematica della economia politica*. Pisa: Tipografia Del Folchetto.

Arrow, K. J. (1964). The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies 31*(2), 91–96. Traducción inglesa de "Le rôle des valeurs boursières pour la répartition la meilleure des risques"(1953).

Arrow, K. J. (1965). *Aspects of the Theory of Risk-Bearing*. Helsinki: Yrjö Jahnssonin Säätiö.

Banks, J., R. Blundell y A. Lewbel (1997). Quadratic engel curves and consumer demand. *The Review of Economics and Statistics 79*(4), 527–539.

Bansal, R. y A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance 59*(4), 1481–1509.

Barten, A. P. (1964). Consumer demand functions under conditions of almost additive preferences. *Econometrica 32*(1/2), 1–38.

Behavioural Insights Team (2011). Behavioural insights team. annual update 2010–11. Informe técnico, Cabinet Office, Londres.

Bernoulli, D. (1738/1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica 22*(1), 23–36.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen y T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark. *Quarterly Journal of Economics 129*(3), 1141–1219.

CNMC (2025). Estudio sobre la economía conductual para una regulación y supervisión eficientes. Informe técnico E/CNMC/002/23, Comisión Nacional de los Mercados y la Competencia, Madrid.

Committee for the Prize in Economic Sciences in Memory of Alfred Nobel (2017). Richard h. thaler: Integrating economics with psychology. scientific background on the sveriges riksbank prize in economic sciences in memory of alfred nobel 2017. Informe técnico, Royal Swedish Academy of Sciences, Estocolmo.

Deaton, A. y J. Muellbauer (1980). An almost ideal demand system. *The American Economic Review 70*(3), 312–326.

 © 2026 Manuel Cabado García

DellaVigna, S. y E. Linos (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica 90*(1), 81–116.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics 75*(4), 643–669.

Epstein, L. G. y S. E. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica 57*(4), 937–969.

Epstein, L. G. y S. E. Zin (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy 99*(2), 263–286.

Executive Office of the President (2015). Executive order 13707 of september 15, 2015: Using behavioral science insights to better serve the american people. Federal Register, 80(181): 56365–56367.

Fabre, J., P. Murphy y R. Matesanz (2010). Presumed consent: A distraction in the quest for increasing rates of organ donation. *BMJ 341*, 922–924.

Friedman, M. y L. J. Savage (1948). The utility analysis of choices involving risk. *Journal of Political Economy 56*(4), 279–304.

Glaeser, E. L. (2006). Paternalism and psychology. *University of Chicago Law Review 73*(1), 133–156.

Hicks, J. R. y R. G. D. Allen (1934). A Reconsideration of the Theory of Value. Part II. A Mathematical Theory of Individual Demand Functions. *Economica 1*(2), 196–219.

Hubble, C. y C. Varazzani (2023). Mapping behavioural insights. OECD Observatory of Public Sector Innovation, 10 de mayo de 2023.

Hurwicz, L. y H. Uzawa (1971). On the integrability of demand functions. En J. S. Chipman, L. Hurwicz, M. K. Richter y H. F. Sonnenschein (eds.), *Preferences, Utility, and Demand*. New York: Harcourt Brace Jovanovich.

Jefatura del Estado (1979). Ley 30/1979, de 27 de octubre, sobre extracción y trasplante de órganos. BOE n.º 266, de 6 de noviembre de 1979.

Johnson, E. J. y D. Goldstein (2003). Do defaults save lives? *Science 302*(5649), 1338–1339.

Kahneman, D. y A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–291.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston and New York: Houghton Mifflin.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics 47*(1), 13–37.

Madrian, B. C. y D. F. Shea (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics 116*(4), 1149–1187.

Maier, M., F. Bartoš, T. D. Stanley, D. R. Shanks, A. J. L. Harris y E.-J. Wagenmakers (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences 119*(31), e2200300119.

Markowitz, H. (1952a). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Markowitz, H. (1952b). The utility of wealth. *Journal of Political Economy 60*(2), 151–158.

Mertens, S., M. Herberz, U. J. J. Hahnel y T. Brosch (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences 119*(1), e2107346118.

Ministerio de Sanidad (2026). España vuelve a superar los 6.300 trasplantes en 2025. Nota de prensa, 16 de enero de 2026.

Ministerio de Sanidad, Servicios Sociales e Igualdad (2012). Real decreto 1723/2012, de 28 de diciembre, por el que se regulan las actividades de obtención, utilización clínica y coordinación territorial de los órganos humanos destinados al trasplante. BOE n.º 313, de 29 de diciembre de 2012.

Nicholson, W. y C. M. Snyder (2017). *Microeconomic Theory: Basic Principles and Extensions*. Boston: Cengage Learning.

OECD (2017). *Behavioural Insights and Public Policy: Lessons from Around the World*. París: OECD Publishing.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica 32*(1), 122–136.

Rothschild, M. y J. E. Stiglitz (1970). Increasing risk: I. a definition. *Journal of Economic Theory 2*(3), 225–243.

Roy, R. (1947). La distribution du revenu entre les divers biens. *Econometrica 15*(3), 205–225.

 © 2026 Manuel Cabado García

Royal Swedish Academy of Sciences (2017). The prize in economic sciences 2017. press release. Nota de prensa, 9 de octubre de 2017.

Rubinstein, A. (2019). *Lecture Notes in Microeconomic Theory: The Economic Agent*. Tel Aviv: Open access edition.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: John Wiley & Sons.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*(3), 425–442.

Shephard, R. W. (1953). *Theory of cost and production functions*. Princeton University Press.

Slutsky, E. (1915). Sulla teoria del bilancio del consumatore. *Giornale degli Economisti e Rivista di Statistica 51 (Anno 26)*(1), 1–26.

Stone, R. (1954). Linear expenditure systems and demand analysis: An application to the pattern of british demand. *The Economic Journal 64*(255), 511–527.

Thaler, R. H. y S. Benartzi (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy 112*(S1), S164–S187.

Thaler, R. H. y C. R. Sunstein (2003). Libertarian paternalism. *American Economic Review 93*(2), 175–179.

Thaler, R. H. y C. R. Sunstein (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Theil, H. (1965). The information approach to demand analysis. *Econometrica 33*(1), 67–87.

Thomas, A. (2022). Who would win from a multi-rate GST in new zealand: Evidence from a QUAIDS model. *New Zealand Economic Papers 56*(2), 141–168.

Thomas, A. (2024). VAT rate structures in theory and practice. Policy Research Working Paper 10677, World Bank, Washington, DC.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies 25*(2), 65–86.

van Oordt, M. (2018). Zero-rating versus cash transfers under the vat. *Fiscal Studies 39*(3), 489–515.

Varian, H. R. (2014). *Intermediate Microeconomics: A Modern Approach* (9 ed.). New York: W. W. Norton.

Von Neumann, J. y O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton university press.

 © 2026 Manuel Cabado García

Tema 5

Teoría de la producción. Caracterización de la tecnología de la empresa en el corto y el largo plazo. El conjunto de posibilidades de producción. La función de producción. Rendimientos locales y globales a escala. Elasticidad de sustitución. Producción conjunta.

5.1. Introducción ⏳ 1'

Toda empresa, sea cual sea su actividad, se enfrenta a una misma pregunta de fondo: cómo transformar sus recursos en bienes y servicios de la manera más eficiente posible. Esta pregunta, aparentemente sencilla, es el punto de partida de la Teoría de la Producción, uno de los pilares del análisis microeconómico de la empresa.

Las empresas son, además, actores centrales de la economía: impulsan la inversión, generan empleo y son una fuente vital de ingresos impositivos, al tiempo que se enfrentan hoy a desafíos como la transición hacia métodos de producción más verdes, eficientes y digitalizados. Comprender sus restricciones técnicas no es solo un ejercicio teórico: es el primer paso para explicar la oferta de las empresas en el mercado y, por extensión, cuestiones tan relevantes como el crecimiento económico o el comercio internacional.

Concretamente, se estudiará esta teoría desde la óptica neoclásica, que concibe la

empresa como una "caja negra". Abstrayéndose de su propiedad, dirección u organización –aspectos propios de la Teoría de la Organización Industrial (TOI)–, se centrará en las combinaciones de insumos técnicamente eficientes y en el nivel máximo de producción alcanzable dada una tecnología. Esta exposición se ceñirá, por tanto, a las restricciones puramente técnicas que enfrenta la empresa[1], dejando a un lado las restantes restricciones a las que también está sujeta. Para ello, se seguirá el siguiente esquema:

1. Construcción de la función de producción, a partir de los axiomas del conjunto de posibilidades de producción.
2. Análisis de la función de producción, distinguiendo entre diferentes horizontes temporales.
3. Producción multiproducto, a partir de una caja de Edgeworth.

5.2. Construcción de la función de producción ⏳6'

Comenzando con la construcción de la función de producción, esta debe reflejar el hecho de que el empresario es un maximizador condicionado, es decir, solo le interesarán las combinaciones técnicamente eficientes.

Para ello, se parte de la representación de la tecnología mediante un vector de producción $y = (y_1, \ldots, y_L) \in \mathbb{R}^L$, donde si $y_i > 0$ es un output y si $y_i < 0$ es un input. A partir de este vector ya será posible construir un conjunto de producción que recoja todas las técnicas posibles, $Y \subset \mathbb{R}^L$.

La teoría de la producción establece una serie de axiomas –implican formas concretas del conjunto de producción y tienen diferentes consecuencias– sobre el conjunto de producción suponiendo que existen dos bienes –Y_1 y Y_2–, que pueden ser inputs o outputs, y que pueden observarse a partir de la representación gráfica del conjunto en la Figura 5.1.

Axiomas sobre el conjunto de producción Dichos axiomas sobre el conjunto son:

- **Axioma 1**: Conjunto no vacío, $Y \neq \emptyset$, es decir, al menos debe existir un vector de producción posible, pues de lo contrario no habría actividad productiva alguna. Y tiene al menos un punto.
- **Axioma 2**: Cerrado[2], es decir, el conjunto incluye su propia frontera –propiedad fundamentalmente técnica–. La curva PP' es la frontera del conjunto de posibili-

[1] Sin embargo, es crucial recordar que, en su actividad, la empresa también está sujeta a una restricción económica –de costes– y a una institucional –relacionada con la estructura de mercados–.

[2] Cerrado y acotado son propiedades topológicas independientes entre sí: cerrado significa que el conjunto incluye todos sus puntos frontera, mientras que acotado significa que no se extiende hasta el infinito, sino que cabe dentro de una región finita del espacio. Compacto equivale a cerrado y acotado a la vez –en $\mathbb{R}^n$, por el teorema de Heine-Borel–, una condición estrictamente más fuerte que exigir solo una de las dos por separado, y no una redundancia. Que Y fuera acotado implicaría un techo físico absoluto a la producción, un nivel máximo de output que no podría superarse por mucho que se incrementaran los

 © 2026 Manuel Cabado García

dades de producción[3], suponiendo que se puede utilizar bien Y_1 para producir Y_2 o viceversa, siendo la superficie rayada del gráfico el conjunto de posibilidades de producción.

- **Axioma 3**: Eliminación gratuita, si un plan de producción es técnicamente posible, también lo es cualquier otro que implique iguales o mayores cantidades aplicadas de inputs e iguales o menores cantidades de output. Esto implica que, por ejemplo, la producción Y_2^0 se puede conseguir con cantidades mayores de inputs, por ejemplo Y_1^1 en lugar de Y_1^0. La implicación es que la empresa puede eliminar las mercancías en exceso.
- **Axioma 4**: Imposibilidad de producción libre, no se pueden producir cantidades positivas de algún output sin emplear ningún input –esto es, no free lunch–. Este axioma impide que algún punto del primer cuadrante que no sea el origen pertenezca al conjunto de producción.
- **Axioma 5**: Posibilidad de inacción: $0 \in Y$. La empresa puede cesar la actividad. Este axioma garantiza que la frontera pasa por el origen. Sin embargo, si existen costes hundidos[4], es decir, si las decisiones de producción ya se han tomado, la inacción no es posible (gráfico adjunto). Esta diferencia es relevante entre el corto y el largo plazo.
- **Axioma 6**: Irreversibilidad, no se puede invertir el proceso productivo, es decir, no se pueden transformar unos outputs en la cantidad de inputs que se utilizaron para su producción. Por lo tanto, no se puede cambiar el papel de los inputs y los outputs.

Axiomas sobre los rendimientos Además, se añaden Axiomas que tienen relación con las características de escala de la tecnología. Aunque no siempre se cumplen simultáneamente por ser algunos incompatibles entre sí, determinan la forma de la frontera del conjunto. De esta forma, estas propiedades determinan los rendimientos:

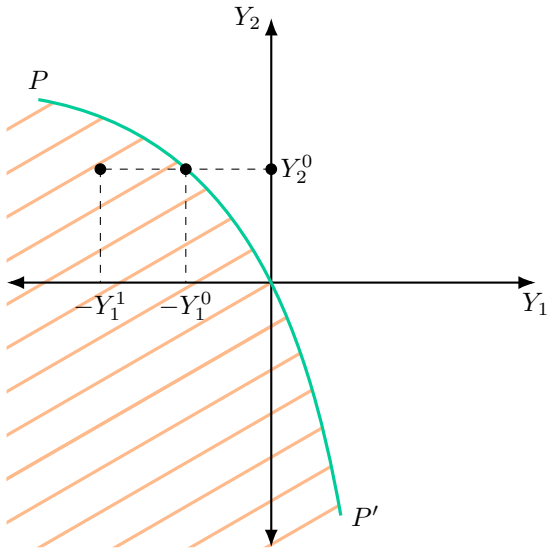
- **Axioma 1:** Aditividad $\forall y_1, y_2 \in Y \Rightarrow y_1 + y_2 \in Y$, es decir, si dos vectores son técnicamente posibles su suma también lo es. También se le llama axioma de libre

inputs, lo que descartaría de raíz cualquier tecnología con rendimientos crecientes a escala sin límite, por ejemplo, una Cobb-Douglas con $\alpha + \beta > 1$, o la función lineal de sustitutos perfectos. Por eso ningún manual estándar lo exige: el axioma de cerrado ya es suficiente para la teoría, pues garantiza que la frontera pertenece al conjunto, y añadir acotación sería una restricción innecesaria y contraria a cómo se modelizan tecnologías reales.

[3]Frontera eficiente o de transformación: Conjunto de puntos de la frontera superior del conjunto de producción. Formada por la combinación técnica eficiente de factores de producción tal que no es posible incrementar la cantidad de X con la misma cantidad de Y ni producir la misma cantidad de X con menor cantidad de Y. Intuitivamente los puntos eficientes se encuentran sobre la frontera del conjunto de posibilidades de producción, aunque ello no es condición suficiente de eficiencia.

[4]O contratos ya firmados con clientes u otro tipo de compromiso que no permita cerrar en el corto plazo.

Figura 5.1: Conjunto de producción y frontera de posibilidades de producción



entrada, ya que si una empresa produce y_1, y entra otra y produce y_2, el conjunto agregado de la producción debe satisfacer la aditividad siempre que exista entrada sin restricciones. Es decir, el plan de producción puede reescalarse hacia arriba. Tiene dos importantes implicaciones:

- Macro: Libertad de entrada, en tanto que el conjunto de producción agregado cumple condiciones aditivas.
- Micro: Posibilidad de montar dos plantas independientes sin que interfieran la una con la otra
- Además, puede implicar rendimientos constantes o crecientes a escala.

- **Axioma 2**: Divisibilidad del proceso productivo, si $y_1 \in Y$ y $\lambda \in [0,1] \Rightarrow \lambda \cdot y_1 \in Y$, es decir, supone la posibilidad de reducir la escala de una combinación productiva factible tanto como se desee. El plan puede reescalarse hacia abajo, lo que implica rendimientos decrecientes o constantes.
- **Axioma 3**: Convexidad, cualquier media ponderada de dos planes de producción factibles es también factible, es decir, si $y_0, y_1 \in Y$ y $\alpha \in [0,1]$, entonces $\alpha y_0 + (1-\alpha)\, y_1 \in Y$[5][6]. Implicaciones:
  - Perfecta divisibilidad de los planes de producción.
  - Rendimientos no crecientes.

[5]La convexidad implica la complementariedad de los factores productivos.

[6]Si se cumplen conjuntamente la aditividad y la divisibilidad, el conjunto de producción es convexo: dados $y_0, y_1 \in Y$ y $\alpha \in [0,1]$, la divisibilidad garantiza que $\alpha y_0 \in Y$ y $(1-\alpha)\, y_1 \in Y$, y la aditividad, que su suma $\alpha y_0 + (1-\alpha)\, y_1 \in Y$. Lo contrario no es cierto: un conjunto convexo con rendimientos decrecientes cumple la divisibilidad pero no la aditividad. La intuición es que, si una empresa sabe producir con dos técnicas, también puede repartir su producción entre ambas en cualquier proporción.

 © 2026 Manuel Cabado García

- Combinaciones equilibradas de inputs son más eficientes que las no equilibradas[7].

- **Axioma 4**: Convexidad estricta, cualquier media ponderada de dos planes factibles distintos queda en el interior del conjunto, de modo que con esa mezcla podría obtenerse más output que la media ponderada de los de cada plan. Esto implica rendimientos estrictamente decrecientes.

Los axiomas anteriores permiten representar la tecnología mediante una función de producción. El conjunto de posibilidades de producción muestra los planes técnicamente posibles y la función de producción incluye aquellos planes técnicamente eficientes.

Función de producción[8]

La función de producción se define como aquella relación técnica que relaciona cantidades de insumos empleados y los bienes y servicios producidos en un momento del tiempo y con una tecnología dada. La función de producción recoge el máximo producto que se puede obtener dado un vector de factores productivos[9]. De esta forma, la función de producción es la frontera del conjunto de posibilidades de producción cuando hay un único output: indica el máximo volumen de producto que puede obtenerse con cada vector de inputs[10].

Dicha función[11] recoge, por tanto, solo métodos de producción técnicamente eficientes, entendiendo por tales aquellas combinaciones de inputs que producen la mayor cantidad de output posible, o que, produciendo igual output, precisan menor cantidad de inputs[12].

[7] En otras palabras, si se consideran dos planes de producción que generan el mismo output con diferentes combinaciones de input, un nuevo plan de producción utilizando la media ponderada de los inputs de ambos planes será posible y el resultado debería ser al menos un output tan grande como el de los planes de producción iniciales.

[8] En este tema se habla de la función de producción en genérico. No se considera el supuesto sobre los rendimientos a escala de la función dado que no es relevante para este tema al no estar considerando la maximización de beneficios. Luego, dependiendo del tema, se concreta una función de producción u otra. Por ejemplo, en comercio internacional se utiliza una función HG1 y en competencia una función con rendimientos decrecientes estrictamente cóncava.

[9] Es preciso distinguir función de producción de método de producción, que es una combinación de insumos factoriales requeridos para la producción de una unidad de output. La función de producción incluye todos los métodos de producción técnicamente eficientes.

[10] Es importante tener presente que no todos los conjuntos de posibilidades de producción son susceptibles de ser representados por medio de una FP.

[11] La función de producción puede ser de:

- Producción conjunta (varios inputs, varios outputs). Distingue vectores input output y atribuye cada input al producto al que está destinado (separable). $f(x_1, \ldots, x_n, y_1, \ldots, y_n)$
- Producción simple (un único output y varios inputs). El análisis se centrará en esta y en la última parte del tema se profundizará en la producción conjunta a partir de un análisis en una caja de Edgeworth. $f(x, y_1, \ldots, y_n)$.

[12] Esta definición es compatible con el supuesto de racionalidad de los agentes, que buscará el mayor aprovechamiento posible de los insumos.

Propiedades Las propiedades genéricas de esta función, derivadas directamente de los axiomas del conjunto de producción, son las siguientes:

1. Cardinal, pues el producto se mide en unidades físicas y su valor numérico tiene significado, a diferencia de la función de utilidad del consumidor, que es meramente ordinal.
2. Pasa por el origen, $f(0) = 0$, por el axioma de imposibilidad de producción libre.
3. No decreciente en los factores, es decir, con productividad marginal no negativa, por el axioma de eliminación gratuita.
4. Continua y dos veces diferenciable[13]. Esto implica perfecta (infinita) divisibilidad de los factores productivos (inputs)[14].

Supuestos adicionales, según la aplicación Más allá de estas propiedades genéricas, según la aplicación concreta se añaden supuestos adicionales sobre la curvatura de la función, sin los cuales no es posible decir nada sobre sus rendimientos a escala:

- Cuasiconcavidad estricta, que implica cierta complementariedad entre los factores, es decir, isocuantas convexas[15].
- Concavidad estricta, más exigente que la cuasiconcavidad, pues combina complementariedad de los factores y rendimientos no crecientes a escala. Es el supuesto habitual en competencia perfecta, donde resulta necesario para que el problema de maximización del beneficio esté acotado.
- Homogeneidad de un grado concreto, por ejemplo, homogénea de grado 1 en comercio internacional.
- Productividad marginal decreciente, un supuesto adicional habitual en la práctica, pero distinto de –y no derivado automáticamente de– la cuasiconcavidad.

Cuando una función es homogénea de grado 1 y tiene productividades marginales positivas pero decrecientes en cada factor, se habla de una función neoclásica "de buen comportamiento"[16].

[13] La diferenciación de la función respecto a un factor proporciona su PMg y garantiza que la pendiente de la función de producción esté bien definida en cualquier punto.

[14] Aunque en la realidad los insumos no son perfectamente divisibles, se asume por comodidad. Si bien existen infinitos insumos que pueden ser considerados, generalmente se consideran dos insumos: Trabajo (L) y Capital (K) → (capital en sentido abstracto, no necesariamente maquinaria, es decir, dinero que se adelanta en el proceso productivo)

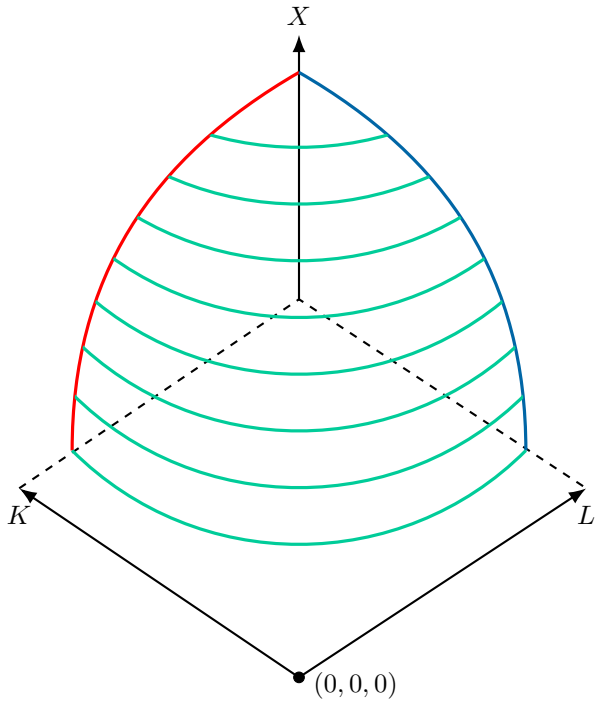
[15] Jehle y Reny (2011, Assumption 3.1) adoptan la cuasiconcavidad estricta como supuesto de partida en su tratamiento genérico de la función de producción, precisamente para no comprometerse con ningún tipo de rendimiento a escala. Toda función cóncava es cuasicóncava, pero no a la inversa. Lo mismo ocurre con la cuasiconvexidad respecto a la convexidad.

[16] Las funciones neoclásicas de buen comportamiento son, por tanto, crecientes, continuas, dos veces diferenciables, homogéneas de grado 1 (HG1) y con PMg positivas pero decrecientes. Son cóncavas, pero no estrictamente, pues la homogeneidad de grado 1 las hace lineales a lo largo de cada rayo vector. La

 © 2026 Manuel Cabado García

Región de producción Gráficamente, en un plano de tres ejes como el de la Figura 5.2, representando dos factores de producción, capital, K, y trabajo, L, y la cantidad producida, X, en el eje vertical, es posible representar la región de producción. Obsérvese que la superficie representada supone, a efectos meramente ilustrativos, concavidad estricta. Tal y como se ha señalado, esta es solo una de las curvaturas posibles de la función de producción, no una propiedad genérica de la misma.

Figura 5.2: Función de producción en tres dimensiones



5.3. Análisis de la función de producción 10'30"

Una empresa tarda en ajustar sus factores para producir con diferentes cantidades de trabajo y de capital[17]. Como las empresas deben preguntarse si pueden alterar o no los factores y, en caso afirmativo, en qué periodo de tiempo, es importante distinguir entre el corto plazo, el largo plazo y el muy largo plazo cuando se analiza la función de producción.

intuición es que duplicar a la vez el trabajo y el capital duplica la producción, mientras que añadir solo más trabajo a la misma maquinaria rinde cada vez menos.

[17]Una fábrica nueva debe planificarse y construirse y la maquinaria y demás equipo de capital debe pedirse y entregarse. Estas actividades tardan en realizarse fácilmente un año o más, por lo que si se analizan las decisiones de producción de un breve periodo de tiempo, como un mes o dos, es improbable que la empresa pueda sustituir mucho trabajo por capital.

Análisis de la función de producción a corto plazo

Se entiende por corto plazo el horizonte temporal en el que al menos un factor de producción, generalmente el capital, es fijo, de modo que la producción solo puede alterarse modificando los factores variables. Se supone, además, que no hay progreso técnico. El horizonte temporal de corto plazo permite introducir el concepto de productividad, que relaciona la producción con las dotaciones factoriales.

Curva de producto total Gráficamente, en un plano producción, X, y trabajo, L, como el de la Figura 5.3, para analizar la función de producción se realizarán cortes verticales sobre la región de producción. De esta manera, se obtiene la curva de producto total, que representa el nivel máximo de producción para cada nivel de trabajo L, dado el capital K.

La curvatura de esta función deriva de la ley de rendimientos decrecientes[18]. El análisis sobre cómo se ve alterada la producción al modificar el factor variable se puede realizar en términos medios o en términos marginales.

Productividad media En términos medios se utilizará la productividad media, PMe, que es la cantidad media de producto por cada unidad del factor productivo

$$PMe = \frac{X}{L} \tag{5.1}$$

y que corresponde gráficamente a la pendiente del rayo vector que une el origen con cada punto de la función de producción.

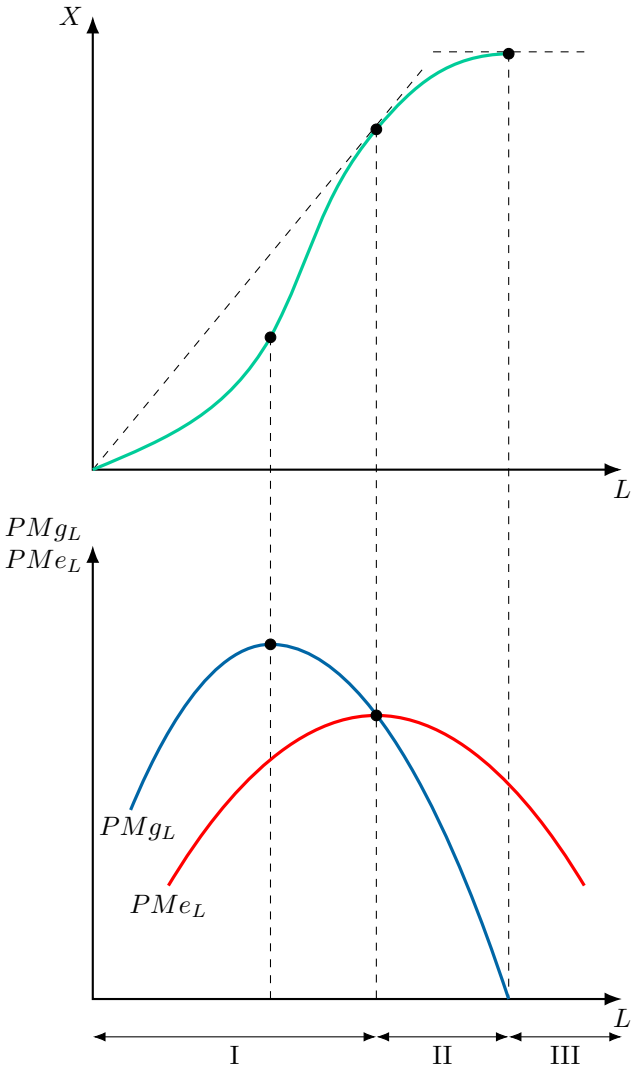
Productividad marginal En términos marginales se aplicará el concepto de productividad marginal, PMg, que es el aumento de la producción por cada unidad adicional

[18]Concretamente, se introduce la ley de las proporciones variables, enunciada por Turgot en 1767 (Turgot, 1767) como respuesta teórica a un ensayo fisiócrata sobre la agricultura francesa —no a partir de observación empírica de explotaciones inglesas—, al buscar la cantidad óptima de trabajo aplicable a una cantidad fija de tierra. De acuerdo con esta ley, dado un factor fijo (por ejemplo, tierra), llegará un punto en que unidades adicionales del factor variable (trabajo) generarán rendimientos marginales decrecientes (es decir, cada unidad adicional de trabajo añadirá menos producto que la anterior). En palabras del propio Turgot, pasado cierto punto, al aumentar las avances "les produits augmenteront encore, mais moins, et toujours de moins en moins". En el caso representado se observa cómo se cumple la ley de rendimientos marginales mixtos: la productividad marginal, es decir, la pendiente, es máxima en el punto de inflexión de la función de producción (pues la productividad antes crece y después decrece).

También se puede dar el caso donde la función de producción tenga forma de U inversa, donde estos rendimientos marginales decrecientes se dan desde el principio. La productividad marginal, es decir, la pendiente, es siempre decreciente.

 © 2026 Manuel Cabado García

Figura 5.3: Producción con un factor variable



del factor variable, en el límite de variaciones infinitesimales[19]

$$PMg = \frac{\partial X}{\partial L} \tag{5.2}$$

y que en la gráfica es la pendiente de la función de producción en cada punto.

Fases de la curva de producto total Gráficamente en un plano adyacente de producto marginal y medio del trabajo, PMg_L y PMe_L, y trabajo se representan las curvas de producto marginal y medio. Esto permite identificar, siguiendo a Cassels (Cassels, 1936), las tres fases en la curva de producto total:

- **Fase I**, o de producción extensiva, donde $PMg > PMe$, por lo que es posible seguir

[19]La intuición es que la productividad marginal responde a la pregunta de cuánto más se produce si se contrata un trabajador más, mientras que la media indica cuánto produce, en promedio, cada trabajador ya contratado.

produciendo ya que el capital está infrautilizado. Se está produciendo extensivamente, por lo que el punto que separa la etapa I de la II, en el que la productividad media es máxima e igual a la marginal, se denomina óptimo técnico o margen extensivo. La producción es ineficiente.

- **Fase II**, donde $PMg < PMe$ y además $PMg > 0$, donde deberán situarse las empresas en el corto plazo (aunque el punto exacto dentro de este tramo dependerá de la estructura de costes de la empresa). En esta fase, la relación de los factores productivos es eficiente, porque se está en la región económica de producción. El punto que separa la etapa II de la III, en el que la producción total es máxima y la productividad marginal se anula, se denomina margen intensivo o máximo técnico.[20]
- **Fase III**, o de producción intensiva, donde $PMg < 0$, es decir, se emplea tanto trabajo para el capital disponible que cada trabajador adicional reduce la producción total. Las empresas deberán reducir el empleo del factor variable, con lo que aumentarían su producción. La producción es técnicamente ineficiente.[21]

Análisis de la función de producción a largo plazo

Por otro lado, en el largo plazo se va a suponer que todos los factores son variables pero sin progreso técnico. Esto implica que no van a existir costes hundidos y por lo tanto va a haber libre entrada y salida de empresas en el mercado, ya que el factor capital también es variable, pudiendo desinvertir. El largo plazo es un horizonte de planificación.

Curvas isocuantas En este horizonte temporal, la Teoría Neoclásica analiza el comportamiento de la empresa valiéndose de las curvas isocuantas que son las curvas de nivel de la función de producción, definidas como el lugar geométrico de las combinaciones de factores que permiten obtener un mismo nivel de producto[22].

Gráficamente, en un plano de factor trabajo, L, y factor capital, K, como el de la Figura 5.4 se realizarán cortes horizontales sobre la región de producción, permitiendo representar las isocuantas.

Propiedades de las isocuantas Poseen las siguientes propiedades:

1. Mayor producción cuanto más alejadas del origen (porque la función de producción es creciente).

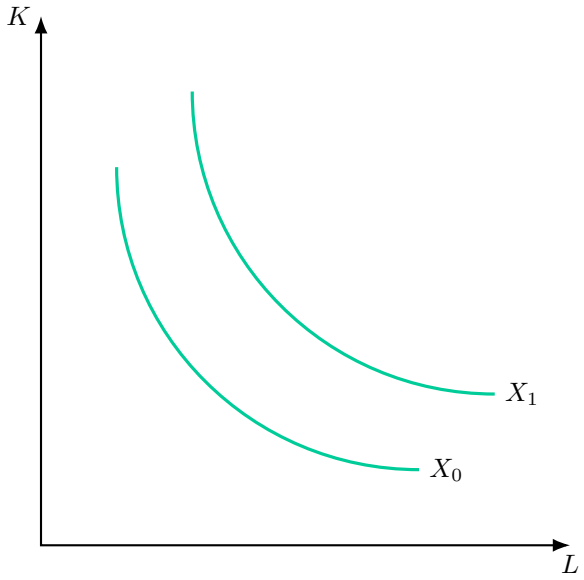
[20] En esta fase es donde se opera en competencia perfecta, aunque en situación de monopolio o en competencia monopolística no tiene que ser así necesariamente.

[21] La intuición es la de una cocina pequeña con demasiados cocineros: a partir de cierto número se estorban y cada cocinero adicional hace que salgan menos platos.

[22] Las isocuantas describen la tecnología en cualquier plazo, pero solo en el largo plazo puede la empresa desplazarse a lo largo de ellas, pues a corto plazo uno de los factores está fijo.

 © 2026 Manuel Cabado García

Figura 5.4: Mapa de curvas isocuantas



2. Convexas, porque se supone que la función de producción es, al menos, cuasicóncava[23].
3. De pendiente negativa.
4. Continuas.
5. Diferenciables.
6. No se cortan.
7. Cardinales.

Relación Marginal de Sustitución Técnica La pendiente de la isocuanta viene dada, cambiada de signo, por la Relación Marginal de Sustitución Técnica, $RMST$, que es la cantidad de capital que puede sustituirse por una unidad adicional de trabajo manteniendo constante la producción, y que equivale al cociente de las productividades marginales de los factores

$$RMST_{L,K} = -\left.\frac{\partial K}{\partial L}\right|_{\bar{X}} = \frac{PMg_L}{PMg_K} \tag{5.3}$$

es decir, gráficamente la $RMST$ es el valor absoluto de la pendiente de la isocuanta.

Grado de sustituibilidad entre los factores: elasticidad de sustitución Sin embargo, para el grado de sustituibilidad de los factores no se utiliza la RMST, pues ésta es una medida absoluta, es decir, está afectada por las unidades de medida de capital y trabajo. Lo que se utiliza es la elasticidad de sustitución, σ, que mide la variación porcentual

[23]De hecho, para que las isocuantas sean convexas basta con que la función de producción sea cuasicóncava (toda función cóncava es cuasicóncava, pero no al revés).

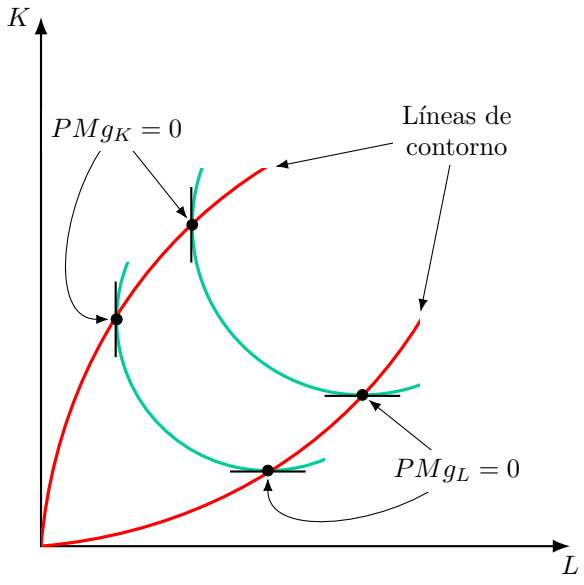
de la proporción capital/trabajo ante una variación porcentual de la $RMST$

$$\sigma = \frac{\partial\,(K/L)}{\partial RMST_{L,K}} \cdot \frac{RMST_{L,K}}{K/L} \tag{5.4}$$

Al ser un cociente de variaciones porcentuales, no depende de las unidades de medida. Cuanto mayor es σ, menos curvada (más recta) es la isocuanta y más fácil resulta sustituir un factor por otro, y su valor permite clasificar las funciones de producción[24].

Región económica de producción A su vez, en el largo plazo también es posible determinar en qué zona deben situarse los productores. Desde un punto de vista exclusivamente de la teoría de la producción, la solución deberá situarse dentro de la región económica de producción, de modo que se cumpla la condición de eficiencia técnica. Dicha área, inscrita entre las líneas de contorno, que se obtienen identificando los puntos en los que las productividades marginales de los factores se igualan a 0[25], como se observa en la Figura 5.5. Además, las líneas de contorno son también isoclinas[26], ya que unen puntos con la misma pendiente.

Figura 5.5: Región económica de producción



[24]Por ejemplo, si la elasticidad de sustitución es igual a dos, ante una variación del 1 % de la RMST, la ratio K/L varía un 2 %.

[25]Nótese que la región económica tiene forma de "almendra" ya que las isocuantas no son circunferencias. Si fueran circunferencias, la línea de contorno superior sería una línea totalmente horizontal (pues la isocuanta tendría pendiente infinita siempre en la misma línea horizontal), y la línea de contorno inferior sería una línea totalmente vertical (pues la isocuanta tendría pendiente cero siempre en la misma línea vertical), siendo la región económica un cuadrilátero.

[26]Isoclinas: Línea a lo largo de la cual la RMST es constante. Puede presentar cualquier forma según la correspondiente función de producción. En el caso de las funciones homogéneas de grado uno (como la CES) o homotética, la isoclina es una recta.

En otras palabras, es la línea que une los puntos donde la RMST de cada isocuanta es constante.

 © 2026 Manuel Cabado García

Rendimientos de escala Finalmente, para analizar los efectos que tiene sobre el producto la variación de los factores se emplea el concepto de rendimientos de escala. Para ello, se supone que todos los factores varían en la misma proporción. Además, será necesario introducir el concepto de elasticidad de escala, e, que se define como la variación porcentual de la producción ante la misma variación porcentual y proporcional de todos los factores de producción, que coincide con la suma de las elasticidades producto y con el grado de una función homogénea, t, si la función fuera homogénea[27]

$$e = \frac{\partial X}{\partial \lambda} \cdot \frac{\lambda}{X} = \varepsilon_{X,K} + \varepsilon_{X,L} = t \tag{5.5}$$

Una función es homogénea de grado t si, al multiplicar todos sus argumentos por un escalar positivo, λ, el valor de la función queda multiplicado por ese escalar elevado a t[28]

$$F(\lambda \cdot L, \lambda \cdot K) = \lambda^t \cdot F(L, K) \tag{5.6}$$

Las funciones homogéneas tienen dos propiedades útiles:

1. Por el teorema de Euler, la suma de las productividades marginales multiplicadas por las cantidades de cada factor es igual a t veces el producto, de modo que, si $t = 1$, remunerar cada factor según su productividad marginal agota exactamente el producto.[29]
2. Además, las productividades marginales son homogéneas de grado $t-1$: si todos los factores se multiplican por S, cada PMg queda multiplicada por S^{t-1} y su cociente, la $RMST$, no cambia.

Ahora bien, el grado de homogeneidad –o elasticidad de escala, e– determina el tipo de

[27] Además, $e = \dfrac{CMe}{CMg}$ si la función de producción es homotética.

[28] La intuición es que el grado indica cuánto crece la producción cuando se amplía la escala: si se duplican todos los factores, la producción se multiplica por 2 si $t = 1$, por más de 2 si $t > 1$ y por menos de 2 si $t < 1$.

[29] La intuición es que, con rendimientos constantes, si cada trabajador y cada máquina cobran exactamente lo que aportan en el margen, el reparto se lleva todo lo producido, sin que sobre ni falte nada.

rendimiento a escala que presenta una función[30][31][32]:

- Si $e = t = 1$ entonces presenta rendimientos constantes a escala, la producción varía en la misma proporción que los factores[33].
- Si $e = t > 1$ entonces presenta rendimientos crecientes a escala, la producción varía proporcionalmente más que los factores.
- Si $e = t < 1$ entonces presenta rendimientos decrecientes a escala, la producción varía proporcionalmente menos que los factores.

Sin embargo, esta elasticidad de escala no tiene por qué ser constante para todos los niveles de producción, por lo que conviene distinguir entre una noción global y una noción local

[30] Diferencia entre productividad y rendimiento:

- Productividad marginal (= rendimientos marginales): cuánto aumenta el producto cuando aumenta la cantidad de un factor en una determinada proporción. Es un concepto propio del corto plazo.
- Rendimiento a escala: cuánto aumenta el producto cuando aumenta la cantidad de todos los factores en la misma proporción. Es un concepto propio del largo plazo.

[31] Posibles causas de los rendimientos de escala:

- Rendimientos crecientes a escala: i) especialización y división del trabajo, ii) indivisibilidades técnicas y gerenciales (por ejemplo, un factor fijo) (por ejemplo, si para dos plantas se necesitan dos ingenieros de mantenimiento, al unirlas y crear una planta grande podría bastar con uno) y iii) cuestiones físicas: por ejemplo, al duplicar el diámetro de un oleoducto, se duplica el material empleado, pero el volumen del tubo se multiplica por 4 (2 vs. 2^2).
- Rendimientos decrecientes a escala: i) actividad gerencial (al aumentar el tamaño de la empresa, la toma de decisiones se vuelve compleja y lenta, por lo que se pierde eficiencia), ii) solapamientos y iii) recursos naturales no renovables: la duplicación de una flota pesquera no hará que se duplique la cantidad de capturas en una misma área (Aunque en realidad este no sería un caso puro de rendimientos decrecientes porque el factor peces es fijo. Del mismo modo, en el caso de las pérdidas gerenciales, no se está duplicando el ff.pp empresario.).

[32] Sin embargo, los rendimientos a escala pueden no ser constantes para todos los niveles de producción. Además, una función de producción puede tener diferentes elasticidades de escala por ejemplo: $X = \ln L + \ln K$

[33] En caso de que los rendimientos sean constantes, la función será homogénea de grado 1, y satisfará una serie de propiedades:

- Los PMg serán independientes de la escala.
- Al igual que el resto de funciones homogéneas, la $RMST$ dependerá únicamente de las proporciones de ffpp, independientemente de la escala de producción.
- Satisfacen el Teorema de Euler: el producto es igual a la suma de los PMg de los inputs multiplicados por su nivel de uso.
$$\frac{dX}{dL} \cdot L + \frac{dX}{dK} \cdot K = X$$
, en consecuencia, en un contexto competitivo los ingresos son repartidos íntegramente entre los ffpp, haciendo que el beneficio sea nulo.

Así

- Si se divide la anterior expresión entre L se obtiene: $PMe_L = PMg_L + PMg_K\left(\frac{K}{L}\right)$.
- Si se divide entre K se obtiene:$PMe_K = PMg_K + PMg_L\left(\frac{L}{K}\right)$.

De aquí se demuestra que, en el óptimo técnico (frontera entre las fases I y II), cuando $PMe_L = PMg_L$, $PMg_K = 0$, y viceversa.

 © 2026 Manuel Cabado García

de los rendimientos a escala (Jehle y Reny, 2011). Los rendimientos globales a escala son los descritos hasta ahora: se cumplen para todo punto y toda proporción de escalado. Las funciones homogéneas son el caso más sencillo, pues su elasticidad de escala es constante e igual al grado de homogeneidad. Los rendimientos locales a escala, en cambio, evalúan la elasticidad de escala en un punto concreto

$$\mu(z) = \lim_{t \to 1} \frac{d \ln f(tz)}{d \ln t} \tag{5.7}$$

de modo que una misma tecnología puede presentar rendimientos crecientes a niveles bajos de producción y decrecientes a niveles elevados, sin ser homogénea en ningún grado.

Las funciones homogéneas tienen además la propiedad de que su $RMST$ solo depende de la proporción de factores, lo que permite introducir un concepto más general, el de función homotética, que es aquella función cuyas isocuantas son ampliaciones radiales unas de otras, de modo que a lo largo de cualquier rayo vector todas tienen la misma pendiente. Por ello, su $RMST$ solo depende de la proporción de factores y, para precios de los factores dados, la senda de expansión es una recta que pasa por el origen[34].

Análisis de la función de producción en el muy largo plazo

Por otro lado, en el muy largo plazo la tecnología no es constante sino que puede producirse progreso técnico. El progreso técnico es una mejora de la tecnología que permite obtener más output con los mismos inputs, o el mismo output con menos inputs. Normalmente eleva la productividad marginal de los factores, aunque no necesariamente de todos en la misma medida. Esto se conoce como innovación de proceso[35].

De esta manera, el progreso técnico puede representarse de dos formas equivalentes: como un desplazamiento hacia arriba de la función de producción —lo que desplaza también hacia arriba, como consecuencia, su curva de productividad marginal—, o como un desplazamiento hacia el origen de las curvas isocuantas, como se observa en la Figura 5.6, ya que con la misma cantidad de factores se obtiene mayor nivel de producción, o para

[34]La intuición es que el mapa de isocuantas es el mismo dibujo a distintas escalas. La homotecia es un concepto más general que la homogeneidad, de forma que todas las funciones homogéneas son homotéticas, pero no todas las funciones homotéticas son homogéneas. O dicho de otra manera: las funciones homogéneas son un caso especial de funciones homotéticas.

Funciones homotéticas: todas las funciones homogéneas serán funciones homotéticas. Una función es homotética si puede ser escrita como una transformación monótona de una función homogénea lineal de los inputs: $g(z) = F(f(z))$. Por ejemplo: $y = z_1^{\alpha} \cdot z_2^{1-\alpha} \to \ln(y) = \alpha \cdot \ln z_1 + (1-\alpha) \ln z_2$.

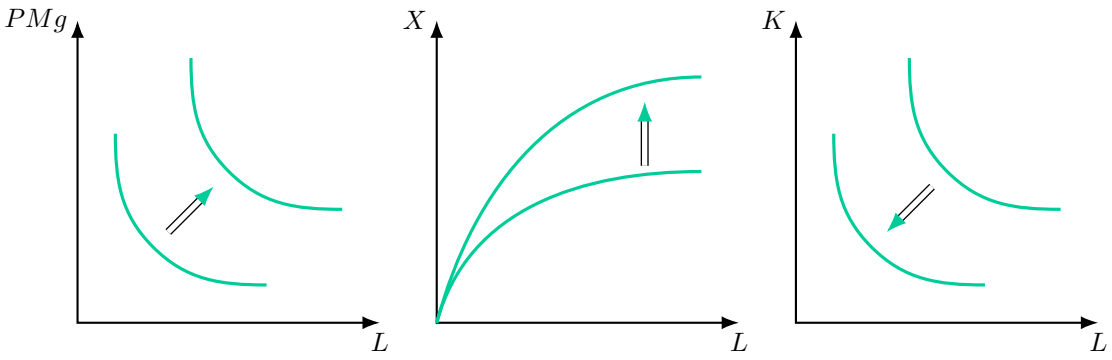
Las funciones homotéticas son útiles porque, a diferencia de las funciones homogéneas, pueden tener retornos a escala variables, pero mantienen una de las propiedades más útiles de las funciones homogéneas: su $RMST$ solo depende de las proporciones de factores, siendo independiente de la escala de producción.

Por tanto, las isoclinas serán un radio-vector. Estas se denominarán líneas de producto, al ser senderos donde teóricamente se puede expandir la producción, siempre y cuando se mantengan constantes los precios relativos de los ffpp.

[35]También se puede dar innovación de producto como que mejore la calidad o se produzcan bienes complementarios nuevos.

obtener el mismo nivel de producto se necesita menos cantidad de factores.

Figura 5.6: Efectos del progreso técnico



Definiciones de progreso técnico A su vez, la preocupación por los efectos del progreso técnico sobre la distribución de la renta dio lugar a distintas definiciones de progreso técnico neutral[36]::

- Neutral a lo Harrod (Harrod, 1939),$Y_t = F\left(K_t, A_t \cdot L_t\right)$, es el progreso técnico que no altera el reparto de la renta entre capital y trabajo cuando se mantiene constante la relación capital-producto, y equivale a que el progreso técnico actúe como un aumento de la eficiencia del trabajo (potenciador de trabajo)[37], mientras que la neutralidad a lo Solow (Solow, 1957),$Y_t = F\left(A_t \cdot K_t, L_t\right)$, es la que no altera ese reparto cuando se mantiene constante la relación trabajo-producto, y equivale a un aumento de la eficiencia del capital (potenciador de capital)[38].[39]
- Hicks (Hicks, 1932) analizó el impacto del mismo manteniendo la ratio $\frac{K}{L}$ constante, utilizando un concepto más amplio de progreso técnico, de forma que se puede manifestar de diferentes formas como se observa en la Figura 5.7:
  - Progreso técnico neutral, la productividad marginal de ambos factores aumenta en la misma proporción, de modo que la $RMST$ a lo largo del rayo vector

[36]El progreso técnico (PT) tal y como se ha introducido se denomina "no incorporado" en el sentido de que cuando aparece una mejora tecnológica, esta afecta a todas las máquinas aumentando todas su producción. Si se dan invenciones que no afectan a todas las máquinas sino a las más nuevas y a las obsoletas no, entonces se denomina "PT incorporado". Para modelizarlo se suele usar una función de producción sin PT y el potenciador multiplica a la inversión en la ecuación de acumulación de capital: $\dot{K} = A_t \cdot (Y_t - C_t)$

[37]Innovación neutral si se mantiene la relación de participaciones en la producción de K y L constante para una relación de K - producto dada (pero cada vez se requiere menos trabajo). Igual en Solow pero con L.

[38]Cabe mencionar que en una función de producción tipo Cobb-Douglas con rendimientos constantes a escala, los dos tipos de PT y el neutral a la Hicks, son equivalentes en un modelo unisectorial.

[39]La intuición es que, con progreso neutral a lo Harrod, es como si cada trabajador valiera por más trabajadores: la tecnología multiplica la eficiencia del trabajo, y lo mismo ocurre con el capital en el caso de Solow.

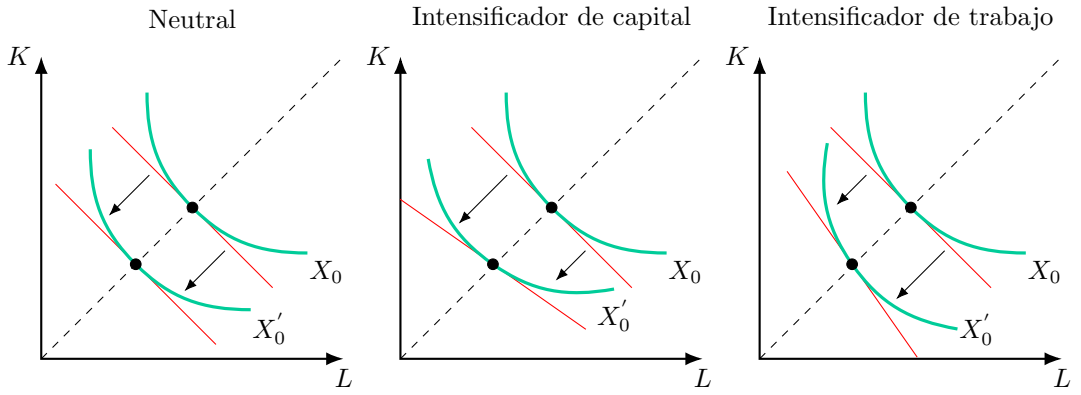
 © 2026 Manuel Cabado García

permanece constante. La isocuanta se desplaza hacia abajo de forma paralela.

- Progreso técnico sesgado, la productividad marginal de un factor aumenta proporcionalmente más que la del otro factor, por lo que la RMST a lo largo del rayo vector cambia. La isocuanta se desplaza hacia abajo de forma sesgada:
  - Progreso técnico intensificador del capital o ahorrador de trabajo, la productividad marginal del factor capital aumenta proporcionalmente más que la del trabajo, de modo que la $RMST$ a lo largo del rayo vector disminuye: $\frac{\uparrow PMg_L}{\uparrow\uparrow PMg_K}$.
  - Progreso técnico intensificador del trabajo o ahorrador del capital, la productividad marginal del factor trabajo aumenta proporcionalmente más que la del capital, de modo que la $RMST$ a lo largo del rayo vector aumenta: $\frac{\uparrow\uparrow PMg_L}{\uparrow PMg_K}$.

Conviene subrayar que Hicks, Harrod y Solow no son una jerarquía en la que uno englobe a los otros, sino tres criterios alternativos de neutralidad, cada uno definido manteniendo constante una relación distinta: Hicks mantiene constante la relación K/L, mientras que Harrod mantiene la relación K/Y y Solow, la relación L/Y.

Figura 5.7: Definiciones de progreso técnico



Habiendo analizado la función de producción en el corto, el largo y el muy largo plazo, se pasa al tercer y último bloque de esta exposición, dedicado a la producción multiproducto.

5.4. Producción multiproducto ⏳ 2'

Por último, considerando más de un producto, que los productores se sitúen sobre su curva isocuanta no es condición suficiente para la eficiencia técnica. En efecto, la eficiencia exigirá una asignación de factores productivos tal que a través de una reasignación de los

mismos no sea posible aumentar la cantidad producida de un bien sin disminuir la del otro. Esto se consigue, como se verá para el caso de dos productos, cuando las isocuantas de cada producto son tangentes (es decir, cuando las RMST son iguales).

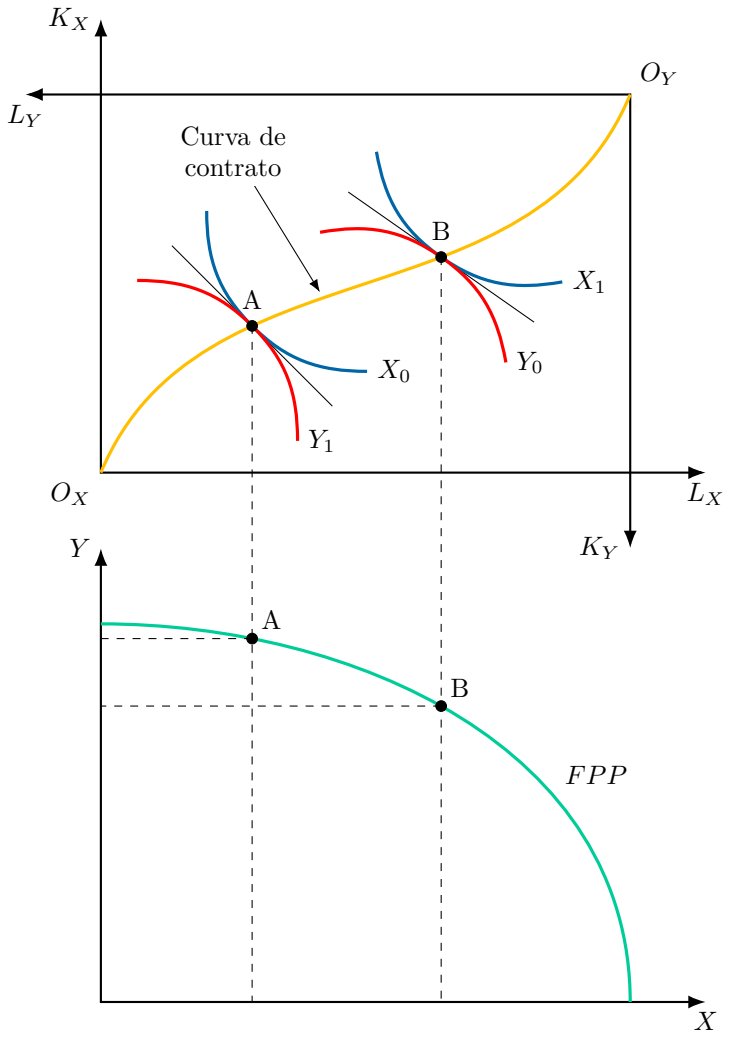
Se supone una empresa que produce dos bienes X e Y utilizando capital y trabajo, con las siguientes funciones de producción

$$X = F_X(K, L) \tag{5.8}$$
$$Y = F_Y(K, L) \tag{5.9}$$

y con las siguientes restricciones de factores disponibles

$$K = K_X + K_Y \tag{5.10}$$
$$L = L_X + L_Y \tag{5.11}$$

Figura 5.8: Caja de Edgeworth y Frontera de Posibilidades de Producción



 © 2026 Manuel Cabado García

Caja de Edgeworth El problema se puede representar gráficamente en una caja de Edgeworth como la de la Figura 5.8:

- Cuyas dimensiones dependen de las cantidades de factores productivos disponibles para la producción de ambos bienes.
- La producción óptima vendrá dada por los puntos en los que las isocuantas son tangentes, en los que se cumple que las relaciones marginales de sustitución técnica de los bienes se igualen, $RMST_X^{K,L} = RMST_Y^{K,L}$.
- El conjunto de puntos en los que la producción es óptima constituye la curva de contrato. En la curva de contrato se cumple la eficiencia en sentido de Pareto en la producción: "No es posible aumentar la producción de un bien, sin disminuir la de otro."[40]

Frontera de Posibilidades de Producción Además, en un plano adyacente de bien X y bien Y:

- A partir de la curva de contrato, uniendo las distintas combinaciones óptimas de producción de los dos bienes se obtiene la frontera de posibilidades de producción[41], FPP, cuya pendiente viene determinada por la relación marginal de transformación[42], RMT_{xy}, es decir, la cantidad de un bien a la que debe renunciarse para producir una unidad adicional del otro, manteniendo constantes la dotación de factores productivos y la tecnología.
- Todos los puntos de la FPP son eficientes en la producción[43]. Por lo tanto, existen infinitos óptimos en la producción. Los puntos interiores a la FPP son ineficientes,

[40] Si se introdujese el problema de maximización del beneficio, en competencia perfecta los productores, para maximizar su beneficio, igualan la Relación Marginal de Sustitución Técnica (RMST) a los precios relativos de los factores, y como en competencia perfecta los precios de los factores son los mismos para todos los productores, la RMST será también la misma para todos los productores, que es la condición del óptimo en la producción (pues, dado que las valoraciones de los factores entre los productores son iguales, no será posible una reasignación de los factores que aumente la producción de un bien sin disminuir la del otro).

[41] La FPP es, dicho de otra manera, el equilibrio de las empresas en el espacio de bienes. Esta traslación es necesaria porque, en el óptimo, las cantidades producidas deben ser iguales a las cantidades consumidas, y los consumidores toman sus decisiones de compra en el espacio de bienes. De ahí que la FPP resulte necesaria para coordinar los planes de producción con los de consumo.

[42] La RMT también puede definirse, en ausencia de fallos de mercado, como el cociente de los costes marginales

$$RMT = \frac{CMg_X}{CMg_Y}$$

La idea es que, a partir del gráfico de la FPP, la distancia en el eje vertical entre los puntos 1 y 2 representa el coste incremental de producir el bien Y, mientras que la distancia horizontal entre esos mismos dos puntos representa el coste incremental de producir el bien X. Por lo tanto, la RMT, en valor absoluto, es igual a CMg_X/CMg_Y.

[43] Si ambos bienes se producen con rendimientos constantes a escala y con la misma intensidad factorial, la FPP es una línea recta, porque trasladar factores de un bien al otro tiene siempre el mismo coste de oportunidad.

y los puntos situados fuera son inalcanzables, a menos que se descubran recursos adicionales y/o una nueva tecnología.

5.5. Conclusiones ⏳ 30"

Como conclusión, la teoría de la producción ofrece una aproximación al comportamiento de los productores, que toman decisiones relativas al conjunto de inputs contratados y outputs producibles sometiéndose a restricciones técnicas, que son las que se han estudiado en este tema.

No obstante, la teoría de la producción presenta ciertas limitaciones[44]:

- Abstracción de otros aspectos importantes (como los costes o los aspectos organizativos), que hace que pierda poder explicativo, pues no se terminan de caracterizar las decisiones de oferta de bienes y de demanda de factores de la empresa.
- Forma funcional determinante. La forma funcional debe definirse cuidadosamente, ya que cambios en la forma de la función pueden afectar notablemente a los resultados (por ejemplo, función de coeficientes fijos vs. función Cobb-Douglas).

A pesar de ello, la teoría de la producción da lugar a conceptos de una importancia económica fundamental, como son:

- La productividad marginal, muy relacionada con cuestiones como la teoría neoclásica de la distribución.
- Los rendimientos a escala, que pueden usarse como un argumento para la justificación económica de los monopolios.
- El progreso técnico, que es un factor determinante del crecimiento económico y en el que se centran las teorías modernas de crecimiento endógeno.

[44] La conjunción de la Teoría de la Producción con otras ramas de la economía será esencial para realizar un análisis certero de la realidad:

- Por ejemplo, la elasticidad de sustitución de factores, aunada con el poder de negociación que tenga la empresa en el mercado de trabajo, explicará en gran medida variables tan importantes como la determinación del salario.
- A su vez, el estudio de las restricciones de variación de factores productivos en distintos plazos del tiempo explicará las distintas elasticidades tanto de los salarios como de la oferta de producción en función del plazo de tiempo que se estudie.
- Los distintos requisitos de la función de producción para producir cierta cantidad pueden constituir a su vez una barrera de entrada que explique la existencia de oligopolios en ciertos mercados.

Por lo tanto, se observa el papel esencial que tiene la teoría de la producción en casi todas las áreas de la economía.

 © 2026 Manuel Cabado García

Apéndice

5.A.1. Tipos de funciones de producción según sustituibilidad

A continuación se presentan diferentes tipos de funciones de producción según su sustituibilidad:

1. Función de producción Cobb-Douglas (propuesta por Wicksell (Wicksell, 1901) hacia 1900-1901[45] y desarrollada por C. Cobb y P. Douglas en 1928), donde:

   a) RMST decreciente. Los factores son imperfectamente sustitutivos.

   $$X = L^{\alpha} \cdot K^{\beta} \longrightarrow RMST_{L,K} = \frac{PMg_L}{PMg_K} = \frac{\alpha \cdot L^{\alpha-1} \cdot K^{\beta}}{\beta \cdot L^{\alpha} \cdot K^{\beta-1}} = \frac{\alpha \cdot K}{\beta \cdot L} \tag{5.12}$$

   b) Elasticidad de sustitución: $\sigma = 1$

   $$X = L^{\alpha} \cdot K^{\beta} \longrightarrow \sigma = \frac{\partial (K/L)}{\partial RMST_{L,K}} \cdot \frac{RMST_{L,K}}{K/L} = \frac{\beta}{\alpha} \cdot \frac{\alpha}{\beta} = 1 \tag{5.13}$$

2. Función de producción de coeficientes fijos, donde:

   a) RMST indeterminada. Los factores son complementarios perfectos.

   $$X = \text{mín}\left\{\frac{L}{a}, \frac{K}{b}\right\} \tag{5.14}$$

   Donde a es la cantidad mínima que se necesita del factor L para producir una unidad de X, y b es la cantidad mínima que se necesita del factor K para producir también una unidad de X.

   b) Elasticidad de sustitución: $\sigma = 0$.

   $$X = \text{mín}\left\{\frac{L}{a}, \frac{K}{b}\right\} \longrightarrow \sigma = 0 \tag{5.15}$$

3. Función de producción lineal, donde

   a) RMST constante. Los factores son perfectamente sustitutivos.

   $$X = a \cdot L + b \cdot K \longrightarrow RMST_{L,K} = \frac{PMg_L}{PMg_K} = \frac{a}{b} \tag{5.16}$$

[45] La atribución del precedente a Wicksteed (Wicksteed, 1894), habitual en muchos manuales, es un error historiográfico: Sandelin (Sandelin, 1976) demostró que esa referencia es infundada y que el verdadero precedente es Wicksell, que ya empleaba implícitamente la función linealmente homogénea de capital y trabajo desde 1895, formalizándola en sus *Lectures on Political Economy* (1901).

b) Elasticidad de sustitución: $\sigma = \infty$.

$$X = a \cdot L + b \cdot K \longrightarrow \sigma = \frac{\partial\,(K/L)}{\partial RMST_{L,K}} \cdot \frac{RMST_{L,K}}{K/L} \xrightarrow{RMST \text{ constante}} \frac{\frac{RMST}{K/L}}{0} = \infty \tag{5.17}$$

4. Función de producción CES (Elasticidad de sustitución constante):

$$X = (L^\rho + K^\rho)^{\beta/\rho} \tag{5.18}$$

donde β es el grado de homogeneidad y ρ el factor de sustitución. Así:

$$RMST = \frac{K^{1-\rho}}{L^{1-\rho}} = \left(\frac{K}{L}\right)^{1-\rho} \tag{5.19}$$

y

$$\frac{\partial K/L}{\partial RMST} = \frac{1}{1-\rho} \cdot \left(\frac{K}{L}\right)^{\rho}$$

por lo que

$$\sigma = \frac{1}{1-\rho}$$

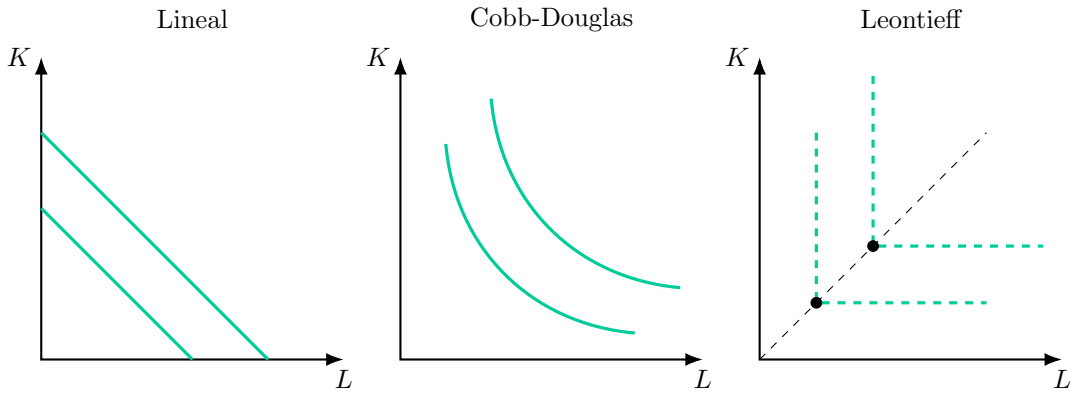
Así:

$\rho \to 0 \Rightarrow \sigma = 1$ Cobb-Douglas

$\rho \to 1 \Rightarrow \sigma = \infty$ Factores perfectamente sustitutivos: lineales

$\rho \to -\infty \Rightarrow \sigma = 0$ Factores perfectamente complementarios: proporciones fijas

Figura 5.9: Tipos de funciones de producción



Mención especial función de proporciones fijas o Leontief Esta función modela situaciones en las que los factores de producción son complementarios perfectos, lo que

 © 2026 Manuel Cabado García

significa que deben usarse en una proporción fija.

Un ejemplo clásico es el de un coche que requiere 4 ruedas y 1 volante. Con 40 ruedas y 10 volantes se pueden producir 10 coches. Si se obtienen 10 ruedas más (50 en total) pero se mantienen los 10 volantes, la producción sigue siendo de 10 coches, porque los volantes son el factor limitante.

La representación gráfica de esto es una isocuanta (la curva que muestra las combinaciones de factores para un mismo nivel de producción) en forma de L. El vértice de la L representa la única combinación eficiente de los factores para ese nivel de producción. Cualquier punto que no esté en el vértice implica un uso ineficiente de los recursos:

- En las ramas horizontales o verticales de la L, hay un exceso de un factor que no contribuye a un aumento de la producción. Aumentar el factor sobrante no aumenta la producción, por lo que se incurriría en un coste innecesario.
- Por lo tanto, el coste mínimo para producir una cantidad dada siempre se alcanza en el vértice, ya que es el punto donde no hay desperdicio de factores.

En resumen, la lógica económica detrás de la función de Leontief es que, dado que los factores no son sustituibles, la única combinación óptima de insumos se encuentra en el vértice de cada isocuanta.

5.A.2. Relación entre la teoría de producción y la teoría de costes

Cabe destacar que el problema de la producción es un problema incompleto si se considera la actividad global de la empresa. Como se ha mencionado en la introducción, la empresa también se encuentra sujeta a una restricción de costes y a una restricción institucional (análisis de la estructura de mercado).

A continuación se introduce el problema de los costes, lo cual permite determinar en qué punto exacto de la isocuanta se sitúa la empresa, dada su estructura de costes. Se adopta el enfoque primal, que parte de que la empresa maximiza su producción sujeta a una restricción de costes.

$$\text{máx} X = F(z) \tag{5.20}$$

$$s.a\ \bar{C} = \sum_{i=1}^{n} w_i \cdot z_i \tag{5.21}$$

La CPO lleva a que la empresa se situará en aquel punto donde se cumpla que la RMST es igual a la pendiente de la recta isocoste. La CSO se cumplirá si la FP es estrictamente cuasicóncava.

$$\frac{PMg_{z_i}}{PMg_{z_j}} = \frac{w_i}{w_j} \tag{5.22}$$

A partir de la resolución del problema, se puede analizar el resultado gráficamente. La empresa se sitúa en aquel punto donde la isocuanta más alejada del origen sea tangente a la recta isocoste. Además, se introduce el concepto de senda de expansión, que es el lugar geométrico de las combinaciones técnica y económicamente eficientes de factores productivos para cada nivel de producción. El punto exacto de la senda de expansión en el que se sitúe la empresa lo mostrará la teoría de los mercados.

5.A.3. La caracterización de las decisiones de la empresa

Nótese que la maximización de beneficios permite superar las tres restricciones a las que se enfrenta la empresa, y caracterizar así sus decisiones de oferta de producto y de demanda de factores productivos:

1. Restricción técnica: la teoría de la producción determina las combinaciones técnicamente eficientes de factores productivos, esto es, aquellas que, para una producción dada, se sitúan en la curva isocuanta.
2. Restricción económica: la teoría de los costes complementa la teoría de la producción, de forma que, de entre todas las combinaciones técnicamente eficientes, determina aquellas que también lo son económicamente, dados una producción y los precios de los factores (esto es, aquellas en las que la isocuanta es tangente a la isocoste).
3. Restricción de mercado: la teoría de los mercados complementa a la teoría de los costes y termina de caracterizar la decisión de la empresa: de entre todas las combinaciones técnica y económicamente eficientes para los distintos niveles de producción, la teoría de los mercados determina el nivel de producción que maximiza el beneficio, dados unos precios del bien y de los factores productivos.

5.A.4. Axiomas de las características de escala y rendimientos de escala

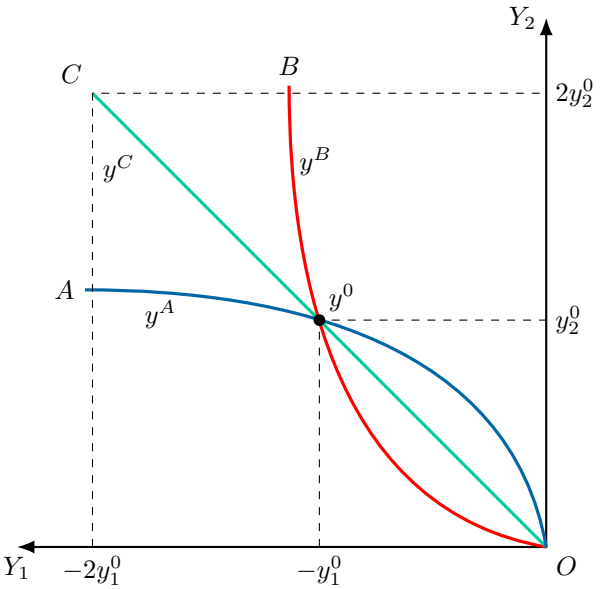
Relación con los rendimientos de escala tal y como se observa en la Figura 5.10:

- Rendimientos a escala decrecientes: Para todo $y \in Y,\ \alpha \cdot y \in Y,\ \alpha \in [0,1]$. Vector de producción que se pueda escalar a una dimensión menor. OA: Se cumple la divisibilidad y la estricta convexidad, pero es incompatible con la aditividad.
- Rendimientos a escala crecientes: Para todo $y \in Y,\ \alpha \cdot y \in Y,\ \alpha > 1$. El vector se puede escalar a una dimensión mayor. OB: Se cumple aditividad, sin postularse divisibilidad. Es incompatible con convexidad y estricta convexidad.
- Rendimientos a escala constantes: Consecuencia de la conjunción de las dos propiedades. Para todo $y \in Y,\ \alpha \cdot y \in Y,\ \forall \alpha > 0$. En este caso la frontera de transforma-

 © 2026 Manuel Cabado García

ción es la recta que pasa por el origen. OC: Se cumplen aditividad y divisibilidad. Convexo, aunque no estrictamente.

Figura 5.10: Axiomas y características de escala



Bibliografía

Cassels, J. M. (1936). On the law of variable proportions. En *Explorations in Economics: Notes and Essays Contributed in Honor of F. W. Taussig*, pp. 223–236. New York: McGraw-Hill.

Harrod, R. F. (1939). An essay in dynamic theory. *Economic Journal 49*(193), 14–33.

Hicks, J. R. (1932). *The Theory of Wages*. London: Macmillan.

Jehle, G. A. y P. J. Reny (2011). *Advanced Microeconomic Theory* (3rd ed.). Harlow: Pearson.

Mas-Colell, A., M. D. Whinston y J. R. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Sandelin, B. (1976). On the origin of the cobb-douglas production function. *Economy and History 19*(2), 117–123.

Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics 39*(3), 312–320.

Turgot, A.-R.-J. (1767). Observations sur le mémoire de m. de saint-péravy.

Wicksell, K. (1901). *Lectures on Political Economy*. London: Macmillan.

Wicksteed, P. H. (1894). *An Essay on the Co-ordination of the Laws of Distribution*. London: Macmillan.

 © 2026 Manuel Cabado García

Tema 6

Teoría de los costes. Análisis de la dualidad en el ámbito de la empresa. Análisis temporal. Otros desarrollos.

6.1. Introducción ⏳ 1'

El Estrecho de Ormuz bloqueado en 2026, el gas disparado por la guerra de Ucrania, los aranceles de EE.UU. encareciendo cadenas de suministro enteras: la última década no deja de recordar que conocer bien la estructura de costes de una empresa no es un mero ejercicio contable, sino una condición para seguir operando.

Por ello, este tema se centra en la teoría de costes[1], que busca identificar, de entre todas las combinaciones técnicamente eficientes, aquellas que lo son económicamente, es decir, que minimizan los costes de producción.

Por consiguiente, el objetivo de la presente exposición será el estudio de los costes económicos[2] a través del siguiente esquema:

1. Teoría neoclásica de los costes, donde se construirá la función de costes y se realizará un análisis temporal.
2. Críticas a la teoría de costes tradicional, desde el ángulo empírico y organizativo.

[1] La empresa se enfrenta a tres tipos de restricciones: las técnicas, que estudia la teoría de la producción, las económicas, que estudia la teoría de costes, y las de mercado, que estudia la teoría de la organización industrial.

[2] Los costes, definidos como el gasto realizado para la adquisición o producción de un bien o servicio, se distinguen entre los

- Costes contables, aquellos obtenidos de la aplicación de convenciones que permiten la obtención de los estados de resultados de la empresa válidos para cuestiones fiscales y de contabilidad de la empresa.
- Costes económicos, ligados al uso de recursos económicos para la producción, que incluyen el coste de oportunidad, es decir, el valor de la mejor alternativa a la que la empresa renuncia al destinar sus recursos a un uso concreto.

3. Otros desarrollos, que amplían el análisis a supuestos más realistas de la empresa.

6.2. Teoría neoclásica de los costes ⏳ 12'

La teoría neoclásica parte de los siguientes supuestos:

1. Racionalidad,
2. Certidumbre,
3. Ausencia de rigideces, se opera en situación de competencia perfecta.
4. Los precios son indicadores de la escasez relativa

Construcción de la función de costes

A partir de la Teoría Neoclásica de la Producción, se obtiene una Función de Producción representativa de una determinada tecnología.

Isocuantas

- De la Función de Producción se pueden obtener las isocuantas, que son las curvas de nivel de la Función de Producción y muestran las combinaciones técnicamente eficientes de factores productivos que producen la misma cantidad de output.
- Gráficamente, en un plano de capital, K, y trabajo, L, como el de la Figura 6.1 se representan las curvas isocuantas q_0 y q_1.

Sin embargo, el concepto de eficiencia técnica no es suficiente para caracterizar la decisión de la empresa, ya que este análisis no muestra en qué punto de la isocuanta se situará la empresa (algo que resuelve la teoría de costes), ni en qué isocuanta se situará (algo que resuelve la teoría de los mercados)[3].

Para ello, se necesita información económica (es decir, los precios de los factores productivos): dentro de un conjunto de combinaciones técnicamente eficientes (es decir, puntos de la isocuanta), resultarán más atractivas para el productor aquellas combinaciones más intensivas en el factor comparativamente más barato.

[3]Este tema se centra en la teoría de los costes (es decir, en las restricciones económicas a las que se enfrenta la empresa), muy importante para:

- Completar las decisiones de producción de la empresa, pues el concepto de eficiencia técnica analizado en la teoría de la producción no es suficiente para caracterizar la decisión de la empresa.
- Determinar la estructura de mercado, pues la estructura de costes es uno de los factores determinantes de la estructura de mercado.

 © 2026 Manuel Cabado García

Problema de la empresa Por tanto, el problema al que se va a enfrentar la empresa es el de elegir aquella combinación de factores productivos que le suponga un menor coste para producir un output dado, o, inversamente, elegir aquella combinación que, dado un coste, proporcione el mayor nivel posible de producto.

Análisis de dualidad Considerando el doble enfoque que presenta el análisis de dualidad en el problema del productor, en ambos casos el problema de la empresa es el de maximizar su beneficio: en el primero, se trata de maximizar el beneficio con un coste dado (por lo que hay que maximizar la producción), mientras que en el segundo se trata de maximizar el beneficio con una producción dada –por ejemplo, un puente, cuyo tamaño viene fijado de antemano por el proyecto y no es una decisión de la empresa– (por lo que hay que minimizar el coste).
Así, el problema del productor se puede plantear como:

- La maximización de la producción dada una restricción de costes. Es decir, tratará de encontrar aquellas combinaciones de factores productivos que le otorguen la máxima producción posible dada una restricción económica.
- La minimización de costes para una determinada producción. Es decir, tratará de encontrar aquellas combinaciones de factores productivos que le produzcan una determinada producción al mínimo coste.

Dicho paralelismo se muestra formalmente a través del análisis de la dualidad que refleja cómo, bajo ciertas condiciones, la combinación de factores productivos será la misma (de la misma manera que en la teoría de demanda neoclásica del consumidor).

Supuestos Para plantear los problemas primal y dual del productor, se parte de los siguientes supuestos:

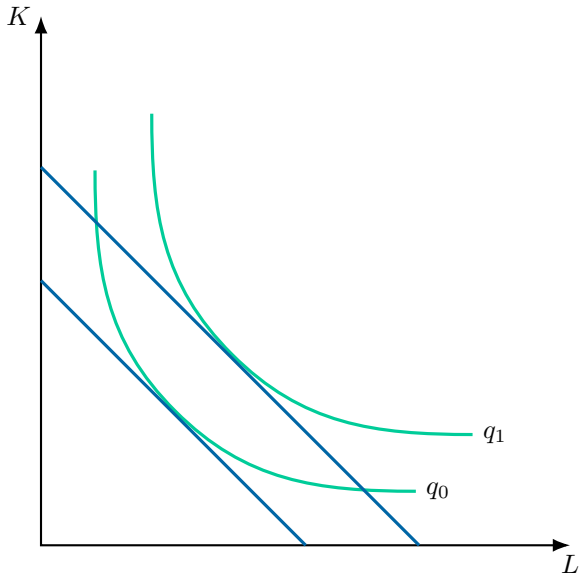
- Contexto de competencia perfecta (es precio-aceptante respecto a los factores productivos).
- Tecnología caracterizada por varios inputs que dan lugar a un solo output, representada a partir de una Función neoclásica $Q = f(X)$.
  Esta función de producción neoclásica es continua y dos veces diferenciable, cardinal, homogénea y estrictamente no negativa, monótona creciente, con PMg de los factores productivos positiva pero decreciente y estrictamente cuasi-cóncava.

Problema primal

El problema primal consiste en elegir la combinación de factores productivos que maximiza la producción

$$\text{máx} \quad Q = f(X) \tag{6.1}$$

Figura 6.1: Curvas isocuantas y rectas isocostes



sujeta a un coste dado

$$\bar{C} = \sum_{i=1}^{n} w_i \cdot x_i \tag{6.2}$$

donde w_i es el precio de los factores y x_i la cantidad de factores empleados.

La condición de eficiencia económica indica que, en el equilibrio, la tasa a la que la empresa puede intercambiar internamente los factores productivos, es decir, la Relación Marginal de Sustitución Técnica, $RMST$, es igual a la tasa a la que puede hacerlo en el mercado, es decir, la ratio de precios de los factores.

Así, el productor demanda factores productivos hasta el punto donde el cociente de las PMg de cada par de factores se iguale a su precio relativo:

$$RMST_{i,j} = \frac{PMg_{x_i}}{PMg_{x_j}} = \frac{w_i}{w_j} \tag{6.3}$$

Gráficamente, como se observa en la Figura 6.1 se representan por orden:

1. Primero se representa la restricción. La curva isocoste, que representa las combinaciones de factores que, dados los precios de los mismos, tienen el mismo coste total. La pendiente de la curva isocoste es igual al cociente de los precios de los factores cambiado de signo ($-w_i/w_j$).
2. En segundo lugar, se representan las curvas isocuantas, antes definidas, cuya pendiente, en valor absoluto, es la $RMST$ y muestra la relación a la que puede intercambiarse un factor por otro para mantener el nivel de producción constante.
3. En tercer lugar, el equilibrio se encuentra en el punto de tangencia de la recta isocoste con la isocuanta más alejada del origen (porque se está maximizando). En

 © 2026 Manuel Cabado García

este punto, $RMST = w_i/w_j$. Por lo tanto, se cumple la condición de eficiencia económica.

A su vez, de la condición de primer orden se obtienen las funciones de demanda ordinarias (marshallianas) de factores productivos, es decir, la demanda de factores productivos en función del precio de los factores y del coste:

$$x_i\left(W, C\right) \tag{6.4}$$

Características de las demandas marshallianas:

- Continuas
- Su respuesta a los precios de los factores no tiene signo determinado a priori, pues combina un efecto sustitución y un efecto renta.
- Crecientes en C si el factor es normal y decrecientes si es inferior.[4]
- Homogéneas de grado 0 en w_i, w_j y C.

Finalmente, sustituyendo las funciones de demanda ordinarias en la función objetivo se obtiene la función indirecta de producción, que refleja la producción máxima alcanzable para cada nivel de precios y coste: $Y(w_i, w_j, C)$

Problema dual

Por otro lado, el problema dual consiste en elegir la combinación de factores productivos que minimiza el coste

$$\text{mín} \quad C = \sum_{i=1}^{n} w_i \cdot x_i \tag{6.5}$$

sujeto a una producción dada

$$\bar{Q} = f\left(X\right) \tag{6.6}$$

La solución es idéntica a la del problema primal, obteniéndose de la condición de primer orden las funciones de demanda condicionadas –o hicksianas– de factores productivos, es decir, la demanda de factores productivos en función del precio de factores y de la producción.

$$h_i\left(W, Q\right) \tag{6.7}$$

Gráficamente, como se observa en la Figura 6.1, para el caso de dos inputs, se invierten los papeles con respecto al problema primal: primero se dibuja la isocuanta que representa el nivel de producción dado (restricción), y después se halla la recta isocoste más cercana al origen (porque se está minimizando) que es tangente a dicha isocuanta.

[4] La intuición es que, si una empresa con más presupuesto pasa de una técnica artesanal a otra automatizada, puede acabar usando menos trabajo poco cualificado aunque su coste total haya aumentado. Ese factor se comporta como inferior.

Teorema de la dualidad De esta forma, el Teorema de la Dualidad indica que los problemas primal y dual tendrán la misma solución siempre que:

- El coste que actúa como restricción en el primal sea el mínimo a alcanzar en el dual, y
- La producción que actúa como restricción en el dual sea la máxima a alcanzar en el primal.

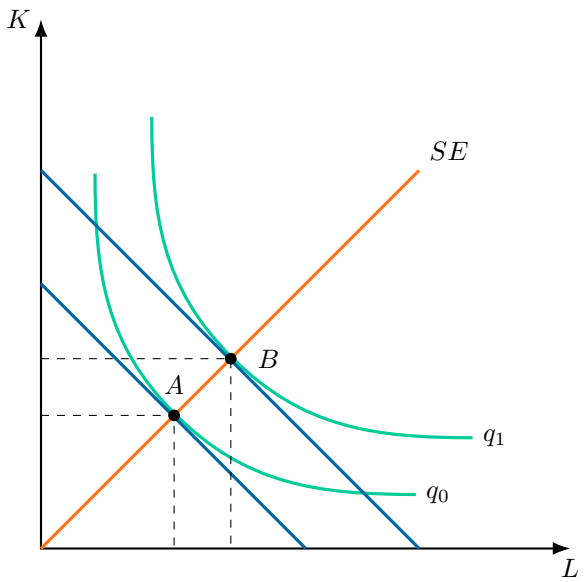
Por consiguiente, el teorema de la dualidad garantiza la existencia (la existencia de la solución primal si existe la solución dual) y la identidad (de ambas soluciones).

Senda de expansión

Recogiendo gráficamente todo lo anterior, como se observa en la Figura 6.2, los puntos de tangencia entre isocuantas e isocostes se recogen en la senda de expansión, SE, que es el lugar geométrico de las combinaciones técnica y económicamente eficientes de factores productivos para cada nivel de producción.

De este modo, expresa el mínimo coste para cada nivel de producción (dual) o, dicho de otra forma, el máximo nivel de producción para cada nivel de costes (primal)[5][6].

Figura 6.2: Senda de expansión



[5]No obstante, el punto exacto de la senda de expansión en el que se sitúe la empresa lo mostrará la teoría de los mercados.

[6]La senda de expansión puede ser una curva o una recta (en las funciones homotéticas –por ejemplo, Cobb-Douglas– es siempre una recta), pero siempre será una isoclina (pues une puntos con la misma pendiente, concretamente con pendiente $-w_i/w_j$).

 © 2026 Manuel Cabado García

Función de costes

Finalmente, de la resolución del problema dual se puede obtener la función de costes (sustituyendo las funciones de demanda condicionadas en la función objetivo), que indica el coste mínimo dados unos precios de los factores y un volumen de producción:

$$C\left(W,Q\right)$$

Posee las siguientes propiedades

1. Continua (para precios y niveles de producción positivos) y estrictamente positiva.
2. No decreciente en precios de los factores, estrictamente creciente en producción[7].

$$\frac{\partial C}{\partial w_i} \geq 0 \text{ con } \frac{\partial C}{\partial Q} > 0$$

3. Homogénea de grado 1 en los precios de los factores, pues si todos los precios aumentan en la misma proporción, para seguir produciendo la misma cantidad de producto el coste debe aumentar en la misma proporción: $t \cdot w_1 \cdot x_1 + t \cdot w_2 \cdot x_2 = t \cdot C$.
4. Cóncava en el precio de los factores, es decir, si el precio de un factor aumenta, el coste mínimo aumenta menos de lo que lo haría si la empresa mantuviera la combinación de factores inicial, puesto que el factor que se ha encarecido será parcialmente sustituido[8].

$$\frac{\partial^2 C\left(w,X\right)}{\partial w^2} \leq 0$$

5. Diferenciable.
6. Se aplica el Lema de Shephard, es decir, la derivada de la función de costes respecto a los precios de los factores será igual a la demanda condicionada de los factores

$$\frac{\partial C}{\partial w_i} = h_i\left(W,Q\right) \tag{6.8}$$

7. Resuelve el problema de la integrabilidad, es decir, es posible asegurar que la función de costes proviene de la optimización de una función de producción. Por consiguiente, la función de costes contiene esencialmente la misma información que la función de producción.

Adicionalmente, si la función es homotética, se cumplirá que la función de costes será multiplicativamente separable, por lo que la función de costes puede ser escrita como:

[7] No decreciente en "w_i, w_j" y estrictamente creciente en "Q" (ya que, si con el mismo coste se pudiese conseguir más producción, no se estarían maximizando beneficios).

[8] El incremento del coste será mayor cuanto menor sea la sustituibilidad de los factores (es decir, cuanto menor sea la elasticidad de sustitución de los factores, σ). Una alta elasticidad de sustitución implica, pues, una mayor concavidad (y por ende menor será el efecto de los cambios en el precio de un factor sobre el coste total).

$C(W, Q) = F(Q) \cdot C(W, 1)$. Es decir, como el resultado de multiplicar una función estrictamente creciente en el output por el coste unitario de producir una unidad.

Variaciones en los precios de los factores

Antes de analizar el efecto del horizonte temporal, cabe preguntarse cómo responde el coste ante variaciones en los precios de los factores. Un incremento en el precio de un factor nunca puede reducir el coste total: si la sustitución de factores que provoca el encarecimiento redujese el coste, la empresa no estaría minimizando costes de partida. El impacto exacto depende de dos elementos: la elasticidad de sustitución entre factores –cuanto mayor sea, más fácil resulta sustituir el factor encarecido por otro más barato–, y el peso del factor en el coste total, medido por la elasticidad coste-precio propio, que aplicando el lema de Shephard resulta ser la proporción del gasto total destinada a dicho factor:

$$\varepsilon_{C,w_i} \equiv \frac{\partial C}{\partial w_i} \cdot \frac{w_i}{C} = \frac{x_i^* \cdot w_i}{C}$$

Así, si el trabajo representa el 10 % del coste total y su precio sube un 1 %, el coste total sube aproximadamente un 0,1 %. Además, por el teorema de la dualidad, en el óptimo las demandas condicionadas del problema dual coinciden con las demandas ordinarias evaluadas en el coste mínimo. Derivando esta identidad respecto al precio de otro factor y aplicando de nuevo el lema de Shephard se obtiene una descomposición formalmente idéntica a la ecuación de Slutsky del consumidor:

$$\underbrace{\frac{\partial x_i\left(w, \bar{C}\right)}{\partial w_j}}_{\text{Efecto total}} = \underbrace{\frac{\partial h_i\left(w, \bar{Q}\right)}{\partial w_j}}_{\text{Efecto sustitución}} - \underbrace{h_j \cdot \frac{\partial x_i\left(w, \bar{C}\right)}{\partial \bar{C}}}_{\text{Efecto renta}}$$

Esta descomposición permite clasificar cada factor como normal o inferior según el signo de la derivada de su demanda ordinaria respecto al coste, exactamente igual que en la teoría de la demanda del consumidor.

Análisis temporal de la función de costes

Una vez analizado el efecto de los precios, la teoría económica distingue además entre coste a corto plazo y a largo plazo, por lo que ahora se realizará un análisis temporal de la función de costes, en el cual sí se consideran constantes los precios relativos de los factores.

 © 2026 Manuel Cabado García

Análisis de corto plazo

Comenzando por el corto plazo, este se define como el horizonte temporal en el que al menos un factor es fijo –normalmente es el capital– y el tamaño de planta se toma como dado. Además en el corto plazo existen:

- Costes fijos, independientes del volumen de producción. Si los costes, además de ser fijos, son irrecuperables, entonces se habla de costes hundidos.
- Costes variables, aquellos que varían al modificar el volumen de producción.

Análisis gráfico Gráficamente, en un plano de costes y cantidad producida como el de la Figura 6.3 se representan las diferentes curvas de costes:

- Costes variables[9], CV, se representan a través de una S invertida[10] que refleja la ley de rendimientos decrecientes o proporciones variables[11].
- Costes fijos[12], CF, son constantes al no depender de la producción, es decir, una línea horizontal.
- Costes totales, CT, la suma vertical de los dos anteriores.

$$CT = CV + CF \tag{6.9}$$

En el gráfico adyacente:

- Coste marginal, CMg, es el aumento del coste total que provoca producir una unidad adicional, y se obtiene en un plano adyacente trazando las tangentes a cada uno de los puntos de la función de costes total[13]. El coste marginal es el cociente entre el

[9] Los costes variables incluyen los costes de materias primas, de mano de obra directa y los gastos de funcionamiento, tales como combustibles, reparaciones ordinarias y mantenimiento rutinario.

[10] Ello se debe a que inicialmente el factor variable genera rendimientos marginales (es decir, PMg) crecientes (es decir, coste marginal decreciente), pero llegado un punto el factor variable genera rendimientos marginales decrecientes (es decir, coste marginal creciente).

[11] La ley de las proporciones variables fue enunciada por Turgot en 1767 como respuesta teórica a un ensayo fisiócrata sobre la agricultura francesa. De acuerdo con esta ley, dado un factor fijo (por ejemplo, tierra), las primeras unidades del factor variable (trabajo) aumentan la producción cada vez más, por lo que el coste marginal es inicialmente decreciente (y la curva de costes, cóncava), pero llegará un punto en que cada unidad adicional añadirá menos producción que la anterior, y el coste marginal pasará a ser creciente (la curva de costes se vuelve convexa).

[12] Los costes fijos incluyen los sueldos de personal fijo –generalmente personal administrativo–, depreciación de maquinaria, gastos de amortización de edificios y reparaciones y gastos de mantenimiento y reparación de los terrenos –si estos existen–. Además, otro elemento que puede ser tratado como un coste fijo es el beneficio normal, que retribuye tanto el capital aportado como la asunción de riesgos.

[13] Como se trata de una derivada, gráficamente se obtiene trazando las tangentes a cada uno de los puntos de la función de coste variable o de coste total (ya que ésta se obtiene desplazando paralelamente el coste variable por el coste fijo). Los ángulos van disminuyendo hasta el punto de inflexión y a partir de ahí empiezan a crecer pasando por el mínimo del coste medio variable y del coste medio total, ya que para esos niveles de producción la tangente (que, como se ha visto, determina el coste marginal) coincide con la secante (es decir, con el rayo vector, que determina el coste medio).

precio del factor variable y su productividad marginal, $CMg = w/PMg$, de forma que el coste marginal es decreciente mientras los rendimientos marginales sean crecientes (mientras la curva de PMg tenga pendiente positiva), y el coste marginal se vuelve creciente cuando los rendimientos marginales se vuelven decrecientes (cuando la curva de PMg pase a tener pendiente negativa).

- Coste fijo medio, $CFMe$, decrece a medida que aumenta la producción pues se va diluyendo en esta.
- Coste variable medio, $CVMe$, decreciente hasta que es intersecado por el coste marginal, es decir, cada unidad de producto adicional tiene a partir de entonces un coste superior a la media, por lo que aumenta ésta. Otra manera de verlo es que el coste variable medio mínimo es el nivel de producción en el que el rayo vector del gráfico superior es tangente a la curva de coste variable[14].
- Coste total medio, es la suma vertical del coste variable y fijo medios. Cortado en su mínimo por CMg, punto en el que coinciden las pendientes de la curva de CT y de la secante desde el origen.
  Nótese cómo el coste variable medio y el coste total medio convergen a medida que aumenta la producción, porque se van diluyendo los CF[15].
  Además, se observa cómo la curva de costes totales medios tiene forma de U. Esto se debe a la ley antes enunciada de las proporciones variables, pues tras los rendimientos marginales crecientes iniciales, los rendimientos marginales se tornarán decrecientes, y llegará un punto en el que el aumento del producto debido a una unidad adicional de factor será tan pequeño, que el coste adicional de ese factor hará crecer el coste medio.

El conocimiento de los costes a corto plazo es especialmente importante para las empresas que producen en un entorno en el que las condiciones de demanda fluctúan constantemente.

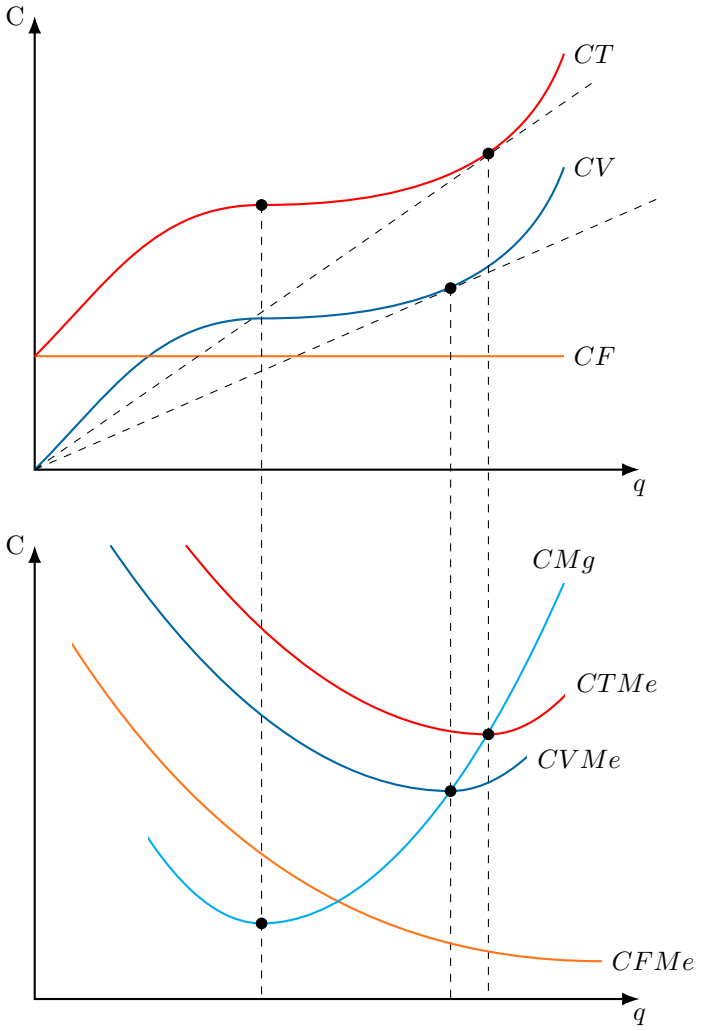
[14] Si, aprovechando el gráfico, se va más allá de la teoría de los costes y se entra en la teoría de los mercados, puede añadirse que el mínimo del coste variable medio se denomina mínimo de explotación, pues si se supone competencia perfecta, en ese punto los ingresos cubren los costes variables, por lo que la empresa estará indiferente entre producir y cerrar ya que en ambos casos pierde los costes fijos. El mínimo de explotación se corresponde con el margen extensivo u óptimo técnico de la teoría de la producción.

Es importante no hablar en el presente tema de "mínimo de explotación", "competencia perfecta", "beneficios", "pérdidas", "oferta", "precios del producto", etc., pues suponen ir un paso más allá ya que todos ellos son términos propios de la teoría de los mercados (y, si se utilizan para aprovechar el gráfico y extender la explicación, dejar claro que eso no es teoría de costes, sino un siguiente nivel de análisis: la teoría de los mercados)

[15] Aprovechando de nuevo el gráfico para adentrarse en la teoría de los mercados, puede añadirse que el mínimo del coste total medio se denomina óptimo de explotación o umbral de beneficio, pues si se supone competencia perfecta, en ese punto los ingresos cubren todos los costes, por lo que a partir de ahí se generarán beneficios positivos. A diferencia del mínimo de explotación, el óptimo de explotación no puede asociarse a un punto concreto de la función de producción, pues el mínimo del coste total medio depende también de los costes fijos. Se sitúa siempre dentro de la fase II.

 © 2026 Manuel Cabado García

Figura 6.3: Curvas de costes de una empresa a corto plazo



Análisis de largo plazo

A diferencia del corto plazo, en el largo plazo se va a suponer que todos los factores son variables –excepto la tecnología que varía en el muy largo plazo– y que las empresas pueden elegir el tamaño de planta. En el largo plazo la función de costes es considerada una función de planificación, ya que para elegir el tamaño de planta se distinguen dos fases:

- Primera fase, las empresas deberán estimar la demanda de producto futura y adaptar en consecuencia su capacidad de producción.
- Segunda fase, decidirán sobre el tamaño de planta instalado.

Por lo tanto, el empresario es libre para decidir la dimensión óptima de la empresa que le permite minimizar los costes.

Análisis gráfico Gráficamente, en un plano de costes y cantidad producida como el de la Figura 6.4, se representan las curvas de costes:

- Costes a largo plazo, $C_{l/p}$, envuelve las curvas de costes a corto plazo, ya que los costes a largo plazo siempre estarán por debajo (o serán iguales) de los costes a corto plazo.
- Curva de coste medio a largo plazo, $CMe_{l/p}$, (ya no se habla de total y variable al no haber costes fijos) se deduce de las curvas de costes medios a corto plazo.
  La tangencia de cada curva de corto plazo con la de largo plazo se produce en la parte decreciente de las curvas de $CMe_{c/p}$ en los puntos a la izquierda del mínimo de la curva de $CMe_{l/p}$, y en la parte creciente de las curvas de $CMe_{c/p}$ en los puntos a la derecha del mínimo de la curva de $CMe_{l/p}$.
  La razón económica es que en la parte izquierda la planta no trabaja a plena capacidad y a la derecha opera con sobrecarga (es decir, se trabajan horas extraordinarias que se pagan más caras, la maquinaria trabaja a velocidad antieconómica y se producen más averías, etc.). Únicamente en el punto mínimo del coste medio a largo plazo, la planta opera óptimamente en el corto plazo.
  Sin embargo, la empresa únicamente observa la curva a corto plazo. La curva de coste a largo plazo es un concepto estratégico, no observable por la empresa. No corresponde con un tamaño físico de planta ni tiene correspondencia con la realidad[16].
- Las curvas de $CMg_{c/p}$ cortan a la curva de $CMg_{l/p}$ en el nivel de producción para el cual sus respectivas curvas de $CMe_{c/p}$ son tangentes a la curva de $CMe_{l/p}$.

Como se observa, la curva de coste medio a largo plazo también tiene forma de U, pero por una razón distinta a la del corto plazo: la empresa disfruta primero de economías de escala y, a partir de cierto tamaño, sufre deseconomías de escala, principalmente de gestión.[17]

Una explicación es que el tamaño de planta está diseñado para producir de manera óptima un determinado nivel de producción, de forma que cualquier desviación de este supuesto implica mayores costes unitarios. La planta es, pues, completamente inflexible, pues no existe ninguna capacidad de reserva.

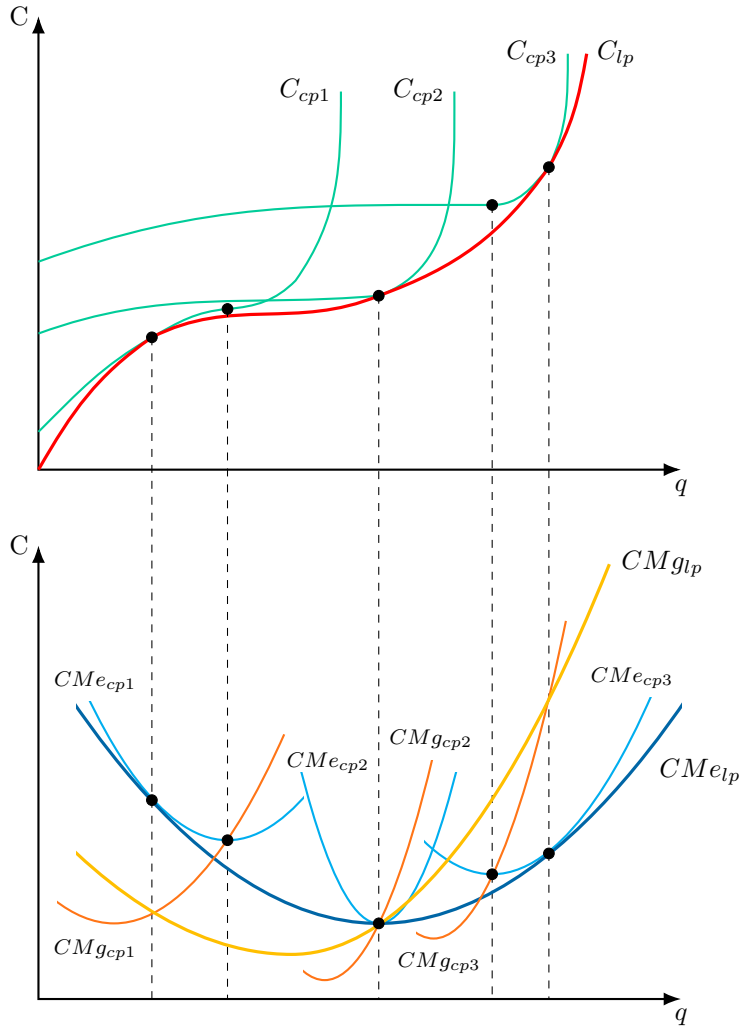
Así, en el punto mínimo de los costes medios a largo plazo, éste es igual al $CMg_{l/p}$,

[16] Cuestiones sobre el largo plazo: Es un horizonte de planificación. La empresa hace su planificación de demanda esperada y elige el tamaño de planta y factores productivos óptimos que va a necesitar para maximizar beneficio. No obstante, una vez tomada la decisión y adquirido el capital, este se vuelve fijo y la actividad se produce en el corto plazo. Es decir, el largo plazo se compone de situaciones del corto plazo que la empresa puede elegir. Supone que la función de producción es continua, implica que los factores productivos son perfectamente divisibles, por consiguiente, hay infinitos tamaños de planta, aunque en realidad se dan solamente unos pocos procesos tecnológicos.

[17] La intuición es que a largo plazo no queda ningún factor fijo con el que "saturar" al factor variable. Lo que encarece la producción de una planta gigantesca no es la falta de espacio, sino la dificultad creciente de coordinarla.

 © 2026 Manuel Cabado García

Figura 6.4: Curvas de costes de una empresa a largo plazo



$CMg_{c/p}$ y $CMe_{c/p}$:

$$CMe_{c/p} = CMe_{l/p} = CMg_{c/p} = CMg_{l/p} \tag{6.10}$$

Relación entre el corto y el largo plazo: teorema de Le Chatelier-Samuelson

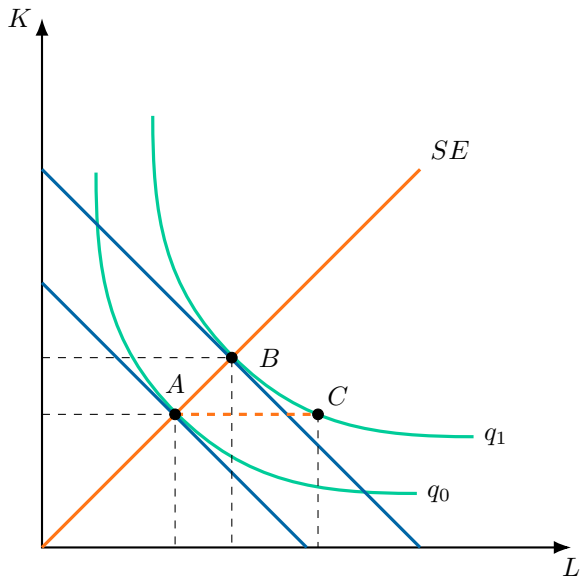
Otra cuestión que se observa a partir del gráfico es que el tramo creciente de las curvas de coste marginal (el único que se ha representado) es más rígido en el caso de los costes a corto plazo que en el largo plazo. Esto es relevante porque en competencia perfecta, el tramo creciente de la curva de coste marginal a partir del umbral de producción es precisamente la curva de oferta.

Este resultado lo recoge el teorema de Le Chatelier-Samuelson (Samuelson, 1947), que establece que las funciones de oferta de la empresa competitiva a corto plazo son más rígidas que las de largo plazo y tanto más rígidas cuanto mayor es el número de factores considerados fijos[18].

[18]La intuición es que en el corto plazo, al ser al menos uno de los factores productivos fijo, la capacidad

Gráficamente, como se observa en la Figura 6.5, partiendo de una asignación como A, si la empresa en el corto plazo busca aumentar la producción de Q_0 hasta Q_1 la empresa deberá aumentar el factor variable, trabajo, hasta una asignación C, con un nivel de costes superior al alcanzado si aumenta la producción a lo largo de su senda de expansión, alcanzando la asignación B, con una curva isocoste asociada más cercana al origen y que supone un menor nivel de costes respecto al largo plazo.

Figura 6.5: Relación entre el largo plazo y el corto plazo: Teorema de Le Châtelier-Samuelson



Economías de escala

Además del horizonte temporal, otro elemento clave para explicar la respuesta de los costes ante cambios en la producción son las economías de escala, que existen cuando el coste medio a largo plazo disminuye al aumentar la producción.[19]

Su medición se lleva a cabo a través de la elasticidad del coste total, $\varepsilon_{c,q}$, que mide la variación porcentual del coste total ante la variación porcentual de la producción:

$$\varepsilon_{c,q} = \frac{\partial C}{\partial q} \cdot \frac{q}{C} = \frac{CMg}{CMe} \tag{6.11}$$

donde, dependiendo de su valor

- Si $\varepsilon_{c,q} = 1$ da lugar a economías de escala constantes –$CMg = CMe$–, donde el coste total varía en la misma proporción que la producción, es decir, el CMe es

de adaptación de la empresa a la nueva estructura de precios es menor ya que el cambio de la cantidad ofertada solo puede realizarse variando alguno de los factores que sí son variables.

[19] La intuición es que imprimir diez mil ejemplares de un libro cuesta mucho menos por ejemplar que imprimir cien, porque la preparación de la imprenta se reparte entre más unidades.

 © 2026 Manuel Cabado García

constante.

- Si $\varepsilon_{c,q} < 1$ da lugar a economías de escala –$CMg < CMe$–, donde el coste total varía en menor proporción que la producción, es decir, el CMe es decreciente.
- Si $\varepsilon_{c,q} > 1$ da lugar a deseconomías de escala –$CMg > CMe$–, donde el coste total varía en mayor proporción que la producción, es decir, el CMe es creciente.

Rendimientos crecientes a escala vs. economías de escala. Ambos conceptos guardan una estrecha relación, probablemente porque la justificación económica de unos y otros es similar (es decir, las indivisibilidades y la especialización en el proceso productivo). Sin embargo, son dos conceptos diferentes ya que los rendimientos crecientes a escala es un concepto de la teoría de la producción y las economías de escala es un concepto de la teoría de los costes.

No obstante, ambos conceptos coinciden en el caso en que la función de producción sea homotética (pues en este caso, los costes y la producción tienen una relación imagen-espejo). En este tipo de funciones, la relación entre rendimientos a escala y economías de escala es inversa[20]:

$$e = \frac{1}{\varepsilon_{c,q}} = \frac{CMe}{CMg} \tag{6.12}$$

donde e es la elasticidad de escala.

Fuera del caso homotético, la senda de expansión deja de ser una recta desde el origen: la empresa que crece cambia las proporciones en que usa los factores, mientras que los rendimientos a escala se definen aumentando todos los factores en la misma proporción. Ambos conceptos describen entonces movimientos distintos, aunque en cada punto de la senda de expansión la elasticidad de escala sigue siendo igual al cociente CMe/CMg (Varian, 1992).[21]

Economías de escala internas y externas Conviene además distinguir las economías de escala internas, que resultan de la propia acción de la empresa al expandir su producción y quedan recogidas en la forma de su curva de costes, de las economías de escala externas (Alfred Marshall), que se producen a nivel de toda la industria –no de la tecnología o los costes de cada empresa individual– cuando el crecimiento del tamaño agregado del sector reduce el coste medio de todas las empresas que lo componen, por ejemplo a través de clústeres industriales, un mercado de trabajo más especializado o *spillovers* de conocimiento entre empresas próximas.

[20] Véase el Apéndice 6.A.1.

[21] La intuición es que una empresa que minimiza costes ya ha elegido la mejor combinación de factores, así que, a partir de ella, crecer escalando todos los factores o hacerlo por la senda óptima tiene, en el margen, el mismo coste.

6.3. Críticas a la teoría tradicional ⏳ 4'

La teoría tradicional ha sido cuestionada tanto por sus fundamentos teóricos como por sus implicaciones empíricas, sobre todo desde los autores de la teoría de la organización industrial. Antes de entrar en esas críticas, conviene explorar qué formas alternativas podría adoptar la curva de costes al margen de la U tradicional, y qué implicaciones tendrían para el mercado.

Curvas de costes sin rendimientos mixtos

Hasta aquí se ha analizado el caso tradicional de curvas en forma de U debido a la existencia de economías de escala primero y deseconomías después. A continuación, se presentan en la Figura 6.6 otros casos en que hay economías de escala constantes, crecientes o decrecientes para cualquier nivel de producción:

- Economías de escala constantes: el coste marginal coincide con el coste medio, por lo que la curva de coste medio será una recta. En competencia perfecta, la oferta a largo plazo sería igual al coste marginal, que a su vez sería igual al coste medio.
- Economías de escala: el coste marginal va por debajo del coste medio. En competencia perfecta el precio será igual al coste marginal, pero éste estará por debajo del coste medio por la existencia de rendimientos crecientes. Por lo tanto, las empresas incurrirán en pérdidas, saldrán del mercado y no habrá producción. La única solución es la existencia de un monopolista natural que sea capaz de influir sobre el precio para elevarlo por encima del coste marginal.
- Deseconomías de escala: el coste marginal va por encima del coste medio. En competencia perfecta, el precio será igual al coste marginal, pero éste estará por encima del coste medio por la existencia de rendimientos decrecientes. Por lo tanto, las empresas obtendrán beneficios, entrarán masivamente en el mercado y en el equilibrio habrá infinitas empresas, produciendo cada una cantidad infinitesimalmente pequeña.

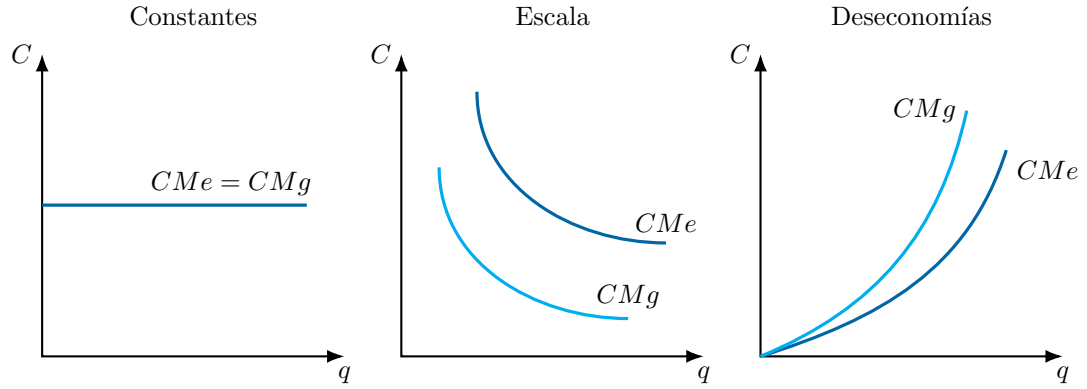
La teoría de los costes de Stigler

Stigler (1939) cuestionó la forma de U del coste medio ya que empíricamente se encuentra que el coste medio a largo plazo no es creciente y que el coste medio variable a corto plazo es constante para los niveles de producción usuales de la empresa y creciente cuando existe sobrecarga de trabajo. Así:

- A corto plazo la curva de coste medio presenta un tramo horizontal.
- A largo plazo la curva de costes medios no es creciente.

 © 2026 Manuel Cabado García

Figura 6.6: Curvas de coste sin rendimientos mixtos



Análisis de corto plazo

En el corto plazo, Stigler señaló que el empresario implementa un tamaño de planta que le otorgue flexibilidad debido a

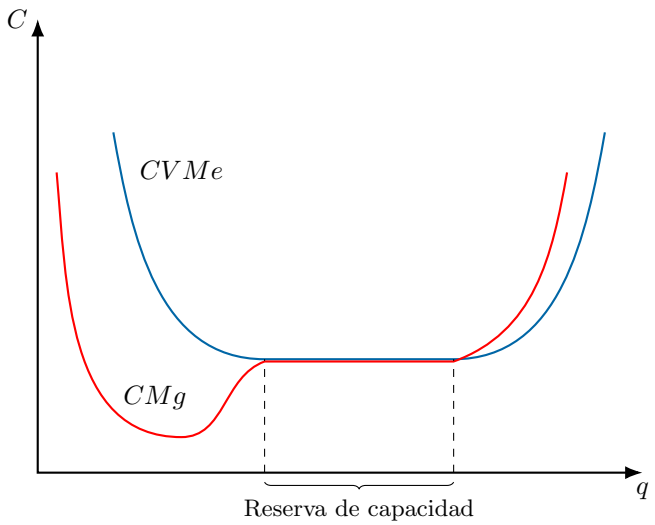
- Razones económicas, relacionadas con la demanda coyuntural y las fluctuaciones cíclicas como las reparaciones de maquinaria o la anticipación a cambios en los gustos de los consumidores.
- Razones técnicas, relacionadas con la indivisibilidad del factor capital y la instalación de una reserva de capacidad que permita a la empresa hacer frente a una posible expansión futura[22].

Gráficamente, en un plano de costes y cantidad producida como el de la Figura 6.7 se representan los diferentes costes:

- En la curva de costes en forma de U se consideraba que el factor fijo no era divisible y, por lo tanto, se utilizaba desde el principio en la producción todo el factor fijo disponible. Así, la relación factor fijo / factor variable iba cambiando, y solamente existía una relación eficiente entre ambos, que era la óptima (mínimo del coste medio).
- En la curva de costes con tramo horizontal se considera que el factor fijo es divisible, de tal forma que hasta que alcanza la capacidad total, se está utilizando una cantidad del factor fijo menor a la disponible. Así, si aumenta o disminuye la producción, el factor fijo y el variable aumentan o disminuyen en la misma proporción, manteniéndose la proporción eficiente entre ambos, de forma que los rendimientos marginales decrecientes no entran en juego hasta que se agota esa reserva de capacidad.

[22]No se debe confundir la reserva de capacidad con el exceso de capacidad de la teoría tradicional. La reserva de capacidad está planificada por el empresario, y el exceso de capacidad lo puede determinar el mercado, pero no está planificado. En la teoría tradicional, se diseña la empresa para producir de forma óptima en el mínimo de los costes medios variables y si se produce antes del mínimo, aparece el exceso de capacidad. Sin embargo, la reserva de capacidad se planifica y supone costes constantes.

Figura 6.7: Teoría de costes de Stigler a corto plazo



Análisis de largo plazo

En el largo plazo, la llamada teoría moderna de costes, sistematizada por Koutsoyiannis (1979) a partir de trabajos empíricos como los de Stigler, sostiene que los costes medios a largo plazo tienen forma de L, pues se dividen en

- Costes de producción, que decrecen a medida que aumenta la producción. La escala óptima puede mejorar mediante innovaciones o produciendo varios productos diluyendo costes[23].
- Costes de administración, a escalas muy grandes serán ligeramente crecientes. No obstante, se argumenta que estas deseconomías de escala gerenciales son insignificantes debido a los métodos perfeccionados de la moderna ciencia de la administración de empresas[24].

Gráficamente, en un plano de costes y cantidad producida como el de la Figura 6.8 se representan los diferentes costes.

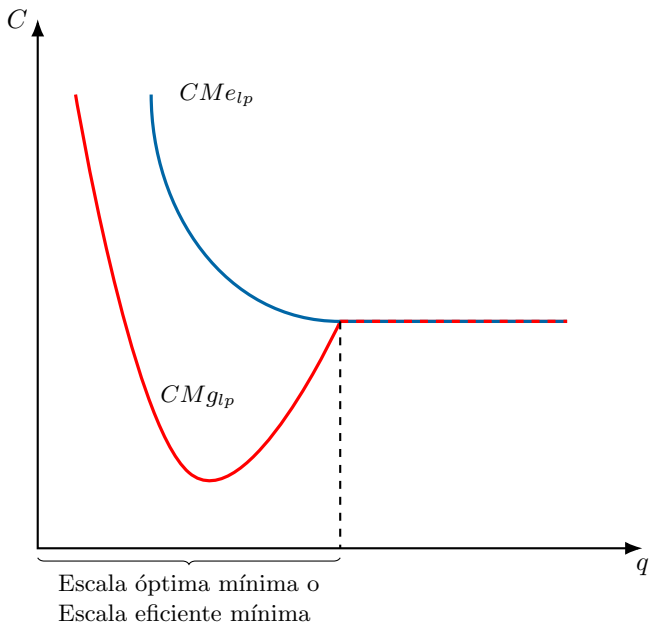
Así, el descenso de los costes de producción compensa con creces el probable aumento de los costes administrativos, de manera que la curva de CMe_{lp} decrece suavemente –algunos autores argumentan así que existirán economías de escala incluso en el largo plazo– o permanece constante –forma de L– para escalas de producción muy grandes.

[23] El hecho de que las CMe disminuyan de forma continua implica la existencia de economías de escala técnicas, pero que se agotan a partir de un punto que se denomina escala mínima eficiente (o escala óptima mínima). Pese a todo, aún podrían disminuir más los costes con nuevas tecnologías, una mejora de la productividad, mayor especialización...

[24] Son decrecientes hasta alcanzar un cierto tamaño de planta pero pueden crecer lentamente para escalas muy grandes. No obstante, aunque existan deseconomías gerenciales pueden compensarse con las economías de escala técnicas.

 © 2026 Manuel Cabado García

Figura 6.8: Teoría de costes de Stigler a largo plazo



La ineficiencia X de Leibenstein

Más allá de cuestionar la forma de la curva de costes,Leibenstein (1966) pone en duda la minimización del coste por parte de la empresa partiendo de los siguientes supuestos:

1. El esfuerzo es endógeno y es un mal para el trabajador.
2. Inercia: los trabajadores eligen su grado de esfuerzo según la costumbre, y solo lo modifican ante una presión externa considerable.
3. Los contratos son incompletos, es decir, no pueden prever todas las contingencias.
4. Existe información asimétrica, es decir, el productor no puede observar el nivel de esfuerzo realizado por el trabajador.

Por todo lo anterior, los trabajadores –que son los que deciden el nivel de esfuerzo a llevar a cabo–, escogerán por lo general un nivel de esfuerzo inferior al que minimizaría el coste medio. De esta manera, el comportamiento optimizador, es decir, la producción que minimiza el coste medio, es una excepción, produciéndose una ineficiencia X, es decir, la diferencia entre el coste en que realmente incurre la empresa y el mínimo que permitiría su tecnología, originada por ese menor esfuerzo.[25]

El grado de competencia juega un papel fundamental en la teoría de Leibenstein:

- Con competencia alta, los trabajadores, en aras de asegurar la supervivencia de la empresa (y conservar, por tanto, su puesto de trabajo), se esforzarán y mejorarán su productividad marginal.

[25]La intuición es que dos fábricas con las mismas máquinas y los mismos salarios pueden tener costes distintos simplemente porque en una se trabaja con más diligencia que en la otra.

- Con competencia baja, los trabajadores tendrán incentivos a relajarse. Así, un aumento de los precios del producto no tiene por qué traducirse en un mayor beneficio, ya que si los trabajadores perciben que ese aumento del precio de venta se debe a un mark-up posibilitado por una menor competencia, entonces el beneficio podría disminuir al aumentar los costes por la ineficiencia X.

Generalmente las empresas se mueven en un contexto de competencia imperfecta, por lo que la presión competitiva no motiva la diligencia en el trabajo.

Implicaciones de la teoría de la Ineficiencia X Así, la aportación de Leibenstein tiene dos implicaciones muy interesantes[26]:

1. Un incremento de los precios de los factores no supone necesariamente un aumento del coste total: La dirección transmitirá a los trabajadores la urgencia de realizar un mayor esfuerzo para reducir costes, con lo que la ineficiencia X disminuye.
2. No existe independencia entre el aumento de precios y el de costes: Si el aumento del precio viene posibilitado en un marco de menor competencia, entonces se producirá un aumento de costes vía una mayor ineficiencia X (a diferencia de la Teoría tradicional, que considera que el aumento del precio no tiene por qué implicar un aumento de los costes).

6.4. Otros desarrollos ⏳2'30"

Otros desarrollos de la teoría de costes serían las economías de alcance, la subaditividad de costes y las curvas de aprendizaje.

Producción multiproductos, economías de alcance

En la práctica, es frecuente que una misma empresa produzca dos o más productos. Existen economías de alcance (también denominadas de gama o de variedad)[27] cuando producir esos bienes en una sola empresa cuesta menos que producirlos por separado en varias

$$C(q_1, q_2) < C(q_1, 0) + C(0, q_2) \tag{6.13}$$

[26] Stigler (1976) argumentó que la teoría de la ineficiencia X estaba infundada, porque es eficiente hacer respetar los contratos, hasta que el valor marginal del incremento en la eficiencia se iguale con el coste marginal de supervisión. Sólo recientemente, los economistas han intentado incorporar la ineficiencia X a los modelos de oligopolio.

[27] Entre los factores que explican la existencia de economías de alcance están: Existencia de inputs públicos, que son aquellos inputs que una vez producido un bien, pueden ser empleados en la producción de otro bien sin coste adicional, y la existencia de inputs que presentan indivisibilidades en la adquisición por parte de la empresa, ejemplo típico es el capital.

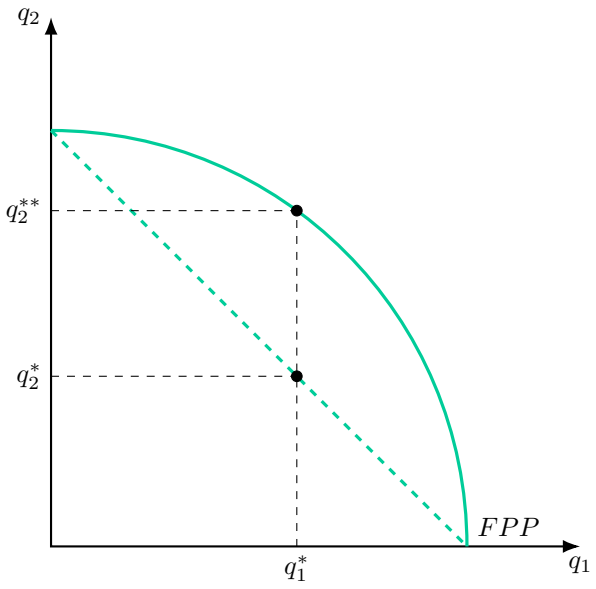
 © 2026 Manuel Cabado García

En muchas ocasiones esto se debe a que los productos son dependientes desde el punto de vista técnico –por ejemplo carne y cuero en el caso del vacuno–, o porque se pueden aprovechar estructuras organizativas o de comercialización comunes.

Gráficamente, en un plano de dos bienes, q_1 y q_2, como el de la Figura 6.9, la restricción tecnológica (es decir, combinaciones técnicamente factibles) viene determinada por la Frontera de Posibilidades de Producción, FPP, que expresa la máxima cantidad de un producto alcanzable, dada una cantidad producida del otro.

La FPP es cóncava ya que presenta economías de alcance, ya que si no las presentase sería una línea recta donde, por ejemplo, dada una cantidad de factores, produciendo q_1^* la empresa sin economías de gama puede producir q_2^*, mientras que con los mismos factores la empresa con economías de gama puede producir más en q_2^{**}.

Figura 6.9: Economías de alcance



Cabe subrayar que no existe una relación directa entre las economías de escala y las de gama.

Relación con el coste de producción de bienes Otra forma de ver las economías de alcance es mediante la relación entre el coste total de producir un bien en términos del otro. Así, se daría una curva de pendiente negativa que en función de la forma mostraría la presencia de economías de alcance:

- Si la representación es cóncava, existen economías de alcance pues el coste de producir dos bienes conjuntamente es menor que el de producirlos por separado.
- Si es una línea recta demuestra la no existencia de economías de alcance.
- Si es convexa entonces muestra la presencia de deseconomías de alcance.

Grado de economías de alcance Para cuantificar el ahorro anterior, el grado de economías de alcance, EA, permite medir el porcentaje de ahorro de costes que se obtiene cuando dos o más productos se producen conjuntamente en lugar de individualmente

$$EA = \frac{C(q_1) + C(q_2) - C(q_1, q_2)}{C(q_1, q_2)} \tag{6.14}$$

Cuando hay economías de alcance, el coste conjunto es menor que la suma de los costes individuales, por lo que EA es mayor que 0. Cuando hay deseconomías de alcance, EA es negativo. En general, cuanto mayor es el valor de EA, mayores son las economías de alcance.

Subaditividad de costes

Cerrando los desarrollos sobre producción multiproducto, se dice que una tecnología presenta subaditividad estricta en costes si resulta menos costoso producir un output dado en una sola empresa que repartiéndolo entre varias

$$C\left(\sum_{i=1}^{n} q_i\right) < \sum_{i=1}^{n} C_i(q_i) \tag{6.15}$$

En esta situación, la subaditividad de costes genera lo que se conoce como monopolio natural[28], ya que cualquier cantidad ofrecida de equilibrio sería menos costosa producirla en una sola empresa, y solo una podría obtener beneficios.

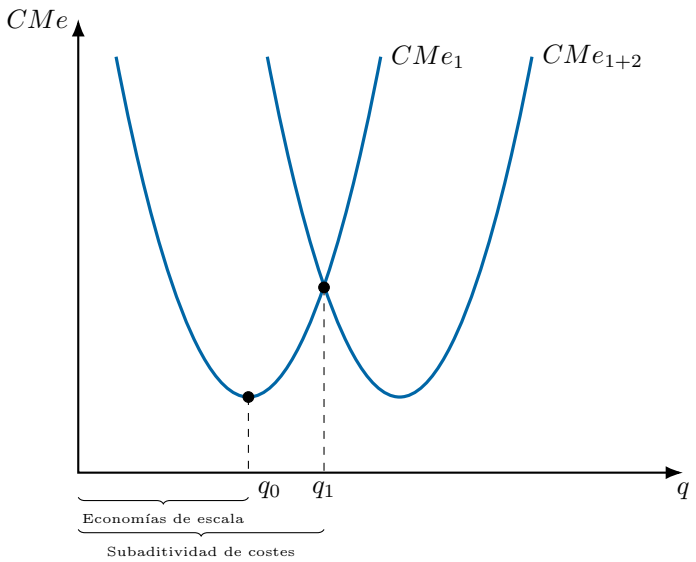
Gráficamente, en un plano de costes medios, CMe, y cantidad producida, q, como el de la Figura 6.10 se representa la curva de costes medios cuando toda la producción se concentra en una empresa, CMe_1, y la de dos empresas que se reparten la producción a medias, CMe_{1+2}. Pero la subaditividad se extiende hasta q_1 a pesar de que existe un tramo con deseconomías de escala. En efecto, aunque en el tramo que va desde q_0 hasta q_1 existen deseconomías de escala, el coste medio de concentrar la producción en una empresa es menor que el de repartirla en dos empresas. A partir de q_1 ya no existe subaditividad de costes y, por lo tanto, no hay monopolio natural.

Por consiguiente, la existencia de Economías de escala es condición suficiente pero no necesaria para la existencia de subaditividad de costes. De hecho, pueden existir deseconomías de escala, y que el coste medio de concentrar la producción en una empresa sea menor que el de repartirla en dos empresas (como sucede entre q_0 y q_1). A partir de q_1 deja de existir subaditividad de costes.

[28] El concepto de subaditividad de costes y su relación con el monopolio natural fue formalizado por Baumol, Panzar y Willig (1982) en su teoría de los mercados contestables, que muestra que la amenaza de entrada puede disciplinar a un monopolio natural incluso sin regulación directa.

 © 2026 Manuel Cabado García

Figura 6.10: Subaditividad de costes



Monopolio natural La subaditividad de costes genera estructuras de mercado con la existencia de un monopolio natural.

Economías de aprendizaje

Por último, la empresa no solo es capaz de reducir los costes medios a largo plazo cuando aumenta el volumen de producción y se van explotando las economías de escala.

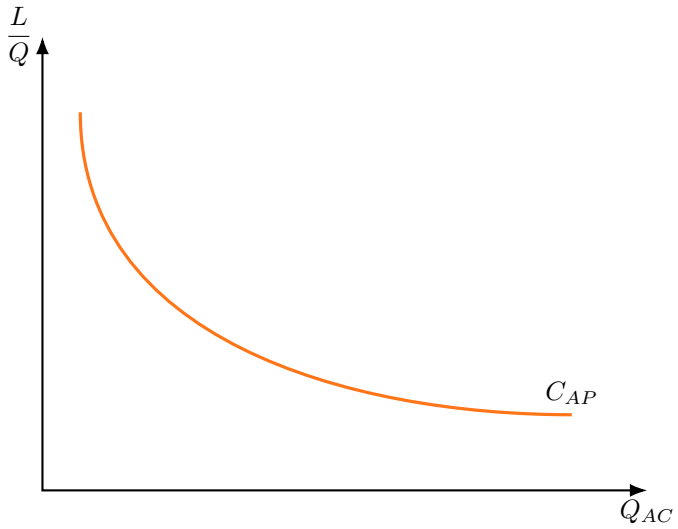
De hecho, la empresa puede ir reduciendo costes gracias a las economías de aprendizaje o de escala dinámicas derivadas del aprendizaje del trabajador y de la gestión, fenómeno que se produce conforme crece la producción acumulada en el tiempo por

- Mayor rapidez y habilidad de los trabajadores, ya que a medida que son más expertos su velocidad aumenta.
- Mayor eficiencia en la gestión, donde los directivos aprenden a programar el proceso de producción más eficazmente, desde el flujo de materiales hasta la organización de la propia producción.
- Los ingenieros que al principio son muy cautos en el diseño de los productos pueden adquirir suficiente experiencia para poder introducir tolerancias en el diseño que ahorren costes sin aumentar los defectos. La mejora y el aumento de las herramientas especializadas y de la organización de la planta también pueden reducir el coste.
- Visión más integrada de la empresa, donde también los proveedores pueden aprender y contribuir a la reducción del coste de la empresa.

Gráficamente, en un plano de trabajo necesario por unidad de producción, L/Q, y de producción acumulada en el tiempo, Q_{AC}, como el de la Figura 6.11, se representa la

curva de aprendizaje, C_{AP}, que muestra el grado en que disminuyen las horas necesarias de trabajo por unidad de producción a medida que aumenta la producción acumulada

Figura 6.11: Curva de aprendizaje



A su vez, dentro de la Teoría del Comercio Internacional las economías de escala dinámicas han servido para justificar la protección temporal frente a la competencia exterior en lo que se conoce como el argumento de la industria naciente.

Economías de escala vs. economías de aprendizaje Gráficamente, en un plano de costes medios y cantidad de producción como el de la Figura 6.12, la diferencia es que las economías de escala se presentarían del punto inicial A al punto B, es decir, reducción de costes sobre la misma curva, mientras que las economías de aprendizaje se presentarían del punto A al C, es decir, el mismo nivel de producción pero con menor coste.

6.5. Conclusiones 30"

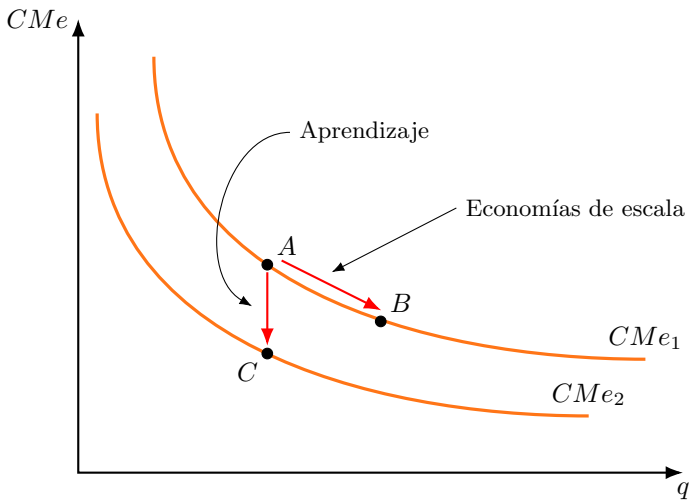
Como conclusión, la teoría de costes es un componente fundamental a la hora de entender el comportamiento de las empresas y los aspectos relevantes de la tecnología de producción. Del mismo modo, la teoría de costes también aporta información sobre cómo va a funcionar un mercado, sirviendo así como una guía para las políticas de competencia.

Así, la estructura de costes de las empresas puede pre-configurar en gran medida la estructura del mercado. Es el caso, por ejemplo, de la subaditividad de costes y el monopolio natural.

Esta lectura neoclásica admite matices: Stigler cuestiona empíricamente la forma de U de la curva de costes, y la ineficiencia X de Leibenstein liga el coste mínimo a la presión

 © 2026 Manuel Cabado García

Figura 6.12: Las economías de escala frente al aprendizaje



competitiva. A ello se suman desarrollos como el alcance, la subaditividad y el aprendizaje, que acercan el análisis a la empresa real.

No obstante, la teoría de los costes no termina de caracterizar las decisiones de oferta de producto y de demanda de factores de la empresa. Para ello es necesario acudir a la teoría de los mercados para ver cuál de todas las combinaciones técnica y económicamente eficientes (es decir, todas las combinaciones que se recogen en la senda de expansión) es la óptima.

Apéndice

6.A.1. Rendimientos crecientes a escala vs. economías de escala

Ambos conceptos guardan una estrecha relación, probablemente porque la justificación económica de unos y otros es similar –es decir, las indivisibilidades y la especialización en el proceso productivo–. Sin embargo, son dos conceptos diferentes:

- Rendimientos crecientes a escala (concepto de la teoría de la producción): al aumentar la dotación de los factores en la misma proporción, la producción aumenta en una proporción mayor[29].
- Economías de escala (concepto de la teoría de los costes): al aumentar la producción, los costes totales aumentan menos que proporcionalmente. Efectivamente, para aumentar la producción se ha tenido que aumentar la dotación de factores, pero no necesariamente en la misma proporción.

No obstante, ambos conceptos coinciden en el caso en que la función de producción sea homotética, pues en este caso los costes y la producción tienen una relación imagen-espejo. En este tipo de funciones, la relación entre rendimientos a escala y economías de escala es inversa:

$$e = \frac{1}{\varepsilon_{c,q}} = \frac{CMe}{CMg}$$

Relación entre rendimientos y economías de escala para funciones NO homotéticas Pero, fuera del caso homotético, la senda de expansión deja de ser un rayo desde el origen, de modo que crecer por ella y escalar todos los factores en la misma proporción son movimientos distintos. Aun así, en cada punto de la senda de expansión la elasticidad de escala coincide con el cociente CMe/CMg, de modo que los rendimientos a escala medidos en la combinación que minimiza costes y las economías de escala siguen yendo de la mano[30]. Para funciones homogéneas de grado HG:

$$\begin{aligned} HG = 1 &\Leftrightarrow \text{Rendimientos constantes} \Leftrightarrow \text{CMe constante} \\ HG > 1 &\Leftrightarrow \text{Rendimientos crecientes} \Leftrightarrow \text{Economías de escala} \\ HG < 1 &\Leftrightarrow \text{Rendimientos decrecientes} \Leftrightarrow \text{Deseconomías de escala} \end{aligned}$$

[29] Los rendimientos a escala se miden a través de la elasticidad a escala, e, que es tanto el grado de homogeneidad si la función es homogénea, como la suma de las elasticidades producto

[30] Además, no hay que confundir rendimientos a escala con los rendimientos marginales de los que se ha hablado en el corto plazo: rendimientos marginales = productividad marginal.

 © 2026 Manuel Cabado García

6.A.2. Funciones de costes resultantes de otras funciones de producción

A continuación se presentan las funciones típicas de coste en el caso de que prevalezcan en el largo plazo rendimientos constantes, crecientes o decrecientes, suponiendo constantes los precios de los factores productivos, ilustrando cada caso con una tecnología distinta.

- Rendimientos constantes de escala a largo plazo —tecnología Leontief, homogénea de grado 1 por construcción, $Q = \text{mín}\{aX_1, bX_2\}$—: la curva de C_{lp} es una línea recta que pasa por el origen: $C_{lp} = Q\left(\frac{w_1}{a} + \frac{w_2}{b}\right)$. Así, el CMe_{lp} al igual que el CMg_{lp} equivalen al coste unitario $\left(\frac{w_1}{a} + \frac{w_2}{b}\right)$, ambas horizontales e iguales entre sí.
- Rendimientos crecientes de escala a largo plazo —tecnología de sustitutos perfectos elevada a una potencia $k > 1$, $Q = (aX_1 + bX_2)^k$, homogénea de grado k y, por tanto, homotética (Varian, 1992)—: el núcleo $Z = aX_1 + bX_2$ sigue siendo de sustitutos perfectos, así que el coste de conseguir Z unidades es $Z \cdot \text{mín}\{w_1/a,\, w_2/b\}$. Sustituyendo $Z = Q^{1/k}$ se obtiene $C_{lp} = Q^{1/k}\,\text{mín}\{w_1/a,\, w_2/b\}$, con exponente $1/k < 1$: el CT_{lp} crece a una tasa decreciente y el CMe_{lp} tiene pendiente negativa. Por ello, el mínimo de los CMe_{cp} nunca coincide con el mínimo del CMe_{lp}, y en mercados competitivos las plantas operan a largo plazo en un nivel de producción inferior a aquel en el que su CMe_{cp} es mínimo.
- Rendimientos decrecientes de escala a largo plazo —tecnología Cobb-Douglas, $Q = AX_1^aX_2^b$ con $a + b < 1$—: la función de costes asociada es $C_{lp} = Ky^{1/(a+b)}$, con exponente $1/(a+b) > 1$, por lo que el CT_{lp} aumenta más que proporcionalmente respecto al nivel de producción. En este caso a la empresa le conviene arrendar una planta con capacidad menor, pues el mínimo a corto plazo se encuentra por debajo del punto de tangencia entre los CMe_{cp} y los CMe_{lp}.

En definitiva, cada tecnología deja una huella distinta en la curva de CMe_{lp} y en la de C_{lp} (Varian, 1992; Koutsoyiannis, 1979):

- Leontief (rendimientos constantes): CMe_{lp} constante —recta horizontal— y C_{lp} recta que pasa por el origen.
- Sustitutos perfectos elevados a $k > 1$ (rendimientos crecientes): CMe_{lp} decreciente y C_{lp} cóncava —crece a tasa decreciente—.
- Cobb-Douglas con $a + b < 1$ (rendimientos decrecientes): CMe_{lp} creciente y C_{lp} convexa —crece a tasa creciente—.

Nótese que las tres funciones comparten una misma lógica: el grado de homogeneidad de la tecnología (1, k o $a + b$, según el caso) determina directamente si C_{lp} crece proporcionalmente, a tasa decreciente o a tasa creciente respecto a Q—y con ello la pendiente

de CMe_{lp}. Sin embargo, ninguna de ellas genera por sí sola los rendimientos mixtos —economías de escala primero, deseconomías después— que dan lugar a la forma de U de la teoría tradicional (véase "Análisis de largo plazo"): al tener una elasticidad de escala constante en todo punto, solo pueden dar una curva monótona o plana. Los rendimientos mixtos surgen, bien como envolvente de curvas de corto plazo con economías de escala primero y deseconomías de gestión después (Koutsoyiannis, 1979), bien de formas funcionales más flexibles —como la función de costes translogarítmica— capaces de representar una elasticidad de escala que varía con el nivel de producción.

6.A.3. Economías de aglomeración

Las economías de aglomeración son los beneficios o reducciones de costes que las empresas y las personas obtienen al concentrarse geográficamente. Este concepto se enmarca dentro de la teoría de costes al explicar por qué la producción es más barata en áreas urbanas o "clusters" industriales, a pesar de los altos precios del suelo y la congestión.

Tipos de economías de aglomeración Existen dos tipos principales que contribuyen a la reducción de costes:

- Economías de localización: Se producen cuando empresas de un mismo sector se concentran en una misma zona. Los beneficios incluyen un mercado laboral especializado, acceso a proveedores específicos y la facilidad para compartir conocimiento e innovación (efectos de derrame tecnológico).
- Economías de urbanización: Ocurren cuando una gran variedad de industrias se concentran en un área metropolitana. Los beneficios provienen del acceso a una infraestructura compartida (carreteras, telecomunicaciones), una gran diversidad de servicios y un mercado laboral amplio y variado.

Costes de transporte y aglomeración La teoría de costes explica que las empresas se localizan en una aglomeración para minimizar los costes de transporte. Al ubicarse cerca de proveedores de materias primas y de un gran mercado de consumidores, las empresas reducen los gastos asociados al traslado de materias primas y productos finales.

La reducción o el bajo nivel de los costes de transporte es un factor clave que impulsa la concentración de la actividad económica. Sin embargo, la aglomeración también puede generar deseconomías de aglomeración, como la congestión del tráfico o la competencia por los recursos, que a su vez limitan el crecimiento de las ciudades.

 © 2026 Manuel Cabado García

6.A.4. Costes hundidos

Los costes hundidos (del inglés, sunk costs) son gastos que ya se han realizado y que no se pueden recuperar de ninguna manera. Son costes retrospectivos e irreversibles[31].

Relación con la teoría de costes La teoría económica de la decisión racional establece que los costes hundidos son completamente irrelevantes para la toma de decisiones futuras. Una decisión solo debe basarse en los costes y beneficios marginales futuros.

El problema surge con la falacia del coste hundido, un sesgo cognitivo que lleva a los individuos o a las empresas a continuar invirtiendo en un proyecto o actividad que no es viable, simplemente porque ya han gastado una cantidad significativa de dinero en él. Por ejemplo, una empresa que invierte millones en una nueva fábrica y, a mitad de la construcción, se da cuenta de que la demanda del producto ha desaparecido, debe abandonar el proyecto. El coste de la construcción ya realizada es un coste hundido y no debe influir en la decisión de continuar o no.

En la teoría de costes, la distinción clave es entre los costes hundidos y los costes de oportunidad. Los costes de oportunidad representan los beneficios perdidos por elegir una alternativa en lugar de otra y son esenciales para una decisión racional, mientras que los costes hundidos, al ser irrecuperables, no lo son.

Implicaciones de política económica El concepto de costes hundidos tiene varias implicaciones importantes para la política y la economía:

- Barreras de entrada: Los grandes costes hundidos pueden actuar como una barrera de entrada al mercado. Si la inversión inicial requerida para entrar en una industria es muy alta y no es recuperable (por ejemplo, la inversión en publicidad de una marca conocida o en I+D), desanimará a los nuevos competidores, lo que puede llevar a una estructura de mercado de monopolio u oligopolio.
- Proyectos públicos y la falacia del coste hundido: La falacia del coste hundido es particularmente relevante en el sector público. Los gobiernos a menudo se sienten presionados a continuar financiando proyectos de infraestructura (como aeropuertos o trenes de alta velocidad) que ya han costado miles de millones, incluso cuando los análisis de coste-beneficio sugieren que el proyecto es un fracaso y que la mejor opción sería abandonarlo. La política económica racional dicta que, si los beneficios futuros esperados de un proyecto son inferiores a los costes futuros esperados, el proyecto debe ser detenido, independientemente de la inversión ya realizada.

[31] Tirole (1988) precisó que la distinción entre coste fijo y coste hundido es de grado, no de naturaleza: un coste fijo es hundido solo durante el periodo de compromiso (por ejemplo, la duración de un contrato de alquiler), lo que conecta directamente con la barrera de entrada que generan los costes hundidos en la teoría de la contestabilidad.

- Regulación: Las autoridades de competencia y los reguladores tienen en cuenta los costes hundidos al evaluar la dinámica de un mercado. Por ejemplo, al considerar la concesión de licencias o la asignación de espectro radioeléctrico, se puede buscar que el coste hundido sea el adecuado para atraer a competidores sin que sea un freno para la entrada de nuevas empresas.

6.A.5. Tipos de economía de escala

Las economías de escala son ventajas que obtiene una empresa al aumentar su volumen de producción, lo que permite reducir el coste medio por unidad. Estas ventajas pueden surgir tanto de factores internos como externos, y se clasifican en varios tipos según su origen y naturaleza.

Economías de escala internas Son aquellas que se generan dentro de la empresa como resultado de su crecimiento:

- Técnicas: Al aumentar la producción, se pueden utilizar máquinas más eficientes o automatizar procesos. Por ejemplo, una fábrica que produce más unidades puede justificar la inversión en tecnología avanzada que reduce el coste por unidad.
- Gerenciales: La especialización de funciones permite una gestión más eficiente. A medida que la empresa crece, puede contratar personal especializado en áreas como logística, finanzas o recursos humanos, lo que mejora la productividad.
- Financieras: Las empresas grandes suelen tener acceso a mejores condiciones de financiación, como tipos de interés más bajos o mayor capacidad de negociación con inversores.
- Comerciales: Al comprar en grandes volúmenes, las empresas pueden obtener descuentos por parte de proveedores. Además, pueden negociar mejores condiciones de distribución o publicidad.
- Administrativas: Algunos costes fijos, como los de oficina o software, se reparten entre más unidades producidas, reduciendo el coste por unidad.

Economías de escala externas Estas provienen del entorno o del crecimiento del sector en el que opera la empresa:

- De concentración geográfica: Cuando varias empresas del mismo sector se agrupan en una región, se benefician de infraestructura compartida, proveedores especializados y un mercado laboral adaptado. Ejemplo: el distrito tecnológico de Silicon Valley.
- Del mercado laboral: En zonas con alta concentración de una industria, es más fácil encontrar trabajadores cualificados, lo que reduce costes de formación y mejora la eficiencia.

 © 2026 Manuel Cabado García

Conviene no confundirlas con los efectos de red, en los que el valor de un producto para cada usuario aumenta con el número de usuarios (redes sociales o plataformas de comercio electrónico). Se trata de economías del lado de la demanda, no de reducciones de costes.

Otras clasificaciones

- Reales vs. pecuniarias:
  - Reales: Reducción de costes físicos, como el uso más eficiente de maquinaria.
  - Pecuniarias: Reducción de costes por obtener mejores precios en materias primas o servicios.
- Estáticas vs. dinámicas:
  - Estáticas: Derivan del tamaño actual de la empresa.
  - Dinámicas: Surgen del aprendizaje acumulado y la mejora continua en procesos.

Ejemplos prácticos

- Una empresa automovilística que produce 100.000 vehículos al año puede negociar precios más bajos por neumáticos que una que produce solo 5.000.
- Una cadena de supermercados puede distribuir sus costes de publicidad entre cientos de tiendas, mientras que una tienda independiente debe asumirlos sola.
- En el sector tecnológico, plataformas como Amazon o Google se benefician de economías de red, donde cada nuevo usuario aporta valor adicional al sistema.

6.A.6. Aplicaciones empíricas: el análisis envolvente de datos (DEA) y la frontera estocástica

Todo lo anterior se ha desarrollado en términos puramente teóricos, a partir de una función de producción supuesta de antemano. Cabe preguntarse cómo se traslada este aparato a la práctica: ¿cómo puede un regulador o un investigador medir, con datos reales, si una empresa concreta está siendo eficiente, sin necesidad de conocer directamente su función de producción o de costes? La literatura utiliza esencialmente dos estrategias:

- El análisis de frontera estocástica (*Stochastic Frontier Analysis*, SFA): estima los parámetros de una función de costes de forma paramétrica –eligiendo previamente una forma funcional concreta– y descompone, mediante técnicas econométricas, la distancia de cada empresa a la frontera eficiente en un componente aleatorio y un componente de ineficiencia.[32] Requiere una notable disponibilidad de datos.

[32] La intuición es que un mal año puede deberse a una helada y no a una mala gestión. El método intenta separar la mala suerte de la ineficiencia.

- El Análisis Envolvente de Datos (*Data Envelopment Analysis*, DEA): una estrategia no paramétrica y más flexible, que consiste en "envolver" las observaciones disponibles mediante la resolución de un problema de optimización lineal, sin necesidad de imponer a priori ninguna forma funcional a la tecnología (Coelli, 1996).

Dada su mayor flexibilidad y menor exigencia de información, el DEA es el método más utilizado en la práctica. Considérese, por simplicidad, un caso con dos factores (x_1, x_2) y un único output Y, de modo que cada punto de un plano $(x_1/Y, x_2/Y)$ representa a una empresa observada. El DEA plantea un problema de optimización lineal para trazar una frontera eficiente que maximice la eficiencia de cada empresa observada y delimite un conjunto convexo que contenga a todas las observaciones.

Las empresas que no se sitúan sobre la frontera son, por definición, técnicamente ineficientes. Considérese una empresa A situada fuera de la frontera: trazando una recta desde el origen hasta A, su intersección con la frontera eficiente define un punto b, y la intersección de esa misma recta con la curva de isocoste define un punto c. A partir de estas distancias se obtienen tres indicadores, acotados entre 0 (máxima ineficiencia) y 1 (máxima eficiencia):

- Eficiencia técnica= Ob/OA: en qué medida la empresa podría reducir sus inputs manteniendo el mismo output.
- Eficiencia asignativa= Oc/Ob: si, conocidos los precios de los factores, la combinación empleada es la óptima.
- Eficiencia económica= eficiencia técnica $\times$ eficiencia asignativa.

El DEA admite además opciones metodológicas relevantes: puede orientarse a minimizar inputs dado el output, o a maximizar el output dados los inputs (equivalentes solo bajo rendimientos constantes a escala). Puede suponer rendimientos constantes o variables a escala y, con datos de panel, puede calcularse un índice de Malmquist de productividad total de los factores para el análisis dinámico[33].

En aplicaciones reales, el DEA se ha utilizado extensamente para medir la eficiencia de operadores de transporte en autobús:Venkatesh y Kushwaha (2018) estudian la eficiencia económica de empresas de autobuses en la India, encontrando ineficiencias significativas. Odeck (2006) estima el impacto de la congestión, el tipo de propiedad, la escala y la localización sobre la eficiencia de operadores noruegos, y Karlaftis (2004) evalúa operadores urbanos en Estados Unidos, encontrando que la mayoría opera bajo rendimientos decrecientes a escala.

[33]El índice de Malmquist compara la eficiencia de una misma empresa entre dos periodos calculando la distancia de cada observación a las fronteras eficientes estimadas en ambos momentos. Su valor se descompone en dos factores: el cambio técnico (desplazamiento de la propia frontera eficiente, por ejemplo, por innovación tecnológica) y el cambio en la eficiencia relativa de la empresa (su acercamiento o alejamiento de la frontera vigente en cada periodo), que a su vez puede subdividirse en cambio en la eficiencia técnica pura y cambio en la eficiencia de escala. Un valor superior a 1 indica ganancia de productividad total de los factores entre ambos periodos.

 © 2026 Manuel Cabado García

6.A.7. La función de costes translogarítmica

Como ya se ha señalado, las funciones de coste vistas hasta ahora —Leontief, sustitutos perfectos, Cobb-Douglas— imponen de antemano una elasticidad de sustitución fija entre factores y una elasticidad de escala constante en todo punto. La función de costes translogarítmica (*transcendental logarithmic*), propuesta por (Christensen et al., 1973), es una forma funcional flexible que no impone ninguna de las dos restricciones a priori: es una aproximación de segundo orden al logaritmo de la función de costes,

$$\ln C(w,y) = \alpha_0 + \sum_i \alpha_i \ln w_i + \alpha_y \ln y + \frac{1}{2}\beta_{yy}(\ln y)^2 + \sum_i \beta_{iy} \ln w_i \ln y + \frac{1}{2}\sum_i \sum_j \beta_{ij} \ln w_i \ln w_j$$

sujeta a las restricciones de simetría $\beta_{ij} = \beta_{ji}$ y de homogeneidad de grado 1 en precios, $\sum_i \alpha_i = 1$, $\sum_j \beta_{ij} = 0 \, \forall i$, $\sum_i \beta_{iy} = 0$ (Varian, 1992).

Por el Lema de Shephard, las participaciones de cada factor en el coste total son lineales en los parámetros,

$$s_i(w,y) = \alpha_i + \sum_j \beta_{ij} \ln w_j + \beta_{iy} \ln y,$$

lo que permite estimar el sistema completo (función de costes + ecuaciones de participación) econométricamente sin imponer la forma de la tecnología subyacente.

Caso particular Si $\beta_{ij} = 0$ para todo i, j, $\beta_{iy} = 0$ y $\beta_{yy} = 0$, la translog se reduce exactamente a una Cobb-Douglas —es la única de las tecnologías vistas en 6.A.2 que queda anidada de forma exacta—. Leontief y sustitutos perfectos no lo están: por su carácter de aproximación local de segundo orden a una tecnología arbitraria, la translog puede aproximar su comportamiento en el entorno de un punto dado, pero no los reproduce de forma exacta y global (para eso existe una forma funcional flexible distinta, la Leontief generalizada de Diewert).

Por qué sí genera rendimientos mixtos A diferencia del caso restringido anterior, cuando β_{yy} y β_{iy} son distintos de cero la elasticidad de escala deja de ser constante y pasa a depender del nivel de producción —precisamente la propiedad que Leontief, Cobb-Douglas y sustitutos perfectos no podían ofrecer en 6.A.2—, permitiendo generar de forma nativa una curva de CMe_{lp} en U sin necesidad de recurrir a la envolvente de curvas de corto plazo.

Aplicación empírica Caves et al. (1984) la emplearon para separar economías de densidad de economías de escala en el sector aéreo estadounidense, encontrando economías

de densidad sustanciales —clave para explicar la eficiencia de las redes en estrella (*hub-and-spoke*)— frente a economías de escala mucho más limitadas a nivel de red completa.

 © 2026 Manuel Cabado García

Bibliografía

Caves, D. W., L. R. Christensen y M. W. Tretheway (1984). Economies of density versus economies of scale: Why trunk and local service airline costs differ. *The RAND Journal of Economics 15*(4), 471–489.

Christensen, L. R., D. W. Jorgenson y L. J. Lau (1973). Transcendental logarithmic production frontiers. *The Review of Economics and Statistics 55*(1), 28–45.

Coelli, T. J. (1996). A guide to DEAP version 2.1: A data envelopment analysis (computer) program. CEPA Working Paper 96/08, Centre for Efficiency and Productivity Analysis (CEPA), University of New England.

Karlaftis, M. G. (2004). A DEA approach for evaluating the efficiency and effectiveness of urban transit systems. *European Journal of Operational Research 152*, 354–364.

Koutsoyiannis, A. (1979). *Modern Microeconomics* (2nd ed.). London: Macmillan.

Leibenstein, H. (1966). Allocative Efficiency vs. "X-Efficiency". *The American Economic Review 56*(3), 392–415.

Nicholson, W. (2005). *Teoría microeconómica. Principios básicos y ampliaciones: principios básicos y ampliaciones*. Editorial Paraninfo.

Odeck, J. (2006). Congestion, ownership, region of operation, and scale: Their impact on bus operator performance in Norway. *Socio-Economic Planning Sciences 40*(1), 52–69.

Pindyck, R., D. Rubinfeld y E. Rabasco (2013). *Microeconomia*. Pearson Italia.

Samuelson, P. A. (1947). *Foundations of Economic Analysis*. Cambridge: Harvard University Press.

Stigler, G. (1939). Production and Distribution in the Short Run. *Journal of Political Economy 47*(3), 305–327.

Stigler, G. J. (1976). The Xistence of X-efficiency. *American Economic Review 66*(1), 213–216.

Varian, H. R. (1992). *Microeconomic Analysis* (3rd ed.). New York: W. W. Norton & Company.

Varian, H. R. (2010). *Microeconomía intermedia: un enfoque actual*. núm. HB171. V37 1996.

Venkatesh, A. y S. Kushwaha (2018). Short and long-run cost efficiency in Indian public bus companies using data envelopment analysis. *Socio-Economic Planning Sciences 61*, 29–36.

 © 2026 Manuel Cabado García

Tema 7

Análisis de mercados (I). El modelo de competencia perfecta. Análisis de equilibrio parcial: corto y largo plazo, dinámicas de ajuste y estabilidad del equilibrio. Análisis de eficiencia y bienestar.

7.1. Introducción ⏳ 1'

En las últimas dos décadas, la concentración empresarial ha aumentado significativamente, con mercados dominados por grandes empresas, llamadas "superestrellas". La globalización y la digitalización han impulsado este fenómeno, con empresas como Amazon y Google aprovechando las economías de escala y los efectos de red[1]. Así, estos factores crean altas barreras de entrada y pueden conducir a mercados donde el ganador se lleva todo.

Esto genera un debate entre economistas sobre si la competencia ha disminuido: mientras algunos argumentan que la menor competencia reduce la innovación, otros sostienen que la concentración es el resultado de una intensa competencia que culmina con unas pocas empresas en la cima.

En el marco de la Teoría Económica, a la hora de analizar las imperfecciones de mercado se toma como punto de referencia el modelo de competencia perfecta en un

[1]En economía, un efecto de red (también llamado externalidad de red o economías de escala del lado de la demanda) es el fenómeno por el cual el valor o la utilidad que un usuario obtiene de un bien o servicio depende del número de usuarios de productos compatibles. Los efectos de red suelen ser sistemas de retroalimentación positiva, por lo que los usuarios obtienen cada vez más valor de un producto a medida que más usuarios se unen a la misma red.

contexto de equilibrio parcial (ceteris paribus)[2].

El modelo lo desarrolla Marshall (1890) a finales del siglo XIX y es de gran importancia tanto desde el punto de vista práctico, ya que se aproxima a la realidad de algunos mercados –como el de las materias primas y los de valores–, como desde el teórico, ya que es el modelo de referencia para comparar con otras estructuras, siendo clave para las autoridades de la competencia.

Por consiguiente, el objetivo de esta exposición será el análisis de este modelo a través del siguiente esquema:

1. Caracterización del modelo de competencia perfecta
2. Equilibrio de la empresa y de la industria.
3. Equilibrio competitivo.
4. Estática comparativa.
5. Relajación de supuestos.

7.2. Caracterización del modelo de competencia perfecta ⏳ 2'

Los supuestos normalmente asociados a la competencia perfecta en el mercado de bienes son:

1. Muchos compradores y muchos vendedores, es decir, atomizado (o pequeño tamaño de la empresa en relación al mercado), lo que implica que todos actúan como precio-aceptantes.
2. Producto homogéneo, por lo que el único factor relevante para la elección por parte de los consumidores será el precio.
3. Información perfecta[3], es decir, todos los agentes que participan en el mercado disponen de toda la información relevante.

[2]Equilibrio parcial: i) El mercado de un bien constituye una pequeña parte de toda la economía, ii) la oferta y demanda de cada bien depende solo del precio del propio bien. Los efectos renta y sustitución cruzados con los demás mercados son despreciables. El gasto en los otros bienes se puede tratar como uno solo, el numerario. La manera más simple de modelar estos supuestos es asumir que en la economía existen dos bienes, y el gasto en todas las mercancías distintas a la que resulta de interés se trata como una única mercancía compuesta (numerario) y las preferencias de los consumidores son cuasilineales respecto al numerario.

El enfoque de equilibrio parcial también es válido para estudiar mercados con varios productos muy similares, como es el caso de los modelos de diferenciación de producto y competencia monopolística.

[3]De los supuestos de producto homogéneo e información perfecta se deriva la Ley de Precio único. La ley del precio único en economía afirma que, en ausencia de barreras comerciales y en una economía con precios que pueden variar libremente en función de la demanda y la oferta (modelo elástico), si se comparan los precios de un mismo bien de dos países diferentes en términos de una misma moneda, entonces esos precios, en equilibrio, tendrán que ser iguales.

 © 2026 Manuel Cabado García

4. Todas las empresas tienen la misma tecnología a largo plazo, igualdad de acceso a factores productivos, y libertad de entrada y de salida[4].
5. Rendimientos mixtos, lo que genera una curva de costes medios en forma de U.
6. Perfecta divisibilidad de bienes y de factores productivos, lo que implica que las funciones de producción y de costes son continuas.
7. Perfecta movilidad de factores productivos, cuyo mercado también se encuentra en competencia perfecta.
8. Costes constantes, es decir, que el precio de los factores productivos se mantiene constante para cualquier nivel de demanda de factor[5].
9. Ausencia de externalidades, de bienes públicos y de intervención del sector público (no hay fallos de mercado)
10. Empresas racionales. Buscan la maximización de beneficios.

Se trata de supuestos bastante restrictivos. Al final de la exposición se levantarán los supuestos de rendimientos mixtos, y el de precio constante de los factores productivos.

Equilibrio de competencia perfecta Así, bajo estos supuestos el equilibrio de competencia perfecta se define como una situación en la que existe un precio de un bien y un nivel de producción tal que los mercados se vacían –Cantidad demandada = Cantidad ofertada–, los consumidores maximizan su utilidad y los productores su beneficio dada una tecnología y precio de los factores de producción.

El papel de los precios A su vez, el precio cumple un papel fundamental en el mercado

- Transmitiendo toda la información relevante
- Mostrando el valor marginal del bien
- Racionando, es decir, son indicadores de escasez.

[4]En el corto plazo puede haber empresas con diferente tecnología, sin embargo en el largo plazo será igual para todas y serán aquellas que sobrevivan en el mercado.

[5]Es decir, se supone que la curva de oferta agregada del factor productivo es totalmente elástica. Se podría decir que se asumen economías de escala externas constantes: los costes de la empresa no varían cuando aumenta la producción de la industria (es decir, los precios de los factores productivos no aumentan cuando la industria, para aumentar su producción, demanda más factores productivos).

Nótese que este supuesto de "costes constantes" no quiere decir "costes marginales constantes". Lo que establece el supuesto es que el precio de los factores productivos se mantiene constante, pero el coste marginal está compuesto por algo más que el precio de los factores: también está afectado por la productividad marginal del factor:

$$CMg = \frac{\bar{W}_i}{PMg_{x_i}}$$

y si hubiese dos factores productivos (por ejemplo, trabajo y capital), entonces el coste marginal sería:

$$CMg = \frac{W}{PMg_L} + \frac{R}{PMg_K}$$

Oferta de las empresas y de la industria Para determinar el equilibrio será preciso determinar la oferta de las empresas individuales y, a partir de ellas, la de la industria. La industria se define como la agregación de todas las empresas que elaboran el mismo bien y la suma de las ofertas de todas las empresas constituye la oferta de la industria

$$S = \sum_{i=1}^{n} s_i \tag{7.1}$$

7.3. Equilibrio de la empresa y de la industria ⏳ 9'

Con todo lo anterior se puede estudiar el equilibrio de la industria y de la empresa en diferentes horizontes temporales. Pero antes se describirán brevemente la economía y la forma de obtener la demanda.

Descripción de la economía

Se considera una economía formada por un conjunto de consumidores (I), de empresas (J) y de mercancías (L).

- Las preferencias del consumidor i sobre las cestas de consumo $x_i = (x_{1i}, \ldots, x_{Li})$ pueden ser representadas por una función de utilidad de buen comportamiento.
- Los agentes disponen inicialmente de una dotación w_{mi} del bien numerario[6] que las empresas podrán tomar y transformar en otros bienes.
- Cada empresa j dispone de un conjunto de posibilidades de producción $Y_j \in \mathbb{R}^{LJ}$. Donde cada elemento de Y_j es un vector de producción $y_i = (y_{1i}, \ldots, y_{Li}) \in \mathbb{R}^L$, donde los elementos con signo negativo representan inputs y los elementos con signo positivo representan el output. Un plan de producción es factible si puede obtenerse transformando las dotaciones iniciales a través del uso de la tecnología.
- El precio del bien numerario queda normalizado a 1. El precio del bien de mercado analizado es p.

Obtención de la curva demanda

El consumidor tiene unas preferencias sobre el bien l y el numerario representables por una función de utilidad cuasilineal respecto del dinero (FU de Marshall cuasilineal). Esta función presenta ER y ES cruzados nulos.

El consumidor maximiza su utilidad sujeto a una restricción presupuestaria. En el punto óptimo la restricción de la renta se satura y la UMg del consumo de un bien se

[6] El bien numerario es el bien cuyo precio se normaliza a 1 para expresar en sus unidades los precios de todos los demás. En el modelo de equilibrio parcial representa el gasto agregado en todos los bienes distintos del analizado.

 © 2026 Manuel Cabado García

iguala al CMg de adquirirlo (su precio) multiplicado por el multiplicador de Lagrange, que representa la UMg de la renta. Esto es, se cumple la segunda Ley de Gossen.

A partir de la CPO se obtiene la cantidad óptima de x_i a consumir. Esta puede ser expresada en función de los parámetros de los que depende (en este caso solo de su propio precio), obteniendo la demanda marshalliana del bien $x_i(p_i)$, que es decreciente y positiva. Esta demanda será independiente de w_{mi} debido al carácter cuasilineal de las preferencias.

Curva de demanda agregada Debido al carácter cuasi-lineal de las preferencias será posible obtener una función de demanda agregada

$$x\left(p\right)=\sum_{i}x_{i}\left(p\right) \tag{7.2}$$

a través de la suma horizontal de las diferentes curvas de demanda individuales, debido a la inexistencia de efectos renta. Por tanto, mantendrá las propiedades de las funciones individuales. Esta función de demanda agregada será continua y no creciente para todo $p>0$.

Función inversa de la demanda de mercado La función inversa de la demanda de mercado se obtiene despejando el precio en la función de demanda agregada, $x\left(p\right)$: en lugar de indicar qué cantidad se demanda a cada precio, indica qué precio corresponde a cada cantidad. Se escribe así:

$$P(x)=x^{-1}(x) \tag{7.3}$$

donde el exponente -1 no es una potencia, sino que denota la función inversa de la demanda $x\left(p\right)$, y la x entre paréntesis es una cantidad concreta del bien. Su interpretación económica es interesante: indica el precio resultante de una demanda agregada de x. Es decir, a este precio los consumidores demandan una cantidad del bien l tal que la demanda agregada es x. Es especialmente relevante que, a nivel individual, el beneficio marginal en términos del numerario se iguala al precio, por lo que el valor de esta función puede ser interpretado como el beneficio marginal social del bien l[7].

A continuación, se pasa a analizar el equilibrio dependiendo del horizonte temporal. La curva de oferta se derivará conforme se vaya resolviendo el problema del productor. La diferencia entre el análisis a cp y a lp está en la capacidad de las empresas de decidir sobre la utilización de todos los ffpp. Los mayores grados de libertad que permite el lp provocan que la curva de costes a lp sea siempre inferior o igual a la de cp.

[7]La eficiencia en la distribución del consumo del bien está asegurada por el hecho de que el beneficio marginal del consumo del bien no numerario es igual para todos los consumidores.

Equilibrio a muy corto plazo

En el muy corto plazo o "periodo de mercado":

1. La dotación de factores es fija, por lo que la producción está dada, es decir, la oferta es totalmente inelástica.
2. No existe libertad de entrada ni salida, por lo que el número de empresas es fijo.

En esta situación, el precio no es más que un mecanismo para racionar la demanda, ya que distribuye la cantidad producida entre los consumidores. En consecuencia, la demanda determina por sí sola el precio, y la oferta, la cantidad de equilibrio de mercado.

Análisis gráfico Gráficamente, en dos planos de precios y cantidad como el de la Figura 7.1 se representa la curva de demanda y oferta de la industria y la empresa:

- En el caso de la industria, la demanda tiene pendiente negativa porque la cantidad demandada aumenta conforme disminuye el precio mientras que la oferta se obtiene mediante la agregación horizontal de las ofertas individuales[8].
- Para la empresa individual, considera la demanda a la que se enfrenta como una línea horizontal pues es precio-aceptante –forma que puede vender toda la cantidad que desee al precio de mercado y nada a un precio superior[9]– y la oferta es fija por ser la dotación de todos los factores fija.

Ejemplos de estos mercados son la lonja de pescado, energía solar, mercado de fruta o productos agrícolas en los que lo único que puede pasar es que baje el precio si baja la demanda.

Equilibrio a corto plazo

El corto plazo es el horizonte temporal en el que:

1. La dotación de algún factor productivo es fija, pero las dotaciones del resto son variables. La cantidad producida ya no es fija –puede variar aumentando o disminuyendo el empleo de factores variables– y se distinguirá entre costes fijos y variables
2. No existe libertad de entrada y de salida, por lo que el número de empresas es fijo y pueden convivir empresas con beneficios y pérdidas.

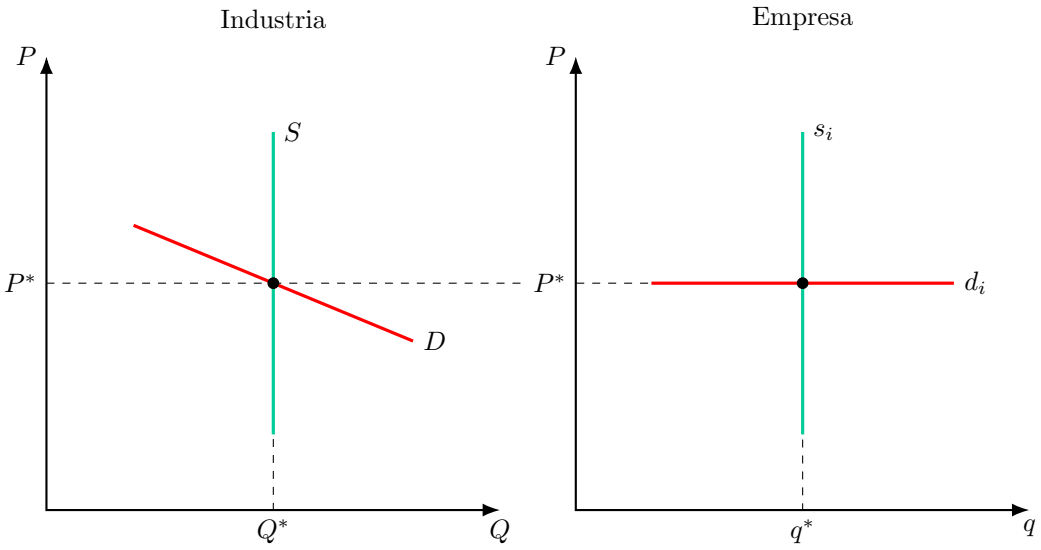
[8] La agregación horizontal de las ofertas individuales solo es un método válido si se asume que, al aumentar la producción, los precios de los factores productivos que emplean las empresas no varían (es decir, si se supone que la oferta de factores es totalmente elástica). Si estos precios aumentaran, no solo lo harían para la última unidad sino para todas las unidades, desplazando la curva de oferta hacia la izquierda e imposibilitando la agregación.

Además el coste de ajuste de todos los factores es infinito, es decir, no se puede variar el nivel de producción que se oferta al mercado variando la cantidad empleada de los factores.

[9] Sí que podría vender a un precio inferior, pero no le interesará porque, como se detallará más adelante, estaría vendiendo por debajo de los costes, lo que le reportaría pérdidas.

 © 2026 Manuel Cabado García

Figura 7.1: Equilibrio competitivo a muy corto plazo



El objetivo de la empresa es el de maximizar beneficios, es decir, la diferencia entre ingresos –valor de mercado de la cantidad vendida– y costes totales –suma de costes fijos y variables–:

$$\text{máx} \quad \pi = P \cdot q - CT \tag{7.4}$$

donde $P(Q)$ es la curva de demanda de la empresa (constante por ser precio aceptante), que, como se verá en la condición de primer orden, es totalmente horizontal (es decir, elasticidad infinita). Es decir, en competencia perfecta, el precio es una variable exógena (mientras que en monopolio es una variable endógena).

Cuya CPO indica que el precio es igual al coste marginal.

$$\frac{\partial \pi}{\partial q} = 0 \Rightarrow P = CMg = IMg \tag{7.5}$$

de esta forma, si $P > CMg$, la empresa puede aumentar su producción en 1 unidad y aumentar su beneficio en P–CMg. Si $P < CMg$, la empresa puede disminuir su producción en 1 unidad y aumentar su beneficio en CMg–P.

Su CSO impone que se maximizarán beneficios en la parte creciente de la curva de costes marginales, CMg.

$$\frac{\partial^2 \pi}{\partial^2 q} < 0 \Rightarrow \frac{\partial CMg}{\partial q} > 0 \tag{7.6}$$

Así, la CSO impone que el CMg sea creciente (es decir, función de producción estrictamente cóncava –rendimientos decrecientes–). Esta condición no es baladí, porque si la producción presenta rendimientos mixtos, la CPO se puede dar en dos puntos: con CMg decreciente y con CMg creciente, como ahora se verá. La CSO debe cumplirse para que el beneficio obtenido sea un máximo y no un mínimo.

© 2026 Manuel Cabado García

Análisis gráfico Gráficamente, el equilibrio se puede analizar desde un enfoque de ingresos y costes totales y un enfoque marginal.

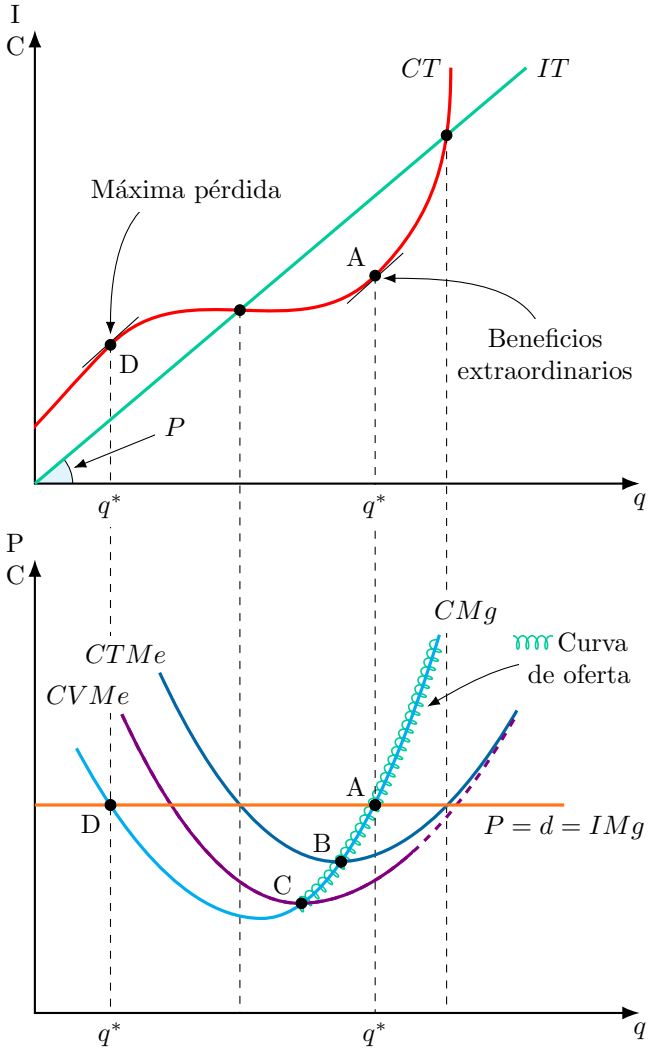
- Ingresos y costes totales: en un plano de ingresos, I,costes, C, y cantidad producida por la empresa, q, como el de la Figura 7.2:
  - Se representa la curva de ingresos totales, IT, y la curva de costes totales, CT, que es siempre creciente, primero a ritmo decreciente y después, por la ley de rendimientos marginales decrecientes, a ritmo creciente.
  - Como el objetivo del empresario es la maximización del beneficio, esto lo logrará en aquel punto de producción en el que la distancia entre los ingresos y los costes es máxima en A, obteniendo beneficios extraordinarios.
  - En los puntos donde ingresos y costes totales se cortan, la empresa obtendría beneficios normales siendo los puntos que están fuera del intervalo entre ellos aquellos donde la empresa conseguiría pérdidas, siendo el punto D aquel donde sería máxima la pérdida.
- Enfoque marginal: desde un enfoque marginal, en un plano adyacente de precios, P, costes, C, y cantidad producida por la empresa, q, se trasladan los puntos obtenidos.:
  - A partir de la función de costes totales y teniendo en cuenta que el coste total medio, $CTMe$, se deriva a partir del valor de la pendiente de los radio vectores a sus distintos puntos, se considerará el volumen de producción donde el coste total medio es mínimo, B.
  - Además, si se considera sobre dicha función el valor marginal, que se estima por el valor de la pendiente de la tangente en sus diferentes puntos y aquel volumen de producción donde el coste marginal es mínimo, se trazará la curva de coste marginal, CMg, que atravesará al coste total medio en su punto mínimo B. A su vez, atravesará a la curva de costes medios variables –la cual se obtiene restando los costes fijos a los totales– en el punto C.
  - Por otro lado, en el mercado competitivo al ser la empresa precio-aceptante, se enfrentará a una curva de precio, que es a su vez la demanda percibida por el empresario, d, su curva de ingreso medio, IMe, y marginal, IMg. Esta será completamente horizontal, estando determinado su valor por la pendiente de la función de ingresos totales, IT, de la empresa.

Umbrales de rentabilidad y cierre A continuación, se comentarán una serie de puntos muy relevantes para el análisis:

- En el punto A los beneficios son extraordinarios, dado que están por encima de los $CTMe$.

 © 2026 Manuel Cabado García

Figura 7.2: Equilibrio competitivo de la empresa a corto plazo



- En el punto B, mínimo del coste total medio, los beneficios son nulos, por lo que se le llama umbral de beneficios u óptimo de explotación. Por encima de ese precio la empresa obtiene beneficios extraordinarios.
- En el tramo del coste marginal de B a C, la empresa tendrá beneficio negativo. No obstante, la empresa cubre el coste medio variable y parte de los costes fijos, por lo que seguirá produciendo cantidades positivas.
- En el punto C, sin embargo, la empresa solo cubre los costes variables pero no los fijos siendo este el punto mínimo de explotación en el que la empresa es indiferente entre cerrar y producir.
- En el punto D el precio también iguala al coste marginal, pero en su tramo decreciente, por lo que no se cumple la condición de segundo orden y la pérdida es máxima. La empresa nunca elegirá producir en ese punto.

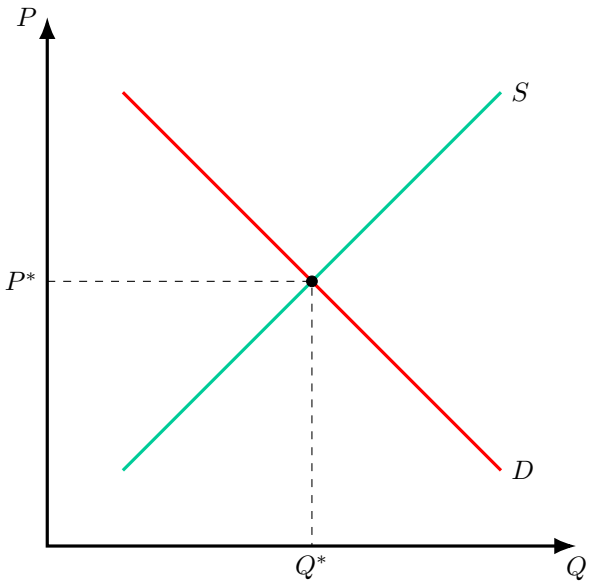
Curva de oferta de la empresa Así, la curva de oferta de la empresa a corto plazo se obtiene por estática comparativa a partir de los sucesivos puntos de intersección entre su curva de CMg y sus sucesivas curvas de demanda, horizontales al nivel de cada precio de mercado. Además, la curva de oferta coincide con el tramo creciente de la curva de costes marginales a partir del mínimo de explotación C. Para precios inferiores la empresa cierra y ofrece una cantidad nula.

Curva de oferta de la industria Por otro lado, la curva de oferta de la industria a corto plazo es la suma horizontal de las curvas de oferta individuales, es decir, la suma de las cantidades que cada empresa desea vender a cada precio, siempre que el precio de los factores no varíe con la producción de la industria[10].

Equilibrio Para determinar el equilibrio se introducen en el análisis los consumidores que son iguales e idénticos por lo que la curva de demanda del mercado es la suma horizontal de las demandas individuales.

Gráficamente, en un plano de precios y cantidades como el de la Figura 7.3, el equilibrio en el corto plazo se produce en la intersección entre oferta y demanda, que será compatible con la existencia de beneficios o pérdidas a corto plazo. Además, el mercado se vacía y el precio transmite toda la información relevante sobre el bien.

Figura 7.3: Equilibrio competitivo a corto plazo



[10]La oferta de mercado es de nuevo la suma horizontal de las ofertas a c/p de las empresas del mercado: es la suma de las cantidades ofrecidas por cada empresario para cada precio, para lo cual es necesario que la oferta de factores sea totalmente elástica (si aumenta la demanda de un determinado factor no varía su precio). Tiene pendiente positiva. Si varían los precios de los factores o la tecnología, se desplaza la curva de oferta a una nueva posición.

 © 2026 Manuel Cabado García

Conclusiones En resumen:

- En el corto plazo pueden convivir empresas que obtengan beneficios extraordinarios (beneficio económico superior a 0) o pérdidas.
- En el equilibrio los consumidores maximizan su utilidad, las empresas maximizan beneficios y el mercado se vacía (esto es las decisiones de consumidores y empresas son compatibles). Consecuencia del equilibrio competitivo es que la restricción presupuestaria se cumple con igualdad[11].
- Las condiciones de viabilidad económica (es decir, las condiciones para que la empresa prefiera seguir produciendo a cerrar) en el corto plazo son:
  - $IT \geq CV$,
  - $P \geq Mín\ CVMe$.
  - Que, si pierde dinero, pierda menos que los costes fijos.

Equilibrio a largo plazo

Por otro lado, en el largo plazo:

1. La dotación de todos los factores es variable, es decir, la cantidad producida puede variar aumentando o disminuyendo el empleo de cualquier factor y no habrá distinción entre costes fijos y costes variables. La empresa podrá modificar el tamaño de planta.
2. Existe libertad de entrada y salida de empresas, lo que implica que el número de empresas ya no es fijo y que no obtendrán ni beneficios ni pérdidas, es decir, los beneficios económicos son nulos.

La dotación variable de todos los factores implica que la combinación de factores productivos elegidos es técnica y económicamente eficiente por parte de la empresa, lo que hará que operen en el tamaño óptimo de planta

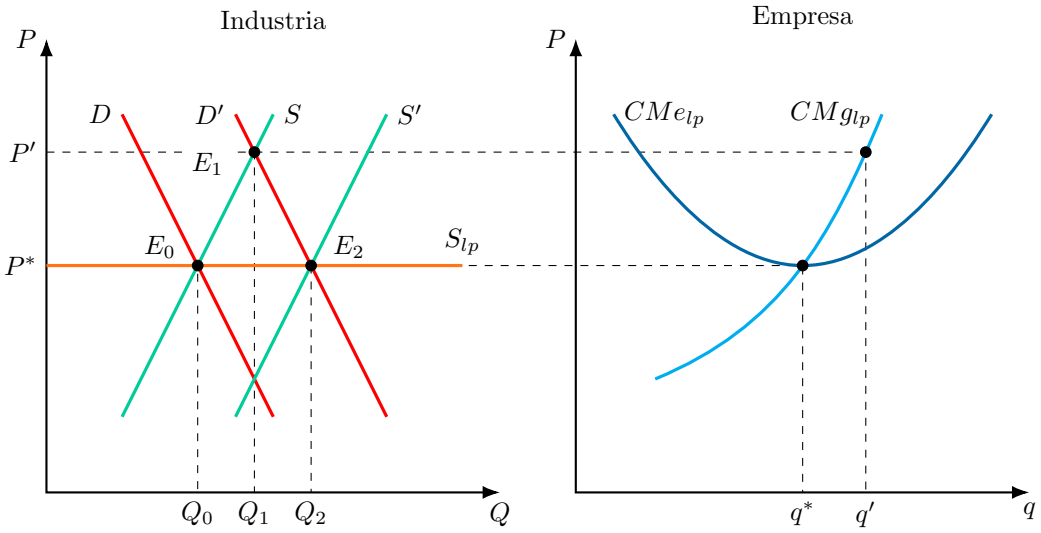
Análisis gráfico Gráficamente en un plano de precios y cantidades como el de la Figura 7.4:

- Dada la dotación variable de todos los factores, la empresa operará con el tamaño óptimo de planta, es decir, en el mínimo de los costes medios a largo plazo, CMe_{lp}, donde estos se igualan a los costes marginales a largo plazo, CMg_{lp}, y ambos coinciden con sus respectivas curvas a corto plazo, todo ello al precio P^*.

[11] Dadas las interpretaciones previas de la curva inversa de demanda agregada, así como de la curva inversa de oferta agregada, el equilibrio parcial competitivo puede ser interpretado como aquel nivel de producción donde el beneficio marginal social del bien l es igual a su coste marginal.

- Respecto a la existencia de libre entrada y salida de empresas, en un plano adyacente de precios y producción se representa la demanda y la oferta de mercado con un equilibrio asociado, E_0.
- Si, por cualquier razón, se incrementa la demanda del mercado, D', y aumenta el precio de equilibrio a P', esto generará beneficios positivos en las empresas instaladas, pues el nuevo precio está por encima de los costes medios, para una producción Q_1 y un equilibrio E_1.
- Pero como hay libertad de entrada, nuevas empresas entrarán en el mercado atraídas por dichos beneficios positivos, aumentando la producción agregada, lo que se refleja gráficamente con un desplazamiento a la derecha de la curva de oferta del mercado, S', hasta que los beneficios se anulan, es decir, hasta que el precio coincida con el mínimo del coste medio y cada empresa vuelva a lanzar al mercado la cantidad óptima, en E_2.
- Por tanto, en el largo plazo todas las empresas operan en la dimensión óptima ($P = CMg_{lp} = CMe_{lp}$), y obtienen beneficios económicos nulos[12].

Figura 7.4: Equilibrio competitivo a largo plazo



Curva de oferta a largo plazo Nótese también que la curva de oferta a largo plazo del mercado es horizontal debido al supuesto de costes constantes, es decir, que cuando la industria demanda más factores productivos para aumentar su producción, el precio de los factores productivos permanece constante.

[12]Nótese que cuando se habla de beneficios nulos, se hace referencia a los beneficios económicos, no solo a los contables. Los beneficios económicos se obtienen restando a los beneficios contables el coste de oportunidad, es decir, restándoles la cantidad que la empresa deja de ganar por dedicar los medios a la producción de dicho bien y no a la mejor alternativa. De ahí que una empresa pueda tener beneficios contables positivos pero beneficios económicos nulos.

 © 2026 Manuel Cabado García

Al ser la curva de oferta del mercado a largo plazo totalmente horizontal, el precio viene determinado por la oferta –es decir, por la tecnología–, y el mantenimiento de este precio queda garantizado por la competencia. Por su parte, la cantidad producida y el número de empresas quedan determinados por la demanda.

Las curvas de oferta y demanda se pueden desplazar por diferentes motivos (aquellas variaciones que no estén en los ejes, es decir, que no sean precio ni cantidad).

Conclusiones En resumen, estas son las implicaciones de que las dotaciones de todos los factores productivos sean variables:

- La combinación de factores productivos elegidos es técnica y económicamente eficiente.
- La curva de CMg es más elástica que en el corto plazo[13].

Se acaba de adelantar que en competencia perfecta la producción en el largo plazo es eficiente, por lo que se pasa a estudiar la optimalidad.

7.4. Equilibrio competitivo ⏳ 2'

Los rasgos generales del equilibrio competitivo al producirse en el mínimo de los costes medios:

1. No hay excesos de capacidad.
2. Los beneficios económicos son cero.
3. Los consumidores adquieren el producto al mínimo precio posible.
4. La maximización de beneficios de las empresas es compatible con la maximización de la utilidad de los consumidores, por lo que los mercados se vacían.
5. Primer teorema del bienestar: en ausencia de fallos de mercado, la competencia perfecta conduce a una asignación que es óptima en el sentido de Pareto, esto es, que ningún agente puede mejorar sin empeorar a otro[14].

A continuación se muestran la eficiencia asignativa y productiva del equilibrio competitivo.

[13] Esto es relevante porque, como se ha visto, en competencia perfecta el tramo creciente de la curva de coste marginal a partir del mínimo de explotación es precisamente la curva de oferta. El fenómeno de la mayor rigidez es una consecuencia de la dotación fija de algún factor productivo en el corto plazo, y es precisamente lo que explica el Teorema de Le Châtelier-Samuelson: las funciones de oferta de la empresa precio-aceptante en el corto plazo son más rígidas que en el largo plazo (y tanto más rígidas cuanto mayor sea el número de factores considerados fijos) debido a que la capacidad de adaptación de la empresa en el corto plazo es menor al tener algunas dotaciones fijas.

[14] El Segundo teorema del bienestar completa este resultado: cualquier asignación óptima de Pareto puede alcanzarse como equilibrio competitivo mediante una redistribución previa de la renta entre los agentes, lo que permite separar los objetivos de eficiencia y de equidad. Su desarrollo completo corresponde al tema de Economía del bienestar.

Eficiencia asignativa Hay eficiencia asignativa cuando se produce la cantidad de cada bien en la que el valor que los consumidores conceden a la última unidad coincide con lo que cuesta producirla[15]. Para demostrar que el equilibrio competitivo la alcanza, se introducen primero los conceptos del excedente del consumidor (EC) y el excedente del productor (EP).

- EC: Diferencia entre lo que el consumidor estaría dispuesto a pagar por una unidad del bien, y lo que paga finalmente[16].
- EP: diferencia entre el precio al que el productor vende finalmente una unidad del bien y el precio mínimo al que estaría dispuesto a venderla, que en competencia perfecta es su coste marginal[17].
- Sumando ambos excedentes se obtendría el excedente social (ES).

Gráficamente, en un plano de precios y cantidades como el de la Figura 7.5, representando las curvas de oferta y demanda, de cuya intersección se derivará el precio y la cantidad de equilibrio. Así:

- EC: El consumidor estaría dispuesto a pagar por una unidad el precio correspondiente al cruce entre la curva de demanda y el eje de precios. No obstante pagará el correspondiente al equilibrio.
- EP: El productor estaría dispuesto a producir una unidad por un importe cercano a 0 (para este caso). No obstante recibirá el precio correspondiente al equilibrio.

Eficiencia del equilibrio: Implicaciones de $P = CMg$:

- El valor que el consumidor concede a la última unidad del bien, medido por su disposición a pagar, se iguala al coste marginal de producirla.
- El excedente social es máximo en competencia perfecta.
- Además es óptimo de Pareto dado que no existe posibilidad de reasignar los recursos de tal forma que se mejore el bienestar de un agente social sin empeorar el de los demás (1TFB).

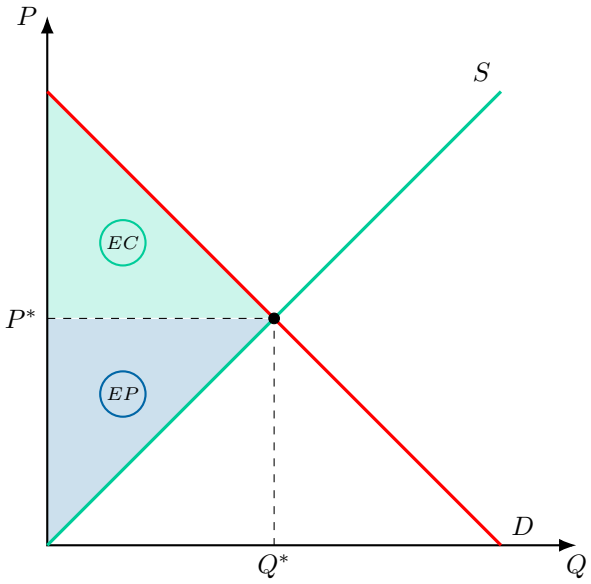
[15] La intuición es que, si la última pizza producida vale para quien la come 12 euros y cuesta 8 producirla, conviene producir otra, y si vale 6 y cuesta 8, sobra. Solo cuando ambas cifras coinciden la cantidad es la adecuada.

[16] El EC es una medida monetaria de la ganancia de utilidad del consumidor, pero solo coincide exactamente con la variación compensatoria y la variación equivalente –las medidas rigurosas del cambio de bienestar– cuando no hay efecto renta sobre el bien, es decir, bajo preferencias cuasilineales (ver Apéndice, 7.A.5). Sin ese supuesto, el EC es solo una aproximación.

[17] La intuición es que, si a un agricultor le cuesta 0,60 euros producir un kilo de naranjas y el mercado se lo paga a 1 euro, su excedente en ese kilo es de 0,40 euros. El EP no debe confundirse con el beneficio: $EP = IT - CVT$, mientras que el beneficio es $\pi = IT - CT = EP - CF$. Ambos coinciden únicamente si no hay costes fijos, lo que ocurre siempre en el largo plazo.

 © 2026 Manuel Cabado García

Figura 7.5: Eficiencia asignativa de la competencia perfecta



Eficiencia productiva Hay eficiencia productiva cuando la cantidad total se obtiene al menor coste posible. En el largo plazo competitivo todas las empresas producen en el mínimo de su curva de CMe_{lp}, que es el mismo para todas porque comparten la misma tecnología, de modo que ninguna otra forma de repartir la producción entre empresas reduciría el coste total.

Propiedades deseables del equilibrio competitivo

Por otro lado, las tres propiedades deseables del equilibrio son: existencia, unicidad y estabilidad. A continuación, se ilustrarán estos tres conceptos con la función de exceso de demanda en un contexto de equilibrio parcial.

Existencia

Comenzando con la existencia, se refiere al concepto de equilibrio, es decir, a que la oferta y la demanda del mercado se corten en algún punto.

El equilibrio existe si hay algún precio al que la cantidad demandada y la ofrecida coinciden. Como las funciones de oferta y demanda son continuas para todo $p > 0$, basta con que haya un precio con exceso de oferta y otro con exceso de demanda para que se crucen en algún punto. Si además la demanda es estrictamente decreciente y la oferta estrictamente creciente, ese cruce es único.

Se define una función de exceso de demanda como

$$Z(P) = D(P) - S(P) \tag{7.7}$$

, de forma que si esta función es continua y toma un valor negativo para algún precio P' y un valor positivo para algún precio P", entonces necesariamente existirá un equilibrio competitivo.

Unicidad

Se refiere a que el equilibrio sea único, es decir, a que las curvas de oferta y demanda se corten en un punto y en ninguno más.

Estabilidad

Se refiere a que el equilibrio sea estable, es decir, a que, ante una perturbación, se vuelva al mismo equilibrio a través de un proceso de ajuste (*tâtonnement*).

El equilibrio será estable si:

- Una vez en el equilibrio no hay fuerzas que lo aparten de él (no hay fuerzas centrífugas).
- Fuera del equilibrio, hay fuerzas internas que lo llevan a él (hay fuerzas centrípetas).

Se distingue entre:

- Estabilidad local, cuando se tiende al equilibrio si la posición tras la perturbación está cercana al equilibrio inicial.
- Estabilidad global, cuando se tiende al equilibrio con independencia de la posición tras la perturbación.

Se pueden contemplar dos formas de analizar la estabilidad del equilibrio:

- La walrasiana, que mide el desequilibrio por el exceso de demanda, es decir, la cantidad demandada menos la ofrecida, $Z(P) = D(P) - S(P)$, siendo el mecanismo de ajuste la variación del precio: este sube cuando hay exceso de demanda y baja cuando hay exceso de oferta. El equilibrio es estable si el exceso de demanda es positivo para precios inferiores al de equilibrio y negativo para precios superiores[18].
- La marshalliana, que mide el desequilibrio por la diferencia entre el precio de demanda y el precio de oferta, $P_D - P_S$, para cada cantidad, siendo el mecanismo de ajuste la variación de la cantidad: los productores la aumentan cuando $P_D > P_S$ y la reducen cuando $P_D < P_S$. El equilibrio es estable si esa diferencia es positiva para cantidades inferiores a la de equilibrio y negativa para cantidades superiores[19].

[18] La intuición es la de una subasta: si a un precio dado faltan unidades, los compradores pujan al alza, y si sobran, los vendedores rebajan el precio para colocarlas.

[19] La intuición es que, si los compradores están dispuestos a pagar por la última unidad más de lo que les cuesta producirla a los vendedores, a estos les compensa producir más, y esa expansión continúa hasta que ambos precios coinciden.

 © 2026 Manuel Cabado García

Hasta aquí se ha visto que el equilibrio existe, es único y es estable. Ahora bien, ¿qué ocurre cuando cambian los costes o se introduce un impuesto? Para responder a esto hace falta un análisis de estática comparativa.

7.5. Estática comparativa ⏳ 2'30"

Una vez establecido el equilibrio competitivo se van a analizar los efectos ante cambios en los costes, que dependerán de si afectan a los costes fijos o a los costes variables.

Incremento de los costes fijos

Este sería el caso del establecimiento de un impuesto de cuantía fija por ejercer la actividad, como el Impuesto sobre Actividades Económicas (IAE). Gráficamente, en dos planos de precio y cantidad, uno para la empresa y otro para el mercado, como los de la Figura 7.6, se representan las curvas de costes de la empresa y la oferta y la demanda de la industria. Así:

- En el corto plazo la oferta no variará, simplemente se producirá un desplazamiento del umbral de beneficios, ya que los costes fijos no afectan a los costes variables medios, que determinan el mínimo de explotación.
- En el largo plazo la curva de costes marginales no se desplaza, al no variar las pendientes de las curvas de costes totales a largo plazo. Lo que sí se producirá será un incremento de la curva de costes medios, CMe', provocando una contracción de la oferta. Como consecuencia, aumentará el precio y la cantidad vendida por cada empresa, y se reducirá el número de empresas de la industria[20].

Incremento de los costes variables

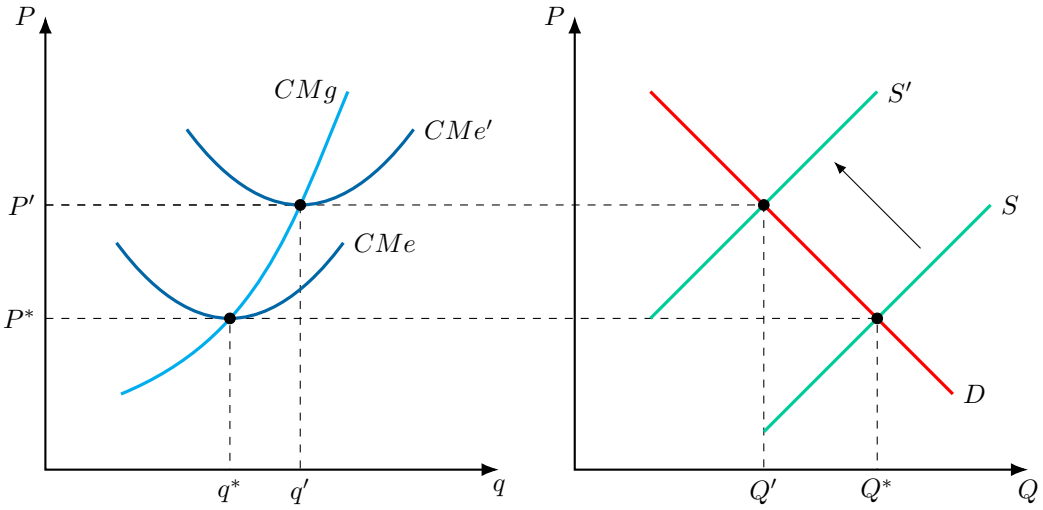
Este sería el caso del establecimiento de un impuesto sobre ventas. En la Figura 7.7, con los mismos dos planos, se representan las curvas de costes de la empresa y la oferta y la demanda de la industria. Así:

- Así, un aumento en los costes variables aumenta el coste marginal lo que a su vez desplaza hacia arriba la curva de coste medio, desplazando la curva de oferta agregada hacia arriba, dando lugar a una menor cantidad de mercado y a un mayor precio[21].

[20] Nótese que la curva de oferta agregada, S, no se dibuja horizontal al estar en el corto plazo (pues se trata de costes fijos).

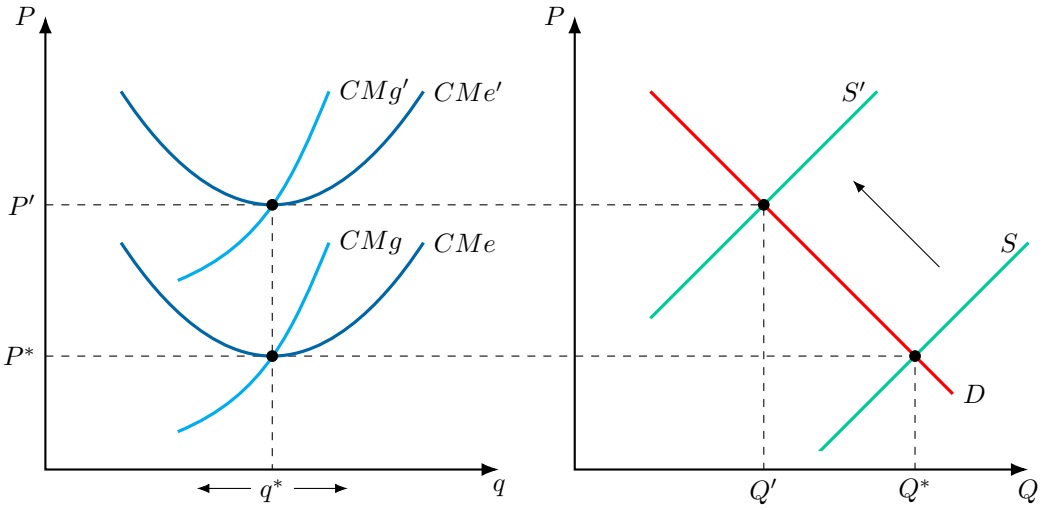
[21] En el corto plazo, se desplaza hacia arriba el $CVMe$ por lo que habrá empresas que cesen su actividad.

Figura 7.6: Aumento de los costes fijos



- Sin embargo, no se puede decir nada sobre la cantidad producida por cada empresa y, por lo tanto, tampoco sobre el número de empresas ya que dependerá del desplazamiento de la curva de coste marginal.

Figura 7.7: Aumento de los costes variables



Variaciones en los impuestos

El análisis anterior tiene una aplicación directa en el estudio de los efectos de los impuestos. En el caso de un impuesto unitario sobre la producción este se considera un coste variable en la medida en que depende del nivel de producción. Por ello, se desplaza hacia arriba la curva de coste marginal.

 © 2026 Manuel Cabado García

Como se observó anteriormente, esto reducirá la cantidad total producida y aumentará el precio, pero en principio no se podría determinar qué ocurriría con la producción de cada empresa y con el número de empresas. Sin embargo, en el caso de impuestos, este se reparte entre productores y consumidores en función de las elasticidades relativas de la oferta y la demanda. Así, cuanto más elástica sea la oferta y más inelástica la demanda, mayor parte pagarán los consumidores[22].

Por ello, como en el largo plazo la curva de oferta del mercado es totalmente elástica, el impuesto se trasladará íntegramente a los consumidores, es decir, íntegramente al precio. Esto hará que la curva de coste marginal se desplace en paralelo hacia arriba en la cuantía del impuesto, por lo que la producción óptima de la empresa individual continuará siendo la misma.

Así, una producción de mercado menor y una producción por empresa constante implican un menor número de empresas.

Cuadro 7.1: Resumen de la estática comparativa

	Traducción		Efectos			
	CF	CV	Q	P	n	q
Aumento de los costes fijos	↑		↓	↑	↓	↑
Aumento de los costes variables		↑	↓	↑	?	?
Impuesto sobre la producción		↑	↓	↑	↓	=

7.6. Relajación de supuestos ⏳ 3'

Hasta el momento se ha supuesto que los costes tienen forma de U y que los precios de los factores están dados, supuestos que, junto con el de información perfecta, garantizan las condiciones de existencia, unicidad y estabilidad del equilibrio competitivo. A continuación, se va a replantear el modelo ya que existen casos en los que estos supuestos no se cumplen.

Curvas de costes sin rendimientos mixtos

Si no existen rendimientos mixtos, entonces los costes a largo plazo no tendrán forma de U. Se analizarán los rendimientos constantes, crecientes y decrecientes a escala como se observa en la Figura 7.8.

[22] Además del reparto de la carga entre oferta y demanda, el impuesto reduce el excedente total respecto del equilibrio sin distorsionar: esta pérdida irrecuperable de eficiencia se representa gráficamente como el triángulo de Harberger, el área comprendida entre las curvas de oferta y demanda entre la cantidad transaccionada con y sin impuesto.

Rendimientos constantes a escala

Los rendimientos constantes a escala se representan mediante una curva de coste marginal constante, lo que da lugar a una curva de coste medio constante –esto es, a economías constantes a escala– y a una curva de oferta horizontal. Existe un precio tal que la oferta agregada es igual a la demanda agregada con beneficio igual a 0. La cantidad quedará totalmente determinada por la demanda y los mercados se vaciarán.

No será posible determinar la producción de cada empresa ni el número de empresas.

Rendimientos crecientes a escala

Los rendimientos crecientes a escala se representan mediante una curva de coste marginal decreciente, lo que da lugar a una curva de coste medio decreciente –esto es, a economías de escala–.

De esta forma no se cumple la condición de segundo orden del problema de optimización, esto es, que el equilibrio se dé en el tramo creciente del coste marginal. Por lo tanto, la regla competitiva de fijar $P = CMg$ generará pérdidas, y las empresas saldrán del mercado hasta que quede un solo productor –monopolio natural– que sea capaz de elevar el precio a, al menos, el nivel del coste medio, para no obtener pérdidas.

La solución sería la existencia de un único productor –monopolio natural–, por lo que no habrá competencia.

Rendimientos decrecientes a escala

Los rendimientos decrecientes a escala se representan mediante una curva de coste marginal creciente, lo que da lugar a una curva de coste medio creciente –esto es, a deseconomías de escala–.

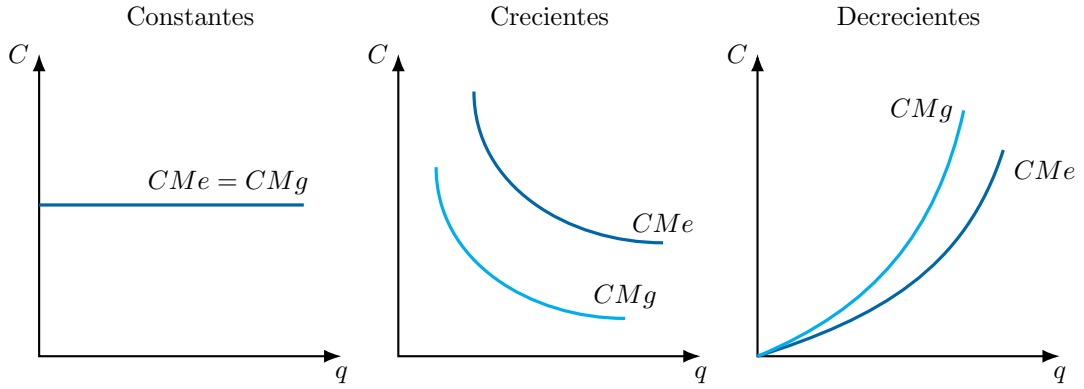
Así, el coste marginal siempre estará por encima del coste medio, por lo que la regla competitiva de fijar $P = CMg$ generará beneficios, y entrarán infinitas empresas al mercado.

Este caso representa una situación poco lógica en la que el equilibrio se produce con un número infinito de empresas y cada una de ellas lanza al mercado una producción infinitesimalmente pequeña.

Conclusión En competencia perfecta los $CMe_{l/p}$ deben tener forma de U. Si no, los resultados no son consistentes con los supuestos del modelo: la forma de U permite obtener un tamaño óptimo de planta lo que, a su vez, permite calcular el número óptimo de empresas en el mercado (lo que determina cuál es la estructura de mercado adecuada). En la Teoría de la Organización Industrial se ha debatido la cuestión de la forma de los $CMe_{l/p}$. Quienes no defienden la forma de U son aquellos autores que creen que hay distintas estructuras de mercado que pueden ser óptimas.

 © 2026 Manuel Cabado García

Figura 7.8: Rendimientos a escala



Economías de escala externas a la empresa pero internas a la industria

Por otro lado, existen economías de escala externas a la empresa pero internas a la industria cuando el coste medio de cada empresa disminuye al aumentar la producción del conjunto de la industria, y no la suya propia. Si el coste medio aumenta, se habla de deseconomías de escala externas.

Las economías de escala externas permiten abandonar el supuesto de costes constantes, es decir, que el precio de los factores productivos es constante –es decir, la oferta agregada de factor productivo es totalmente elástica–. Levantar este supuesto permite obtener una curva de oferta del mercado del bien a largo plazo que no sea horizontal.

Cuasiofertas Ahora la oferta agregada del bien ya no puede calcularse sumando ofertas individuales independientes, sino sumando las cuasiofertas de las empresas, es decir, lo que cada empresa ofrecería a cada precio dado un nivel de producción de la industria. Según su origen, las economías externas pueden ser:

- Pecuniarias: se deben a la interacción que puede existir entre el nivel de producción de la industria y el nivel de precios de los factores productivos (si la oferta de factores no es perfectamente elástica, al aumentar la demanda de ff.pp aumenta su precio).
- Tecnológicas: aparecen cuando el rendimiento de una tecnología dada se ve afectado por el nivel de producción de la industria. Explican por qué un grupo de empresas puede ser más eficiente que una sola, y sus fuentes, ya señaladas por Marshall, son:
  - Aprovechamiento de proveedores especializados que al abastecer a varias empresas pueden especializarse en bienes o factores productivos más específicos, por ejemplo, *just in time*, aumenta el número de proveedores...
  - Creación de un mercado laboral especializado: Aumenta la flexibilidad laboral. Si el trabajador pierde su puesto de trabajo tiene otras oportunidades.

- Externalidades tecnológicas: Spill-overs, difusión informal de know how, acceso rápido a nuevas tecnologías y difusión de conocimiento.

Deseconomías de escala externas

Ante deseconomías de escala externas la industria es de coste creciente, es decir, el precio de los factores productivos aumenta cuando aumenta la producción agregada. Este caso es frecuente, y puede explicarse porque, por ejemplo, la industria utiliza factores productivos escasos (no ilimitados).

Gráficamente, como se observa en la Figura 7.9, se va a representar en un plano de precios y cantidad producida la oferta y la demanda en el corto plazo, determinando su intersección el equilibrio E_0:

- En este contexto si la industria presenta costes crecientes un aumento en la demanda desplaza la demanda hacia la derecha, D', el precio aumentará hasta P', produciéndose beneficios extraordinarios que atraerán a nuevas empresas al mercado.
- Sin embargo, en este caso los precios de los factores aumentarán cuando su demanda aumente por lo que la curva de costes medios a largo plazo de todas las empresas, tanto instaladas como entrantes, se desplazará hacia arriba así como la curva de costes marginales que también se desplazará hacia arriba, cambiando pues la estructura de costes de la empresa[23].
- Así, la entrada de nuevas empresas se producirá hasta que el precio coincida con el mínimo del nuevo coste medio (esto es, la curva de oferta agregada se desplazará a la derecha, pero no tanto, por lo que cortará a la demanda a un precio mayor que el inicial), y cada empresa lanzará al mercado una nueva cantidad óptima.
- Es decir, de nuevo, en el largo plazo todas las empresas operan en la nueva dimensión óptima ($P = CMg_{lp} = CMe_{lp}$), y vuelven a obtener beneficios nulos.
- Esto hará que la curva de oferta agregada a largo plazo tenga pendiente positiva.

Economías de escala externas

Se trata de una industria de coste decreciente, en la que el precio de los factores productivos disminuye cuando aumenta la producción agregada.

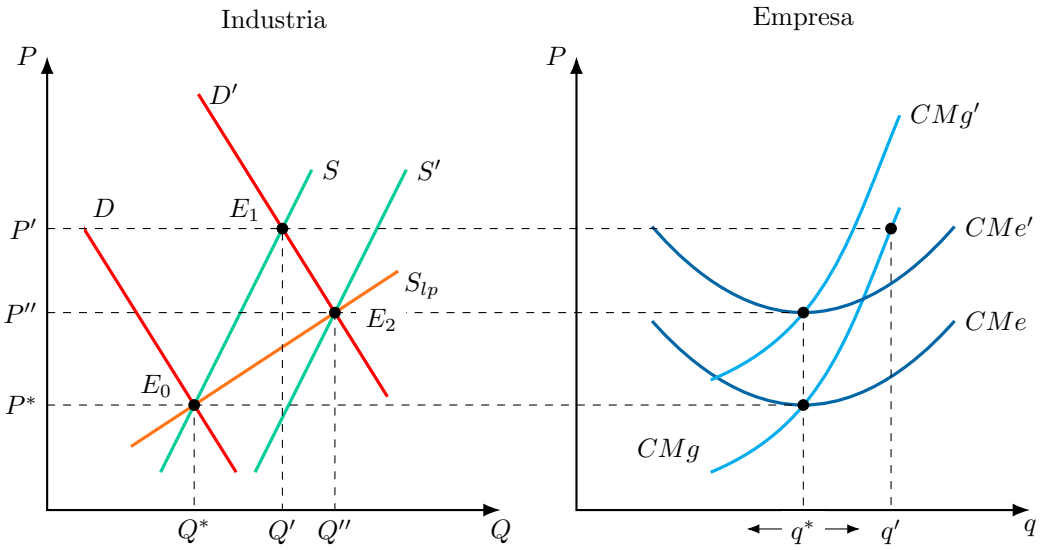
- Este caso es menos frecuente, y podría darse en la práctica por la existencia de un clúster de empresas[24] donde se dan economías de aglomeración (especialización

[23] Nótese que la nueva dimensión óptima de la empresa puede ser menor, igual o mayor que la anterior, en función de cómo se desplace hacia arriba la curva de coste marginal.

[24] Un clúster empresarial es una concentración geográfica de compañías, organizaciones e instituciones relacionadas en un campo en particular que puede estar presente en una región, estado o nación.

 © 2026 Manuel Cabado García

Figura 7.9: Oferta a largo plazo de una industria con coste creciente

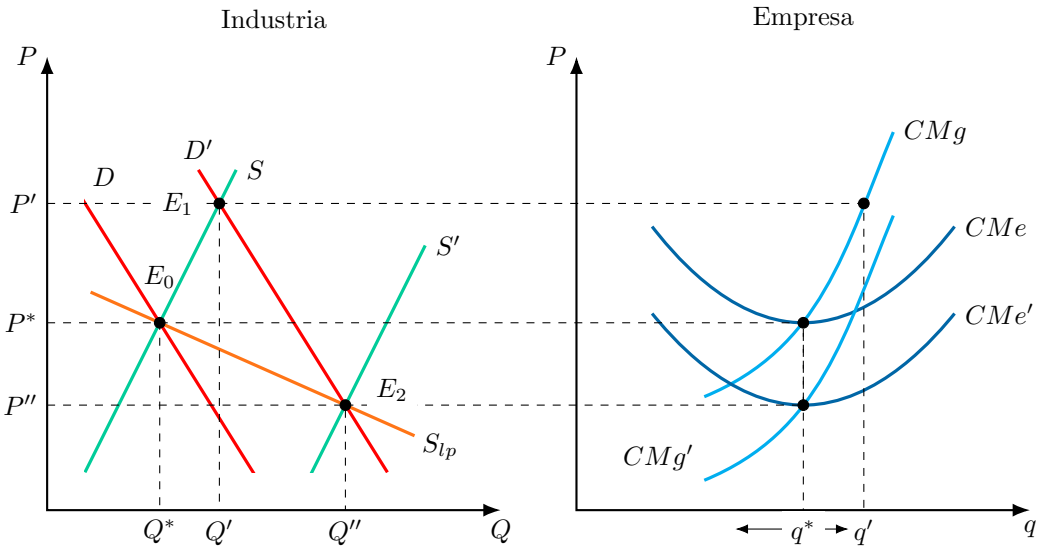


del trabajo, menores costes de transporte y de almacenamiento, y externalidades positivas).

- Como ejemplos pueden mencionarse Silicon Valley, Hollywood o Wall Street en Estados Unidos o, en el caso español, el clúster de la cerámica de Castellón.

El razonamiento es igual que el anterior, solo que ahora la curva de coste marginal se desplaza hacia abajo, lo que genera una curva de oferta agregada a largo plazo con pendiente negativa, como se observa en la Figura 7.10.

Figura 7.10: Oferta a largo plazo de una industria con coste decreciente



Además, las economías de escala externas se han utilizado frecuentemente como argumento a favor de la protección de la industria naciente: sería deseable proteger la industria

nacional para que ésta aumente la producción y, con ella, disminuyan los costes medios de las empresas hasta que sean competitivas internacionalmente.

El modelo de la telaraña

Por último, hasta ahora se ha supuesto que la información es perfecta en el sentido de que consumidores y productores reaccionan simultánea e inmediatamente a los shocks. Sin embargo, en algunos mercados los agentes pueden no conocer de inmediato los nuevos precios –o conocerlos pero no actuar hasta saber si se trata de variaciones transitorias o permanentes en el precio–, por lo que el ajuste puede no ser instantáneo, y los productores pueden necesitar más tiempo para ajustar su capacidad productiva a los nuevos precios –es el caso, por ejemplo, de la agricultura o la construcción–. En estos casos, la oferta no dependerá del precio en curso, sino del precio del período anterior.

Este modelo recibe el nombre de telaraña[25] porque conduce a oscilaciones del equilibrio que serán convergentes, divergentes o constantes en función de las elasticidades relativas de la oferta y la demanda.

El supuesto clave de este modelo es que los productores forman expectativas ingenuas sobre el precio: esperan que el precio en t sea igual al del periodo anterior, $t-1$. El problema en este modelo es el ajuste ante shocks de demanda, dado que no será instantáneo. Este mecanismo no presenta problema si el mercado ya se encuentra en equilibrio, de tal forma que

$$D(p_t) = s(p_t^e) = s(p_t^*)$$

No obstante, si se produce un shock de demanda, comenzará un proceso de ajuste hasta converger al nuevo precio de equilibrio.

Diferentes casos:

- En el caso convergente, los precios tienden al equilibrio en el largo plazo. Gráficamente, en un plano de precios y cantidades como el de la Figura 7.11, se representa el equilibrio inicial entre una oferta y una demanda, con un precio P_0. Si se desplaza la demanda hacia la derecha, con esa cantidad ofertada el precio aumenta a P_1, y a ese precio la cantidad ofertada aumentará a Q_1, pero cuando Q_1 sale al mercado el precio baja a P_2. Las fluctuaciones de precios se van reduciendo progresivamente y en el largo plazo se acaba alcanzando el equilibrio.

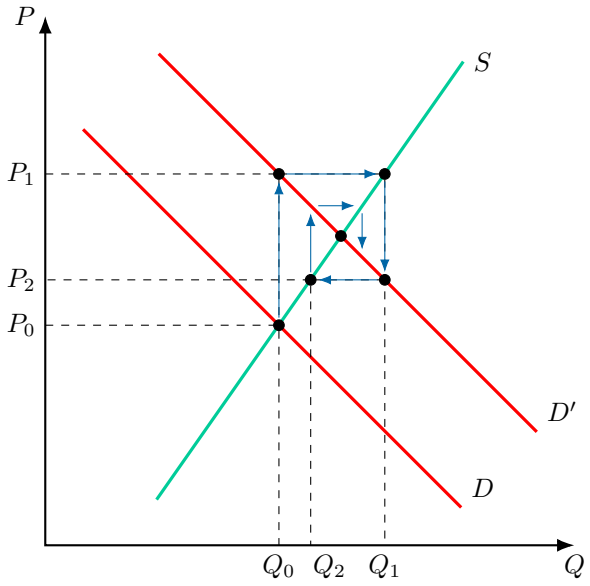
  Esto ocurre cuando la oferta es menos elástica que la demanda (es decir, es algo más vertical que la demanda).

- Si el modelo es divergente entonces los precios se desvían del equilibrio en el largo plazo. El razonamiento sería inverso al anterior y ocurre cuando la oferta es más elástica que la demanda (es decir, es algo más horizontal que la demanda).

[25] El término "telaraña" (cobweb) fue acuñado por Kaldor (1934) y popularizado por Ezekiel (1938), quien lo formalizó a partir de los trabajos independientes de Ricci, Tinbergen y Schultz en torno a 1930.

 © 2026 Manuel Cabado García

Figura 7.11: Modelo de la telaraña con fluctuaciones convergentes



- Por otro lado, en el caso de oscilaciones constantes los precios oscilan alrededor del equilibrio con una amplitud que no cambia, sin llegar a alcanzarlo. Esto ocurre cuando la oferta es igual de elástica que la demanda.

7.7. Conclusiones ⏳ 30"

Como conclusión, la competencia perfecta es la estructura de mercado de referencia en términos de eficiencia porque, en ausencia de fallos de mercado, su equilibrio es óptimo en el sentido de Pareto, tal como establece el Primer Teorema de la Economía del Bienestar. Una vez alcanzado el equilibrio competitivo, ningún agente tiene incentivos a modificar su comportamiento, lo que convierte a este modelo en el referente normativo por excelencia dentro del análisis económico.

No obstante, en la práctica, son escasos los mercados que cumplen con los estrictos supuestos de este modelo. A pesar de ello, se considera deseable por sus implicaciones en eficiencia asignativa.

Sin embargo, desde una perspectiva dinámica, presenta limitaciones importantes: al no existir beneficios extraordinarios en el largo plazo, los incentivos a innovar y diferenciarse se ven reducidos, lo que puede frenar el progreso tecnológico y la evolución del mercado.

Esta restricción se supera mediante estructuras de mercado más flexibles, como la competencia monopolística, donde la diferenciación de productos y la posibilidad de obtener beneficios extraordinarios en el corto plazo fomentan la inversión en innovación y creatividad empresarial.

Apéndice

7.A.1. Levantamiento de otros supuestos

Aparte de estos supuestos, la teoría económica se ha interesado en el levantamiento de otros:

- Pocos productores. Los modelos de monopolio y oligopolio muestran cómo las empresas pueden tener poder de mercado si no existen muchos productores, lo que conduce a una situación socialmente ineficiente.
- Producto heterogéneo. El modelo de competencia monopolística estudia el caso de productos heterogéneos.
- Información imperfecta. Los modelos de Economía de la información permiten estudiar los casos en que los compradores o vendedores no disponen de información completa o simétrica.
- Presencia de externalidades y de bienes públicos. Los modelos de la Economía del bienestar permiten ver cómo varían las conclusiones cuando se introducen externalidades y bienes públicos.

7.A.2. Eficiencia con incertidumbre (Arrow y Hahn, 1971)

De acuerdo con el resultado de Arrow y Debreu, recogido en el manual de Arrow y Hahn (1971), el mecanismo competitivo produce asignaciones eficientes de recursos también en un mundo con incertidumbre, es decir, con información incompleta pero simétrica sobre el estado de la naturaleza, siempre que se cumplan las siguientes condiciones:

- Si los agentes se comportan como maximizadores de la utilidad esperada.
- Si asignan a cada estado de la naturaleza probabilidades subjetivas e independientes de sus propios actos (no existen comportamientos estratégicos)
- Y existe un sistema completo de mercados para cubrirse de todas las posibles contingencias.

La intuición del resultado consiste en redefinir los propios bienes en función del estado de la naturaleza en que se consumen –una manzana hoy y una manzana mañana si llueve son, a efectos del modelo, dos bienes distintos–, lo que permite tratar la incertidumbre dentro del mismo marco de bienes contingentes que sustenta el Primer Teorema del Bienestar. La tercera condición, el sistema completo de mercados, es la más restrictiva: exige que exista un mercado para cada combinación posible de bien y estado de la naturaleza, de forma que los agentes puedan asegurarse frente a cualquier contingencia mediante los llamados títulos contingentes (Arrow, 1964). Si ese mercado es incompleto, el resultado deja de garantizarse.

 © 2026 Manuel Cabado García

7.A.3. Ventajas de la competencia perfecta

Pese a que ciertamente muy pocos mercados pueden considerarse como perfectamente competitivos debe quedar claro que el análisis del equilibrio (parcial) en competencia perfecta es especialmente útil por:

1. Ser la base de la teoría de mercados, pues sirve como marco de referencia ante situaciones más realistas en las que existen fallos de mercado –la pérdida de eficiencia asociada al poder de mercado de un monopolio, por ejemplo, se mide precisamente como desviación respecto de este óptimo competitivo–,
2. Se aproxima a la realidad de algunos mercados, como los de commodities y valores –el trigo, el petróleo o los mercados bursátiles, donde numerosos participantes negocian un producto prácticamente homogéneo–.
3. Permite introducir argumentos gráficos e inequívocos para la medición de las ganancias de bienestar generadas por el intercambio: al cumplirse exactamente $P = CMg$ en el equilibrio, el excedente del consumidor y del productor pueden leerse directamente como áreas del gráfico de oferta y demanda, sin necesidad de instrumentos analíticos más sofisticados.
4. El enfoque de equilibrio parcial permite aislar cualquier fenómeno existente en un mercado de la interrelación con otros mercados, permitiendo una interpretación más clara de los efectos de fallos de mercados, y de las medidas que podrían ayudar a solventarlos.

7.A.4. Controles legales de precios

El equilibrio competitivo desarrollado a lo largo del tema presupone que los precios pueden ajustarse libremente hasta vaciar el mercado. En la práctica, sin embargo, es frecuente que la autoridad económica fije por ley un rango admisible de precios $\bar{P}$, ya sea mediante un precio mínimo –$\bar{P} = \{p \geq p_{min}\}$– o un precio máximo –$\bar{P} = \{0 \leq p \leq p_{max}\}$–, motivados típicamente por presiones de grupos de interés o por objetivos redistributivos.

Una cantidad y un precio $(\bar{x}, \bar{p})$ constituyen un equilibrio con controles de precios si: el precio satisface el control legal, $\bar{p} \in \bar{P}$, el precio minimiza la diferencia entre oferta y demanda agregadas dentro del rango permitido y la cantidad de equilibrio es la menor entre la cantidad ofertada y la demandada al precio $\bar{p}$.

Cuando el control no es vinculante –el precio de equilibrio competitivo ya se encuentra dentro del rango permitido–, el resultado coincide exactamente con el del equilibrio competitivo sin restricciones. Cuando sí es vinculante, el precio de mercado coincide con el propio límite legal: por ejemplo, ante un precio máximo inferior al de equilibrio competitivo, cualquier precio ligeramente inferior al máximo seguiría generando exceso de demanda, y el proceso de puja entre consumidores insatisfechos llevaría el precio hasta el

propio límite legal, momento en el que la restricción impide seguir realizando intercambios mutuamente beneficiosos. Dado que el intercambio sigue siendo voluntario, la cantidad finalmente transaccionada es la menor entre la que desean adquirir los consumidores y la que desean ofrecer las empresas al precio controlado. El razonamiento simétrico se aplica a un precio mínimo vinculante, superior al de equilibrio.

Más allá de la pérdida de excedente total respecto del óptimo competitivo, cuando el control de precios es vinculante aparece además un problema de racionamiento: al no ser la cantidad ofertada –con un precio máximo– suficiente para satisfacer toda la demanda potencial, no queda determinado, dentro del propio modelo, cómo se reparte la producción disponible entre los consumidores dispuestos a adquirirla al precio controlado. Esa asignación final depende de detalles institucionales adicionales –colas, racionamiento administrativo, mercados paralelos–, que en la práctica suelen generar pérdidas de bienestar todavía mayores que las derivadas del mero desajuste entre oferta y demanda[26].

7.A.5. Relación entre el equilibrio parcial y el equilibrio general

A lo largo de este tema el análisis se ha centrado en el enfoque de equilibrio parcial, que aísla la determinación de precios y cantidades en un único mercado. Conviene precisar, para cerrar el tema, que este enfoque no es una alternativa completamente distinta del equilibrio general walrasiano (Walras, 1874), sino un caso particular de este bajo dos supuestos adicionales: la existencia de un bien numerario y preferencias cuasilineales respecto de él.

En un marco de equilibrio general con numerario, la renta inicial de cada consumidor se entiende como su dotación de numerario más su participación en los beneficios de las empresas de las que es propietario. El equilibrio general determina simultáneamente el precio del bien de interés, p_x (el del numerario, p_m, está normalizado a 1), el consumo de cada agente de ambos bienes, y la producción del bien no numerario, de modo que cada consumidor maximiza su utilidad sujeto a su restricción presupuestaria completa –incluyendo su participación en los beneficios empresariales–, cada empresa maximiza beneficios, y los mercados de ambos bienes se vacían simultáneamente. Cuando las preferencias son cuasilineales, resolver este sistema completo resulta equivalente a resolver únicamente el problema de equilibrio parcial del mercado del bien x desarrollado en el cuerpo del tema: la demanda de x carece de efecto renta, por lo que no depende de la riqueza de los consumidores, y el mercado del numerario se vacía automáticamente por la ley de Walras.

La elección entre uno u otro enfoque no es, por tanto, una cuestión de rigor –ambos son formalmente correctos bajo sus respectivos supuestos–, sino de conveniencia analítica según el fenómeno que se quiera estudiar. El equilibrio parcial resulta preferible cuando el mercado analizado representa una proporción reducida del gasto total de los consumidores

[26] El control de alquileres en numerosas ciudades europeas es el ejemplo de manual más citado de precio máximo vinculante con racionamiento no resuelto por el propio mecanismo de mercado.

 © 2026 Manuel Cabado García

y sus conexiones con el resto de la economía son débiles –por ejemplo, en el análisis de medidas antimonopolio en un sector concreto–. El equilibrio general resulta necesario, en cambio, cuando los efectos de la política analizada se transmiten con fuerza a otros mercados –por ejemplo, los efectos de una política impositiva sobre el mercado de trabajo, con repercusiones directas sobre el mercado de la vivienda–.

La justificación habitual del equilibrio parcial –que el mercado analizado es "pequeño" respecto del resto de la economía– no es, sin embargo, la única posible. Existe una justificación alternativa, de naturaleza puramente tecnológica y válida con independencia del tamaño relativo del mercado: el teorema de no sustitución. Si el numerario es el único factor primario –no producido– de la economía, si el resto de bienes producidos distintos del bien x se obtienen con rendimientos constantes a escala a partir del numerario y de otros bienes producidos, y si no existe producción conjunta, entonces los precios del resto de bienes de la economía permanecen constantes con independencia de lo que ocurra en el mercado de x –no por ser este pequeño, sino porque la propia estructura tecnológica de la economía impide que las perturbaciones en x se transmitan a los precios relativos de los demás bienes–[27].

7.A.6. Formalización de la eficiencia: el excedente total como problema de maximización

La eficiencia del equilibrio competitivo demostrada gráficamente en el cuerpo del tema admite una formalización analítica equivalente. La asignación óptima de consumo y producción se obtiene resolviendo el problema de maximización del excedente total,

$$\max_{x} \int_0^x P(s)\,ds - C(x)$$

cuya condición de primer orden es $P(x^*) = C'(x^*)$ –exactamente la condición que caracteriza al equilibrio competitivo ($P = CMg$)–. Esto confirma, por vía analítica, que el equilibrio de mercado y el óptimo social coinciden: el precio de demanda inversa actúa como el beneficio marginal social de cada unidad adicional, y el coste marginal como su coste marginal social, de modo que igualarlos maximiza el excedente conjunto de consumidores y productores.

[27] El resultado fue descubierto de forma independiente por Samuelson y Georgescu-Roegen en 1949 y publicado por ambos en 1951 en el mismo volumen editado por Koopmans, en el contexto de la teoría de Leontief de insumo-producto: Samuelson demostró el caso de dos bienes, Koopmans el de tres, y la generalización al caso de N bienes se debe a Arrow (Arrow, 1951).

7.A.7. Una generalización: mercados disputables y algoritmos de precios

El desarrollo posterior de la teoría de los mercados disputables (Baumol, 1982) ha reforzado el interés por los resultados del modelo de competencia perfecta, al mostrar que no es necesario que se cumplan literalmente todos sus supuestos –en particular, la existencia de un número elevado de empresas– para que los beneficios de una industria se mantengan en el nivel competitivo. Lo relevante no es la competencia efectiva, sino la competencia potencial: la amenaza de que nuevas empresas puedan entrar rápidamente y sin costes hundidos disciplina el comportamiento de las empresas ya instaladas, aunque estas sean pocas.

Esta generalización mantiene vigencia incluso en la economía digital de plataformas y los algoritmos dinámicos de fijación de precios: si los consumidores tienen acceso a información perfecta sobre precios, las empresas deben fijar el precio competitivo para poder vender. Sin embargo, esos mismos algoritmos también tienen capacidad de incrementar el grado de colusión tácita entre empresas y de discriminar perfectamente el precio pagado por cada consumidor (Ezrachi y Stucke, 2016), lo que podría acabar constituyendo, en sí mismo, una nueva barrera de entrada a la industria.

 © 2026 Manuel Cabado García

Bibliografía

Arrow, K. J. (1951). Alternative proof of the substitution theorem for Leontief models in the general case. En T. C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, pp. 155–164. New York: Wiley.

Arrow, K. J. (1964). The role of securities in the optimal allocation of risk-bearing. *The Review of Economic Studies 31*(2), 91–96. Traducción al inglés de "Le rôle des valeurs boursières pour la répartition la meilleure des risques" (1953).

Arrow, K. J. y F. H. Hahn (1971). *General Competitive Analysis*. San Francisco: Holden-Day.

Baumol, W. J. (1982). Contestable markets: An uprising in the theory of industry structure. *The American Economic Review 72*(1), 1–15.

Ezekiel, M. (1938). The cobweb theorem. *The Quarterly Journal of Economics 52*(2), 255–280.

Ezrachi, A. y M. E. Stucke (2016). *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*. Cambridge, MA: Harvard University Press.

Jehle, G. A. y P. J. Reny (2011). *Advanced Microeconomic Theory* (3rd ed.). Harlow: Pearson.

Koutsoyiannis, A. (1979). *Modern Microeconomics* (2nd ed.). London: Macmillan.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Mas-Colell, A., M. D. Whinston y J. R. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Nicholson, W. (2005). *Teoría microeconómica. Principios básicos y ampliaciones*. Editorial Paraninfo.

Pindyck, R. S. y D. L. Rubinfeld (2009). *Microeconomía* (7 ed.). Madrid: Pearson Educación. Traducción y revisión técnica: Esther Rabasco y Luis Toharia.

Varian, H. R. (2010). *Microeconomía intermedia: un enfoque actual*.

Walras, L. (1874). *Éléments d'économie politique pure ou théorie de la richesse sociale*. Lausanne: L. Corbaz.

 © 2026 Manuel Cabado García

Tema 8

Análisis de mercados (II). Teoría del monopolio. Discriminación de precios. Monopolio natural. Análisis de eficiencia y bienestar. La regulación del monopolio.

8.1. Introducción ⏳ 1'

¿Qué tienen en común Google y la empresa que suministra el agua en tu ciudad? En sentido estricto, ambas son monopolios, pero la teoría económica las trata de forma radicalmente distinta, dando lugar a diferente regulación.

Al agua se la regula precisamente porque es un monopolio necesario: se acepta un único proveedor por razones de eficiencia, pero se vigila que no abuse de esa posición. A Google, en cambio, se la persigue por lo contrario: no por ser eficiente tener un único buscador, sino por haber usado su posición dominante para bloquear la competencia allí donde sí sería posible. Así se explican el proceso antimonopolio contra Google en Estados Unidos —iniciado en 2020 y resuelto en 2024, con sentencia que declaró ilegal esa posición de monopolio— y la Digital Markets Act europea, plenamente aplicable desde marzo de 2024.

Este segundo tipo de poder de mercado no solo preocupa por su efecto sobre los precios: como muestra Autor et al. (2020), la creciente concentración de mercado en manos de estas empresas explica en buena medida la caída del porcentaje del PIB destinado a la remuneración del trabajo, al desplazar la actividad hacia empresas con márgenes muy superiores al coste marginal y con una participación del trabajo en sus ingresos mucho menor. El poder de mercado, por tanto, no solo distorsiona los precios: también condiciona

cómo se reparte la renta entre capital y trabajo.

Por ello, el objetivo de esta exposición será el análisis del monopolio, el impacto en el bienestar y su regulación a través del siguiente esquema:

1. Caracterización del monopolio.
2. Equilibrio del mercado.
3. Tipos de monopolio.
4. Regulación y control.

8.2. Caracterización del monopolio ⏳ 30"

Comenzando con la caracterización, la estructura de monopolio se basa en[1]

1. Un único oferente[2] que, con un comportamiento racional, elige el precio o la cantidad que maximiza su beneficio, sabiendo que la demanda determinará la otra variable. Además, la demanda a la que se enfrenta la empresa es la propia demanda de mercado.
2. Producto homogéneo y ausencia de sustitutivos cercanos (lo que determina su poder de mercado y con frecuencia, su ineficiencia y necesidad de regulación).
3. Barreras de entrada y salida en la industria, lo que posibilita beneficios extraordinarios.
4. Numerosos compradores con una demanda lineal –lineal para simplificar el análisis–.
5. Competencia perfecta en el mercado de factores.
6. No existirán bienes públicos ni externalidades[3].

Como consecuencia de lo anterior, el monopolista tendrá poder de mercado, es decir, capacidad para fijar un precio superior a su coste marginal, porque se enfrenta a una curva de demanda con pendiente negativa y no perfectamente elástica[4].

[1] Para un desarrollo más completo de la caracterización del monopolio, véase (Segura, 1993).

[2] Razones por las que existe un único oferente:

- La existencia de ventajas absolutas en costes.
- La existencia de barreras a la entrada (podrán existir beneficios extraordinarios en el largo plazo) ya sea por causa de: Comportamientos estratégicos, barreras legales, control de los recursos y de la tecnología necesaria para la producción por parte de una sola empresa.

Además, a diferencia del análisis de competencia perfecta, en este caso no es necesario distinguir entre la demanda de la empresa y la de la industria. Al haber una sola empresa, ambas coinciden. El monopolista no toma paramétricamente los precios ya que son endógenos.

[3] Y los consumidores van a disponer de información perfecta, aunque no en todos los casos.

[4] El monopolista percibe una curva de demanda inversa $D^{-1}(x) = P(x)$ como la curva de IMe, como se verá. A su vez, a la hora de maximizar el beneficio prestará especial atención a la curva de IMg, que tendrá una mayor pendiente que la curva de demanda inversa (derivado de la existencia de poder de mercado, el productor internaliza el efecto del aumento de la producción sobre la disponibilidad a pagar de la demanda). Este efecto quedará sintetizado por la elasticidad-precio de la demanda.

 © 2026 Manuel Cabado García

8.3. Equilibrio del mercado ⏳ 8'

No obstante, el monopolista no fijará cualquier precio, sino aquel precio que maximice sus beneficios que es la diferencia entre los ingresos totales, $P(Q) \cdot Q$, y los costes, $C(Q)$, donde $P(Q)$ es la curva de demanda con pendiente negativa –elasticidad de la demanda finita– y el precio es una variable endógena –mientras que en competencia perfecta era exógena–:

$$\text{máx} \quad \pi = P(Q) \cdot Q - C(Q) \tag{8.1}$$

Las CPO del problema establecen que el ingreso marginal es igual al coste marginal

$$\frac{\partial \pi}{\partial Q} = 0 \Rightarrow IMg = CMg \tag{8.2}$$

donde el ingreso marginal, es decir, la variación del ingreso total al vender una unidad más, es inferior al precio, que coincide con la demanda y con el ingreso medio

$$IMg = P\left(1 - \frac{1}{|\varepsilon_Q|}\right) < P = D = IMe, \qquad \varepsilon_Q \neq \infty \tag{8.3}$$

Para vender una unidad más, la empresa debe bajar el precio: gana el precio de esa unidad adicional (efecto positivo), pero pierde la rebaja aplicada a todas las que ya vendía (efecto negativo), de ahí que el ingreso marginal quede por debajo del precio[5]. Por su parte, si el IMg fuera menor que el CMg, la empresa reduciría la producción, y viceversa.

Diferencia respecto a Competencia perfecta En competencia perfecta también se cumple que $IMg = CMg$, pero el IMg es igual al precio, mientras que en un monopolio el IMg es menor que el precio. Esto implica que:

- La valoración privada del bien, IMg, –realizada por el monopolista– es menor que la valoración social de dicho bien, P o D, –realizada por el mercado–. En consecuencia, el monopolista valorará menos que la sociedad el bien y producirá una cantidad inferior a la óptima desde el punto de vista del mercado.
- En competencia perfecta, el productor debe hacer frente a una demanda perfectamente elástica (es decir, un aumento infinitesimal del precio le hace perder todas sus ventas), por lo que $|\varepsilon_{Q,P}| = \infty$, lo que en la fórmula (8.3) antes vista permite que $IMg = P$[6].

[5] La intuición es la de un quiosco que vende 100 periódicos a 2 euros: para vender el 101 tiene que ponerlos todos a 1,98. Ingresa 1,98 por el nuevo, pero deja de ingresar 2 céntimos en cada uno de los 100 anteriores, así que el ingreso adicional es $1{,}98 - 100 \times 0{,}02 = -0{,}02$ euros: vender uno más le hace ingresar menos.

[6] ¿Por qué el IMg es menor que la Demanda?: $IT = P(Q) \cdot Q$, $IMg = P(1–(1/|\varepsilon Q, P|)) < P$, porque la elasticidad es finita, $IMe = P = D$ y $IMg < IMe$

Índice de Lerner Reorganizando la CPO se obtiene el índice de Lerner, IL, (Lerner, 1934), que mide el poder de mercado como la parte del precio que excede al coste marginal. En el óptimo del monopolista, ese margen es igual a la inversa del valor absoluto de la elasticidad-precio de la demanda

$$IL = \frac{P - CMg}{P} = \frac{1}{|\varepsilon_Q|}, \quad IL \in (0, 1) \tag{8.4}$$

siendo necesario que la elasticidad sea mayor que uno ya que en caso contrario el ingreso marginal sería negativo no pudiéndose igualar a los costes marginales que siempre son positivos[7]. Por ende, el precio depende positivamente del coste marginal y negativamente de la elasticidad demanda-precio.

A su vez, la CSO, que debe cumplirse para que el beneficio obtenido sea un máximo y no un mínimo, establece que la pendiente del ingreso marginal es menor que la del coste marginal

$$\frac{\partial^2 \pi}{\partial^2 Q} = 0 \Rightarrow \frac{\partial IMg}{\partial Q} < \frac{\partial CMg}{\partial Q} \tag{8.5}$$

por lo que el monopolista puede maximizar tanto en el tramo creciente de los costes marginales –como exigía la CSO en competencia perfecta–, como en el tramo decreciente –siempre y cuando, en este caso sea más plana que la curva del ingreso marginal–.

Gráficamente, en un plano de precios y cantidades como el de la Figura 8.1:

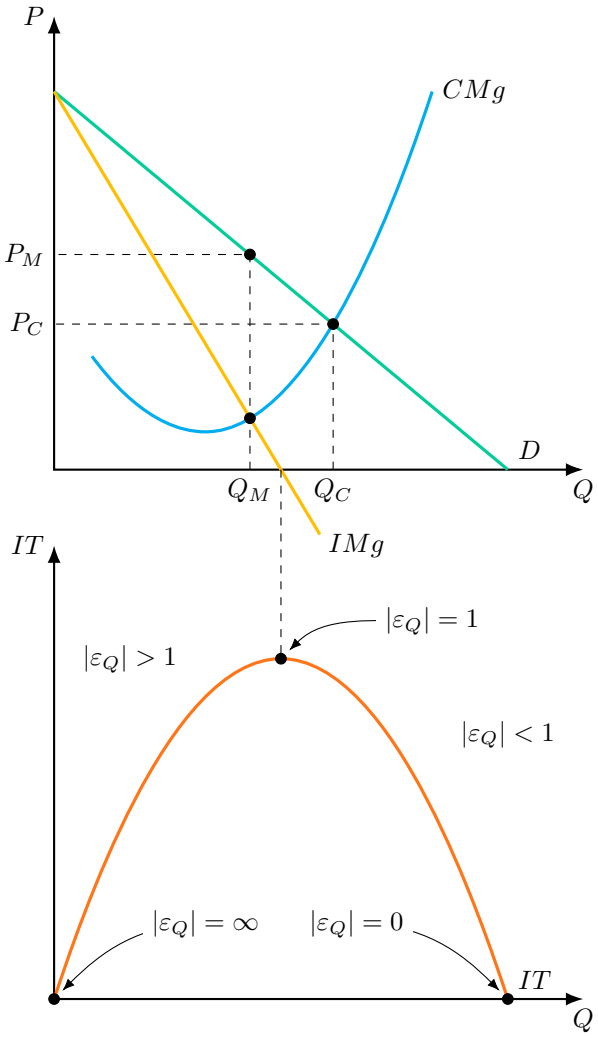
- Se representa la curva de demanda a la que se enfrenta un monopolista, D, su respectiva curva de ingreso marginal, IMg, y la curva de coste marginal, CMg.
- El monopolista fijará un precio, P_M, donde el coste marginal se iguale al ingreso marginal, dando lugar a una cantidad Q_M. En competencia perfecta, donde el coste marginal se iguala a la demanda, el precio es menor, P_C, y la cantidad producida mayor, Q_C.
- En un plano adyacente de ingresos y cantidades se representa la curva de ingresos totales, que nacerá en el origen y alcanzará el máximo cuando el IMg es nulo, es decir, cuando la elasticidad sea igual a 1. Se verifica así que el monopolista produce en el tramo elástico de la demanda.

Finalmente, la asociación única que existe en competencia perfecta entre precios y cantidades no se da en el monopolio, ya que en éste el precio es una variable endógena para el monopolista. Así, no hay una curva de oferta propiamente dicha. El precio y la cantidad no solo dependerán de la tecnología, sino también de la demanda (a menor elasticidad, mayor precio). Es decir, no existirá una relación unívoca entre precio y cantidad.

[7] De otra manera, en el tramo inelástico una reducción de la producción aumenta el ingreso y disminuye el coste, por lo que aumentaría el beneficio. Por consiguiente, en este tramo no se pueden maximizar beneficios.

 © 2026 Manuel Cabado García

Figura 8.1: Equilibrio del monopolio



Por consiguiente, la consecuencia de la inexistencia de una curva de oferta del monopolista será que no se podrán hacer ejercicios de estática comparativa con validez general (el monopolista es esclavo de su demanda).

Equilibrio a corto plazo

En cuanto al análisis en diferentes horizontes temporales, el corto plazo se define como el horizonte temporal en el que la dotación de al menos un factor productivo –normalmente capital– es fija.

La condición de viabilidad económica en el corto plazo es que el nivel de precios cubra los costes variables

$$P = IMe \geq Mín\, CVme$$

es decir, puede obtener pérdidas pero solo si son menores que los costes fijos.

Por ello, el monopolista tendrá beneficios o pérdidas en función de donde se sitúe la

curva de CMe con respecto al precio.

Equilibrio a largo plazo

Por otro lado, el largo plazo se define como aquel horizonte temporal en el que la dotación de todos los factores productivos es variable.

La condición de viabilidad económica en el largo plazo es que el nivel de precios cubra los costes medios

$$P \geq Mín\, CMe$$

es decir, que en el largo plazo el monopolista solo producirá si cubre todos los costes, similitud que comparte con la competencia perfecta.

Monopolio vs Competencia Perfecta No obstante presenta una serie de diferencias con la competencia perfecta:

- El monopolista podría obtener ganancias, ya que las barreras de entrada impiden el acceso de nuevas empresas al mercado.
- La maximización de beneficios no la realiza el monopolista necesariamente en el mínimo de la curva de costes medios.

Barreras de entrada A largo plazo existe monopolio si existen barreras a la entrada de nuevos competidores. Estas a su vez pueden ser:

1. Técnicas: Por conocimiento específico de una técnica productiva de bajo coste, o CMg y CMe decrecientes para un gran intervalo de niveles de producción.
2. Legales por razones de legislación de patentes, potenciar la innovación o por interés nacional.
3. Estratégicas: obstáculos que las empresas instaladas levantan deliberadamente para reducir o eliminar la probabilidad de entrada. Se distinguen de las barreras "inocentes", que surgen como efecto secundario, no buscado, de la propia maximización del beneficio.

Eficiencia del equilibrio del monopolio

Por otro lado, el equilibrio bajo monopolio es ineficiente bajo varios puntos de vista.

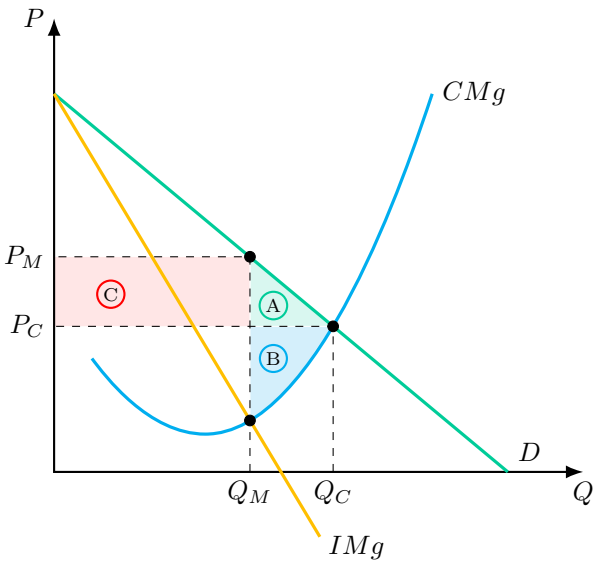
Ineficiencia asignativa

Desde un punto de vista asignativo, el monopolio es ineficiente respecto a la situación de competencia perfecta pues se produce menor cantidad a mayor precio. Es decir, no se maximiza el excedente total (excedente del consumidor + excedente del productor).

 © 2026 Manuel Cabado García

Gráficamente como se observa en la Figura 8.2, la ineficiencia corresponde a las áreas A y B que es el denominado triángulo de Harberger (Harberger, 1954)[8].

Figura 8.2: Ineficiencia asignativa del monopolio



La ineficiencia implica que se podría aumentar el bienestar de algunos agentes del mercado sin reducir el de los demás. Esto se debe a que hay una gama de unidades no producidas para las cuales los consumidores estarían dispuestos a pagar un precio superior al coste de su producción.

El origen de esta ineficiencia se encuentra en que para producir y vender una unidad más el monopolista debería bajar el precio de todas las unidades vendidas, no solo la de la última.

Ineficiencia de factores productivos

Por otro lado, una producción menor que la de competencia perfecta implica una menor demanda de factores productivos –lo que podría generar desempleo–. Además, el monopolista retribuirá menos a los factores productivos, ya que estos se retribuyen según el ingreso del producto marginal –$IMg \cdot PMg$– y en monopolio este es inferior al valor de la productividad marginal –$P \cdot PMg$– ya que $IMg < P$. Por lo tanto se emplearán menos factores productivos y se retribuirán menos.

[8] Concretamente:

- El área A representa una pérdida irrecuperable de eficiencia por el lado del consumo (pues, en competencia perfecta, este área formaría parte del excedente del consumidor).
- El área B representa una pérdida irrecuperable de eficiencia por el lado de la producción (pues, en competencia perfecta, este área formaría parte del excedente del productor).
- El área C es excedente que pierde el consumidor en favor del productor.

Ineficiencia X de Leibenstein

A su vez, la ausencia de competencia reduce los incentivos del monopolista a reducir sus costes, produciéndose la ineficiencia X de Leibenstein (Leibenstein, 1966). Además, la existencia del monopolio favorece las actividades de búsquedas de rentas, para mantener la posición de monopolio mediante la creación de barreras de entrada.

¿Regular o no regular?

Sin embargo, hay que señalar que la mera presencia de estas ineficiencias no es suficiente para requerir una intervención o regulación de los poderes públicos. En algunos casos, se permiten los monopolios, como se verá: Patentes o utilities (o monopolio natural).

En este caso habría que ponderar la eficiencia asignativa frente a la eficiencia productiva y la eficiencia dinámica[9]. Por tanto, la idea básica es que la ineficiencia del monopolio en sentido asignativo por sí sola no lo descalifica. Esto será materia de la política de Defensa de la Competencia y, como se verá después, lleva a la teoría de la Regulación y Control en monopolio. Además, a los fallos de mercado se contraponen los fallos del sector público y de la regulación.

Por consiguiente, de cara a regular el monopolio y determinar la política óptima, es fundamental considerar y estudiar de qué tipo de monopolio se trata y de dónde procede su fuente de poder de mercado.

8.4. Tipos de monopolio ⏳ 4'

Como se acaba de señalar, normalmente el monopolio genera una ineficiencia respecto a la situación de competencia perfecta. No obstante, se pueden dar escenarios donde:

- El monopolio no genere ineficiencias –por ejemplo, el monopolio discriminador de primer grado– pero no tenga en cuenta consideraciones de reparto de la renta (reparto del excedente social).
- El monopolio genere una ineficiencia asignativa pero asegure la eficiencia productiva –por ejemplo, el monopolio natural–.

De esta forma, la existencia de un monopolio no justifica su desmantelamiento en sí mismo, sino que dependerá del tipo concreto de monopolio.

[9] Debate en torno a las pérdidas reales de eficiencia:

- Harberger (1954) estimó para la industria manufacturera de Estados Unidos una pérdida irrecuperable en torno al 0,1 % de la renta nacional (antes del gran desarrollo de las grandes tecnológicas) y concluyó que el coste de bienestar del monopolio era mucho menor de lo que se suponía.
- Williamson: el monopolio tiene su razón de ser, permite reducir costes.

 © 2026 Manuel Cabado García

Monopolio discriminador

Comenzando con el monopolio discriminador, es aquel en el que el monopolista no establece un único precio, sino que cobra precios distintos en función del consumidor o de las unidades vendidas, con el objetivo de aumentar sus beneficios respecto a la situación de precio uniforme. Así, separando los mercados podrá aumentar sus beneficios[10].

Para que se dé la discriminación se requiere que

1. El monopolista pueda distinguir a consumidores o grupos con distinta disposición a pagar (en la discriminación de tercer grado, con distinta elasticidad-precio de la demanda).
2. No exista posibilidad de arbitraje, es decir, de reventa entre consumidores con distinta disposición a pagar.

Según la clasificación ofrecida por Pigou (1920), se distinguen tres tipos de discriminación de precios.

Discriminación de primer grado o perfecta

El monopolista observa perfectamente la valoración de cada consumidor. Cada unidad se vende al consumidor que más la valora, al precio máximo que estaría dispuesto a pagar. Por lo tanto, hay tantos precios como consumidores.

Hay pocos ejemplos en el mundo real, pero pueden citarse los zocos o el médico de un pueblo. No obstante, la tarificación dinámica mediante la utilización de algoritmos y datos para segmentar el mercado de manera muy granular, se acerca a este tipo de discriminación[11].

En este caso el ingreso marginal se iguala a la demanda, $IMg = D$, por lo que el monopolista producirá hasta que la última unidad se venda a un precio igual al CMg y se apropiará de todo el excedente del consumidor. Es decir, el excedente del consumidor será nulo, y el monopolista se apropiará de todo el excedente.

El resultado es eficiente en el sentido de Pareto, por lo que la intervención estatal no está justificada por motivos de eficiencia, aunque sí por motivos de equidad.

Gráficamente, la curva de IMe iría por encima de la curva de demanda, D, es decir, de IMg, ya que en el ingreso medio se tiene en cuenta, no solo la última unidad vendida,

[10] Un ejemplo real de esta tensión entre precio único y discriminación, aunque del lado del comprador, es el mercado eléctrico mayorista español (y del resto de la UE), donde rige el llamado sistema marginalista: todas las centrales ofertan su coste y la última en entrar para cubrir la demanda —la más cara, normalmente de gas— fija el precio que cobran todas las demás, incluidas las renovables o la nuclear, con costes mucho más bajos. Es, en esencia, un precio único no discriminador. La alternativa debatida, el sistema pay-as-bid (cada generador cobra su propio coste ofertado), equivaldría a aplicar al comprador la misma lógica de la discriminación de precios de primer grado. La crisis de precios del gas de 2021-2022, que disparó el precio marginal y generó beneficios extraordinarios a las centrales de bajo coste, reavivó este debate y llevó a medidas como el tope al gas (la llamada «excepción ibérica»).

[11] Véase el apartado 8.A.7 del Apéndice.

sino también todas las anteriores (que se vendieron a un mayor precio)[12]. Por lo tanto, $IMg = D < IMe$.

Discriminación de segundo grado o fijación no lineal de precios

La discriminación de precios de segundo grado consiste en cobrar un precio por unidad que depende de la cantidad comprada, por lo que también se denomina fijación no lineal de precios. Todos los consumidores se enfrentan al mismo menú de precios, de modo que quienes compran la misma cantidad pagan lo mismo[13].

Gráficamente, como se observa en la Figura 8.3, se considera la posibilidad de establecer tres precios diferentes, P_1, P_2 y P_3. La curva de ingreso marginal asociada, IMg_{d2}, tendría forma escalonada dado que en cada tramo de producción al mantener el precio constante el ingreso marginal es el mismo. El monopolista, al buscar la maximización del beneficio, establecerá el nivel de producción máximo aquel donde el ingreso marginal es igual al coste marginal, es decir, Q_3.

Ahora el excedente del consumidor respecto a la situación de monopolio convencional es menor, ampliando el excedente en manos del monopolista. Por lo tanto, si el monopolista practica este tipo de discriminación de precios perjudica a los consumidores, se beneficia ampliando su excedente y se incrementa la cantidad transaccionada en el mercado[14].

Ejemplos[15]:

- Descuentos por grandes cantidades, que hacen que el precio unitario sea menor que cuando se compran cantidades más pequeñas .
- Tarifas en dos partes, que hacen que el precio unitario sea menor cuanto mayor es la cantidad comprada.

[12] Además, en discriminación de primer grado es preciso que el efecto renta de la demanda individual sea nulo. Si no, la curva de demanda variaría con el precio. Es decir, para que la curva de demanda no se vaya desplazando al adquirir sucesivamente las distintas unidades, es preciso que las deducciones de renta no afecten.

[13] La intuición es la de la oferta "3×2" del supermercado: el vendedor no sabe quién valora más el producto, pero deja que cada cliente se clasifique solo según cuántas unidades se lleve.

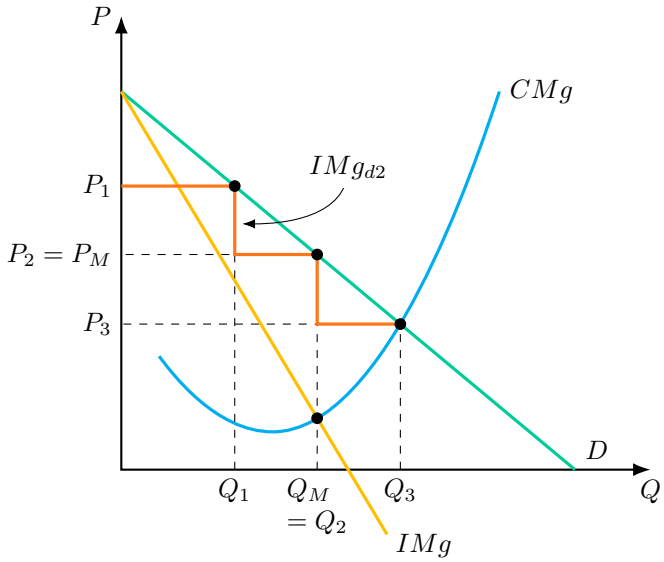
[14] Cada triángulo es un bloque. Se trata de una estructura habitual en los servicios de suministro, como el agua o la electricidad. Los bloques deben estar diseñados con precisión para evitar trasvases entre bloques que perjudiquen al monopolista (restricción de autoselección). Además, sería crucial tener en cuenta el efecto renta (y esto también se puede plantear a modo de crítica) puesto que vendiendo las primeras unidades más caras, esto podría empobrecer al consumidor y reducir la demanda de las siguientes unidades.

[15] Otros:

- Elección de la calidad de producto (Michael Mussa y Sherwin Rosen): el monopolista ofrece distintas calidades de producto que incitan a los consumidores a autoseleccionar el tipo óptimo.
- Bundling y Tying: el monopolista ofrece varios productos en una cesta única a un precio agregado inferior que si lo hiciera por separado.

 © 2026 Manuel Cabado García

Figura 8.3: Discriminación de precios de segundo grado



Discriminación de tercer grado

El monopolista es capaz de observar características de los consumidores a partir de las cuales puede inferir información en cuanto a su valoración, permitiéndole discriminar entre cada segmento de mercado, tratándolos como mercados independientes.

La discriminación de precios de tercer grado significa que el monopolista vende a cada persona (o grupo de personas) el bien a precios distintos, pero cobra el mismo precio por todas las unidades del bien que vende a esta persona. En otros términos, se hablaría de una única planta –un único CMg– y varias demandas –varios IMg–.

A su vez, es aplicable la regla simple de la elasticidad inversa analizada en el caso del monopolio, según la cual se cargarán en todos los mercados precios superiores a los costes marginales y tanto mayores cuanto menor sea la elasticidad de la demanda de cada mercado[16].

Un ejemplo son los abonos de transporte con precios más bajos para la tercera edad[17]. Se cobra un precio diferente a cada grupo, pero un mismo precio dentro del grupo.

Otro ejemplo serían los cupones: Permiten a las empresas segmentar el mercado entre clientes sensibles al precio (demanda elástica) que usan los cupones, y clientes menos sensibles (demanda inelástica) que pagan el precio completo, maximizando así los benefi-

[16] Tirole (1988) demuestra que un aumento de la producción total es condición necesaria para que la discriminación de tercer grado mejore el bienestar respecto al precio uniforme: si la discriminación no incrementa la producción total, necesariamente reduce el bienestar agregado.

[17] A menudo se citan también los abonos de transporte con precio reducido para jóvenes. Formalmente también son discriminación de tercer grado, pero no responden a la lógica de un monopolista maximizador de beneficios, que cobraría menos al grupo más elástico: los jóvenes no tienen una elasticidad alta (ya que tienen necesidades importantes de movilidad y, sin embargo, no suelen tener una renta elevada que les permita afrontar el transporte privado, por lo que su elasticidad es baja). La razón de los abonos a los jóvenes es política, no económica.

cios[18].

Monopolio natural

Por otro lado, existe monopolio natural cuando una sola empresa puede abastecer toda la demanda del mercado a un coste menor que cualquier reparto de esa producción entre varias empresas, es decir, cuando hay subaditividad de costes[19]

$$C(Q) < \sum_{i=1}^{n} C(q_i), \quad Q = \sum_{i=1}^{n} q_i \tag{8.6}$$

pero las economías de escala son una condición suficiente pero no necesaria para que exista un monopolio natural.

Gráficamente, en un plano de precios, costes medios y cantidades como el de la Figura 8.4:

- Se representa la curva de costes medios derivada de la concentración de la producción en una empresa, CMe_1, y la correspondiente a dos empresas que se reparten la producción total a medias, CMe_2.
- En el gráfico se ve cómo en el tramo decreciente de la curva de coste medio para una empresa existe subaditividad de costes. Pero la subaditividad se extiende hasta Q_1 a pesar de que existe un tramo con deseconomías de escala. En efecto, aunque en el tramo que va desde Q_0 hasta Q_1 existen deseconomías de escala, el coste medio de concentrar la producción en una empresa es menor que el de repartirla en dos empresas.
- A partir de Q_1 ya no existe subaditividad de costes y, por lo tanto, no hay monopolio natural.
- De esta manera, el monopolio natural será deseable si la demanda corta a la curva de CMe_1 en algún punto hasta Q_1. Más allá, el monopolio natural no será deseable, pues el reparto de la producción entre más empresas genera menores costes.

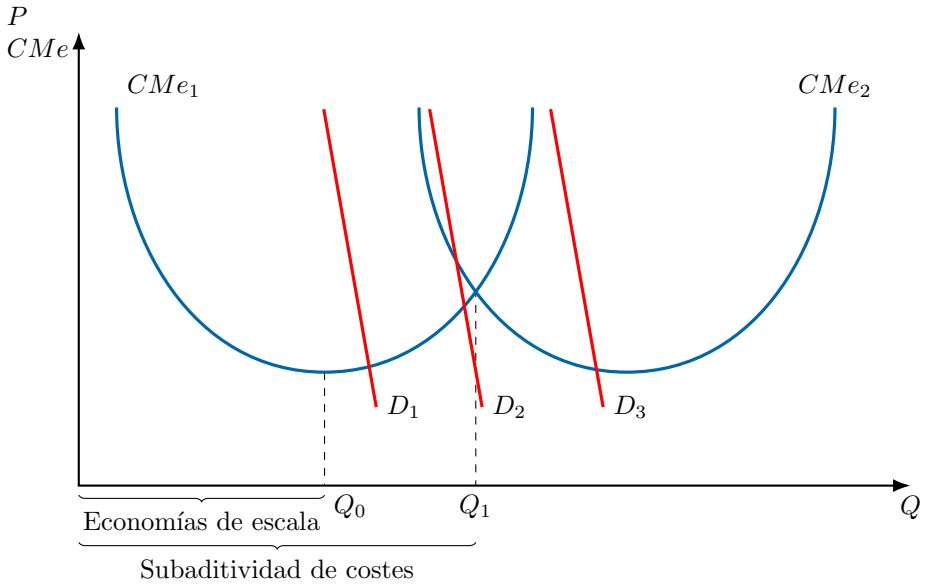
En general, los mercados monopolísticos naturales surgirán cuando el tamaño mínimo eficiente sea grande en relación con la demanda del mercado. Esto explica en parte por qué antes había muchos monopolios naturales y ahora hay menos: conforme aumenta la demanda de productos (por ejemplo, por la apertura al comercio internacional), el tamaño mínimo eficiente se hace menor en relación a la demanda de mercado.

[18] En el comercio minorista, se ofrecen cupones a través de aplicaciones móviles para segmentar clientes por historial de compra. Las telecomunicaciones los usan para atraer nuevos clientes o retener a los existentes con descuentos en dispositivos. El sector alimentario sigue usando cupones para productos específicos, dirigidos a consumidores sensibles al precio.

[19] La intuición es la red de agua de una ciudad: tender dos redes de tuberías paralelas para que compitan dos empresas duplicaría el coste sin añadir nada al servicio.

 © 2026 Manuel Cabado García

Figura 8.4: Sostenibilidad del monopolio natural



Sostenibilidad Por otro lado, su sostenibilidad va a depender de dónde se sitúe la demanda:

- Si la demanda se sitúa en algún punto anterior a Q_0, el monopolio natural será deseable y sostenible, pues un outsider no podrá entrar ya que, para entrar, tendría que colocar un precio por debajo del precio del incumbente, y eso, para el tramo decreciente de CMe_1, es sinónimo de $P < CMe_1$, por lo que obtendría pérdidas.
- Si la demanda se sitúa entre Q_0 y Q_1, el monopolio natural será deseable, pero su sostenibilidad dependerá de si la demanda se sitúa o no por debajo de la curva CMe_2[20].

Esta lógica de sostenibilidad enlaza con la teoría de los mercados atacables o contestables (contestable markets), desarrollada por Baumol (Baumol, 1982): bajo los supuestos anteriores, la mera amenaza de entrada disciplina al monopolista, que se ve obligado a fijar un precio igual al coste medio y a renunciar a los beneficios extraordinarios, incluso sin regulación explícita.

Junto a estos dos escenarios de eficiencia y equidad, conviene considerar también cómo cambia el análisis cuando el monopolista opera con una estructura de producción o de mercado distinta a la estándar de una única planta y un único producto frente a un único comprador.

Otros tipos de monopolio

Otros tipos de monopolio serían:

[20] Véase el desarrollo completo en el apartado 8.A.5 del Apéndice.

- Monopolio multiplanta: se trata de la producción de un mismo producto en varias plantas con estructuras de costes diferentes[21].
- Monopolio multiproducto: la solución dependerá de la relación que exista entre las demandas de los bienes producidos por el monopolista, y también de la relación que haya entre los costes de producción.
- Monopolio bilateral: o monopolista vs monopsonista, donde existe un único vendedor y un único comprador (por ejemplo, un trabajador de cualidades únicas y una única empresa que puede necesitar sus conocimientos). En este caso el resultado está indeterminado: solo puede acotarse un intervalo de precios y cantidades, y el punto final depende del poder negociador de cada parte, por lo que se estudia con la teoría de juegos[22].

8.5. Regulación y control ⏳ 6'

Por último, se va a analizar el aspecto de regulación y control de los monopolios, ya que la asignación generada es ineficiente. La regulación tiene como objeto evitar esas distorsiones y maximizar el bienestar, entendido como la suma del excedente del consumidor y el excedente del productor. Se va a diferenciar entre regulación fiscal, de precios y otros instrumentos[23].

Regulación fiscal

Comenzando con la regulación fiscal, se pueden establecer tres medidas, cuyos efectos se representan, en tres planos de precio, P, y cantidad, Q, en la Figura 8.5:

[21] Así, la producción se reparte entre plantas de forma que:

- El coste de la última unidad en ambas plantas coincide: $CMg(q_1) = CMg(q_2)$. La planta que tenga ventaja en costes: i) producirá más si los costes marginales son crecientes y ii) producirá todo si los costes marginales son constantes.
- El coste de la última unidad producida coincide con el ingreso que genera ($CMg(q_i) = IMg(Q)$).
- Si los costes marginales son diferentes y constantes, produce TODO en la planta con menor coste marginal. Si los costes marginales son distintos pero crecientes, produce menos pero no nada en la planta con mayor Cmg.

[22] En este caso, se puede llegar a diferentes conclusiones:

- Soluciones no negociadas, que dependerán por tanto del poder de negociación de ambas partes.
- Soluciones negociadas, que conseguirán maximizar el beneficio conjunto, generando así resultados óptimos.

[23] En términos generales, la regulación se puede realizar:

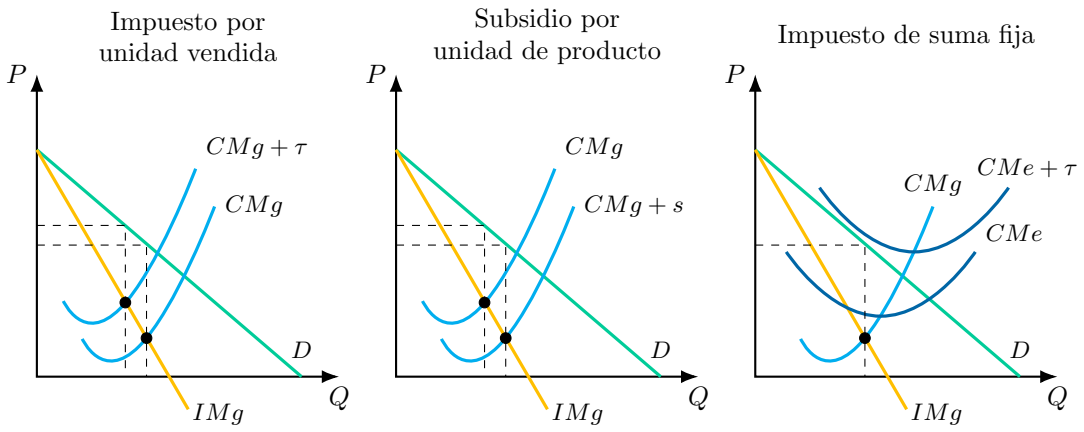
- Por el lado de la oferta, instaurando mecanismos de precios y de entrada y salida.
- Por el lado de la demanda, fomentando la agrupación de consumidores para mejorar el reparto del excedente. Este tipo de regulación, sin embargo, presenta grandes inconvenientes por: i) el elevado coste de organización de los consumidores, ii) el posible desacuerdo entre ellos y iii) la imposibilidad de evitar free-riders. Por eso esta exposición se va a centrar en la regulación por el lado de la oferta.

 © 2026 Manuel Cabado García

1. Impuesto por unidad vendida: como el monopolio tiene poder de mercado podría trasladar el impuesto a los precios aumentando estos y disminuyendo la producción. Gráficamente, se trasladaría la curva de costes marginales, $CMg + \tau$, estableciendo un precio mayor.
2. Subsidio por unidad de producto, lo que trasladaría gráficamente la curva de costes marginales hacia la derecha. Se transfiere renta al monopolista, la cual podría recuperarse técnicamente mediante un impuesto de suma fija. Sin embargo, es una medida más difícil de justificar aunque podría conseguirse un aumento de las cantidades producidas y la disminución de los precios.
3. Impuesto de suma fija, con el que solo variarían los costes medios por lo que la cantidad y el precio de equilibrio no cambian. El gobierno podría con este impuesto disminuir o controlar las ganancias del monopolio sin afectar al precio o a la producción.

Sin embargo, en general la regulación fiscal no es efectiva ya que se reducen los beneficios pero no se alcanza una asignación de competencia perfecta.

Figura 8.5: Regulación fiscal del monopolio



Regulación de precios

Por otro lado, otro conjunto de técnicas son las que tienen que ver con la regulación de precios:

- Óptimos de primer orden, que tratan de restablecer la eficiencia asignativa.
- Óptimos de segundo orden, que buscan el máximo bienestar posible con la condición de que la empresa cubra sus costes.
- Óptimos de tercer orden, que tratan de limitar los beneficios extraordinarios.

Óptimo de primer orden

El regulador dispone de toda la información disponible acerca de los costes. Su objetivo es, como se ha dicho, restablecer la eficiencia asignativa (tratando de lograr una solución first best) (es decir, imponer $P = CMg$), y cubrir pérdidas o eliminar beneficios extraordinarios:

- Si la condición competitiva se produce en el tramo decreciente de los costes medios, $CMe > CMg = P$, por lo que se producirán pérdidas. Para cubrir estas pérdidas el gobierno puede otorgar una subvención.
- Si la condición competitiva se produce en el tramo creciente de los costes medios, $CMe < CMg = P$, por lo que se producirán beneficios. Para eliminar estos beneficios el gobierno puede exigir un impuesto de suma fija por el valor del área sombreada y luego repartirlo entre los consumidores (sin embargo, esta solución lleva asociada una elevada complejidad práctica e incentiva a que el monopolista no revele sus costes verdaderos).

El problema de esta solución de primer orden es que, si se produce en el tramo decreciente de los costes medios (que es lo más común), es necesario recaudar vía impuestos[24] la cuantía de la subvención a otorgar al monopolista, lo que puede producir una distorsión que genere ineficiencia.

Por eso se suelen aplicar soluciones de segundo orden (óptimos de solución Second-Best).

Óptimo de segundo orden

El óptimo de segundo orden busca el mayor bienestar social compatible con que el monopolista cubra sus costes, lo que suele implicar eliminar sus beneficios extraordinarios. Existen diferentes métodos: criterio del coste medio, tarifas con restricciones de capacidad, la tarifa en dos partes, en el caso del monopolio multiproducto la regla de la elasticidad inversa o precios de Ramsey, y el mecanismo de Loeb y Magat.

Criterio del coste medio

El Gobierno impone $P = CMe$, asegurando beneficios nulos al monopolista. El problema es que la eficiencia puede resentirse, ya que la garantía de que los precios se revisarán para cubrir costes anula el incentivo de las empresas a reducirlos. Se produce un trade-off eficiencia-rentabilidad.

[24] Los impuestos progresivos producen distorsiones vía modificación de los precios relativos, y los de suma fija son demasiado regresivos y difíciles desde un punto de vista social.

 © 2026 Manuel Cabado García

Precio según utilización de capacidad: demandas pico-valle

Algunas actividades se caracterizan por estar sujetas su demanda a fluctuaciones sistemáticas. Esto afecta principalmente a empresas de servicios públicos, como la provisión de energía eléctrica, agua, transporte público, llamadas telefónicas, donde hay momentos del día en que la demanda aumenta considerablemente –demanda pico– en relación a otros momentos –demanda valle–. También afecta a otros bienes de provisión privada, como la demanda turística o los taxis.

Las características principales de estos sectores son:

1. La demanda es estacional u horaria.
2. El output no es almacenable, por lo que no se pueden trasladar los bienes y servicios de un momento a otro, por lo que no es posible un arbitraje temporal.
3. Todas las fases de la demanda son atendidas con una misma capacidad, por lo que la capacidad existente no se utiliza de manera uniforme, es decir, es un caso de producción conjunta. Por tanto, para satisfacer la demanda en los períodos pico se necesitan inversiones con un largo período de maduración que imponen un exceso de capacidad instalada en los períodos valle.

La regulación óptima en estos casos pasa por[25]:

- Cobrar Tarifa, $T = CMg \cdot q$ en los momentos valle. Por tanto, se cubren los costes de uso.
- Cobrar Tarifa $T = CF/n + CMg \cdot q$ en los momentos pico, donde n es el número de consumidores. Por tanto, se cubre el coste de financiar la capacidad instalada y se cubren los costes de uso.

Es decir, se asegura que el beneficio del monopolista va a ser cero haciendo que los costes fijos sean soportados por los consumidores que consumen en períodos de demanda pico, mientras que los consumidores que consumen en períodos de demanda valle deben soportar exclusivamente los costes variables[26].

Tarifa en dos partes

Consiste en asegurar el nivel de producción competitivo (es decir, eficiente) al tiempo que se asegura un beneficio normal –beneficio nulo–[27].

[25] En algunos libros presentan la tarifa pico-valle como $P = CVMe$ en valle y $P = CF/n + CVMe$ en pico

[26] Por ejemplo, la reforma eléctrica española en verano de 2021. Entre los cambios, se incluye variabilidad en precios y discriminación en función de la franja horaria. Los diferentes tipos de discriminación horaria que existían hasta entonces se reforman, dando paso a una única discriminación horaria de la Tarifa 2.0TD. Las franjas horarias de la nueva tarifa eléctrica se dividen en tres periodos: i) Punta: horas más caras, ii) Llano: horas con precio medio, iii) Valle: horas con precio reducido

[27] Esta estrategia de precios fue analizada por primera vez por Oi (1971) en un artículo clásico sobre el precio de las entradas a un parque de atracciones —de ahí el título del propio artículo, “A Disneyland

Así, para configurar la tarifa, T, se establece una cuota fija por el derecho a consumir los servicios del monopolista natural (la cuota de enganche, por ejemplo la de conexión telefónica o eléctrica), por un importe A, y un precio constante por unidad consumida, p, de modo que el pago variable es $p \cdot q$, de forma que la función de pagos del consumidor sería

$$T = A + p \cdot q \tag{8.7}$$

donde, A serían los costes fijos a los que tiene que hacer frente el monopolista para atender la demanda, y se distribuirían uniformemente entre los consumidores, $A = CF/n$ y el precio variable por unidad se representará mediante los costes marginales $p = CMg$.

Problemas La tarifa en dos partes, sin embargo, tiene tres problemas:

1. Demanda constante: supone que la demanda es constante e independiente de la cuantía que tome la parte fija de la tarifa, supuesto no realista, ya que es previsible que sí exista cierta sensibilidad de la demanda ante variaciones en la cuota fija.
2. Regresividad: podría dar lugar a desigualdad, ya que los consumidores de menores ingresos podrían quedar expulsados completamente del mercado. Una posible solución a este punto podría ser no fijar la cuota entre el número de consumidores, sino que la cuota sea diferente para los consumidores en función de la renta: los consumidores con rentas más bajas pagarían cuotas más bajas, que se verían compensadas con las mayores cuotas del resto de consumidores.
3. Maximización de eficiencia: Podría no estar maximizándose la eficiencia, ya que se está aplicando la tarifa de manera uniforme o en función de la renta de los consumidores, pero no en función de la elasticidad. La eficiencia sería mayor si la cuota fija fuese mayor para los consumidores de demanda inelástica y menor para los de demanda elástica.
   No obstante, al igual que la cuota fija uniforme, podría ocasionar problemas de equidad si los de demanda inelástica son los más pobres –por ejemplo, porque se trata de un bien de primera necesidad–. Sería una solución similar a la de la Regla de Ramsey del monopolio multiproducto.

Precios de Ramsey

Por otro lado, si el monopolio es multiproducto la regulación puede adaptar la regla de elasticidad inversa o precios de Ramsey. Esta regla es un óptimo de segundo orden ya que se trata de cubrir los costes de la empresa, $CMe > CMg$.

Para ello, reparte el coste fijo mediante un margen sobre el coste marginal inversamente proporcional a la elasticidad-precio de cada demanda, de modo que los productos con una

Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly"—: una cuota fija de entrada más un precio por cada atracción, el mismo ejemplo que ilustra el criterio de tarifa en dos partes desarrollado a continuación.

 © 2026 Manuel Cabado García

demanda más rígida soportan un margen mayor. Ese margen es tanto mayor cuanto más costoso resulte, en términos de bienestar, cumplir la restricción de beneficio, recogida en el multiplicador λ[28][29]:

$$\frac{P - CMg}{P} = \frac{\lambda}{1+\lambda} \cdot \frac{1}{|\varepsilon|} \tag{8.8}$$

De esta forma, el nivel de bienestar es mayor al que se obtendría si los costes fijos se repartieran por igual entre todos los consumidores.

Sin embargo, los principales inconvenientes son:

1. Puede generar desigualdad pues los productos con baja elasticidad precio suelen tener un peso mayor en los individuos con poca renta.
2. Exige conocer los costes y la demanda de todos los bienes, por lo que su aplicación se limita a la determinación de los precios en las empresas públicas.
3. No son precios necesariamente libres de subsidios cruzados –es decir, la producción

[28] Esto sería para bienes independientes, ya que para bienes complementarios o sustitutivos la elasticidad se sustituye por la superelasticidad (variación de la demanda conjunta de los bienes ante la variación del precio del bien i):

$$\frac{P_i - CMg_i}{P_i} = \frac{\lambda}{1+\lambda} \cdot \frac{1}{\varepsilon_i} + \frac{\varepsilon_{ij}}{\varepsilon_i} \cdot \frac{p_j q_j}{p_i \cdot q_i}$$

Tal y como recoge la anterior ecuación, el mencionado mark-up vendrá determinado por el trade-off al que se enfrenta el regulador (recogido en el número Ramsey) y por la interdependencia entre los bienes que produce. En particular, si los bienes son sustitutivos ($\varepsilon_{ij} > 0$), la menor demanda de un bien por el aumento de su precio se ve compensada en el otro mercado, con lo que el precio óptimo será mayor que si los bienes son independientes. En sentido contrario si los bienes son complementarios, el efecto de la variación del precio de un bien se verá amplificado en el segundo mercado, con lo que el precio óptimo será inferior al caso de los bienes independientes.

[29] La intuición es que conviene cargar más margen sobre aquello que se seguirá comprando casi igual aunque suba el precio: encarecer el suministro de agua apenas reduce su consumo, mientras que encarecer un servicio fácil de sustituir ahuyenta a los usuarios y destruye más bienestar. Concretamente:

- λ mide la dificultad de cumplir la restricción de viabilidad del monopolista. De esta manera, si $\lambda = 0$, no existe razón para incumplir la eficiencia asignativa y $P = Cmg$. Por el contrario, a medida que λ tiende a ∞, la restricción de viabilidad se vuelve cada vez más condicionante y el precio tiende al precio de monopolio.
  También puede interpretarse como cualquier otro multiplicador Lagrangiano. Indica cuánto varía la función objetivo cuando se relaja la restricción. En este caso, relajar la restricción implica permitir al monopolio imponer un nivel general de precios mayor y obtener así unos beneficios mayores, por lo que la función objetivo (excedente social) se reduciría.
- $\lambda/(1+\lambda)$ es el "número Ramsey" y representa el nivel general de precios de la empresa regulada . $\lambda/(1+\lambda)$ determina cual es el mark up óptimo.
  Además, la Regla de la elasticidad inversa de la demanda, según la cual el margen precio-coste debe ser proporcionalmente inverso a la elasticidad-precio de la demanda y mayor cuanto mayor sea el valor de λ, que es el coste en términos de bienestar de la restricción de beneficio. La «regla de Ramsey-Boiteux», que fue desarrollada en EDF por Marcel Boiteux (1956) (uno de los ingenieros economistas franceses más célebres, que luego fue presidente de EDF y padre de su programa nuclear) y que presenta numerosas analogías con las propuestas sobre fiscalidad presentadas por Frank Ramsey (1927), alumno de Keynes, matemático, filósofo y economista que murió a los 26 años.

Por tanto, a menor elasticidad mayor será el precio establecido del bien. Es decir, mayor participación tendrá dentro del nivel general de precios de la empresa. El porcentaje de desviación de cada precio respecto a su CMg multiplicado por la elasticidad demanda-precio ha de ser igual para cada bien.

del bien 1 genera pérdidas, pero se mantiene porque recibe un 'subsidio' procedente del bien 2–, por lo que pueden dar lugar a una configuración del mercado no sostenible[30].

4. Su implementación práctica tendería a ir acompañada por otras medidas, por ejemplo, de política fiscal.

Mecanismo de Loeb y Magat (Loeb y Magat, 1979)[31]

El regulador permite al monopolista fijar el precio que desee, comprometiéndose a entregarle un subsidio equivalente a todo el excedente del consumidor que genere dicho precio. De esta forma:

- Ventaja: el monopolista tendrá incentivos a fijar un precio igual al coste marginal y producir la cantidad de competencia perfecta para maximizar el excedente del consumidor y, por tanto, sus beneficios.
- Desventajas: i) requiere que el regulador conozca con precisión la demanda para fijar la cuantía del subsidio y ii) implica una redistribución a favor del monopolista difícil de justificar políticamente.

Óptimo de tercer orden

Por último, existe un último tipo de regulación que es el óptimo de tercer orden. Así, si la información no es completa para restablecer la eficiencia el regulador intentará al menos limitar los beneficios extraordinarios.

Regulación de la tasa de beneficios

Para ello se puede limitar la tasa de beneficios derivada de la rentabilidad del capital invertido. Este método tiene

[30] Para evitar los subsidios cruzados, la clave es la transparencia y la separación.

- Separación contable: Las empresas deben llevar registros separados para cada servicio, asignando los costes de forma precisa. Esto evita que los ingresos de una actividad rentable financien a otra.
- Precios por costes: Las tarifas deben reflejar los costes reales de producción de cada servicio, sin inflar unos para subsidiar otros.
- Subsidios directos: En lugar de subsidios cruzados, el gobierno puede dar ayudas directas a los consumidores que lo necesiten, pagadas con impuestos.
- Fiscalización: Los entes reguladores deben auditar las cuentas y sancionar cualquier práctica anticompetitiva.

[31] Otro mecanismo similar es el de Baron y Myerson (Baron y Myerson, 1982): Este mecanismo tiene una cierta similitud con el de Loeb y Magat, pero no es tan poco equitativo. La idea es la siguiente: El regulador ofrece a la empresa la posibilidad de revelar sus costes. Si lo hace, le concederá una "renta informacional", que se manifiesta en que se permite a la empresa elevar su precio en un determinado margen por encima del CMg, de modo que obtenga unos ciertos beneficios extraordinarios. Además, la renta informacional concedida será mayor cuanto más eficiente sea la empresa, de modo que se incentive una revelación honesta de información.

 © 2026 Manuel Cabado García

- Ventaja: fácil observabilidad de los datos de beneficios (al menos en sentido contable) y del capital invertido.
- Desventaja: pueden conducir a una ineficiencia por una sobrecapitalización o efecto Averch-Johnson (Averch y Johnson, 1962). Por lo tanto, una empresa sometida a este tipo de regulación elegirá una combinación de capital y trabajo ineficiente, lo que se traducirá en mayores costes y, por ende, en precios más elevados.

Analíticamente,

$$\text{máx} \quad \pi = IT - w \cdot L - r \cdot K \tag{8.9}$$

sujeto a la tasa de rendimiento permitida.

$$\frac{IT - w \cdot L}{K} = r \leq s \tag{8.10}$$

Si la tasa de rendimiento permitida es superior al coste del capital (lo cual es plausible pues si no, no se invertiría nada), $\pi = (s - r) K > 0$[32], implica que los beneficios se incrementarán aumentando K, lo que le llevará a una sobrecapitalización.

Gráficamente, en un plano de beneficios y capital como el de la Figura 8.6 se representa la curva de evolución de los beneficios, π, en función del capital y una línea recta que muestra la restricción del regulador, es decir, los máximos beneficios permitidos para cada unidad de capital. Todos los puntos en la recta o por debajo de ella serán aceptables por los reguladores, ya que se cumplirá la restricción

Si la empresa maximizara beneficios elegiría un nivel de capital K^*. Sin embargo, dada la restricción estos no son alcanzables pues exceden a los permitidos por lo que elegirá el nivel K' con unos beneficios π'.

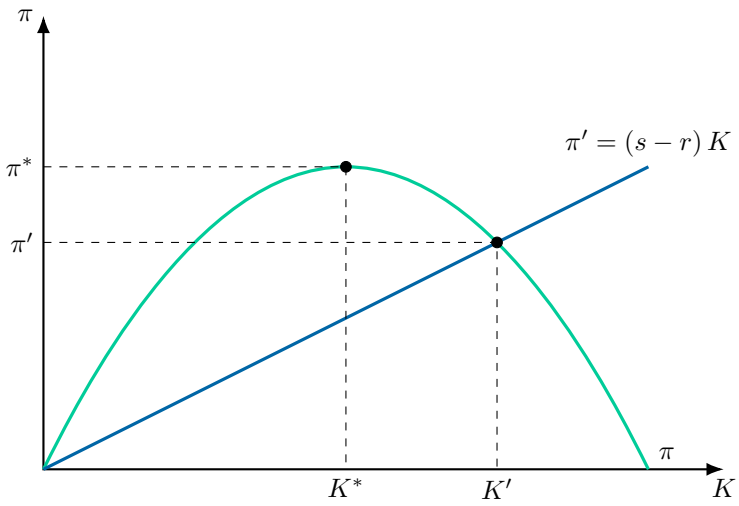
Sin embargo, todos los sistemas de fijación de precios vistos hasta ahora presentan los siguientes problemas:

- Problemas de información: el regulador requiere un conocimiento relativamente preciso de las funciones de coste, de las funciones de demanda, de las preferencias sociales, etc.
- Problemas de incentivos: si el regulador utiliza la información de costes que le proporcionan las empresas, éstas tendrán incentivos a comunicarle costes superiores a los reales y a no esforzarse en rebajar los costes.

[32] Operando en la ecuación de tasa máxima de retorno:

$$\begin{aligned} (IT - w \cdot L)/K &= s \\ (IT - w \cdot L) &= K \cdot s \\ IT - w \cdot L - r \cdot K + r \cdot K &= K \cdot s \\ \pi &= (s - r) K \end{aligned}$$

Figura 8.6: Efecto Averch-Johnson



En consecuencia, en ocasiones es preferible introducir cierta flexibilidad, generando incentivos para que las empresas actúen de forma eficiente, que serán las soluciones que se verán a continuación.

Menú de contratos: Modelo de Laffont y Tirole (1986)

Consiste en un modelo que endogeniza la función de costes basándose en la teoría de la agencia, donde el gobierno es el principal y el empresario el agente.

Parte del supuesto de que el gobierno, que busca maximizar el excedente social, obtiene un bien público de la empresa privada que financia a través de una transferencia, TR. El regulador puede observar los costes, C, que ha generado la empresa durante el periodo pasado pero ignora el esfuerzo necesario para reducirlos, siendo dicho esfuerzo completamente conocido por la empresa. Así, la empresa, al iniciar cada periodo, anunciará su previsión de costes, C^a.

Laffont y Tirole (Laffont y Tirole, 1986) proponen que el regulador ofrezca a la empresa una transferencia de la forma

$$TR = a + b \cdot (C - C^a) \tag{8.11}$$

donde a es un pago fijo garantizado con independencia del resultado, y b es la proporción de la desviación entre el coste real, C, y el anunciado, C^a, que el regulador se compromete a cubrir. En general, a y b no son constantes fijas, sino que el regulador los define en función del propio coste anunciado —$a\,(C^a)$ y $b\,(C^a)$—, de modo que a cada posible anuncio le corresponde un contrato distinto: cuanto peor se declare la empresa (mayor C^a), menor será el pago fijo que el regulador esté dispuesto a garantizarle.

- Si $b = 0$, el regulador paga solo la cantidad fija a y nada por el déficit: la empresa

 © 2026 Manuel Cabado García

asume todo el riesgo de coste, por lo que tiene un gran incentivo a reducirlo, pero a cambio exigirá una parte fija elevada.

- Si $b = 1$, el regulador cubre íntegramente cualquier desviación de coste: la empresa no asume riesgo ni tiene incentivos para reducir costes, por lo que el regulador puede pagar una parte fija menor.

Ofreciendo varios pares (a, b) a la vez — un menú de contratos—, el regulador consigue que sea la propia empresa quien revele su estructura de costes al elegir. Si la empresa prevé que puede reducir costes, preferirá un contrato con $b = 0$ — el contrato "Fixed price"—, ya que le compensa asumir el riesgo a cambio de quedarse con todo el ahorro. Si, en cambio, considera que ya no puede reducir mucho los costes, preferirá asegurarse por completo con $b = 1$ — el contrato "Cost Plus"—.

Price cap, IPC-X o techo de precios

Se trata de un procedimiento interesante para abordar el trade-off entre eficiencia asignativa y eficiencia productiva[33]. Se aplicó por primera vez en 1984 a British Telecom. El mecanismo de regulación es el siguiente:

1. El regulador establece un nivel inicial de ingresos que permite recuperar los llamados "costes razonables" de la empresa regulada.
2. A partir de esos ingresos, y para un determinado nivel de demanda, se define un precio medio.
3. Se establece una regla de evolución de dichos precios. Concretamente, se permite que los precios crezcan a una tasa previamente establecida: $IPC–X$, donde el IPC mide la inflación (la tasa de crecimiento de los precios) y X es la tasa de crecimiento de productividad de la industria fijada en el momento t_0, de forma que cuanto mayor sea la productividad de la industria, menor será el precio (es decir, las ganancias de productividad se trasladan al consumidor)[34]:

$$P_t = P_{t-1}\,(1 + IPC - X) \tag{8.12}$$

Nótese que la ecuación fija la evolución del precio con la X pactada de antemano, no con la productividad que la empresa efectivamente alcance: por eso, si el crecimiento

[33] Los antecedentes de este método de regulación se encuentran en el informe Littlechild (Littlechild, 1983) sobre la regulación de las telecomunicaciones en Reino Unido. Del sector de las telecomunicaciones se extendió a todos los servicios públicos (agua, gas, aeropuertos, electricidad).

[34] Cuando se habla aquí de "productividad de la industria", ello no implica necesariamente que haya varias empresas. Puede seguir habiendo un único monopolista: piénsese, por ejemplo, en el titular de una autopista, en el que la productividad podría medirse por el número de vehículos que la utilizan en el momento t0. El monopolista podría aumentar su productividad (y, por tanto, sus beneficios) consiguiendo que circulen más vehículos para un nivel de costes determinado, o consiguiendo reducir los costes para un tráfico determinado.

real de su productividad supera esa X ya fijada, sus costes caen más rápido que el precio, obteniendo así beneficios adicionales. Si crece menos, sus costes no bajan lo suficiente para seguir el ritmo del precio, y se ve penalizada.

Este sistema proporciona incentivos a la empresa para mejorar su eficiencia productiva, pues la reducción de costes incrementará su nivel de beneficios. Para un valor dado de X, el incentivo a reducir costes será mayor cuanto más largo sea el paréntesis regulador (pues durante más tiempo estará obteniendo beneficios extraordinarios).

La ventaja es que se incentiva la eficiencia productiva, es decir, mínimos costes posibles. No obstante, también tiene desventajas:

- Se requiere mucha información para fijar el parámetro X.
- Conforme se acerque el final del período regulado, el interés de la empresa por reducir costes disminuye. El problema puede reducirse aumentando el paréntesis regulador, pero esto puede, a su vez, dar lugar a beneficios excesivos para las empresas reguladas.
- Se deja la calidad como variable propensa a ser manipulada para aumentar los beneficios (endogenización de la calidad).

Performance-based regulation

No obstante, ante las limitaciones de la regulación por tasa de rendimiento y del *price cap*, se ha extendido un sistema híbrido de reparto de beneficios (*earnings sharing* o escala móvil), una de las formas de la regulación por incentivos o *performance-based regulation*. Se fija una banda para la tasa de rendimiento del capital. Si la empresa la supera, el exceso se comparte con los consumidores o revierte al sector público, y si cae por debajo del límite inferior, puede solicitar la revisión de la banda.

De esta forma se permite la negociación entre ambas partes y que ambos consigan parcialmente su objetivo. A su vez, este mecanismo ofrece protección al regulado en caso de que, ante un shock externo, r caiga súbitamente[35].

[35] En el sector eléctrico, la Performance-based regulation se ha ido integrando progresivamente, especialmente en la regulación de las empresas de distribución eléctrica (monopolios naturales). La Comisión Nacional de los Mercados y la Competencia (CNMC) establece ciertos criterios de calidad de servicio que las empresas deben cumplir para recibir su retribución.

- Calidad del suministro: La CNMC usa indicadores como el TIEPI (Tiempo de Interrupción Equivalente de la Potencia Instalada) y el NIEPI (Número de Interrupciones Equivalente de la Potencia Instalada) para evaluar el desempeño de las distribuidoras. Si una empresa mejora estos índices y reduce las interrupciones del servicio a los consumidores, puede obtener una bonificación en su retribución. Por el contrario, si empeora, se enfrenta a penalizaciones económicas. El objetivo es incentivar a las empresas a invertir de forma eficiente en el mantenimiento y modernización de la red para garantizar un servicio más fiable.
- Pérdidas de energía: La regulación también establece un objetivo de pérdidas de energía en la red de distribución. Las empresas que logran reducir estas pérdidas por debajo del valor de referencia son recompensadas, ya que esto se traduce en una mayor eficiencia para el sistema en su conjunto.

 © 2026 Manuel Cabado García

Otros instrumentos

Otros instrumentos son el método *yardstick* o las subastas.

Regulación por comparación: Yardstick competition (óptimo de tercer orden)

Cuando existen notables problemas para recoger información (costes), una opción es la regulación por comparación: la retribución de cada monopolista se fija a partir de los costes de otras empresas similares que operan en mercados parecidos, de modo que ninguna puede mejorar su tarifa inflando sus propios costes[36].

- Ventajas: Este tipo de regulación reduce los problemas de información asimétrica.
- Inconvenientes: Método intrínsecamente subjetivo. El regulador debe tomar también como modelo la calidad ya que si solo se preocupa de los costes, la empresa regulada podría controlar sus costes a costa de reducir la calidad de sus productos. Existe riesgo de colusión (que las empresas vecinas mientan sobre sus costes), que puede arreglarse con un conocimiento técnico del sector por parte del regulador o con una muestra suficientemente grande[37].

Subasta de explotación de monopolios

La subasta del monopolio, también conocida como competencia a la Demsetz (Demsetz, 1968), consiste en conceder la explotación de un monopolio durante un tiempo determinado mediante una subasta entre posibles operadores, de modo que se compite por el mercado en lugar de en el mercado[38]. La idea es que la subasta produzca competencia ex ante, y la periodicidad de la misma, competencia ex post.

Así, la subasta de la explotación del monopolio permitiría sustituir a la regulación de acceso y de precios, porque la propia subasta origina la competencia.

La ventaja es que este tipo de subastas es una manera de abordar el problema de información asimétrica inicial, incitando a los posibles entrantes a revelar su precio de reserva a cambio de la prestación de servicios.

La desventaja reside en que este mecanismo puede no ser apropiado en caso de que sea difícil especificar los estándares de cantidad y calidad, o de monitorizar al regulado. En este caso, la subasta podría conducir a un problema de riesgo moral.

Los principales problemas de la subasta del monopolio son:

1. Posible colusión de los licitadores.

[36] La intuición es la de un profesor que pone la nota según la media de la clase: nadie puede subir su calificación rebajando el nivel del examen, solo esforzándose más que los demás.

[37] Por ejemplo, establecer estándares de puntualidad en los trenes de UK, comparando con los mejores operadores de la UE o realizar comparaciones a nivel regional en el servicio.

[38] La intuición es la de la concesión de una cafetería en una estación: solo habrá una, pero varias empresas pujan por gestionarla ofreciendo los precios más bajos al viajero.

2. Ineficiencia: los competidores harán ofertas en función de los costes medios más un margen de beneficio razonable, por lo que los precios serán subóptimos.
3. Duración de la concesión si la prestación del servicio exige inversiones en activos específicos que tardan mucho tiempo en amortizarse[39]. Dificultad para especificar el contrato y todas las posibles contingencias.
4. Efecto aprendizaje si la prestación del servicio exige inversiones en activos que generan efecto aprendizaje[40].

8.6. Conclusiones ⏳ 30"

Como conclusión, se ha analizado el monopolio y su regulación para solucionar la ineficiencia que produce en la asignación. La regulación persigue el bienestar conjunto, pero es necesario también tener en cuenta los efectos de la propia regulación como, por ejemplo, la captura del regulador.

La visión estática adoptada lleva a determinar que las prácticas monopolistas distorsionan la asignación de recursos y por lo tanto queda justificada la política antimonopolio.

Sin embargo, autores como Schumpeter[41] subrayan las ganancias que los beneficios del monopolio pueden traer al desarrollo económico, relacionadas con los fondos excedentes que generan recursos para invertir en I+D o con las innovaciones que se utilizan para mantenerse por delante de los competidores y conservar la posición de poder de mercado.

[39] Duración de la concesión: Si la concesión es corta (es decir, por debajo del plazo de amortización de los activos específicos), va a haber problemas con independencia de la titularidad de los activos:

- Si es el sector público el que hace frente a la inversión, la empresa no tendrá incentivos a utilizar eficientemente el equipo (conservación, tasa media de uso, etc.).
- Si es la empresa la que hace frente a la inversión, entonces la empresa incorporará un margen sobre el precio para cubrirse del riesgo de que la concesión no sea renovada.

Si la concesión es larga (es decir, igual o por encima del plazo de amortización del activo específico), se suavizarán los problemas señalados, pero surgirán otros:

- Al ser tan largo el período de la concesión, las ventajas de la competencia potencial desaparecerán.
- A tan largo plazo la incertidumbre sobre las condiciones de demanda y de costes (progreso técnico) es tan grande que resultará imposible diseñar, poner en práctica y exigir el cumplimiento de un contrato óptimo.

[40] Efecto aprendizaje: Si los activos incorporan tecnologías muy especializadas, el primer concesionario tendrá ventajas sobre sus competidores en la segunda subasta por el aprendizaje realizado, de forma que el grado de competencia ex post puede verse mermado

[41] Visiones monopolio: i) Hicks: Visión negativa del monopolio. ii) Stigler: Escuela de Chicago: las fuerzas de mercado serán suficientes para romper cualquier barrera de entrada. iii) Schumpeter: La innovación en productos requiere incentivos. iv)Neoaustriacos: La regulación del monopolio es la que lo hace sostenible paradójicamente.

 © 2026 Manuel Cabado García

Apéndice

8.A.1. Ejemplos de las estructuras analizadas

Un monopolio es un mercado con un solo vendedor y muchos compradores:

1. Google: En el mercado de los buscadores de internet, Google tiene una cuota de mercado dominante. Aunque hay otras opciones, su posición es tan fuerte que actúa como un monopolio de facto.
2. Microsoft: Durante décadas, el sistema operativo Windows ha mantenido una posición monopólica en el mercado de PC.
3. AENA: En España, esta empresa pública es el único gestor de la mayoría de los aeropuertos.
4. Meta: Con la propiedad de Facebook, Instagram y WhatsApp, Meta domina el mercado de redes sociales y mensajería en muchas regiones.
5. De Beers: Históricamente, esta empresa controló una gran parte del mercado mundial de diamantes, lo que le permitió influir en su precio y suministro.

Un monopolio natural surge cuando es más eficiente que una sola empresa provea el servicio, debido a los altos costes fijos de la infraestructura:

1. Redes de agua potable y alcantarillado: Es ineficiente y costoso que varias empresas construyan sus propias redes de tuberías en una misma ciudad.
2. Servicios de electricidad: La infraestructura de postes, cables y subestaciones eléctricas tiene costes fijos muy altos.
3. Redes de gas natural: La red de gasoductos es una infraestructura masiva que solo es rentable para un único proveedor.
4. Líneas ferroviarias: El coste de construir y mantener las vías férreas es tan alto que un solo operador es más eficiente.
5. Redes de telecomunicaciones de fibra óptica: Aunque varios proveedores de servicios utilicen la red, la infraestructura física a menudo es propiedad de una sola entidad en una zona específica.

Un monopolio multiproducto se da cuando una sola empresa vende varios productos relacionados entre sí.

1. Apple: La empresa vende hardware (iPhone, Mac), software (iOS, macOS) y servicios (App Store, iCloud) que están interconectados, incentivando a los usuarios a permanecer en su ecosistema.
2. Compañías de telecomunicaciones: Muchas ofrecen paquetes que incluyen telefonía móvil, internet y televisión en un solo contrato.

3. Microsoft: La suite Office 365 (Word, Excel, PowerPoint) es un ejemplo clásico, donde la empresa domina el mercado de software de productividad.
4. Amazon: Controla tanto el comercio electrónico como los servicios de streaming (Prime Video) y los dispositivos inteligentes (Alexa), creando un ecosistema de productos y servicios.
5. Google: Además del buscador, ofrece servicios como Gmail, Drive, Maps y YouTube, que se refuerzan mutuamente.

Un monopolio multiplanta es una empresa con monopolio que produce su bien en varias plantas, cada una con su propia estructura de costes:

1. Empresas eléctricas: Una sola compañía puede generar electricidad en varias plantas (hidroeléctricas, nucleares, solares) y luego distribuirla, optimizando la producción entre ellas para minimizar costes.
2. Grandes farmacéuticas: Una farmacéutica con una patente exclusiva puede producir su medicamento en múltiples fábricas en diferentes países, adaptándose a las regulaciones y costes de producción de cada región.
3. Empresas petroleras: Una compañía de petróleo puede tener el monopolio en la distribución de combustible en un país, produciéndolo en varias refinerías con diferentes capacidades y tecnologías.
4. Cementeras: Una gran cementera puede operar varias plantas para abastecer distintas regiones de un país, minimizando los costes de transporte y producción.
5. Fábricas de automóviles: Una empresa con una marca dominante puede tener varias fábricas que producen el mismo modelo de coche, pero con diferentes tecnologías o costes.

Un monopsonio es un mercado con un solo comprador y muchos vendedores:

1. El gobierno como comprador de armamento: El Estado es el único cliente para las empresas que fabrican armamento especializado como misiles o aviones de combate.
2. Una fábrica en una ciudad pequeña: En una comunidad rural, si una fábrica es la única fuente de empleo, actúa como el único comprador de la mano de obra local.
3. Grandes cadenas de supermercados: Para los pequeños agricultores, un supermercado grande puede ser el único comprador de sus productos en la región, dándole poder para fijar precios.
4. Franquicias de ligas deportivas: Las ligas profesionales como la NBA o la NFL actúan como monopsonios, ya que son los únicos compradores del talento de los jugadores a nivel profesional.
5. Empresas mineras: En las regiones donde la minería es la principal actividad económica, una empresa minera puede ser el único empleador de mano de obra cualificada.

 © 2026 Manuel Cabado García

Un monopolio bilateral es un mercado donde hay un solo vendedor (monopolio) y un solo comprador (monopsonio).

1. Negociación entre un sindicato y una empresa: Un sindicato que representa a todos los trabajadores de una empresa actúa como un monopolio de la fuerza laboral, y la empresa es el único comprador de esa mano de obra.
2. Proveedor único de componentes y un único fabricante: Una empresa que produce un componente patentado esencial (monopolio) solo se lo vende a una única fábrica de aviones (monopsonio).
3. Gobierno negociando con una empresa de servicios: Cuando el gobierno necesita un servicio muy específico, como la construcción de un sistema de defensa único, puede negociar con una sola empresa que sea la única capaz de proveerlo.
4. Una empresa de electricidad estatal y un gran consumidor: Si una empresa eléctrica estatal es el único proveedor de energía (monopolio) y hay una sola gran fábrica en la región que es la única consumidora (monopsonio), se crea un monopolio bilateral.
5. Venta de una patente exclusiva: Una empresa que tiene una patente única para una tecnología avanzada (monopolio) puede negociar su venta o licencia con una sola empresa interesada en usarla (monopsonio).

8.A.2. Tipo de regulación y origen de monopolio

El tipo de regulación va a depender del origen del monopolio (partiendo de la base de que los monopolios son indeseables, a no ser que se trate de monopolios fruto de la I+D (patentes) o monopolios naturales (por ventajas en costes, como se ha visto)):

- Comportamientos estratégicos: la medida adecuada será introducir la competencia en el mercado a través del organismo de defensa de la competencia, cuyo objetivo debería ser la disolución de monopolios.
- Posesión en exclusiva de recursos (patentes, licencias, etc.): la medida adecuada será evitar alzas repentinas en los precios mediante el control de los mismos.
- Estrechez del mercado, que da lugar a un monopolio natural: este es el caso que se va a analizar con detenimiento, pues es el que resulta necesario regular:
    - No es aconsejable desmantelarlo porque, como se ha visto, el monopolio natural es deseable desde el punto de vista de la eficiencia productiva (es decir, mínimos costes posibles).
    - Pero tampoco hay que dejarlo desregulado, ya que el monopolista podría aumentar mucho el precio, resintiéndose aún más la eficiencia asignativa.
    - Pero, además, el regulador no debe excederse en su exigencia de eficiencia asignativa, ya que en el monopolio natural se produce un trade-off entre eficiencia

asignativa y viabilidad financiera, pues cuanto más se obligue al monopolista natural a comportarse de manera competitiva (es decir, a alcanzar la eficiencia asignativa, $P = CMg$), mayores serán sus pérdidas.

En conclusión, la actuación gubernamental frente a un monopolio natural no debe ser desmantelarlo, sino encontrar un equilibrio entre eficiencia asignativa y rentabilidad financiera. Además, a los fallos de mercado se contraponen los fallos del sector público en la regulación.

8.A.3. Discriminación de precios de cuarto grado (intertemporal)

Este grado de discriminación adicional a los tres que distinguió Pigou se refiere a la discriminación de precios en el tiempo: dividir a los consumidores en grupos según su impaciencia y cobrarles precios distintos según el momento en que compran. La estrategia habitual consiste en fijar un precio alto al principio —capturando a los consumidores con mayor disposición a pagar y menor paciencia— y bajarlo progresivamente con el tiempo, para atraer después al grueso del mercado, más paciente y con una demanda más elástica. Es el caso, por ejemplo, de los móviles o los videojuegos[42].

En general, esta estrategia es rentable y no plantea ningún problema para el monopolista mientras pueda ir renovando el bien —como ocurre, de hecho, con los propios móviles y videojuegos mencionados, que quedan obsoletos con cada nuevo modelo o versión—. El problema aparece cuando el bien es duradero y no existe forma de renovarlo: en ese caso, es el monopolista el que crea su propia competencia consigo mismo en el futuro, ya que cada unidad vendida hoy compite con las unidades que él mismo venderá mañana a un precio menor. Relacionada con esta cuestión, se encuentra la conjetura de Coase (Coase, 1972), la cual enuncia que, si los consumidores son lo suficientemente pacientes y prevén que, tras el primer periodo de ventas, el monopolista reducirá el precio, esperarán a que esto ocurra y por tanto el monopolista no podrá llevar a cabo discriminación.

Según Coase, si la depreciación de un bien es nula y su vida infinita, el precio acabará convergiendo al CMg, ya que para este tipo de bienes la demanda de mañana no es otra cosa que la demanda residual de hoy: a esa demanda de mañana se la puede atraer bajando el precio, pero si los agentes tienen expectativas racionales y anticipan esa bajada, no comprarán hoy. Se discrimina en función de la impaciencia temporal.

[42] Son numerosos los bienes que siguen este patrón. Un smartphone recién lanzado sale al mercado con un precio elevado que va reduciéndose progresivamente a medida que atrae a los consumidores más pacientes. Lo mismo ocurre con los videojuegos, cuyo precio de lanzamiento (70-80€ en el caso de los grandes títulos) suele caer a la mitad en 6-12 meses y aún más en rebajas posteriores. El ejemplo clásico de la literatura, sin embargo, es el de los libros: las editoriales publican primero la edición de tapa dura, cara, y solo 12-18 meses después la edición de bolsillo, mucho más barata —el propio origen histórico de la Conjetura de Coase, pensada precisamente para bienes duraderos como este—. Un último ejemplo, ya fuera del ámbito tecnológico, es la ropa de temporada: la colección nueva se vende a precio completo y, al final de temporada, se liquida con rebajas.

 © 2026 Manuel Cabado García

Al final, el monopolista baja el precio hoy y además el precio seguirá decreciendo conforme pasa el tiempo. Todos los consumidores saben la senda que seguirá el precio, pero cada uno compra el bien en un momento u otro en función de su impaciencia por el consumo de dicho bien.

Esto lleva a la paradoja de Coase: si la vida de un bien es infinita y los consumidores son lo suficientemente pacientes, el resultado tiende al de CP ($P = Cmg$). Por ende el monopolista perdería su poder. Solución (del monopolista): Obsolescencia programada o leasing.

8.A.4. Captura del regulador

La captura del regulador se produce cuando el organismo encargado de vigilar a una industria acaba actuando en defensa de los intereses de esa industria en lugar de en defensa del interés general[43].

Esto ocurre por varias razones:

- Asimetría de la información: Las empresas reguladas tienen un conocimiento mucho más profundo y detallado de su industria que los reguladores. Esto les da una ventaja significativa en las negociaciones y en la toma de decisiones.
- Puertas giratorias: Es común que exdirectivos de empresas reguladas pasen a ocupar puestos en los organismos reguladores, y viceversa. Esto crea un conflicto de intereses y una mentalidad pro-industria dentro del ente regulador.
- Presión política y lobbying: Las grandes empresas invierten enormes cantidades de dinero en presionar a los políticos y a los reguladores para que se adopten normativas que les beneficien, ya sea limitando la entrada de nuevos competidores o estableciendo precios favorables.
- Recursos limitados: Los organismos reguladores suelen tener menos presupuesto y personal que las empresas a las que deben supervisar, lo que limita su capacidad para investigar y fiscalizar de forma exhaustiva.

En suma, la regulación se vuelve menos efectiva, protegiendo a las empresas establecidas en lugar de a los consumidores. Los precios pueden ser más altos de lo necesario, se bloquea la innovación y la competencia, y se perpetúa un statu quo que beneficia a unos pocos a expensas del interés general.

[43] La teoría de la captura del regulador fue formulada por el economista estadounidense George Stigler en su influyente artículo "The Theory of Economic Regulation" (1971).
Stigler, que más tarde ganaría el Premio Nobel de Economía, desafió la idea predominante de que la regulación siempre busca el interés público. Su teoría postula que las industrias no son simplemente víctimas pasivas de la regulación, sino que activamente la buscan y la moldean a su favor. Argumentaba que los reguladores son un recurso político que las empresas pueden "comprar" para obtener beneficios, como limitar la entrada de competidores o fijar precios favorables, a expensas de los consumidores y del interés general.

Algunos ejemplos:

- UE: La regulación bancaria post-2008 fue criticada por su influencia de la industria financiera.
- España: Exministros del gobierno han ocupado puestos en grandes empresas energéticas, lo que sugiere una regulación favorable.
- EEUU: La FAA delegó la certificación de seguridad del Boeing 737 MAX a la propia compañía, con trágicas consecuencias.

A continuación, se proponen diferentes soluciones para mitigar la captura del regulador:

- Transparencia Radical: Exigir a los reguladores que publiquen todas sus reuniones, comunicaciones y decisiones con las empresas reguladas. Ejemplo: La UE exige a los grupos de presión que se registren y declaren con quién se reúnen en la Comisión Europea, aunque el sistema aún tiene fallos.
- Restricción de las "Puertas Giratorias": Implementar períodos de enfriamiento más largos y estrictos, prohibiendo que los exreguladores trabajen para las empresas que supervisaban. Ejemplo: En Estados Unidos, la ley de ética en el servicio público impone restricciones a los funcionarios que dejan el servicio gubernamental, aunque su aplicación varía.
- Aumento de Recursos y Competencias: Dotar a los organismos reguladores de mayor presupuesto, personal especializado y autonomía para que puedan investigar y fiscalizar a las empresas de forma efectiva sin depender de la industria.
- Promoción de la Competencia: Fomentar la entrada de nuevos competidores para reducir el poder de mercado de las empresas dominantes. Un mercado competitivo, por su propia naturaleza, reduce el atractivo de la captura del regulador.

Estas medidas buscan reequilibrar el poder entre los reguladores y las empresas, garantizando que el interés público prevalezca sobre el interés privado.

8.A.5. Condiciones de sostenibilidad y existencia del monopolio natural

Concretamente habría que diferenciar para industrias que producen un solo bien e industrias que produzcan varios bienes.

Un solo bien

La condición de existencia sería la subaditividad de costes, tal y como se ha visto anteriormente.

Condiciones de sostenibilidad (es decir, condiciones para que el monopolio pueda sobrevivir sin que la aparición de nuevos competidores acabe con él):

 © 2026 Manuel Cabado García

- Que la industria tenga características de monopolio natural no significa que el monopolio sea por sí solo sostenible.
- Como ocurría con la deseabilidad, que el monopolio natural sea sostenible o no va a depender de dónde se sitúe la demanda:
    - Si la demanda se sitúa en algún punto anterior a Q_0, el monopolio natural será deseable y sostenible, pues un outsider no podrá entrar ya que, para entrar, tendría que colocar un precio por debajo del precio del incumbente, y eso, para el tramo decreciente de CMe_1, es sinónimo de $P < CMe_1$, por lo que obtendría pérdidas.
    - Si la demanda se sitúa entre Q_0 y Q_1, el monopolio natural será deseable, pero su sostenibilidad dependerá de si la demanda se sitúa o no por debajo de la curva CMe_2:
        - Si la demanda queda por debajo de la curva de CMe_2 (por ejemplo D^*), el monopolista estará produciendo Q^* al precio P^*, donde el beneficio es nulo. El monopolio será sostenible, ya que si entra una empresa (lo que hará con un precio entre P_0 –precio mínimo– y P^*), la incumbente no puede acomodarse (es decir, repartirse la producción con la entrante) ya que, aun haciéndolo, la curva de CMe_2 quedará siempre por encima de la demanda –es decir, del precio–, por lo que ambas obtendrán pérdidas. Por lo tanto, a la incumbente sólo le queda como opción luchar bajando el precio (y la cantidad producida) hasta llegar a P_0 y Q_0, donde no solo el beneficio de la incumbente es nulo (como ya ocurría antes), sino también el de la outsider. Ésta anticipa esta secuencia y decide no entrar, por lo que, con D^*, el monopolio natural será sostenible y se producirá Q^* a P^*.
        - Si la demanda, para algún tramo, queda por encima de la curva de CMe_2 (por ejemplo D^*), el monopolista estará produciendo Q^* al precio P^*, donde el beneficio es nulo. El monopolio no será sostenible, ya que si entra una empresa (lo que hará con un precio entre P_0 –precio mínimo– y P^*), la incumbente inicialmente tratará de bajar el precio (y la cantidad), pero a partir de Q_A (punto de corte con CMe_2) se dará cuenta de que le sale más rentable acomodar la entrada, repartirse la producción con la entrante y colocar el precio por encima del CMe_2, que seguir luchando. En efecto, si se acomoda obtendrá beneficios positivos ($P > CMe_2$), y si lucha seguirá obteniendo beneficios nulos ($P = CMe_1$). La empresa outsider anticipa esta secuencia y decide entrar, por lo que, con D^*, el monopolio natural no será sostenible para el tramo creciente de los CMe, y entrará, al menos, otra empresa, y producirán conjuntamente donde maximicen los beneficios conjuntos (es decir, allí donde la distancia entre CMe_2 sea máxima, por

ejemplo, Q^{**} y P^{**}).

Varios bienes

Condiciones de existencia:

Complementariedad de costes: el coste marginal (o incremental) de producir un bien disminuye cuando aumenta la producción de los demás bienes de la empresa. En el monopolio multiproducto, las economías de escala dejan de ser condiciones suficientes para la existencia de subaditividad de costes[44]. Lo que se exige es que haya complementariedad de costes:

$$C\left(Q_1, Q_2\right) - C\left(Q_1\right) \geq C\left(Q_1, Q_2, Q_3\right) - C\left(Q_1, Q_3\right)$$

Economías de alcance o de gama: Las economías de alcance se refieren a la reducción del coste promedio de una empresa en la producción de dos o más productos o servicios de forma conjunta. El coste de producir en una fábrica varios bienes es menor a producir dichos bienes por separado.

$$C\left(X, Y\right) < C\left(X, 0\right) + C\left(0, Y\right)$$

Condiciones de sostenibilidad: ausencia de subsidios cruzados.La condición de sostenibilidad del monopolio natural multiproducto es que los precios fijados por el monopolista sean precios que no den lugar a subsidios cruzados.

Hay subsidios cruzados cuando los ingresos de un bien no cubren su coste incremental y esa diferencia se financia con los ingresos de otro bien, que supera su coste de producción aislada[45].

El monopolio no será sostenible porque un competidor podría entrar en el mercado del bien "que subsidia" (el que genera ingresos superiores a lo que costaría producirlo por separado) y vender a un precio inferior.

Los precios estarán libres de subsidios si se cumplen las dos siguientes pruebas:

1. Prueba del coste incremental: $P_i \cdot Q_i \geq C(Q)$–$C(Q_{-i})$, es decir, que el ingreso generado por el bien cubra al menos el coste incremental de producir el bien. Esto asegurará que el bien no recibe subsidio.
2. Prueba de no aislamiento: $P_i \cdot Q_i \leq C(Q_i)$, es decir, que el ingreso generado por el bien no supere el coste de producción de dicho bien de forma aislada. Esto asegurará que el bien no subsidia a otro bien.

[44] Esto se debe a que, en el monopolio multiproducto, podría haber deseconomías de gama que descompensasen las economías de escala.

[45] La intuición es que un bien "vive" a costa de otro: si la línea rentable paga el déficit de la deficitaria, sus clientes están pagando más de lo que costaría atenderlos por separado, y un entrante podría ofrecerles solo ese servicio más barato.

 © 2026 Manuel Cabado García

8.A.6. Estudios empíricos sobre la ineficiencia del monopolio

De Loecker, Eeckhout y Unger (2020) Este estudio (De Loecker et al., 2020), titulado "The Rise of Market Power and the Macroeconomic Implications" y publicado en el Quarterly Journal of Economics, es uno de los más influyentes sobre el poder de mercado. Los autores, Jan De Loecker, Jan Eeckhout y Gabriel Unger, estimaron los márgenes de las empresas cotizadas estadounidenses desde 1955.

- Metodología y hallazgos: Su principal hallazgo fue que los márgenes de beneficio (markups), es decir, la diferencia entre el precio de venta de un bien y su coste marginal de producción, han aumentado significativamente en Estados Unidos desde la década de 1980. En promedio, el margen agregado sobre el coste marginal pasó del 21 % en 1980 al 61 % en 2016. Este aumento no se debe a un incremento en la eficiencia, sino a un mayor poder de mercado de las empresas más grandes.
- Consecuencias: El estudio concluye que este aumento del poder de mercado tiene efectos macroeconómicos perjudiciales:
  - Distribución de la renta: Al tener mayores márgenes, las empresas se quedan con una porción más grande de la riqueza, lo que explica en gran medida la caída de la participación de los salarios en el PIB.
  - Ineficiencia asignativa: El aumento de los precios por encima de los costes genera una pérdida de bienestar social, ya que se producen menos bienes de los que serían óptimos.
  - Baja inversión: Las empresas con poder de mercado no tienen incentivos para invertir en nuevos proyectos, lo que frena el crecimiento económico.

El estudio de De Loecker, Eeckhout y Unger proporciona una base empírica sólida para argumentar que el poder de mercado es un problema económico central en la actualidad.

Autor, Dorn, Katz, Patterson y Van Reenen (2020) Este estudio (Autor et al., 2020), titulado "The Fall of the Labor Share and the Rise of Superstar Firms", es fundamental para entender la conexión entre el poder de mercado y el mercado laboral. Los autores, incluyendo al reconocido economista David Autor, se centran en el fenómeno de las "firmas superestrella" y sus efectos.

- Metodología y hallazgos: Utilizaron datos a nivel de empresa en Estados Unidos para demostrar que la concentración del mercado ha aumentado significativamente. La cuota de mercado de las empresas más grandes en varios sectores (minorista, financiero, servicios) se ha disparado. Los autores argumentan que esto ha provocado la caída del porcentaje del PIB destinado a los salarios, un fenómeno que ha desconcertado a muchos economistas.

- Mecanismo: el estudio muestra que la caída se debe sobre todo a una reasignación entre empresas. La tecnología y la globalización desplazan las ventas hacia las empresas más productivas de cada sector, que tienen márgenes altos y una participación del trabajo en el valor añadido muy baja. Al ganar peso estas empresas superestrella, cae la participación del trabajo en el conjunto de la economía, aunque no caiga dentro de cada empresa.

Este estudio empírico refuerza la idea de que la creciente concentración de mercado, impulsada a su vez por la tecnología y la globalización, es un factor clave detrás de los cambios en la distribución de la renta en las últimas décadas.

8.A.7. Tarificación dinámica

La tarificación dinámica, o precios dinámicos, es una estrategia que ajusta automáticamente y en tiempo real el precio de un bien o servicio según la demanda, la capacidad disponible y otros factores, como la hora, la temporada o los datos del propio cliente. Su objetivo principal es maximizar los ingresos y la eficiencia, optimizando el uso de la capacidad disponible.

La tarificación dinámica encaja perfectamente en la teoría del monopolio y de la discriminación de precios. En un mercado perfectamente competitivo, ningún vendedor podría ajustar sus precios de esta manera porque la competencia lo obligaría a vender al precio del mercado. Sin embargo, un monopolista, al tener poder de mercado, puede explotar la distinta disposición a pagar de los consumidores para maximizar sus ganancias.

La tarificación dinámica es, en esencia, un intento de acercarse a la discriminación de precios de primer grado (o perfecta), donde el monopolista cobra a cada cliente el precio máximo que está dispuesto a pagar. Aunque lograr la discriminación de primer grado perfecta es casi imposible, la tarificación dinámica utiliza algoritmos y datos para segmentar el mercado de manera muy granular, acercándose a ese ideal.

- Segmentación del mercado: Un monopolista que aplica tarificación dinámica divide el mercado en microsegmentos basados en su comportamiento. Por ejemplo, en el transporte compartido, una persona que necesita un viaje con urgencia durante una tormenta tiene una demanda inelástica y está dispuesta a pagar un precio más alto. El sistema de tarificación dinámica de la empresa lo detecta y aumenta el precio, capturando ese excedente del consumidor.
- Aprovechamiento del excedente del consumidor: La teoría del monopolio explica que un monopolista, al vender a un precio único, pierde ventas de aquellos que estarían dispuestos a pagar un precio más bajo que el precio de monopolio, pero superior al coste marginal. La tarificación dinámica le permite vender a precios más bajos en momentos de menor demanda, captando a esos clientes y aumentando sus beneficios globales.

 © 2026 Manuel Cabado García

En resumen, la tarificación dinámica es una estrategia de precios avanzada para monopolios y empresas con poder de mercado que les permite explotar la heterogeneidad de la demanda y acercarse a la maximización de beneficios a través de una sofisticada discriminación de precios.

La tarificación dinámica se usa en muchos sectores con demanda variable.

- Transporte: Empresas como Uber o Cabify ajustan sus tarifas en tiempo real (tarifa dinámica o surge pricing). El precio sube en horas punta o con mal tiempo, cuando la demanda aumenta, y baja en horas de menor demanda.
- Turismo: Aerolíneas y hoteles modifican los precios de billetes y habitaciones constantemente, basándose en la disponibilidad, la demanda y la antelación de la reserva.
- Entretenimiento: Las entradas para eventos deportivos o conciertos a menudo varían de precio según la demanda, llegando a ser muy caras en el momento del evento si la venta es alta.

8.A.8. Otras dudas y preguntas

¿Es mejor regular precios o cantidades? En empresas innovadoras sería más conveniente regular en precios frente a las más tradicionales donde sería conveniente regular cantidades.

¿Por qué el monopolista produce en el tramo elástico siempre? Lógica: En el tramo elástico si baja el precio la demanda aumenta más que proporcionalmente. En el tramo inelástico aún tiene margen para subir el precio puesto que la demanda aumentará menos que proporcionalmente dado que es menos sensible. Además, si solo existen costes fijos entonces los costes marginales serán 0, CMg=0.

Monopsonio: S es la oferta = Gme. A es la ganancia de excedente del consumidor y B+C la pérdida de eficiencia. Se compra menos a un precio menor por el poder de mercado.

Beneficios = Ingresos totales Para que el máximo de beneficios se iguale al máximo de ingresos totales en un monopolio, debe cumplirse una condición específica: los costes de la empresa deben ser cero.

El máximo beneficio se logra cuando el ingreso marginal (IMg) es igual al coste marginal (CMg).

El máximo ingreso total se logra cuando el ingreso marginal (IMg) es igual a cero.

Para que ambas condiciones ($IMg = CMg$ e $IMg = 0$) se cumplan simultáneamente, el coste marginal (CMg) también debe ser igual a cero. Si el coste marginal es cero, significa que la producción de una unidad adicional no incurre en costes, por lo que el coste total de la empresa es cero o muy cercano a cero.

Situaciones en las que podría darse: es muy raro que los costes marginales sean exactamente cero, pero hay situaciones en las que se acercan mucho, especialmente en las

industrias digitales.

Por ejemplo:

- Software y contenido digital: Una vez que un software o una película se ha desarrollado, el coste de producir y distribuir una copia adicional (ya sea mediante una descarga o streaming) es prácticamente nulo. El coste de desarrollo (coste fijo) es muy alto, pero el coste marginal (el coste de vender una copia más) es casi cero.
- Patentes y propiedad intelectual: Una empresa con una patente puede haber invertido mucho en investigación, pero una vez que el producto está listo, el coste de producir una unidad adicional puede ser mínimo.
- Datos y servicios en la nube: El coste de permitir que un usuario adicional acceda a un servicio de almacenamiento en la nube o a una base de datos es muy bajo una vez que la infraestructura ya está en funcionamiento.

En la siguiente tabla se recogen las principales funciones utilizadas. Se puede ver aquí cómo el IMg tiene el doble de pendiente que la demanda.

Cuadro 8.1: Fórmulas básicas

	Caso general	Caso lineal
Demanda inversa		$p(x) = a - b \cdot x$
Ingreso total	$IT = p(x) \cdot x$	$a \cdot x - b \cdot x^2$
Ingreso medio	$IT/x = p(x)$	$a - b \cdot x$
Pendiente IMe	$p'(x)$	$-b$
Ingreso marginal	$p(x) + x \cdot p'(x)$	$a - 2 \cdot b \cdot x$
Pendiente IMg	$2 \cdot p'(x) + x \cdot p''(x)$	$-2 \cdot b$

Derivación matemática de las fórmulas básicas

La fila del ingreso marginal de la tabla anterior se obtiene aplicando la regla del producto, no la regla de la cadena: si el ingreso total es $IT = P(Q) \cdot Q$, el ingreso marginal es su derivada respecto a Q:

$$IMg = \frac{d(IT)}{dQ} = \frac{d[P(Q) \cdot Q]}{dQ} = P'(Q) \cdot Q + P(Q) \cdot 1 = P(Q) + Q \cdot P'(Q) \qquad (8.13)$$

El término $P(Q)$ que "sale fuera" no procede de la regla de la cadena, sino que es exactamente el segundo sumando de la regla del producto ($P(Q) \cdot 1$).

El ingreso medio es, por definición, $IMe = IT/Q = P(Q)$ — coincide siempre con la propia curva de demanda inversa.

Derivando de nuevo, la pendiente de IMe es simplemente $P'(Q)$. La pendiente de IMg, por su parte, se obtiene aplicando de nuevo la regla del producto, esta vez sobre

 © 2026 Manuel Cabado García

$P'(Q) \cdot Q$:

$$\frac{d(IMg)}{dQ} = \frac{d[P(Q) + Q \cdot P'(Q)]}{dQ} = P'(Q) + [P'(Q) \cdot 1 + Q \cdot P''(Q)] = 2 \cdot P'(Q) + Q \cdot P''(Q) \tag{8.14}$$

Estas cuatro expresiones son exactamente las de la columna "Caso general" de la tabla anterior.

Para el caso lineal, $P(Q) = a - b \cdot Q$, se cumple $P'(Q) = -b$ y $P''(Q) = 0$. Sustituyendo estos valores en las expresiones generales anteriores se obtiene:

- Ingreso total: $IT = a \cdot Q - b \cdot Q^2$
- Ingreso marginal: $IMg = a - 2 \cdot b \cdot Q$
- Ingreso medio: $IMe = a - b \cdot Q$
- Pendiente de IMe: $-b$
- Pendiente de IMg: $-2 \cdot b$

Estos resultados coinciden exactamente con la columna "Caso lineal" de la tabla, y confirman el resultado clásico de que la curva de IMg tiene el doble de pendiente que la curva de demanda[46].

Por último, reexpresando $IMg = P(Q) + Q \cdot P'(Q)$ en términos de la elasticidad-precio de la demanda $\left(\varepsilon = \frac{dQ}{dP} \cdot \frac{P}{Q}\right)$ se obtiene $IMg = P(1 - 1/|\varepsilon|)$.

En el óptimo, $IMg = CMg$, por lo que $CMg = P(1 - 1/|\varepsilon|)$ y, despejando, se obtiene el propio Índice de Lerner ya presentado en "Equilibrio del mercado":

$$\frac{P - CMg}{P} = \frac{1}{|\varepsilon|}$$

8.A.9. Teoría del monopsonio

El monopsonio es el caso polar de poder de mercado en el lado de la demanda: un único comprador (por ejemplo, de trabajo) que se enfrenta a toda la curva de oferta del mercado, en vez de tomarla como dada al salario vigente. Para contratar una unidad adicional, el monopsonista debe subir el salario no solo al nuevo trabajador marginal, sino también a todos los ya contratados, por lo que el gasto marginal de contratar supera al salario de mercado[47].

Formalmente, si el gasto total de contratar l unidades de trabajo a un salario w es wl, el gasto marginal de contratar una unidad adicional es

$$GMg_l = \frac{\partial(wl)}{\partial l} = w + l \cdot \frac{\partial w}{\partial l} \tag{8.15}$$

[46] Véase (Pindyck y Rubinfeld, 2010), capítulo 10; (Varian, 1992), capítulo 14; y (Koutsoyiannis, 1979), capítulo 2, para las tres derivaciones completas.

[47] Véase (Nicholson y Snyder, 2017), capítulo 16, "Monopsony in the Labor Market".

que excede al salario siempre que la curva de oferta de trabajo tenga pendiente positiva ($\partial w/\partial l > 0$), a diferencia de un comprador competitivo, que toma el salario como dado y para el que $GMg_l = w$.

El monopsonista maximiza beneficios contratando hasta el punto en que el ingreso marginal del trabajo, IMg_l, iguala al gasto marginal, $IMg_l = GMg_l$, en lugar de $IMg_l = w$, como ocurriría si tomase el salario como dado. Como la curva de GMg_l se sitúa siempre por encima de la curva de oferta, el monopsonio contrata una cantidad de trabajo menor que la que se contrataría en competencia perfecta, a un salario también menor. El resultado es formalmente análogo al del monopolio: así como el monopolista fija un precio superior al competitivo produciendo por debajo del óptimo social, el monopsonista fija un salario inferior contratando también por debajo del óptimo social[48].

8.A.10. Monopolio multiproducto

En este caso, la solución dependerá de la relación que haya entre las demandas de los bienes producidos por el monopolista, y también de la relación que haya entre los costes de producción.

Demandas independientes y costes separables

Cada bien se tratará como un mercado separado.

Demandas dependientes y costes separables

El Índice de Lerner sería en este caso:

$$IL = \frac{1}{|\varepsilon_{ii}|} + \frac{P_j - CMg_j}{P_i} \cdot \frac{Q_j}{Q_i} \cdot \frac{\varepsilon_{ij}}{|\varepsilon_{ii}|}$$

Entonces:

- Si los bienes son sustitutivos, $\varepsilon_{ij} > 0$, y, por lo tanto, $IL > 1/|\varepsilon_{ii}|$, es decir, aumenta el poder de monopolio. Esto ocurre porque el monopolista tiene más dominio sobre los consumidores –estos no pueden "escapar"–.
- Si los bienes son complementarios, $\varepsilon_{ij} < 0$, y, por lo tanto, $IL < 1/|\varepsilon_{ii}|$, es decir, disminuye el poder de monopolio.

Demandas independientes y costes no separables

Esta estructura genera una situación de learning by doing, es decir, que los costes de la empresa disminuyen con la producción. De ahí que el monopolista tenga incentivos a

[48] Para el desarrollo gráfico y un ejemplo numérico completo, véase (Nicholson y Snyder, 2017), capítulo 16, figura 16.4 y ejemplo 16.4.

 © 2026 Manuel Cabado García

aumentar sus ventas reduciendo el precio de un bien, pues esta reducción se compensa con unos menores costes para el bien 2. Si el efecto aprendizaje es muy intenso, puede incluso ocurrir que el monopolista decida fijar en el bien 1 un precio que no cubra su coste incremental. Éste sería un ejemplo de subsidios cruzados (es decir, la producción del bien 1 genera pérdidas, pero se mantiene porque recibe un 'subsidio' procedente del bien 2).

8.A.11. Monopolio bilateral

Se habla de monopolio bilateral cuando existe un único vendedor y un único comprador. Por ejemplo, un trabajador de cualidades únicas y una empresa que puede necesitar sus conocimientos. En este caso, se puede llegar a diferentes soluciones:

- Soluciones no negociadas, que dependen del poder negociador de ambas partes:
  - Solución del monopolista: el monopolista fija p y q y el monopsonista lo acepta.
  - Solución del monopsonista: el monopsonista fija p y q y el monopolista lo acepta.
- Soluciones negociadas:
  - Maximización conjunta de beneficios.
  - Integración vertical.

A diferencia de las soluciones no negociadas, las soluciones negociadas conducirán a resultados óptimos.

Bibliografía

Autor, D., D. Dorn, L. F. Katz, C. Patterson y J. Van Reenen (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Averch, H. y L. L. Johnson (1962). Behavior of the Firm Under Regulatory Constraint. *The American Economic Review 52*(5), 1052–1069.

Baron, D. P. y R. B. Myerson (1982). Regulating a Monopolist with Unknown Costs. *Econometrica 50*(4), 911–930.

Baumol, W. J. (1982). Contestable Markets: An Uprising in the Theory of Industry Structure. *The American Economic Review 72*(1), 1–15.

Boiteux, M. (1956). Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica 24*, 22–40.

Coase, R. H. (1972). Durability and Monopoly. *The Journal of Law and Economics 15*(1), 143–149.

De Loecker, J., J. Eeckhout y G. Unger (2020). The Rise of Market Power and the Macroeconomic Implications. *The Quarterly Journal of Economics 135*(2), 561–644.

Demsetz, H. (1968). Why Regulate Utilities? *The Journal of Law and Economics 11*(1), 55–65.

Harberger, A. C. (1954). Monopoly and Resource Allocation. *The American Economic Review 44*(2), 77–87.

Koutsoyiannis, A. (1979). *Modern Microeconomics* (2 ed.). London: Macmillan.

Laffont, J.-J. y J. Tirole (1986). Using Cost Observation to Regulate Firms. *Journal of Political Economy 94*(3), 614–641.

Leibenstein, H. (1966). Allocative Efficiency vs. "X-Efficiency". *The American Economic Review 56*(3), 392–415.

Lerner, A. P. (1934). The Concept of Monopoly and the Measurement of Monopoly Power. *The Review of Economic Studies 1*(3), 157–175.

Littlechild, S. C. (1983). Regulation of British Telecommunications' Profitability: Report to the Secretary of State. Informe técnico, Department of Industry, London.

Loeb, M. y W. A. Magat (1979). A Decentralized Method for Utility Regulation. *Journal of Law and Economics 22*(2), 399–404.

 © 2026 Manuel Cabado García

Nicholson, W. y C. M. Snyder (2017). *Microeconomic Theory: Basic Principles and Extensions* (12 ed.). Boston, MA: Cengage Learning.

Oi, W. Y. (1971). A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly. *The Quarterly Journal of Economics 85*(1), 77–96.

Pigou, A. C. (1920). *The Economics of Welfare*. London: MacMillan and Co.

Pindyck, R. S. y D. L. Rubinfeld (2010). *Microeconomía* (7 ed.). Madrid: Pearson Educación.

Ramsey, F. P. (1927). A Contribution to the Theory of Taxation. *The Economic Journal 37*(145), 47–61.

Segura, J. (1993). *Teoría de la Economía Industrial*. Madrid: Civitas.

📝 *Buena referencia para entender algunas partes del tema como la caracterización del monopolio o la discriminación de precios.*

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell Journal of Economics and Management Science 2*(1), 3–21.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Varian, H. R. (1992). *Microeconomic Analysis* (3 ed.). New York: W. W. Norton & Company.

Williamson, O. E. (1979). Transaction-Cost Economics: The Governance of Contractual Relations. *The Journal of Law & Economics 22*(2), 233–261.

 © 2026 Manuel Cabado García

Tema 9

Análisis de mercados (III). Teoría del oligopolio: soluciones no cooperativas y soluciones cooperativas. Análisis mediante la aplicación de la teoría de juegos.

9.1. Introducción ⏳ 1'

En las economías avanzadas, numerosos estudios muestran un aumento de la concentración industrial medida por indicadores como el Índice Herfindahl-Hirschman[1], mark-ups empresariales, porcentaje de beneficios de las empresas sobre el PIB, etc. Así, existe un creciente debate público sobre si el poder de mercado se ha vuelto "demasiado grande", en especial en el sector tecnológico: ¿cuáles son, si es así, sus implicaciones a nivel microeconómico y macroeconómico?

En este nuevo panorama, la interdependencia estratégica es mayor entre las empresas que dominan el mercado, estudiada por la teoría del oligopolio.

Así, la Teoría de la Organización Industrial en general (y la Teoría del Oligopolio en particular) se va a nutrir de la Teoría de Juegos[2] como instrumento vertebrador de sus teorías. Esta se define como aquella rama de las matemáticas que estudia el comportamiento de los agentes (jugadores) que toman decisiones que les afectan mutuamente. Esta interdependencia obligará a los jugadores a considerar las posibles acciones del resto a la hora de desarrollar sus estrategias.

Por consiguiente, el objetivo de la presente exposición será el estudio de la teoría del

[1] Véase el apartado 9.A.3 del Apéndice.

[2] A lo largo del tema hay que hacer referencia explícita a la Teoría de Juegos cada vez que se define un concepto de dicha rama: "En Teoría de Juegos, el concepto X se refiere a ..."

oligopolio y de sus determinaciones de precios y cantidades a través del siguiente esquema:

1. Caracterización del oligopolio.
2. Oligopolios no cooperativos.
3. Oligopolios cooperativos.
4. Sostenibilidad del oligopolio, analizando las barreras de entrada.

9.2. Caracterización del oligopolio ⏳ 1'

En primer lugar, cabe definir qué es lo que se consideraría un juego: se trata de una representación formal en la que un determinado número de individuos interactúa en un entorno de interdependencia estratégica[3].

El oligopolio es una estructura de mercado con pocos vendedores, cada uno consciente de que sus decisiones afectan a los demás. Sus supuestos característicos son los siguientes:

1. Pocos oferentes y muchos demandantes.
2. Conocimiento común: todos los jugadores conocen las reglas del juego, saben que los demás también las conocen, saben que los demás saben que las conocen, y así sucesivamente.
3. Producto homogéneo y ausencia de sustitutivos cercanos.
4. Barreras de entrada al mercado.
5. Interdependencia estratégica y consciencia de ella por parte de las empresas. La consciencia de la interdependencia es muy importante, ya que implica que las estrategias de las empresas se basarán en variaciones conjeturales (es decir, creencia de la empresa 1 sobre cómo cambiará la variable de decisión de la empresa 2 –cantidad o precio– ante cambios en la variable de 1).

La consecuencia de los puntos 1-3 es que el oligopolista tendrá poder de mercado, es decir, se enfrentará a una curva de demanda que no será perfectamente elástica y que cumple con los supuestos de la Teoría de la Demanda. A su vez, se debe incorporar al análisis el concepto de variación conjetural, definida como la creencia que una empresa tiene sobre la forma en que sus competidores pueden reaccionar si varía su producción o precio.

A continuación, se analizarán dos tipos de oligopolios: los no cooperativos y los cooperativos.

[3]Elementos de un juego: i) Jugadores (racionales, maximizan utilidad, beneficios, etc) ii) reglas (relativas a la información y el horizonte temporal) iii) resultados iv) pagos.

 © 2026 Manuel Cabado García

9.3. Oligopolios no cooperativos ⏳ 7'30"

Se comenzará estudiando los oligopolios no cooperativos a través de análisis para dos empresas.

Variable de decisión: cantidades

Cuando la variable de decisión son las cantidades, la empresa 1 maximiza su beneficio

$$\max_{q_1} \quad \pi_1 = p\left(q_1 + q_2^e\right) \cdot q_1 - c_1\left(q_1\right) \tag{9.1}$$

tomando como dada la producción esperada de la empresa 2. La condición de primer orden resultante tiene la lectura habitual de todo problema de maximización de beneficios: la empresa debe producir hasta que su ingreso marginal iguale su coste marginal. La particularidad del oligopolio es que el ingreso marginal de la empresa 1 depende de cómo espera que reaccione la empresa 2 ante sus propias decisiones —lo que se conoce como variación conjetural—, de modo que la condición se escribe

$$IMg = CMg \Rightarrow p\left(Q\right) + q_1 \cdot \frac{\partial p}{\partial q_1}\left(1 + \underbrace{\frac{\partial q_2^e}{\partial q_1}}_{\substack{\text{Variación} \\ \text{conjetural}}}\right) = CMg \tag{9.2}$$

donde $\frac{\partial q_2^e}{\partial q_1}$, la variación conjetural, recoge precisamente esa conjetura: cómo cree la empresa 1 que cambiará la producción de la 2 ante una variación infinitesimal de su propia producción. Será precisamente el valor que se asuma para esta variación conjetural lo que distinga el caso de Cournot del de Stackelberg, como se verá a continuación.

Modelo de Cournot

Comenzando por el modelo de Cournot (1838), este sigue los siguientes supuestos:

1. Dos empresas que toman sus decisiones simultáneamente y producen unas cantidades q_1 y q_2.
2. Existe miopía en la toma de decisiones, es decir, la variación conjetural es nula.
3. La demanda es lineal: $P\left(q_1, q_2\right) = a - b\left(q_1 + q_2\right)$ y $Q = q_1 + q_2$.
4. Pendiente finita de la demanda lo que demuestra que hay cierto poder de mercado.
5. Costes marginales idénticos y constantes, es decir, rendimientos constantes a escala.
6. Producto homogéneo.
7. Subastador. Elige el precio al que se vacía el mercado para la oferta y la demanda.

A partir de las condiciones de primer orden, CPO, se pueden obtener las funciones de reacción de las empresas: en teoría de juegos, la función de reacción, $\varphi_1\left(q_2\right)$, indica el

nivel de producción, en este caso de la empresa 1, q_1, que maximiza su beneficio dado el nivel de producción de la empresa 2, q_2^e

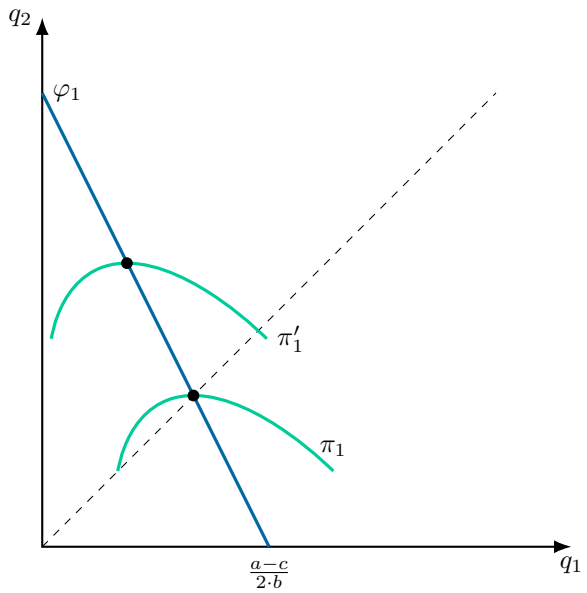
$$\varphi_1(q_2) = q_1 = \frac{1}{2} \cdot \frac{a - Cmg_1}{b} - \frac{1}{2} \cdot q_2^e \tag{9.3}$$

Las funciones de reacción de ambas empresas son simétricas ya que los costes marginales son iguales y tienen pendiente negativa, pues un aumento de q_1 implica una disminución de q_2 ya que ambas empresas se reparten el mercado.

Gráficamente, en un plano de q_1 y q_2 como el de la Figura 9.1

- Las funciones de reacción pueden obtenerse a partir de las curvas isobeneficio: lugar geométrico de combinaciones del nivel de producción de ambas empresas que proporcionan el mismo nivel de beneficio a una de las empresas.
- Cada curva isobeneficio de la empresa 1 alcanzará un máximo en el punto de intersección con su función de reacción y representarán un mayor beneficio según se desplacen hacia su eje de producción.
- Concretamente, la empresa 1 obtiene su mayor beneficio cuando la empresa 2 produzca 0, $q_2 = 0$, es decir, cuando se sitúe sobre su eje de producción[4], produciendo una cantidad de monopolio $\frac{1}{2} \cdot \frac{a-c}{b}$.

Figura 9.1: Obtención gráfica de la curva de reacción



Analíticamente, el equilibrio se obtiene igualando las funciones de reacción, obtenién-

[4] Aclaración: que se sitúe sobre su eje de producción quiere decir que la curva isobeneficio se sitúe sobre su eje de producción, no que el punto máximo de la curva isobeneficio se sitúe sobre su eje de producción.

 © 2026 Manuel Cabado García

dose así la cantidad de producción para cada empresa, donde c es el coste marginal

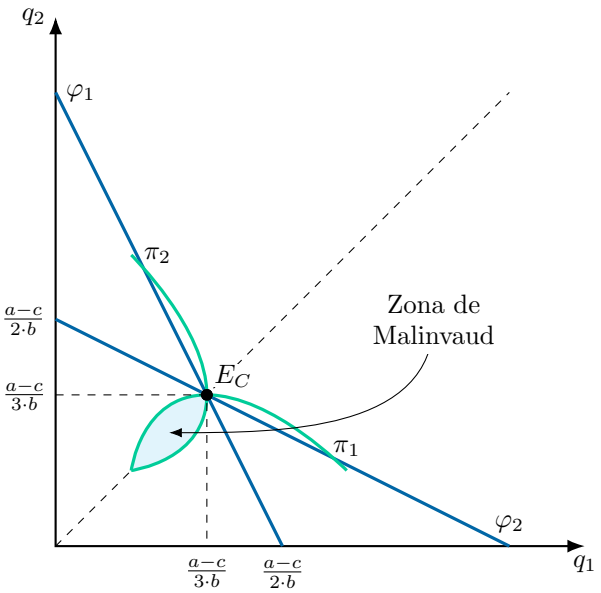
$$q_i = \frac{1}{3} \cdot \frac{a-c}{b}, \quad i = \{1, 2\} \tag{9.4}$$

por lo que la cantidad total producida en Cournot será

$$Q_C = \frac{2}{3} \cdot \frac{a-c}{b} \tag{9.5}$$

Gráficamente, como se observa en la Figura 9.2, el equilibrio coincidirá con la intersección de ambas funciones de reacción, ya que ambas empresas deciden simultáneamente, y la intersección coincidirá con la bisectriz del plano, pues ambas empresas tienen la misma estructura de costes.

Figura 9.2: Equilibrio en el modelo de Cournot



Equilibrio de Cournot Por otro lado, el equilibrio de Cournot tiene una serie de implicaciones

1. Es un equilibrio de Nash, que en terminología de teoría de juegos es una situación en la que cada jugador juega su estrategia mejor respuesta, dada la estrategia de los demás, por lo que no tiene incentivos a modificar su estrategia[5].
2. Será estable si, en el plano de la figura, con q_1 en abscisas, la pendiente de la función de reacción de la empresa 1 es mayor, en valor absoluto, que la de la empresa 2, ya que de lo contrario las curvas no convergerían al equilibrio mediante el proceso de ajuste dinámico entre ambas empresas.

[5]La intuición es que nadie se arrepiente de su decisión al ver lo que han hecho los demás, como dos vecinos que, sabiendo a qué hora sale el otro, no cambiarían su hora de salida para evitar el atasco.

3. No es un equilibrio eficiente en el sentido de Pareto, ya que ambas empresas estarían mejor cooperando. Esto se puede observar en el hecho de que las curvas isobeneficio de ambas empresas no son tangentes. En la lente formada por las curvas isobeneficio π_1 y π_2, denominada zona de Malinvaud (1972)[6], ambas empresas obtendrían mayores beneficios.
4. Misma producción, si los costes marginales de ambas empresas son iguales.

Modelo de Stackelberg

Por otro lado, en el modelo de Stackelberg (von Stackelberg, 1934) se parte de los mismos supuestos de Cournot, pero con una asimetría en el orden de las decisiones: existe una empresa líder[7], que decide primero, y una empresa seguidora, que decide después observando la decisión de la líder. La empresa líder conoce la función de reacción de la seguidora, la cual mantiene una variación conjetural nula por lo que es un juego secuencial (análisis dinámico) donde la empresa líder decidirá primero. La variación conjetural de la líder, en cambio, es distinta de cero y negativa, ya que anticipa que cada unidad adicional que produzca reducirá la producción de la seguidora[8].

Analíticamente, el problema maximizador de beneficios de la empresa líder sería el mismo que el de Cournot, solo que ahora estaría sujeto a una restricción: la función de reacción de la empresa seguidora, la cual es conocida para la empresa líder

$$q_2 = \varphi_2 (q_1) \tag{9.6}$$

Las condiciones de primer orden incorporan esa variación conjetural negativa, que con demanda lineal vale $\frac{\partial q_2}{\partial q_1} = -\frac{1}{2}$, y, sustituyendo en función de q_1, se obtiene la cantidad

[6] Cap. 6 ("Imperfect competition and game situations"), sección 3 ("Duopoly"), pp. 150-153. Malinvaud demostró geométricamente que el equilibrio de Cournot cae dentro del arco de tangencias de las curvas isobeneficio, nunca sobre él: "the Cournot equilibrium C is represented by a point which lies inside RS" (p. 153).

[7] La clave del equilibrio de Stackelberg es la credibilidad y el compromiso del líder: para que su elección de producción resulte creíble para la seguidora, la empresa líder debe respaldarla con decisiones visibles e irreversibles —inversión en capacidad, contratos a largo plazo— que le impidan desviarse de lo anunciado una vez la seguidora ha reaccionado (Tirole, 1988). Mantener esa posición de liderazgo en el tiempo exige, además, vigilar el comportamiento de los seguidores y reforzar las barreras de entrada que protegen la posición adquirida.

[8] Liderazgo en precios: al contrario que el resultado obtenido en liderazgo en cantidades, en este caso, las empresas preferirán ser las seguidoras en la fijación de precios. La razón es simple: una vez la líder fija el precio, la seguidora puede actuar como free-rider y produce tanto como desee dado el precio. A cambio, la líder se ha de acomodar a la selección de producción de la seguidora para asegurar que el precio es el establecido.

 © 2026 Manuel Cabado García

óptima de producción de la empresa líder, que será la cantidad de monopolio[9]

$$q_1 = \frac{1}{2} \cdot \frac{a-c}{b} \tag{9.7}$$

Sin embargo, el beneficio no será el de monopolio ya que habrá que sumar la producción de la seguidora, lo que reduce el precio respecto a la situación de monopolio. Sustituyendo la cantidad que produce la líder en la función de reacción de la seguidora se obtiene

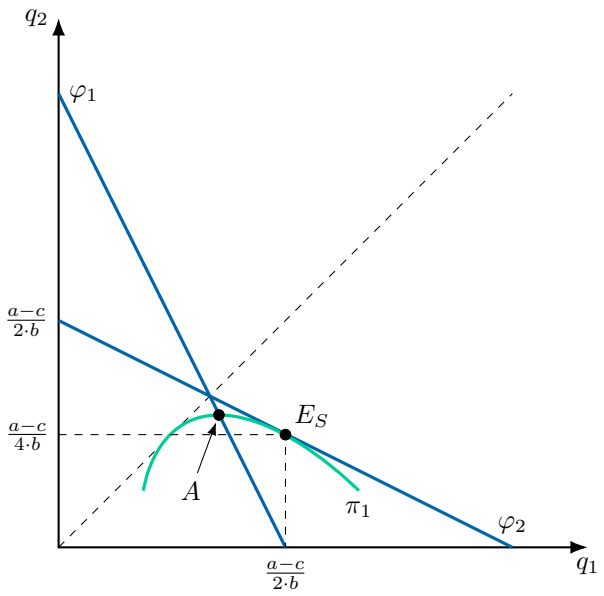
$$q_2 = \frac{1}{4} \cdot \frac{a-c}{b} \tag{9.8}$$

Por consiguiente, sumando (9.7) y (9.8) la cantidad producida en Stackelberg será

$$Q_S = \frac{3}{4} \cdot \frac{a-c}{b} \tag{9.9}$$

Gráficamente, en un plano de cantidad producida por la empresa 1 y la 2 como el de la Figura 9.3, el equilibrio no se produce en la intersección de las curvas de reacción ya que la líder, al elegir primero, se situará en la curva de isobeneficio más baja que sea tangente a la curva de reacción de la seguidora.

Figura 9.3: Equilibrio en el modelo de Stackelberg



Equilibrio de Stackelberg En cuanto a las implicaciones del equilibrio de Stackelberg

1. El equilibrio de Stackelberg, E_S, es un equilibrio de Nash, pero no es eficiente en

[9] La intuición es que la líder sabe que, cuanto más produzca, menos espacio dejará a la seguidora, igual que el primer comercio que abre en un barrio con un gran local sabe que el que llegue después abrirá uno más pequeño.

el sentido de Pareto porque la seguidora podría obtener mayores beneficios a través de la cooperación sin disminuir los beneficios de la líder en el punto A.

2. Pérdida de eficiencia menor que en Cournot porque la cantidad global producida es mayor: $Q_S > Q_C$.

Conclusión de los modelos de Cournot y Stackelberg

En suma, en ambos modelos analizados, como consecuencia del comportamiento miope y de que son juegos de una sola vez (no repetidos), no se maximizan los beneficios conjuntos. Por lo tanto, a largo plazo es poco probable que sigan compitiendo y no colaboren.

Variable de decisión: precios

Por otro lado, dentro de los oligopolios no cooperativos la variable de decisión de las empresas también puede ser los precios. Para ello, se analizará el duopolio simultáneo de Bertrand (1883). Concretamente, uno de los supuestos menos realistas de los modelos anteriores era el tipo de competencia ya que, a corto plazo, es más lógico competir en precios que en cantidades.

Se van a asumir los mismos supuestos que en el caso de Cournot pero ahora las empresas toman sus decisiones sobre los precios tomando como dados los precios de las demás empresas[10].

La principal implicación es que en el equilibrio si ambas empresas tienen la misma estructura de costes los precios de ambas empresas deben ser idénticos e iguales a su coste marginal:

$$p_1 = p_2 = CMg_i, \quad i = \{1, 2\} \tag{9.10}$$

dando lugar al equilibrio competitivo.

Esto ocurre ya que, como el producto es homogéneo, si una empresa fijase un precio más bajo se haría con toda la demanda. Sin embargo, la otra empresa reaccionaría también bajando el precio. Este proceso se repetiría hasta que el precio de 1 se igualase al precio de la empresa 2 y se igualase al coste marginal.

Paradoja de Bertrand Por lo tanto, se da la denominada como paradoja de Bertrand con un equilibrio equivalente al de competencia perfecta dando lugar a un oligopolio competitivo, siendo la cantidad producida

$$Q_B = \frac{a - c}{b} \tag{9.11}$$

[10] Además, se van a suponer dos oligopolistas que ofrecen un producto homogéneo y presentan una misma estructura de costes, es decir, el coste medio será igual al marginal asumiendo también la existencia de rendimientos constantes de escala.

 © 2026 Manuel Cabado García

Equilibrio de Bertrand Concretamente, el equilibrio de Bertrand es un equilibrio de Nash, ya que ninguno de los oligopolistas puede incrementar su beneficio modificando su precio, pero no es un óptimo de Pareto, ya que las empresas podrían obtener mayores beneficios cooperando.

Críticas No obstante, su resultado es muy dependiente de los supuestos de partida, ya que la variación de cualquiera conduciría a una situación radicalmente distinta[11]:

- Si el coste marginal de ambas empresas fuera diferente, la empresa con menores costes acapararía todo el mercado pudiendo incluso fijar precios de monopolio dependiendo de donde se sitúen los costes marginales de la empresa competidora: si el coste marginal de la rival es lo bastante bajo, la empresa 1 solo puede capturar todo el mercado fijando un precio justo por debajo de ese coste, obteniendo un margen igual a la diferencia de costes. Pero si el coste marginal de la rival es tan alto que supera el propio precio de monopolio de la empresa 1, esta no necesita preocuparse por la rival en absoluto y puede fijar directamente su precio de monopolio sin riesgo de que la rival le rebaje el precio.
- Si el producto no fuera homogéneo sino ligeramente diferenciado, las funciones de reacción dejarían de ser discontinuas y pasarían a tener pendiente positiva: los precios se convierten en complementos estratégicos, de modo que si la rival sube su precio, a la empresa también le conviene subir el suyo, justo lo contrario de lo que ocurre con las cantidades en Cournot, que son sustitutos estratégicos (Tirole, 1988)[12].
- Si no existen rendimientos constantes a escala
  - Ante rendimientos crecientes, se podría generar una guerra de precios no pudiendo existir equilibrio ya que conllevaría pérdidas.
  - Con rendimientos decrecientes, por ejemplo, derivado de límites de capacidad, la demanda residual no atendida por la empresa con menores costes que acapara el mercado, podría ser atendida por otras empresas a un precio mayor.

[11] La solución más satisfactoria sería disponer de un modelo teórico de oligopolio que, considerando que los oligopolistas compiten vía precios, condujera a un equilibrio final en el que el output total fuese inferior al competitivo. Así, Edgeworth supuso que si dos empresas instaladas compiten entre sí vía precios lo hacen con dimensiones dadas y estas dimensiones provocan restricciones de capacidad que hacen que el equilibrio final de la conjetura de Bertrand no sea el competitivo. Si la empresa 1 aumenta el precio significativamente, la empresa 2 se enfrentará a una demanda que no podrá satisfacer, por lo que la demanda residual acudirá a 1 y esto generará beneficios positivos e incitará a la empresa 2 a cambiar su precio.

No obstante, este modelo no tiene solución, aunque resuelve la paradoja de Bertrand.

[12] La intuición es que, si la gasolinera de enfrente sube el precio, a la de al lado le conviene subirlo también porque apenas perderá clientes, mientras que si una empresa rival produce más, a la otra le conviene producir menos para no hundir el precio.

Cabe preguntarse si es posible reconciliar los resultados de Cournot y Bertrand, que predicen equilibrios radicalmente distintos a partir de supuestos de comportamiento aparentemente similares. Kreps y Scheinkman (1983) ofrecieron una respuesta mediante un juego en dos etapas: en la primera, las empresas eligen simultáneamente su capacidad productiva, igual que en el modelo de Cournot. En la segunda, compiten en precios tomando esa capacidad como dada, con un mecanismo de racionamiento eficiente de la demanda no atendida. El resultado es que el equilibrio de precios de la segunda etapa conduce exactamente a las cantidades y al precio del equilibrio de Cournot: la competencia en precios, cuando está limitada por decisiones previas de capacidad, reproduce el resultado en cantidades. Esta síntesis explica por qué el supuesto de capacidad ilimitada de Bertrand es crucial para su resultado competitivo.

Teniendo en cuenta modelos dinámicos y considerando las estrategias de los competidores, el resultado será el de una situación de oligopolio cooperativo, que se analizará a continuación.

Cuadro 9.1: Resultados de oligopolio

	Producción	Beneficio	Nash	Pareto
Competencia perfecta o Bertrand	$\frac{a-c}{b}$	0	✓	×
Oligopolio de Stackelberg	$\frac{3}{4}\cdot\frac{a-c}{b}$	>0	✓	×
Oligopolio de Cournot	$\frac{2}{3}\cdot\frac{a-c}{b}$	$\gg 0$	✓	×
Monopolio	$\frac{1}{2}\cdot\frac{a-c}{b}$	$\ggg 0$	×	✓

9.4. Oligopolios cooperativos ⏳ 6'

Hasta el momento se han estudiado situaciones en las que las empresas competían entre sí en el mercado. No obstante, se ha demostrado que las empresas estarían en mejor situación si cooperasen, tomando sus decisiones de fijación de precios y cantidades de forma conjunta con el fin de actuar en la práctica como un único monopolista y maximizando los beneficios conjuntos a repartir. Cabe distinguir entre colusión explícita y tácita.

 © 2026 Manuel Cabado García

Colusión explícita: cártel

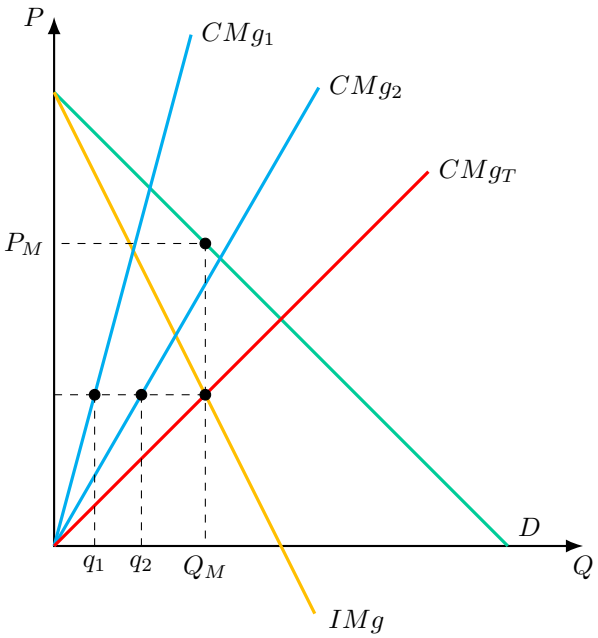
Cuando la colusión es explícita, se constituye lo que se denomina como un cártel. Ejemplos de estos van desde los gremios medievales y la OPEP hasta los cárteles del narcotráfico[13].

Dentro de un cártel, los oligopolistas renuncian a su autonomía delegando la toma de decisiones en una agencia central de cártel que establece el reparto de cuotas del mercado. El problema del cártel es análogo al del monopolio multiplanta ya que existirán varias estructuras de costes.

Gráficamente, en un plano de precios y cantidades como el de la Figura 9.4 se representan:

- Una única demanda, D, e ingreso marginal, IMg, y, suponiendo dos oligopolistas, las curvas de costes marginales de las empresas de cuya suma se determina la curva de costes marginales total, CMg_T.
- De la intersección de los costes marginales con los ingresos marginales y proyectando hasta la demanda, se establecerá el precio de equilibrio del mercado, P_M, y proyectando el equilibrio hacia las estructuras de costes de las empresas cada una de las cuotas de mercado, q_1 y q_2.

Figura 9.4: Cártel como monopolio multiplanta



Sin embargo, será necesario estudiar la sostenibilidad de la cooperación, ya que no

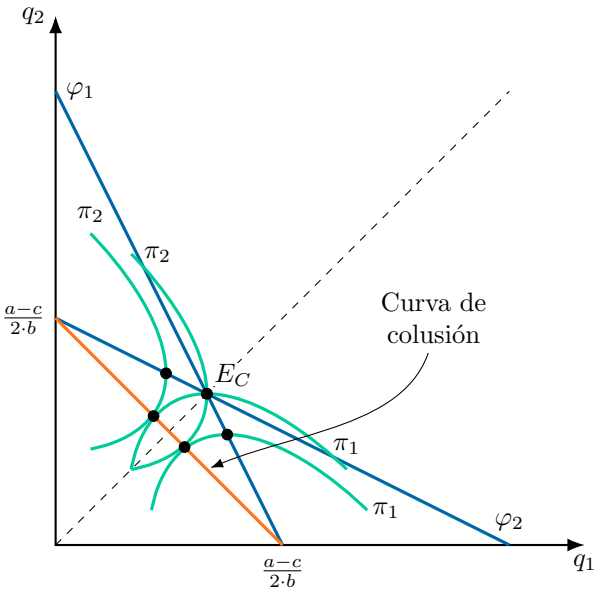
[13]Las organizaciones criminales controlan la producción, distribución y precio de drogas ilícitas. Además, realizan acuerdos explícitos sobre territorios, cuotas de producción, precios y cooperación para eliminar rivales o evadir a las autoridades. Finalmente, utilizan la violencia como mecanismo de cumplimiento.

siempre será posible llegar a pactos que permitan llegar a un cártel sostenible ya que hay una serie de factores que dificultan la colusión como

1. Reparto de beneficios, donde la solución solo es sostenible si ambas empresas mejoran.

    - Gráficamente, en un plano de cantidades de 1 y 2 como el de la Figura 9.5 se representan las curvas isobeneficio de cada una de las empresas. La curva de colusión es el lugar geométrico de los puntos en que una isobeneficio de la empresa 1 es tangente a una de la empresa 2, y une las cantidades de monopolio de ambas empresas[14].
    - El cártel solo maximizará beneficios en cualquiera de los puntos sobre dicha recta mientras que fuera de la misma el cártel podría ser inestable.

Figura 9.5: Curva de colusión



2. Superioridad absoluta en costes de una empresa, donde la empresa con menores costes producirá la cantidad de monopolio y el resto de empresas no producirá nada.

    - La empresa productora debería otorgar pagos compensatorios al resto de empresas por su cooperación.
    - Sin embargo, esta solución es poco creíble ya que pocas empresas estarán dispuestas a cerrar para aceptar después pagos compensatorios.

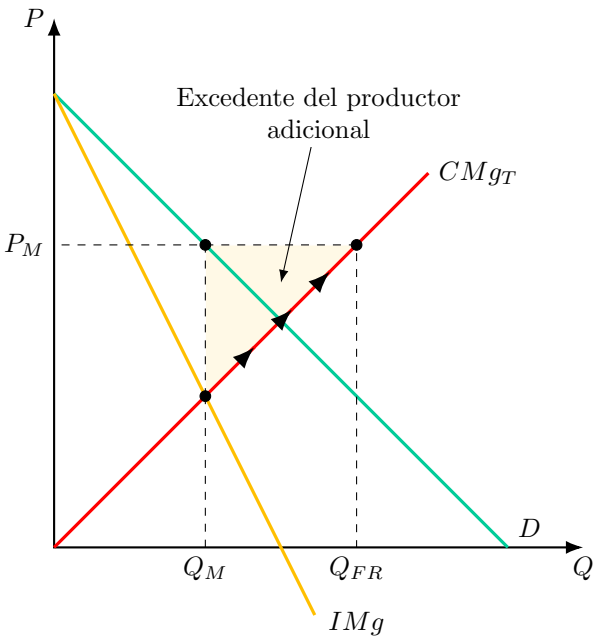
[14] La intuición es que en esos puntos ya no es posible mejorar a una empresa sin empeorar a la otra: cualquier reparto sobre la curva agota las ganancias de cooperar.

 © 2026 Manuel Cabado García

3. Número elevado de empresas, ya que cuantas más empresas participen en el oligopolio más costosa será la cooperación.
4. Tentación a la violación de la colusión, ya que la cooperación vía cártel no es un equilibrio de Nash, por lo que existirán incentivos a desviarse del acuerdo.

    - Gráficamente como se observa en la Figura 9.6 la empresa actuará como monopolista sobre un n-ésimo de la demanda.
    - Sin embargo, encontrará más beneficioso actuar como un free-rider, es decir, aumentar la producción hasta que se iguale el precio al coste marginal del resto del monopolio en Q_{FR} y apropiándose del excedente del productor adicional[15].
    - Por lo tanto, la empresa preferirá incumplir el acuerdo tanto si el resto de empresas lo cumplen como si no, dándose lo que en teoría de juegos se conoce como estrategia dominante. Una estrategia dominante es aquella que proporciona a un jugador los mayores pagos independientemente de lo que hagan sus rivales.
    - Así, la no cooperación sería el equilibrio de Nash, pero no un óptimo de Pareto, ya que todas las empresas podrían obtener mayores beneficios cooperando. Este último punto es lo que se conoce en teoría de juegos como dilema del prisionero.

Figura 9.6: No cumplimiento del acuerdo de colusión



[15]Esto se puede dar ya que, a diferencia de lo que ocurre en monopolio, el aumento de la producción no reduce tanto el precio, debido a que la empresa solo concentra una parte de la producción

Dilema del prisionero[16]

El dilema del prisionero se puede representar a través de una matriz de pagos de teoría de juegos como la del Cuadro 9.2 donde se representan las dos empresas y sus estrategias sobre cooperar o no cooperar. La estrategia dominante de ambas empresas es la no cooperación, dando lugar a una situación similar a la de Cournot[17].

Cuadro 9.2: El dilema del prisionero

		Empresa 2	
		Cooperar	No cooperar
Empresa 1	Cooperar	5,5	1,10 ✓
	No cooperar	✓10,1	✓2,2 ✓

Soluciones El cártel solo perdurará en el tiempo si se aplica alguna de las soluciones al dilema del prisionero, como

1. Apelar a un Agente externo para que refuerce el acuerdo y que pueda imponer soluciones
2. Altruismo o forjar una relación de confianza entre empresas, es decir, si los agentes se ponen en el lugar del otro
3. Interrelaciones prolongadas en el tiempo indefinidamente o superjuego[18], de modo que, siempre que el futuro se valore lo suficiente, respetar el acuerdo puede sostenerse como equilibrio.

Estrategias trigger o gatillo Esto último se dará si en el acuerdo se recoge una estrategia tipo trigger o gatillo que en teoría de juegos es aquella estrategia en la que las empresas cooperarán mientras todos cooperen y dejarán de hacerlo cuando alguna no coopere:

[16] Se observa como la estrategia dominante en ambos es la de no cooperar (hacer trampas) y el resultado final será el correspondiente a la solución de Cournot, reflejando el dilema que supone que los óptimos no sean estables. En base a estos resultados, la Teoría de la Organización Industrial señala la poca transcendencia de las coaliciones con el objetivo de restringir la competencia en el mercado por la dificultad de sostenerse en el tiempo (Escuela de Chicago).

[17] La lógica de la matriz de pagos es la siguiente: desde la óptica de la empresa 1, si la empresa 2 decide cooperar, entonces 1 no coopera. Sin embargo, si la empresa 2 decide no cooperar, entonces 1 decide no cooperar. En el caso de la empresa 2 el razonamiento es análogo.

[18] Aquí el comportamiento de las empresas puede estar guiado por dos tipos de estrategias:

a) Estrategias incondicionadas: No muestran dependencia histórica. En consecuencia, son la repetición del EN estático en cada periodo.

b) Estrategias condicionadas: Muestran una dependencia histórica, por lo tanto, las acciones podrán cambiar en base a lo observado en los anteriores periodos.

 © 2026 Manuel Cabado García

- Así, la empresa cooperará –y el cártel será sostenible– cuando el flujo de beneficios en el tiempo de la estrategia a cooperar, π^C, sea mayor que el beneficio de romper el acuerdo, π^{FR}, más el flujo de beneficios en el tiempo de no cooperar, π^{NC}, donde β representa el factor de descuento

$$\frac{\pi^C}{1-\beta} \geq \pi^{FR} + \beta \cdot \frac{\pi^{NC}}{1-\beta} \tag{9.12}$$

  es decir, las empresas solo respetarán el acuerdo si las ganancias perpetuas de cooperar son mayores que las ganancias perpetuas de no hacerlo[19].

- Este umbral de paciencia se generaliza fácilmente a un cártel de n empresas simétricas: la colusión es sostenible si el factor de descuento cumple

$$\beta \geq 1 - \frac{1}{n} \tag{9.13}$$

  (Nicholson y Snyder, 2017)—cuantas más empresas formen el cártel, mayor debe ser su paciencia para que el acuerdo se mantenga, ya que cada una se queda con una porción menor de los beneficios colusorios y el incentivo relativo a desviarse crece.

- De esta manera, cuanto mayor sea el factor de descuento, que representa el grado de paciencia de las empresas, más posibilidades habrá de que cooperen y de que el cártel sea sostenible. No obstante, para que la estrategia tipo trigger funcione es necesario que la amenaza de no cooperación sea creíble, es decir, que cuando una empresa haya actuado como free rider el resto obtengan mayores ganancias no cooperando que cooperando.

El resultado anterior asume que las empresas observan directamente si un rival ha incumplido el acuerdo, lo que garantiza que el castigo sea inmediato y la amenaza, creíble. Sin embargo, en muchos mercados reales las empresas solo observan el precio de mercado, sujeto además a perturbaciones de demanda ajenas a su comportamiento. Green y Porter (1984) analizaron cómo sostener la colusión en ese contexto de información imperfecta: si el precio cae por debajo de un umbral determinado, las empresas no pueden distinguir si se debe a una caída de la demanda o a que un rival ha bajado precios de forma encubierta, por lo que desencadenan igualmente una guerra de precios de duración limitada como castigo, tras la cual se retoma la cooperación. La colusión sigue siendo sostenible en equilibrio, pero a costa de guerras de precios periódicas que se producen incluso cuando nadie ha incumplido realmente el acuerdo.

Frente a esta fragilidad interna de los cárteles, las autoridades de defensa de la com-

[19] Este resultado, analizado por James Friedman (1971), es una aplicación concreta del teorema folk, según el cual, con horizonte temporal infinito y un factor de descuento suficientemente alto, cualquier pago medio factible superior al del equilibrio de Nash estático puede sostenerse como equilibrio del juego repetido. En particular, si se cumple la condición anterior, las estrategias trigger constituyen un equilibrio de Nash perfecto en subjuegos cuyo pago es Pareto superior al del equilibrio de Nash estático.

petencia han desarrollado un mecanismo que la explota desde fuera: los programas de clemencia, que premian a la empresa participante en un cártel que lo denuncie —y aporte pruebas de su existencia— con la exención del pago de una eventual sanción.[20] Al ofrecer a cada miembro del cártel un incentivo adicional para desviarse que compite directamente con el incentivo a mantener el acuerdo, estos programas debilitan aún más la ya de por sí precaria sostenibilidad de la colusión analizada en este bloque.

Colusión tácita

Por otro lado, dentro de los oligopolios cooperativos la colusión puede ser tácita, es decir, una coordinación de precios o cantidades que las empresas alcanzan sin ningún acuerdo, solo observando y anticipando la conducta de sus rivales. Surge a largo plazo, cuando deja de existir miopía, y complica el control de las autoridades por la dificultad de probar su existencia[21].

[20] Algunos de los cárteles más importantes desarticulados gracias a programas de clemencia:

- Cártel de camiones (DAF, Daimler, Iveco, MAN, Renault/Volvo, Scania), Comisión Europea, 2016-2017: fijaron precios y retrasaron tecnología anticontaminante durante 14 años, con 3.810 M€ en multas combinadas. MAN, primera en acogerse a la clemencia, obtuvo inmunidad total y evitó una multa estimada en 1.200 M€. Afectó a más de 100.000 compradores de camiones en España.
- Cártel Líbor/Euríbor, Comisión Europea, 2013: manipularon coordinadamente los tipos de interés de referencia interbancarios, con 1.710 M€ en multas entre ocho bancos. Barclays, con inmunidad total, evitó alrededor de 690 M€.
- Cárteles de divisas "Three Way Banana Split" y "Essex Express" (Barclays, RBS, Citigroup, JPMorgan, MUFG), Comisión Europea, 2019: coordinaron operaciones de compraventa de divisas vía chats privados, con 1.068 M€ en multas combinadas. UBS, con inmunidad total, evitó alrededor de 285 M€.
- Cártel de tecnología de limpieza de emisiones diésel (BMW, Volkswagen/Audi/Porsche), Comisión Europea, 2021: acordaron limitar el desarrollo de tecnología anticontaminante, con 875 M€ en multas. Daimler obtuvo inmunidad total —primer cártel de la historia de la UE sancionado solo por restringir el desarrollo técnico.
- Cártel de divisas "Sterling Lads" (Barclays, RBS, HSBC, Credit Suisse), Comisión Europea, 2021: compartieron información y coordinaron órdenes de divisas, con 344 M€ en multas. UBS, con inmunidad total, evitó alrededor de 94 M€.
- Cártel del automóvil español (21 empresas fabricantes y distribuidoras, más dos consultoras), CNMC, 2015: intercambiaron información comercial sensible entre fabricantes y distribuidores, con 171 M€ en multas, la mayor de la CNMC hasta dicha fecha. SEAT, Volkswagen Audi España y Porsche Ibérica quedaron exentas por clemencia.

[21] Algunos ejemplos de colusión tácita son:

- Gasolineras: Es un caso clásico. Varias gasolineras situadas cerca unas de otras suelen mantener precios muy similares. Si una sube el precio, las demás la siguen rápidamente. Esta sincronización, sin acuerdo explícito, evita una "guerra de precios" que erosionaría los márgenes de beneficio de todas las empresas. Cada una asume la respuesta de sus rivales para mantener la rentabilidad colectiva.
- Aerolíneas: En rutas específicas con pocos competidores dominantes, las aerolíneas a menudo ajustan sus tarifas de manera coordinada. Observan los sistemas de reservas y precios de sus rivales. Si una aerolínea "líder" aumenta los precios, las otras pueden seguirle para evitar la competencia agresiva, lo que demuestra una comprensión implícita del mercado y el deseo de no desencadenar una espiral de bajadas de precios.

 © 2026 Manuel Cabado García

9.5. Sostenibilidad del oligopolio ⏳ 4'

Frente a la inestabilidad interna de la colusión analizada en el bloque anterior —donde cada empresa tenía incentivos a desviarse del acuerdo—, la sostenibilidad del oligopolio en el largo plazo depende de un mecanismo distinto: las barreras de entrada. Hasta aquí se han presentado los modelos cooperativos y los no cooperativos, que tenían en común la existencia de barreras de entrada que impedían la irrupción de nuevas empresas. Y es que en oligopolio, los beneficios extraordinarios atraerán a nuevas empresas que, en principio, eliminarán dichos beneficios. Sin embargo, si existe algún tipo de barrera a la entrada, puede que esto no ocurra. Formalmente, una barrera de entrada es un coste de producción que debe soportar la empresa entrante pero no las empresas ya instaladas (Stigler, 1968), una asimetría que las incumbentes pueden explotar precisamente porque no todas las empresas parten de la misma posición.

En este apartado se va a analizar cómo ha evolucionado el papel de las barreras de entrada en la teoría del oligopolio:

- En una primera etapa de la Teoría de la Organización Industrial, la Escuela de Chicago argumentó que la generación de beneficios extraordinarios incentivaría la entrada de nuevas empresas y, por tanto, un oligopolio nunca sería sostenible en el largo plazo. La competencia acabaría imponiéndose.
- Sin embargo, teorías posteriores refinarían los planteamientos de la Escuela de Chicago, indicando que su argumentación sólo es válida si no existen barreras de entrada.

Tipos de barreras de entrada:

- Legales: licencias, patentes, etc.
- Acceso privilegiado a factores productivos (por ejemplo, río, mina).
- Técnicas: ventajas en costes (por ejemplo, economías de escala, de experiencia, o de gama).
- Estratégicas: son las creadas por las propias empresas ya presentes en la industria.

- Banca Minorista: En mercados con un oligopolio bancario, las grandes entidades suelen ofrecer tasas de interés parecidas para depósitos o hipotecas. Monitorean las ofertas de la competencia y realizan ajustes que mantienen una uniformidad en el mercado sin necesidad de acuerdos directos. El objetivo es preservar márgenes saludables, evitando una competencia de tasas que disminuiría la rentabilidad general del sector.
- Refrescos y Bebidas: Empresas como Coca-Cola y PepsiCo compiten fuertemente en marketing y distribución, pero rara vez en precios directos a nivel mayorista. Los aumentos de precios de una suelen ser replicados por la otra, lo que permite a ambas mantener altos sus márgenes de beneficio. Existe una comprensión no hablada de que una competencia de precios agresiva sería mutuamente destructiva, beneficiando solo al consumidor.
- Grandes Cadenas de Supermercados: En regiones donde pocas cadenas dominan el sector minorista, los precios de productos básicos o marcas blancas a menudo convergen. Las cadenas monitorean constantemente los precios de sus competidores y ajustan los suyos para no quedarse atrás, pero evitan iniciar una competencia de precios total que reduciría los márgenes de todas. Es una forma de mantener la estabilidad del mercado.

Barreras técnicas

Existen barreras técnicas o naturales que suelen estar relacionadas con la posesión de determinadas materias primas o recursos y hacen referencia a economías de escala, derivadas de rendimientos crecientes de escala que permiten operar a menores costes medios. Las entrantes podrían instalarse con su tamaño óptimo, pero ello provocaría una bajada de los precios del mercado, y si eligen instalarse con un tamaño menor que el óptimo, dificultarían su supervivencia.

Otras barreras técnicas pueden estar relacionadas con las ventajas en costes, requerimientos de capital inicial o diferenciación del producto.

Centrando ahora la atención en las barreras estratégicas, creadas deliberadamente por las empresas ya instaladas, a continuación se desarrolla el modelo del precio límite.

Modelo del precio límite

El modelo del precio límite de Bain (1956); Sylos-Labini (1962); Modigliani (1958) parte de la hipótesis de que las empresas instaladas (incumbentes) no maximizan sus beneficios con un precio de monopolio, sino que establecen un precio tal ("precio límite") que la entrada de un competidor adicional con su capacidad mínima provocaría una caída del precio que eliminaría todo el beneficio extraordinario, tanto para la entrante como para las ya instaladas

Supuestos del modelo:

- El mercado está dominado por un monopolista o cártel, y existe una empresa que es una entrante potencial.
- Tanto la empresa establecida como la potencial tienen la misma tecnología.
- La tecnología presenta inicialmente rendimientos crecientes a escala y, a partir de un cierto nivel de producción, rendimientos constantes.
- La incumbente amenaza con mantener constante la producción en caso de entrada de la potencial.

Gráficamente, en un plano de precios y cantidades como el de la Figura 9.7, se representan tres pares de precio y cantidad: el de monopolio (P_M, Q_M), el límite (P_L, Q_L) que la incumbente elige realmente para disuadir la entrada, y el de competencia (P_C, Q_C), al que se llegaría en un mercado perfectamente competitivo:

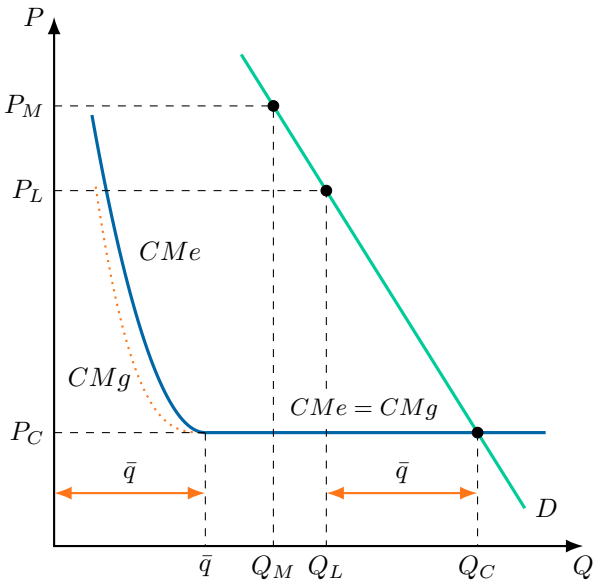
- Se representan la curva de demanda, D, y las curvas de coste medio, CMe, y de coste marginal, CMg, comunes a la incumbente y al entrante. El coste medio decrece hasta la escala mínima eficiente, $\bar{q}$, y a partir de ahí es constante e igual al coste marginal, P_C. Para cualquier precio mayor que el precio límite, P_L, el incumbente obtiene

 © 2026 Manuel Cabado García

mayores beneficios a corto plazo, pero a largo plazo entrarán nuevas empresas en el mercado y los beneficios extraordinarios se reducirán.

- La incumbente elige Q_L de forma que $Q_C - Q_L = \bar{q}$, donde $\bar{q}$ es la escala mínima eficiente del entrante, la misma distancia que marcan las dos flechas de la figura: si este produjera exactamente $\bar{q}$, la cantidad total del mercado alcanzaría Q_C y el precio caería hasta P_C, justo donde su coste medio también toca su mínimo —beneficio nulo—. Si en cambio produce algo menos que $\bar{q}$, su CMe todavía no ha bajado del todo hasta P_C, mientras que el precio de mercado, al quedar la cantidad total algo por debajo de Q_C, solo sube ligeramente por encima de P_C: el coste crece más deprisa de lo que sube el precio, y la entrada genera pérdidas. Si produjera más que $\bar{q}$, la cantidad total superaría Q_C y el precio caería por debajo de P_C, que es su coste medio, de modo que también tendría pérdidas.
- Así, el coste de la barrera estratégica para la incumbente es un coste de oportunidad: al producir $Q_L > Q_M$ para cerrar esa ventana de entrada, cobra un precio $P_L < P_M$ y renuncia a la diferencia de beneficio que obtendría actuando como monopolista puro.

Figura 9.7: Modelo del precio límite



La amenaza de mantener la producción de la incumbente no es creíble, ya que si la nueva empresa entra, no sería racional que la incumbente mantuviera la producción inicial: lo lógico sería que acomodara la entrada de la nueva empresa adaptando su producción para maximizar beneficios. La única manera de que la amenaza de mantener el nivel de producción sea creíble es que:

- El juego dure un número infinito de períodos (o que no se sepa cuándo va a acabar)

y los jugadores sean suficientemente pacientes[22].

- La empresa incumbente incurra en costes hundidos que fuerzan a la empresa a ejecutar su amenaza. Esta última cuestión se analiza en un juego de tres etapas.

Se plantea el juego extensivo en tres etapas con dos empresas: la instalada (monopolista o cártel) y la entrante, siguiendo el modelo de inversión en capacidad de Dixit (1979; 1980):

1. Etapa: la incumbente puede mantener una actitud pasiva (es decir, esperar a la entrante), o activa (incurriendo en un coste hundido –por ejemplo, ampliando su capacidad productiva–).
2. Etapa: la empresa entrante decide si entrar o no.
3. Etapa: si la entrante ha entrado, la incumbente puede luchar o acomodarse en función de su rentabilidad.

Formalmente, sea π^M el beneficio de monopolio, π^D el beneficio de duopolio de la incumbente si se acomoda y π^L su beneficio si lucha contra la entrante. Si la incumbente ha incurrido en el coste hundido C_0, ese coste no penaliza el beneficio de luchar —la capacidad ampliada se emplea precisamente para ello—, pero sí el de acomodarse ($\pi^D - C_0$) o el de quedarse sola sin que haya entrado nadie ($\pi^M - C_0$). Resolviendo por inducción hacia atrás:

1. Tercera etapa: la incumbente luchará solo si $\pi^D - C_0 < \pi^L$.
2. Segunda etapa: anticipándolo, la entrante no entrará si prevé que la incumbente luchará, porque en ese caso incurriría en pérdidas.
3. Primera etapa: a la incumbente solo le convendrá incurrir en el coste hundido si con ello logra quedarse sola en el mercado y $C_0 < \pi^M - \pi^D$.

Mercados contestables o impugnables

Con su teoría, Baumol (Baumol et al., 1982) dio un giro al debate sobre la necesidad de regular los mercados de competencia imperfecta: para él, el número de empresas presentes en una industria es irrelevante, pues lo que importan son las condiciones de entrada y de salida.

En efecto, si en un mercado los outsiders pueden entrar a competir sin desventajas respecto a los operadores instalados, dicho mercado es impugnable, y en la medida en que los precios de la industria estén por encima del coste medio, se podrían dar en él estrategias de hit and run: un outsider entra, vende a un precio menor, obtiene beneficios extraordinarios, y sale del mercado.

[22] Con horizonte finito, la amenaza solo puede volverse creíble si los entrantes no están seguros del tipo de incumbente al que se enfrentan (por ejemplo, si creen posible que sea de los que luchan siempre), lo que permite construir una reputación (Kreps y Wilson, 1982; Milgrom y Roberts, 1982).

 © 2026 Manuel Cabado García

Por ello, las empresas que operan en mercados contestables tenderán a comportarse de manera competitiva ($P = CMe$), y no habrá beneficios extraordinarios. De esta manera, si se consigue que el mercado sea atacable, entonces se podrá llegar a una solución competitiva sin que se cumplan los requisitos especialmente restrictivos de la competencia perfecta.

Condiciones de mercado impugnable Las condiciones que se tienen que dar para que un mercado sea impugnable son:

1. No puede haber ventajas absolutas en costes (es decir, igualdad de acceso a la tecnología).
2. Puede haber rendimientos crecientes a escala, e incluso un monopolio natural, siempre que entrar y salir no suponga costes hundidos[23].
3. Puede haber costes fijos, pero no hundidos (es decir, no hay barreras a la salida).
4. Existen costes de modificar los precios –costes de menú–.
5. La demanda se adapta instantáneamente a los precios[24].
6. No hay diferenciación del producto.

Implicaciones de política económica. Las implicaciones de política económica de los mercados contestables son muy potentes: la política de defensa de la competencia debe interesarse más por las barreras que por el número de empresas en un mercado:

- Esto es muy importante desde el punto de vista práctico, ya que eliminar las barreras a la entrada es más fácil que aumentar el número de competidores.
- El sistema español de defensa de la competencia parece guiarse por la teoría de los mercados impugnables. En efecto, la Ley 15/2007, de Defensa de la Competencia, señala en su preámbulo que el objetivo debe ser la defensa de una "competencia efectiva", es decir, que persigue que el mercado se comporte de manera competitiva, con independencia del número de empresas presentes en él.
- Los sectores donde más impacto ha tenido esta concepción de la defensa de la competencia han sido las industrias de red (esto es, aquellas que prestan su servicio a través de una infraestructura en red, como vías, tendidos o cables, cuya construcción exige una fuerte inversión inicial, mientras que el coste marginal de atender a un usuario más es muy bajo, por ejemplo, ferrocarriles, energía o telecomunicaciones).

[23] El primer supuesto y la ausencia de costes hundidos garantizan la no existencia de barreras de entrada.
[24] El cuarto y quinto supuesto garantizan que a la incumbente no le dé tiempo a reaccionar.

9.6. Conclusiones ⏳ 30"

Como conclusión, pese a los avances teóricos en el estudio del oligopolio es muy difícil tener un cuadro completo respecto a la formación de precios debido a la multitud de factores que intervienen en su análisis (por ejemplo, información asimétrica, existencia de equilibrios múltiples, posibilidad de renegociación de un acuerdo dinámico y cambiante, la presencia de avances tecnológicos, etc.).

Por este motivo, una de las formas más prometedoras de luchar contra la colusión es incentivar su propia inestabilidad interna, como hacen los programas de clemencia ya analizados.

Finalmente, cabe destacar que la teoría del oligopolio (o competencia imperfecta) es susceptible de extensión mediante la consideración de bienes diferenciados.

 © 2026 Manuel Cabado García

Apéndice

9.A.1. Ejemplos de oligopolios

Existen diferentes ejemplos de oligopolios tanto en precios (modelo de Bertrand) como en cantidades (Modelo de Cournot):

- Competencia en precios: Sectores como las telecomunicaciones (especialmente en ofertas de tarifas móviles o de fibra), la distribución minorista (supermercados, grandes almacenes en productos básicos) y, en cierta medida, algunas aerolíneas (en rutas muy competidas).
- Competencia en cantidades: Sectores como la automoción, la industria química pesada, la fabricación de aviones o la producción de acero. En estos casos, las empresas invierten en capacidad de producción a largo plazo y luego compiten para vender lo que han producido, influyendo en el precio de mercado a través de la cantidad total ofrecida. En el caso de las bebidas como Coca-Cola y Pepsi, compiten por la cantidad de producto vendido y por la omnipresencia en los puntos de venta.

Sin embargo, es importante notar que en la realidad muchos oligopolios exhiben una mezcla de ambos tipos de competencia, o pueden alternar entre una y otra dependiendo de las condiciones del mercado y la estrategia de las empresas. Además, la diferenciación del producto, la publicidad, la innovación y los servicios postventa también son herramientas clave de competencia en los oligopolios.

Finalmente, también cabe destacar algunos ejemplos de líder-seguidora (modelo de Stackelberg):

- Industria Automotriz (Vehículos Eléctricos): Tesla actúa como líder, invirtiendo masivamente y fijando la pauta en vehículos eléctricos e infraestructura de carga. Los fabricantes tradicionales (seguidores) observan su estrategia y éxito, luego adaptan sus propios planes de producción y desarrollo de VE. Tesla toma la iniciativa en un mercado emergente, obligando a los demás a reaccionar y seguir su dirección para no quedarse atrás en la transformación del sector.
- Tecnología y Electrónica de Consumo: Apple suele ser el líder al lanzar productos innovadores como el iPhone. Establece tendencias de diseño, características y precios premium. Competidores como Samsung o Xiaomi (seguidores) analizan el impacto de Apple y luego lanzan sus propias ofertas, buscando diferenciarse o competir con precios más bajos. Apple da el "primer golpe" en la innovación, y los demás ajustan sus estrategias en función de esa jugada inicial.
- Telecomunicaciones (Inversiones en Infraestructura): Grandes operadores como Movistar a menudo actúan como líderes en el despliegue de nuevas redes, como el 5G. Su enorme inversión inicial en infraestructura y cobertura establece un precedente.

Los competidores más pequeños o con menos recursos (seguidores) observan esta escala y deciden sus propias estrategias de despliegue, adaptándose a la hoja de ruta que el líder ha marcado en el mercado de telecomunicaciones.

- Comercio Minorista (Logística): Amazon es un claro líder en el comercio electrónico y la logística de entrega. Sus inversiones en almacenes automatizados y redes de distribución establecen un estándar de eficiencia. Otros minoristas (seguidores como Walmart o El Corte Inglés) deben adaptar y mejorar sus propias cadenas de suministro y servicios de entrega para intentar igualar la conveniencia y rapidez que Amazon ha impuesto en el sector.
- Industria Farmacéutica (Lanzamiento de Medicamentos): Una farmacéutica como Pfizer que invierte en I+D para un medicamento revolucionario actúa como líder. Su decisión de introducir el fármaco y su precio inicial marcan la pauta. Otras compañías (seguidores) con tratamientos similares o genéricos observan el éxito del líder y ajustan sus propias estrategias de desarrollo o comercialización. El que llega primero al mercado con una innovación define las reglas iniciales del juego.

9.A.2. Los cárteles de la droga como caso extremo de oligopolio

Los cárteles de la droga pueden entenderse como un caso extremo de oligopolio y, más concretamente, de cártel (Fiorentini, 1995): el mercado ilícito de ciertas drogas suele estar dominado por un número reducido de grandes organizaciones criminales que, aunque convivan con actores menores, concentran la producción, el transporte y la distribución a gran escala —la misma condición de número reducido de vendedores que define cualquier oligopolio.

Esta concentración genera una fuerte interdependencia estratégica: las decisiones de un cártel sobre rutas, volumen de producción, precios al por mayor o expansión territorial afectan directamente a los demás, y una acción agresiva de uno puede desencadenar una represalia similar del resto, degenerando en conflictos armados por el control del mercado (Dell, 2015).

Las barreras de entrada, además, son extraordinariamente altas: entrar en el narcotráfico a gran escala exige

- Grandes inversiones en infraestructura, logística, corrupción y armamento.
- Control de rutas y territorios, a menudo conseguido con violencia.
- Redes de contactos que permitan operar con impunidad.
- Conocimiento especializado en producción y distribución clandestina.
- El riesgo, siempre presente, de ser detenido o asesinado.

Los cárteles son, además, colusión explícita en su forma más literal: se constituyen precisamente mediante acuerdos para repartirse territorios ("plazas"), fijar precios muy por

 © 2026 Manuel Cabado García

encima de los costes de producción, restringir la oferta para mantenerlos altos, y colaborar frente a rivales o autoridades (Poret y Tejedo, 2006). Junto a esos acuerdos explícitos, también existe colusión tácita en la forma en que cada organización observa y responde a las acciones de las demás para mantener la estabilidad del mercado y evitar guerras internas que reduzcan los beneficios de todos. Como cualquier empresa en un oligopolio, su objetivo último es maximizar el beneficio manipulando la oferta y el precio.

Existen, no obstante, diferencias notables respecto a un oligopolio legal: al operar en la ilegalidad total, las barreras de entrada y las sanciones por incumplir los acuerdos son mucho más brutales (violencia, muerte) que cualquier sanción de un mercado legal. Además, al no estar sujetos a leyes de competencia, la colusión puede llevarse a extremos que la regulación antimonopolio impediría en un mercado legítimo.

9.A.3. Índice de Herfindahl-Hirschman

El Índice de Herfindahl-Hirschman (IHH) es una herramienta clave para medir la concentración de un mercado, propuesto originalmente por Herfindahl (1950) y, de forma independiente, por Hirschman (1964). Indica si un mercado es competitivo o si está dominado por pocas empresas, lo que es crucial para entender un oligopolio.

Se calcula sumando los cuadrados de las cuotas de mercado de todas las empresas, expresadas en porcentaje:

$$IHH = s_1^2 + \ldots + s_n^2$$

donde s_i es la cuota de mercado, en porcentaje, de la empresa i.

Por ejemplo, en un mercado con cuatro empresas de 25 % de cuota, el IHH sería: $25^2 + 25^2 + 25^2 + 25^2 = 2500$.

Los valores se interpretan de la siguiente manera:

- IHH < 1000: Mercado no concentrado (competitivo).
- $1000 \leq$ IHH ≤ 1800: Mercado moderadamente concentrado. Aquí se sitúan muchos oligopolios.
- IHH > 1800: Mercado altamente concentrado. Indica un oligopolio con gran poder de mercado, o un monopolio si el valor se acerca a 10.000.

Una de las mayores ventajas del IHH es que otorga más peso a las empresas grandes, reflejando de forma más precisa el poder de mercado. Esto lo hace indispensable para las autoridades antimonopolio, que lo usan para evaluar el impacto de fusiones y adquisiciones. Si una fusión aumenta el IHH por encima de un umbral aceptable, podría ser bloqueada para proteger la competencia. Aunque es muy útil, su principal limitación es que no considera otros factores, como las barreras de entrada o la elasticidad de la demanda.

9.A.4. Concepto de teoría de juegos

La Teoría de Juegos es una rama de las matemáticas que analiza el comportamiento estratégico de individuos o agentes (llamados jugadores). Se utiliza para estudiar situaciones en las que las decisiones de un jugador afectan directamente a los demás, y viceversa. Debido a esta interdependencia estratégica, se ha convertido en una herramienta esencial en el análisis económico, ya que muchas decisiones de mercado (como la fijación de precios o la producción en un oligopolio) encajan en este marco.

Elementos de un juego Un juego es una representación formal de una interacción estratégica. Para describirlo, se utilizan los siguientes elementos:

- Jugadores: individuos o empresas que toman decisiones de forma racional para maximizar su beneficio o utilidad.
- Reglas: definen las acciones permitidas, el orden de los movimientos y la información disponible para cada jugador.
- Resultados: los posibles desenlaces del juego según las acciones de los jugadores.
- Pagos: el valor o utilidad que cada jugador recibe por cada resultado posible. Estos pagos reflejan las preferencias de los jugadores.

Representación de un juego Para representar un juego, se utilizan dos formas principales:

- Forma extensiva: se representa visualmente con un árbol de juego que muestra el orden de las decisiones, las acciones posibles y los pagos finales.
    - Información perfecta: cada jugador conoce todos los movimientos anteriores de sus rivales.
    - Información imperfecta: los jugadores no conocen todos los movimientos previos al momento de tomar su decisión.
- Forma normal (o estratégica): se presenta en una tabla o matriz que muestra las estrategias de cada jugador y los pagos correspondientes a cada combinación de estrategias. Una estrategia se define como un plan de acción completo y contingente, que especifica la decisión de un jugador en cada posible situación del juego.

Tipos de juegos Los juegos se clasifican según sus características principales:

- Cooperativos vs. no cooperativos: en los juegos cooperativos, los jugadores pueden negociar acuerdos vinculantes. En los no cooperativos, no es posible la negociación.

 © 2026 Manuel Cabado García

- Simultáneos vs. secuenciales (también llamados estáticos y dinámicos, respectivamente): en los juegos simultáneos, los jugadores toman sus decisiones al mismo tiempo. En los secuenciales, hay un orden de turnos (un líder mueve primero, y el seguidor reacciona).
- Repetidos vs. de una sola vez (one-shot): en los juegos repetidos, los mismos jugadores se enfrentan varias veces, lo que genera un historial que permite condicionar la estrategia al comportamiento pasado del rival, como ocurre con las estrategias trigger analizadas en el cuerpo del tema. En los juegos de una sola vez, cada decisión se toma sin ese historial. Esta distinción es independiente de si el juego es estático o dinámico: el modelo de Stackelberg, por ejemplo, es un juego dinámico pero de una sola vez, no repetido.

Conceptos clave de equilibrio La Teoría de Juegos busca predecir el resultado de una interacción estratégica a través de conceptos de equilibrio.

- Equilibrio de Nash: es la combinación de estrategias donde ningún jugador tiene incentivos para cambiar unilateralmente su decisión, dada la estrategia de los demás. En otras palabras, cada jugador está haciendo lo mejor que puede, considerando lo que hacen sus rivales.
  - Un juego puede tener uno, múltiples o ningún equilibrio de Nash (si no se permiten estrategias mixtas).
  - No siempre es el resultado óptimo para todos los jugadores en conjunto (no es necesariamente óptimo de Pareto).
- Estrategia dominante: es la mejor estrategia para un jugador, sin importar lo que hagan sus rivales. Si una empresa tiene una estrategia dominante, es muy probable que la elija, ya que siempre maximizará su pago. Sin embargo, no son comunes, ya que las decisiones óptimas a menudo dependen de lo que hagan los competidores.
- Funciones de reacción: describen cómo una empresa oligopolista ajustará su producción (o precio) en respuesta a cada posible nivel de producción (o precio) de su competidor. El punto donde las funciones de reacción de los jugadores se cruzan es el equilibrio de Nash de ese juego, ya que en ese punto, la estrategia de cada empresa es la mejor respuesta a la de su rival.
- Racionalidad secuencial: un equilibrio de Nash puede apoyarse en una amenaza que no es creíble. Considérese una empresa incumbente que amenaza con mantener su producción constante si una entrante decide entrar: si la entrada llega a producirse, a la incumbente ya no le convendrá cumplir esa amenaza, por lo que la estrategia deja de ser una predicción razonable —precisamente el problema señalado más arriba, al estudiar el modelo del precio límite, donde se indicó que la amenaza de mantener la

producción de la incumbente no es creíble—. Se exige entonces que las estrategias de equilibrio satisfagan la racionalidad secuencial: el comportamiento prescrito debe ser óptimo en cada punto del juego, se alcance o no ese punto en el desarrollo del equilibrio.

- Inducción hacia atrás y Equilibrio de Nash Perfecto en Subjuegos: en juegos con información perfecta, la racionalidad secuencial se aplica mediante la inducción hacia atrás: se resuelve primero el comportamiento óptimo en los últimos nodos de decisión del árbol del juego y, anticipando esa respuesta, se determina el comportamiento óptimo en el nodo anterior, y así sucesivamente hasta el nodo inicial. Un perfil de estrategias que supera esta prueba en todo subjuego se denomina Equilibrio de Nash Perfecto en Subjuegos (ENPS, Selten (1965)). El propio modelo de Stackelberg desarrollado en este tema es un caso de ENPS: al sustituir la función de reacción de la seguidora en el problema de maximización de la líder, esta anticipa correctamente la respuesta óptima de la seguidora antes de decidir su propia producción, exactamente la lógica de la inducción hacia atrás.
- Juegos repetidos, horizonte temporal y sostenibilidad de la colusión: cuando un juego se repite un número finito y conocido de veces, la inducción hacia atrás obliga a que, en el último periodo, la estrategia dominante sea desviarse del acuerdo cooperativo, ya que no quedan periodos futuros que disciplinen el comportamiento. Anticipándolo, lo óptimo es desviarse también en el penúltimo periodo, y así sucesivamente hasta el primero. Es la misma lógica que la paradoja de la cadena de tiendas de Selten (1978), en la que un monopolista que se enfrenta a una serie finita y conocida de entrantes no logra disuadirlos con amenazas de guerra de precios[25]. La conclusión cambia radicalmente con horizonte infinito o final incierto, donde deja de aplicarse la inducción hacia atrás: es precisamente en ese contexto donde las estrategias trigger de Friedman, analizadas en el cuerpo del tema, constituyen un ENPS sostenible.

Información incompleta: el enfoque bayesiano En muchas situaciones económicas los jugadores no conocen toda la información relevante del juego, incluidos los pagos de sus rivales —por ejemplo, dos empresas que desconocen mutuamente su estructura de costes—. Se habla entonces de juegos de información incompleta.

- Transformación de Harsanyi: el enfoque de Harsanyi (1967) reinterpreta la información incompleta como un juego de información imperfecta: la naturaleza realiza un primer movimiento, asignando a cada jugador un tipo aleatorio que determina sus preferencias, y solo el propio jugador observa la realización de su tipo, aunque la distribución de probabilidad de la que se extrae es conocimiento común. Un juego construido de este modo se denomina juego bayesiano.

[25] La intuición es que un castigo que no puede llegar nunca no disciplina a nadie: si todos saben cuál es el último periodo, en él nadie teme represalias, y esa certeza se propaga hacia atrás.

 © 2026 Manuel Cabado García

- Equilibrio bayesiano de Nash: es el concepto de equilibrio asociado a un juego bayesiano: un perfil de reglas de decisión —una elección estratégica para cada tipo posible— que constituye un equilibrio de Nash del juego en el que cada jugador maximiza su pago esperado sobre las realizaciones de los tipos de los demás.

Teoría de juegos cooperativa A diferencia de la teoría de juegos no cooperativa, cuya unidad de análisis es el jugador individual, la teoría de juegos cooperativa toma como unidad de análisis las coaliciones —grupos de jugadores capaces de alcanzar acuerdos vinculantes y repartirse el excedente conjunto—. Conviene no confundir este uso técnico con el sentido en que este mismo tema emplea la palabra "cooperativo" al hablar de oligopolios cooperativos (colusión, cárteles): ese es un uso económico informal del término, mientras que la teoría de juegos cooperativa formal analiza qué reparto del excedente total generado por una coalición es estable o equitativo.

- El core: conjunto de repartos del excedente conjunto tales que ninguna coalición, desde un jugador aislado hasta la formada por todos ellos, puede obtener por su cuenta un resultado mejor que el que le corresponde en ese reparto —es, por tanto, el análogo cooperativo del concepto de estabilidad frente a desviaciones que en la teoría no cooperativa aporta el equilibrio de Nash.
- El valor de Shapley: propuesto por Shapley (1953), es un reparto del excedente conjunto que asigna a cada jugador el promedio de su contribución marginal a todas las posibles coaliciones a las que podría unirse, satisfaciendo un conjunto de axiomas de equidad (simetría, eficiencia, aditividad).

Diseño de mecanismos y subastas El diseño de mecanismos invierte la lógica habitual de la teoría de juegos: en lugar de predecir el resultado de un juego dado, se diseñan las reglas del juego —el mecanismo— para inducir un resultado deseado por quien lo diseña, incluso cuando los participantes disponen de información privada. El campo, iniciado por Hurwicz y refinado por Maskin y Myerson —los tres compartieron el Premio Nobel de 2007 por ello— (Hurwicz, 1973; Maskin, 1999; Myerson, 1979), tiene como objetivo principal construir mecanismos incentivo-compatibles: reglas del juego en las que a cada participante le convenga revelar honestamente su información privada.

- Las subastas son la aplicación más desarrollada de este enfoque, cuyo tratamiento óptimo formalizó Myerson (1981). En una subasta al mejor postor, cada pujador conoce su propia valoración del bien pero no la de sus rivales, por lo que su comportamiento óptimo se resuelve mediante un equilibrio bayesiano de Nash.
- Subasta de segundo precio a sobre cerrado: propuesta por Vickrey (1961), es aquella en la que el bien se adjudica al pujador que ofrece más, pero este paga la segunda puja más alta, no la suya propia. Este diseño hace que la estrategia óptima de cada

pujador sea pujar exactamente su valoración real del bien, sin necesidad de especular sobre las pujas de sus rivales[26]. Por su autor, se la conoce también como subasta de Vickrey.

9.A.5. Generalización del modelo de Cournot a N empresas

El modelo de Cournot desarrollado en el cuerpo del tema se ha presentado para dos empresas, pero su lógica se extiende de forma directa a un número cualquiera de empresas, lo que permite además estudiar cómo se comporta el equilibrio a medida que aumenta el número de competidores. Sea un mercado con N empresas que compiten en cantidades, con una función de demanda inversa lineal $P = a - bQ$, donde $Q = \sum_{i=1}^{N} q_i$, y con costes marginales constantes c_i, potencialmente distintos entre empresas.

Cada empresa i maximiza su beneficio tomando como dada la producción del resto:

$$\max_{q_i} \pi_i = (a - b \cdot Q - c_i)\, q_i \tag{9.14}$$

cuya condición de primer orden es

$$a - c_i - b \cdot Q_{-i} - 2 \cdot b \cdot q_i = 0, \tag{9.15}$$

donde Q_{-i} es la producción conjunta del resto de empresas. Resolviendo el sistema de N condiciones de primer orden se obtiene la producción de equilibrio de cada empresa,

$$q_i^* = \frac{a - N \cdot c_i + \sum_{j \neq i} c_j}{(N+1)\, b}, \tag{9.16}$$

la producción total,

$$Q^* = \frac{N \cdot a - \sum_{i=1}^{N} c_i}{(N+1)\, b}, \tag{9.17}$$

y el precio de equilibrio,

$$P^* = \frac{a + \sum_{i=1}^{N} c_i}{N+1}. \tag{9.18}$$

Cuando todas las empresas tienen el mismo coste marginal $c_i = c$, las expresiones anteriores se simplifican al resultado ya conocido para dos empresas, generalizado a N:

$$q_i^* = \frac{a - c}{(N+1)\, b}, \qquad Q^* = \frac{N \cdot (a - c)}{(N+1)\, b}, \qquad P^* = \frac{a + N \cdot c}{N+1}. \tag{9.19}$$

Este resultado permite además analizar el comportamiento del equilibrio a medida que aumenta el número de competidores: cuando $N \to \infty$, el precio de equilibrio converge al

[26] La intuición es que, como lo que se paga no depende de la propia puja sino de la del segundo, pujar menos de lo que se valora el bien solo sirve para arriesgarse a perderlo, y pujar más, para arriesgarse a pagar de más.

 © 2026 Manuel Cabado García

coste marginal,

$$\lim_{N\to\infty} P^* = \lim_{N\to\infty} \frac{a + N \cdot c}{N+1} = c, \tag{9.20}$$

es decir, el resultado de Cournot converge al equilibrio de competencia perfecta a medida que el mercado se atomiza, lo que confirma que el poder de mercado del oligopolio depende crucialmente del número de competidores (Tirole, 1988).

Bibliografía

Bain, J. S. (1956). *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries*. Harvard University Press.

Baumol, W. J., J. C. Panzar y R. D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

Bertrand, J. (1883). Théorie mathématique de la richesse sociale. *Journal des Savants 67*, 499–508.

Cabral, L. M. B. (2000). *Introduction to Industrial Organization*. Cambridge, MA: The MIT Press.

Cournot, A. A. (1838). *Recherches Sur les Principes Mathématiques de la Théorie des Richesses*. L. Hachette.

Dell, M. (2015). Trafficking Networks and the Mexican Drug War. *American Economic Review 105*(6), 1738–1779.

Dixit, A. (1979). A Model of Duopoly Suggesting a Theory of Entry Barriers. *The Bell Journal of Economics 10*(1), 20–32.

Dixit, A. (1980). The Role of Investment in Entry Deterrence. *The Economic Journal 90*(357), 95–106.

Fiorentini, G. (1995). Oligopolistic Competition in Illegal Markets. En G. Fiorentini y S. Peltzman (eds.), *The Economics of Organised Crime*, pp. 274–288. Cambridge, UK: Cambridge University Press.

Green, E. J. y R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica 52*(1), 87–100.

Harsanyi, J. C. (1967). Games with Incomplete Information Played by "Bayesian"Players, I. The Basic Model. *Management Science 14*(3), 159–182.

Herfindahl, O. C. (1950). *Concentration in the U.S. Steel Industry*. Tesis doctoral, Columbia University.

Hirschman, A. O. (1964). The Paternity of an Index. *American Economic Review 54*(5), 761–762.

Hurwicz, L. (1973). The Design of Mechanisms for Resource Allocation. *American Economic Review 63*(2), 1–30.

 © 2026 Manuel Cabado García

Jehle, G. A. y P. J. Reny (2011). *Advanced Microeconomic Theory* (3rd ed.). Harlow: Financial Times Prentice Hall.

Koutsoyiannis, A. (1979). *Modern Microeconomics* (2nd ed.). London: Macmillan.

Kreps, D. M. (2023). *Microeconomic Foundations II: Imperfect Competition, Information, and Strategic Interaction*. Princeton University Press.

Kreps, D. M. y J. A. Scheinkman (1983). Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes. *The Bell Journal of Economics 14*(2), 326–337.

Kreps, D. M. y R. Wilson (1982). Reputation and imperfect information. *Journal of Economic Theory 27*(2), 253–279.

Lipczynski, J., J. Goddard y J. O. S. Wilson (2017). *Industrial Organization: Competition, Strategy and Policy* (5th ed.). Harlow: Pearson.

Malinvaud, E. (1972). *Lectures on Microeconomic Theory*. Amsterdam: North-Holland.

Mas-Colell, A., M. D. Whinston y J. R. Green (1995). *Microeconomic Theory*. New York: Oxford University Press.

Maskin, E. (1999). Nash Equilibrium and Welfare Optimality. *The Review of Economic Studies 66*(1), 23–38.

Milgrom, P. y J. Roberts (1982). Predation, reputation, and entry deterrence. *Journal of Economic Theory 27*(2), 280–312.

Modigliani, F. (1958). New Developments on the Oligopoly Front. *Journal of Political Economy 66*(3), 215–232.

Myerson, R. B. (1979). Incentive Compatibility and the Bargaining Problem. *Econometrica 47*(1), 61–73.

Myerson, R. B. (1981). Optimal Auction Design. *Mathematics of Operations Research 6*(1), 58–73.

Nicholson, W. y C. M. Snyder (2017). *Microeconomic Theory: Basic Principles and Extensions* (12th ed.). Cengage Learning.

Perloff, J. M. (2022). *Microeconomics: Theory and Applications with Calculus* (5th, Global ed.). Pearson.

Poret, S. y C. Tejedo (2006). Law Enforcement and Concentration in Illicit Drug Markets. *European Journal of Political Economy 22*(1), 99–114.

Segura, J. (1993). *Teoría de la Economía Industrial*. Madrid: Civitas.

Selten, R. (1965). Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit. *Zeitschrift für die gesamte Staatswissenschaft 121*, 301–324, 667–689.

Selten, R. (1978). The chain store paradox. *Theory and Decision 9*(2), 127–159.

Shapley, L. S. (1953). A Value for n-Person Games. En H. W. Kuhn y A. W. Tucker (eds.), *Contributions to the Theory of Games II*, Annals of Mathematics Studies 28, pp. 307–317. Princeton, NJ: Princeton University Press.

Shy, O. (1996). *Industrial Organization: Theory and Applications*. Cambridge, MA: The MIT Press.

Stigler, G. J. (1968). *The Organization of Industry*. Homewood, IL: Richard D. Irwin.

Sylos-Labini, P. (1962). *Oligopoly and Technical Progress*. Harvard University Press.

Tirole, J. (1988). *The Theory of Industrial Organization*. Cambridge, MA: The MIT Press.

Varian, H. R. (1992). *Microeconomic Analysis* (3rd ed.). New York: W. W. Norton & Company.

Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance 16*(1), 8–37.

von Stackelberg, H. (1934). *Marktform und Gleichgewicht*. Springer.

 © 2026 Manuel Cabado García

Tema 10

Análisis de mercados (IV). Diferenciación de productos. La competencia monopolística y otros desarrollos.

10.1. Introducción ⏳ 1'

La Teoría microeconómica neoclásica se construyó sobre el supuesto de la existencia de bienes homogéneos para modelizar el comportamiento de los agentes económicos. No obstante, en la práctica se observa que muchas industrias producen un elevado número de bienes diferenciados en muchas características o dimensiones, tales como la localización geográfica, la calidad, el diseño o el grado de servicios proporcionado con el producto entre otros.

Esta insatisfacción teórica llevó a autores como Sraffa, Robinson o Chamberlin a desarrollar modelos de Competencia Imperfecta, entre los que se encuentran los de competencia monopolística, donde el supuesto fundamental es la existencia de productos no homogéneos, es decir, diferenciados, entendidos como sustitutivos cercanos entre sí desde el punto de vista del consumo.

Así, la diferenciación puede ser:

- Horizontal, los bienes se diferencian por sus características, sin que haya unanimidad entre los consumidores sobre qué característica es mejor (por ejemplo, un jersey rojo o negro).
- Vertical, todos los consumidores coinciden en qué variedad es mejor, de modo que a igualdad de precio todos elegirían la de mayor calidad, y solo difieren en cuánto están dispuestos a pagar por ella[1].

[1] Un ejemplo de diferenciación vertical sería la compra de un televisor. En general, todo consumidor

Por consiguiente, el objetivo de esta exposición es el análisis de la competencia monopolística a través del siguiente esquema:

1. Justificación de la diferenciación de productos, con el modelo de Lancaster.
2. Diferenciación horizontal espacial, donde se desarrollará el modelo de Hotelling.
3. Diferenciación horizontal no espacial, donde se analizará el modelo de Chamberlin y el modelo de preferencia por la variedad de Dixit-Stiglitz.
4. Diferenciación Vertical, que se analizará modificando el modelo de Hotelling.

10.2. Justificación de la diferenciación de productos ⏳ 1'

Comenzando con la justificación, la diferenciación de productos hace que el concepto de industria o mercado sea más ambiguo[2]. Así, cabe plantear el por qué las empresas buscan diferenciarse:

- Por el lado de la oferta, la diferenciación supone que cada empresa ofrece su producto en exclusiva, por lo que la empresa evita una demanda totalmente elástica, logrando enfrentarse a una demanda con pendiente negativa que le otorga poder de mercado, es decir, puede influir en el precio de su variedad y aumentarlo sin perder a todos sus clientes.
- Por el lado de la demanda, se justifica por el modelo de Lancaster de demanda de características[3].

Lancaster parte del enfoque de que lo que demandan los consumidores no son bienes en sí mismos, sino las características presentes en ellos, que son las que en realidad les permiten satisfacer sus necesidades. Así, dichas características coexisten en un mismo bien en distintas proporciones.

Gráficamente, como se observa en la Figura 10.1, partiendo de que hay dos características, a_1 y a_2, se dibujan los vectores correspondientes a cada uno de los bienes, X, Y, Z que dependiendo de la pendiente incorporarán más de una característica u otra. Además, sobre cada vector se marca el punto que alcanzaría el consumidor si gastase toda su renta

prefiere una mejor calidad de imagen y sonido, pero cada uno tiene distinta disposición a pagar por un televisor con una determinada imagen y sonido.

[2] Una "industria" se puede definir desde dos puntos de vista:

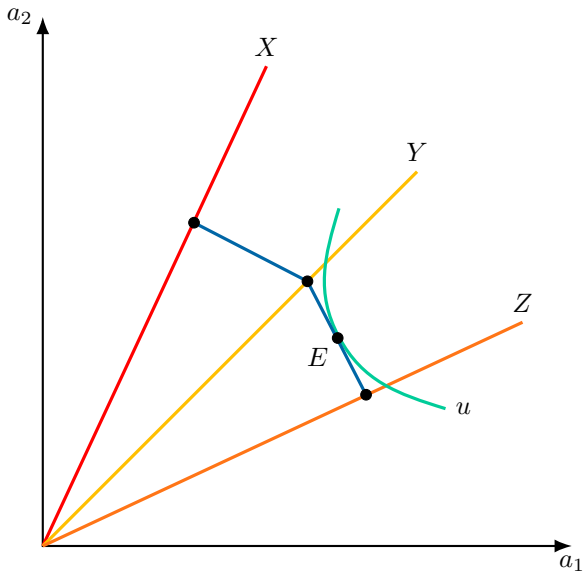
- Criterio de oferta: una industria estará formada por aquellas empresas que utilizan tecnologías similares o producen un mismo bien o servicio. Éste sería el criterio utilizado por la Clasificación Nacional de Actividades Económicas 2009 (CNAE 2009).
- Criterio de demanda: una industria estará formada por aquellas empresas cuyos productos son considerados razonablemente sustitutivos por los consumidores.

[3] Véase el desarrollo completo en el apartado 10.A.2 del Apéndice.

 © 2026 Manuel Cabado García

en ese bien, de modo que la distancia al origen depende del precio, y la unión de esos puntos forma la frontera de posibilidades de consumo. Si baja el precio de un bien, su punto se aleja del origen y la frontera se amplía. No obstante, el consumidor puede elegir consumir dos bienes como hace en E, para una curva de indiferencia, u.

Figura 10.1: Modelo Lancaster simplificado



Así, esta nueva teoría de la demanda permite explicar tres cuestiones que la teoría neoclásica no era capaz de explicar:

1. Ante bajadas del precio de un bien, el consumidor puede no variar su comportamiento dada su preferencia por una determinada característica (a no ser que baje mucho), explicando la fidelidad a las marcas.
2. Si una nueva marca desarrollase un producto H con una combinación equilibrada de características entre los bienes Y y Z, como se observa en el gráfico, se haría con la demanda del consumidor, explicando el desarrollo de nuevos productos.
3. Además, esta teoría justifica la importancia de dar a conocer las características de un producto, explicando la publicidad.

Una vez introducida la justificación de diferenciación de productos, a continuación se analizan los modelos de diferenciación horizontal.

10.3. Diferenciación horizontal espacial 6'30"

Comenzando con los modelos de diferenciación horizontal espacial, estos estudian un enfoque de direcciones mediante modelos de oligopolio. En ellos las preferencias de los individuos se definen sobre el espacio de características, es decir, hay un solo bien con distintas

características[4], y los individuos demandan un determinado nivel de características –por ejemplo, el nivel de azúcar en la Coca-cola: los consumidores demandan Coca-cola normal, o light, o zero–.

Concretamente, los bienes se diferencian entre sí por su localización geográfica, que es una metáfora del grado que poseen de una determinada característica. Así, en estos modelos, cuando una empresa decide su localización, en realidad lo que está decidiendo es el grado de la característica que va a producir[5].

Para su estudio se analizará el modelo de Hotelling.

Modelo de Hotelling

El modelo de Hotelling o de competencia espacial (Hotelling, 1929) explica la diferenciación horizontal con un producto homogéneo utilizando la competencia espacial, es decir, los bienes se diferencian entre sí por su situación geográfica, que funciona como una metáfora del grado que posee de una determinada característica, por lo que el enfoque va a ser de direcciones, es decir, las preferencias de los consumidores se definen sobre un espacio de características.

Supuestos

1. Calle lineal de longitud uno y densidad uno.
2. Consumidores uniformemente distribuidos a lo largo de ella.
3. Dos empresas, (A) y (B), que venden un producto idéntico con los mismos costes de producción, de forma que en lo único en que se diferencian es en la localización (una en cada extremo)[6].
4. Existe un coste de transporte, t, de los consumidores. Los costes de transporte pueden interpretarse como la heterogeneidad de gustos de los consumidores: si, por ejemplo, los costes de transporte del consumidor son nulos, ello significa que no le supone ningún coste pasar de consumir una variedad a otra, es decir, significa que los gustos son homogéneos, por lo que los individuos no valoran la diferenciación ofrecida

[4]Uno de los supuestos cruciales detrás de la Paradoja de Bertrand es la homogeneidad del producto. De esta forma, el precio es la única variable de interés para los consumidores, por lo que con solo dos empresas que compitan en precios, el resultado será el competitivo. No obstante, este supuesto rara vez se da en la realidad: los productos se encuentran diferenciados por sus características o existe una distancia espacial entre los puntos de venta que provoca que los consumidores tengan cierta preferencia por una de las variedades ofertadas.

En consecuencia, las empresas podrán fijar un precio por encima del Cmg. El precio de las diferentes variedades quedará determinado por la elección del grado de diferenciación por parte de las empresas del oligopolio.

[5]La tecnología de la empresa presenta CMe decrecientes debido a la presencia de CF. Esto provocará que el número de incumbentes sea finito y, debido a la condición de libre entrada de las empresas, los beneficios tenderán a 0.

[6]Decisiones empresas: Precio (corto plazo), Localización (largo plazo).

 © 2026 Manuel Cabado García

por las empresas (es decir, les da igual que las empresas estén situadas en lugares distintos). Lo contrario ocurre si los gustos de los consumidores son heterogéneos ($t > 0$).

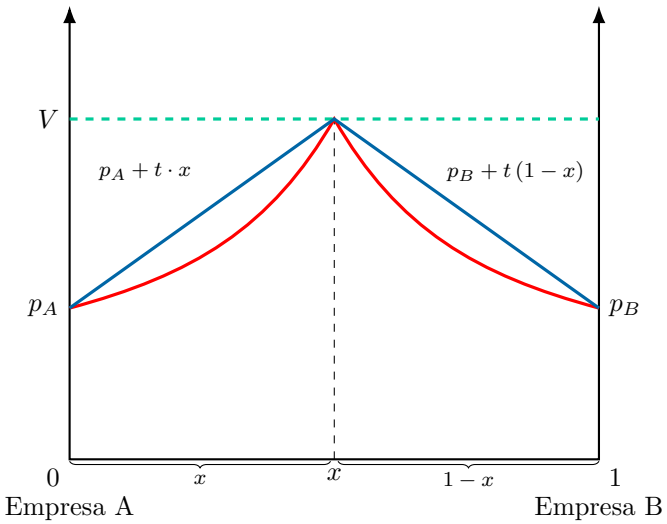
5. Todos los consumidores consumen, pues se asume que su excedente, V, es positivo tras pagar el precio y los costes de transporte.

La utilidad del consumidor, u, vendrá dada por su excedente, V, o "predisposición a pagar", menos el precio pagado, P, menos los costes de transporte, t, que pueden ser lineales, $t \cdot x$, –líneas rectas– o cuadráticos, $t \cdot x^2$, –curvas convexas–, como se verá.

$$u = V - P - t \cdot x \tag{10.1}$$

Gráficamente, en un plano de ciudad lineal, 0 y 1, y de excedente del consumidor como el de la Figura 10.2, si el consumidor se sitúa en x y cada empresa se sitúa en un extremo, la utilidad que obtiene el individuo de consumir 0 es la misma que la que obtiene de consumir 1: ninguna, pues en los extremos –que es donde están las tiendas–, $P + t = V$, pues el excedente del consumidor se agota por el precio pagado y por los costes de transporte por el camino recorrido hasta cualquiera de las tiendas. Se pueden definir los costes cuadráticos o lineales (costes cuadráticos en rojo, lineales en azul en el gráfico).

Figura 10.2: Ciudad lineal



Una vez establecidos los supuestos, se analizan los tipos de competencia en la ciudad lineal.

Competencia con localización exógena

Se asume que las localizaciones de las empresas son fijas, de forma que se mantienen en los extremos. Las empresas deciden, pues, únicamente sobre los precios (equilibrio de Nash

en precios).

Un consumidor que esté indiferente estará situado (es decir, tendrá unos gustos) allí donde su excedente neto de comprar el bien en la tienda A (a una distancia x) sea igual a su excedente neto de comprar el bien en la tienda B (a una distancia 1–x): $U_A = U_B$.

$$p_A + t \cdot x = p_B + t\,(1 - x) \tag{10.2}$$

La demanda de cada empresa será toda la que esté entre ella y el consumidor indiferente, de forma que, operando:

$$D_A\,(p_A, p_B) = x = \frac{p_B - p_A + t}{2 \cdot t} \tag{10.3}$$

$$D_B\,(p_A, p_B) = 1 - x = \frac{p_A - p_B + t}{2 \cdot t} \tag{10.4}$$

Dados los supuestos de arriba (localización en los extremos y excedente neto positivo), las demandas obtenidas (y, por tanto, el resto de conclusiones que siguen) serán las mismas con independencia de la forma que adquieran los costes de transporte (lineales –ejemplo visto– o cuadráticos –con x^2 y $(1–x)^2$–). Con máxima diferenciación (localización exógena) da igual que los costes sean lineales o cuadráticos.

Los beneficios de una empresa serán:

$$\pi_A = (p_A - c) \cdot D_A \tag{10.5}$$

donde $(p_A - c)$ es el margen unitario de la empresa A y c es el coste marginal.

Así, el programa de optimización de una empresa será la maximización de los beneficios

$$\max_{\{p_A\}} \quad \pi_A\,(p_A - c) \cdot D_A \tag{10.6}$$

de cuya CPO tanto para la empresa A como para la empresa B se obtienen las funciones de reacción:

$$FR\,(A) = \frac{p_B + t + c}{2}; FR\,(B) = \frac{p_A + t + c}{2} \tag{10.7}$$

y usando la simetría del problema, $p_A = p_B$, se obtienen los precios y los beneficios de las empresas[7]:

$$\pi_A^* = \pi_B^* = \frac{t}{2} \tag{10.8}$$

$$P_A^* = P_B^* = c + t \tag{10.9}$$

Por tanto, se observa que cuanto mayor sea la heterogeneidad de gustos entre los

[7]La elasticidad de la demanda cruzada de la empresa es positiva pero no tiende a infinito como en competencia perfecta. Por esta razón si una empresa reduce su precio acapara la demanda pero no toda por la diferenciación.

 © 2026 Manuel Cabado García

consumidores (es decir, mayor sea t), más cautivos serán de la empresa más cercana físicamente (es decir, de la empresa más cercana a sus gustos), y, por lo tanto, las empresas podrán exigir unos precios mayores y obtener beneficios más altos. Así, si los gustos de los consumidores son homogéneos, $t = 0$, se dará el resultado de Bertrand ($p = c$, $\pi = 0$), mientras que si son heterogéneos, $t > 0$, el precio tendrá un mark-up y los beneficios serán positivos.

Competencia con localización endógena

En el caso de que la localización sea endógena, el juego se define en dos etapas: primero, las empresas eligen simultáneamente su localización y, luego, dada su localización, deciden simultáneamente sus precios. Así, si una empresa decide moverse para el centro, experimentará 2 tipos de efectos contrapuestos: i) efecto demanda (+), pues aumenta su mercado objetivo, y ii) efecto competición (–), pues aumenta la competencia en precios (las empresas se diferencian menos).

Así, ante:

- Costes cuadráticos: el efecto competición será mayor que el efecto demanda[8], por lo que las empresas se situarán en los extremos de la ciudad (es decir, diferenciación máxima) y los precios serán, por tanto, máximos (d'Aspremont et al., 1979)[9].
  Si no hay competencia en precios (por ejemplo, porque el Gobierno fija un precio), entonces el efecto competición desaparece y las empresas se localizarán en el centro de la ciudad (es decir, diferenciación mínima).
- Costes lineales: puede no haber equilibrio. Ello se debe a que, si en una primera etapa las empresas deciden localizarse en el interior del intervalo en lugar de en los extremos, entonces una reducción en el precio de, por ejemplo, la empresa B, atraerá a consumidores situados entre las dos empresas, pero también a consumidores situados a la izquierda de la empresa A. Esto hace que las funciones de demanda sean discontinuas, y que las funciones de beneficios sean discontinuas y no cóncavas, por lo que puede no existir un equilibrio en precios.

[8] Los costes de transporte cuadráticos se representan, como se ha visto en el gráfico, mediante curvas convexas. Eso quiere decir que, a medida que la empresa se acerca al consumidor los costes de transporte disminuyen más que proporcionalmente. Es decir, la heterogeneidad de gustos se atenúa mucho, el consumidor es menos cautivo de la empresa y ésta tiene menos poder de mercado, aumentando mucho la competencia en precios. Por eso se dice que con costes cuadráticos el efecto competición es muy intenso (más que el efecto demanda).

[9] No obstante, trabajos posteriores como el de Neven (1985) matizaron este resultado: en el juego en dos etapas de localización y precios, el equilibrio nunca es de diferenciación mínima, pero no siempre es de diferenciación máxima, dependiendo de la curvatura de las preferencias de los consumidores.

Implicaciones de bienestar

En suma, cuando las empresas toman decisiones sobre su localización, no toman decisiones óptimas desde un punto de vista social: adoptan una diferenciación excesiva. Un planificador central podría aumentar el bienestar global minimizando los costes de transporte a los que se enfrentan los individuos localizando las empresas en los puntos ¼ y ¾[10].

Así, la estructura de precios no afecta al bienestar, debido a las demandas unitarias de los consumidores. El mark-up solo produce una redistribución de excedentes.

De esta forma, el modelo de Hotelling permite estudiar la competencia en precios con productos diferenciados en el contexto de duopolio. No obstante, deja sin responder el número de empresas que decidirán entrar en el mercado, así como su grado de diferenciación, bajo la presencia de costes de entrada. El modelo de ciudad circular de Salop (1979) permite comprender esta cuestión, en un contexto de oligopolios solapados[11].

Por consiguiente, en los modelos de diferenciación espacial las empresas desean diferenciarse para explotar la heterogeneidad de gustos de los consumidores que se expresa por los costes que estos afrontan por cambiar el consumo de una variedad por otra (es decir, los costes de transporte)[12].

Una idea básica pero crucial de estos modelos es que no hace falta que existan beneficios extraordinarios para que en un mercado se dé una competencia imperfecta. El grado de diferenciación será el resultado de dos fuerzas que operan en sentidos opuestos, el efecto demanda y el efecto competición.

[10] El planificador social escoge las localizaciones de las dos empresas, minimizando el coste de transporte medio de los consumidores, ya que el consumo es fijo. Esta forma de proceder es válida ya sea cuando las empresas: i) Ejercitan su poder de mercado como en el caso estudiado (máxima diferenciación) ii) se ven forzadas a fijar un precio igual al coste marginal (mínima diferenciación). Para unas ubicaciones dadas, y siempre que el mercado esté cubierto, la estructura de precios no altera la suma del excedente del consumidor y los beneficios en este modelo de demanda inelástica.

Por la simetría del problema, el planificador social elegirá las ubicaciones de las dos empresas en el medio de cada mitad del segmento, cuando la densidad de consumidores es uniforme. Por lo tanto, en el resultado de mercado existe demasiada diferenciación desde el punto de vista social.

[11] Véase el apartado 10.A.1 del Apéndice: la ciudad circular puede ser una analogía al mercado de centros comerciales alrededor de un anillo circular del centro urbano. La novedad de este modelo es que el número de empresas es endógeno y la localización exógena. Por lo tanto, lo que se estudia es la entrada de empresas dado que la localización ya ha sido tratada en el modelo de ciudad lineal.

[12] Sin embargo, en la vida real no siempre se da la diferenciación. Tirole (1988) señala tres razones:

1. Las empresas tienen cierta tendencia a situarse donde está la demanda.
2. La aglomeración de las empresas presenta ventajas: las economías de aglomeración permiten la especialización, la generación y aprovechamiento de externalidades positivas, y la disminución de los costes de almacenamiento y transporte empresariales para los bienes intermedios.
3. Ausencia de competencia en precios: la diferenciación de un producto tiene por objetivo suavizar la competencia en precios. Si no se compite en precios –por ejemplo, podrían estar regulados–, entonces la diferenciación será mínima.

 © 2026 Manuel Cabado García

10.4. Diferenciación horizontal no espacial ⏳9'

Por otro lado, ahora se analizará la diferenciación horizontal no espacial mediante modelos de competencia monopolística, donde las preferencias se definen en un espacio de bienes y no en un espacio de características.

El Modelo de Chamberlin

El estudio tradicional de la Competencia Monopolística comienza con el modelo de Chamberlin (1933) o modelo de grupo numeroso. Parte de la existencia de un conjunto de productos percibidos por los consumidores como sustitutivos cercanos con una elasticidad de sustitución alta, por lo que se utiliza un enfoque de no direcciones.

Estructura intermedia De esta forma, la competencia monopolística es una estructura de mercado intermedia entre la competencia perfecta y el monopolio, en la que muchas empresas, con libre entrada y salida, venden variedades diferenciadas de un mismo producto[13].

- Similitudes con la competencia perfecta: i) elevado número de empresas (el suficiente como para que individualmente no sean capaces de influir sobre el precio de la industria) y ii) libertad de entrada y de salida.
- Diferencia con la competencia perfecta: producto diferenciado, lo que otorga a cada empresa poder de mercado sobre su variedad, aunque está limitado a la existencia de competencia.
- Existen interdependencias no reconocidas.

Concretamente, en el enfoque no direcciones las preferencias de los individuos se definen sobre el espacio de bienes, es decir, hay varios bienes que son sustitutivos imperfectos, y los individuos diversifican su consumo entre dichos bienes (por ejemplo, Coca-cola vs. Pepsi). Que los bienes sean sustitutivos imperfectos significa que la elasticidad cruzada de su demanda será alta, pero no infinita.

Ahora no se pretende estudiar aspectos estratégicos de los productores como el posicionamiento de productos o competencia en precios. Ahora se estudiarán las consecuencias de la estructura de competencia monopolística sobre la eficiencia de producción y el bienestar.

[13] La intuición es que cada cafetería de un barrio es la única que ofrece su café, su local y su trato, y puede cobrar algo más que las demás sin perder a todos sus clientes, pero si abusa del precio estos se irán a la de enfrente.

Supuestos

Los supuestos del modelo son básicamente los mismos que en competencia perfecta, excepto por la no existencia de un producto homogéneo:

1. Diferenciación horizontal, es decir, competencia monopolística, de manera que cada empresa lanzará al mercado una variedad de producto diferenciada, por lo que habrá tantas variedades como empresas.
   Así, cada empresa será monopolista de su variedad, lo que le otorgará poder de mercado, de manera que ya no se enfrentará a una curva de demanda totalmente elástica sino con pendiente ligeramente negativa, pudiendo pues fijar un precio por encima de sus costes marginales.
   No obstante, su poder de mercado será limitado (es decir, la pendiente de su demanda será sólo ligeramente negativa), ya que cada empresa sigue sujeta a la competencia de las otras empresas en la medida en que los productos, al ser meras variedades, son sustitutivos imperfectos los unos de los otros.
2. Supuesto heroico[14]: todas las empresas del grupo se enfrentan a curvas de demanda idénticas, porque las preferencias de los consumidores se reparten por igual entre las variedades, y tienen curvas de costes idénticas (con forma de U), es decir, la diferencia de los productos no está en los costes. De ello se deriva que existe un precio común para toda la industria.
3. Maximización miope de beneficios. Las empresas maximizan beneficios, pero tienen un comportamiento miope: piensan que si ellas modifican los precios, las demás mantendrán constantes los suyos, es decir, las empresas tienen una variación conjetural nula. Por tanto, no existe interacción estratégica.

Equilibrio en el corto plazo

En este contexto, se define el corto plazo como el horizonte temporal donde viene dado el nivel de capital, $\bar{K}$, el tamaño de planta y el número de empresas, es decir, no hay entrada de empresas.

El equilibrio resultante es similar al de monopolio, es decir, los monopolistas se enfrentan a una curva de demanda con pendiente negativa por la heterogeneidad de productos, aunque ligeramente negativa ya que existe un gran número de bienes sustitutivos imperfectos.

Miopía: demanda planeada y efectiva Además, la miopía de los agentes a la hora de tomar decisiones da lugar a:

[14] Heroic assumption: An assumption that the speaker considers very unlikely to be true but temporarily accepts in order to make a hypothetical argument.

 © 2026 Manuel Cabado García

1. Demanda planeada, d, que refleja variación en las ventas sin tener en cuenta la reacción de los competidores. Esta curva es más elástica que la de demanda efectiva porque la empresa, por su miopía, cree que ante una disminución del precio va a vender más cantidad de la que realmente vende. Esta curva se desplaza (la ajusta la empresa) cuando el nivel de precios varía.
2. Demanda efectiva, D, que refleja variación en las ventas cuando todos los competidores ajustan sus precios suponiendo que la empresa mantiene su cuota de mercado. Solo se desplaza cuando cambia el número de empresas.

Análisis gráfico Gráficamente, en un plano de precios y cantidades como el de la Figura 10.3,

- La pendiente de la demanda efectiva, D, es mayor que la pendiente de la demanda planeada, d. Esto significa que ante una bajada de precios el aumento de las ventas real será menor que el esperado.
- Partiendo de un punto de equilibrio como el A, se supone que la empresa baja el precio de P_0 a P_1, creyendo que nadie lo seguirá, aumentando su demanda de x_0 a x_1.
- Sin embargo, el resto de empresas bajarán los precios para no perder su cuota, lo que hace que las ventas solo pasen de x_0 a x_1' [15].
- Por consiguiente, la auténtica demanda de la empresa estaría quebrada, y se correspondería con la demanda planeada, d, para precios superiores a P_0 y con la demanda efectiva, D, para precios inferiores a P_0. Aunque esta curva quebrada surge del cruce de d y D en el análisis de Chamberlin, él no la empleó como herramienta de análisis: fue Sweezy (1939) quien la convirtió en un modelo de oligopolio para explicar la rigidez de los precios.

Por otro lado, el objetivo de la empresa será la maximización de los beneficios determinado por los ingresos totales, IT, menos los costes totales, CT, asumiendo que los costes totales tienen una forma tradicional de U[16]:

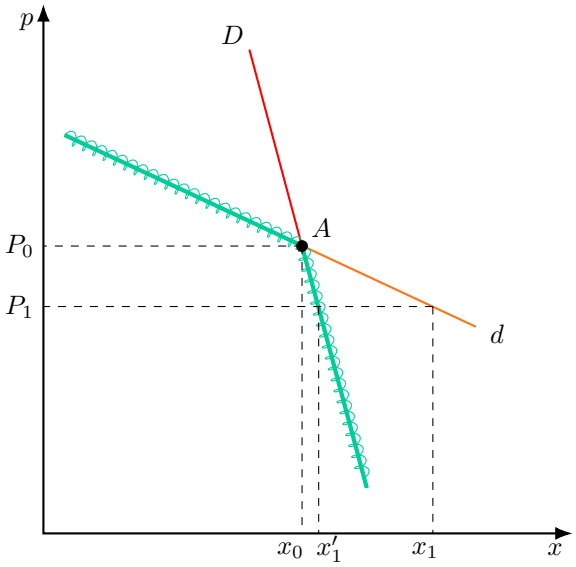
$$\text{máx} \quad \Pi = IT - CT = P(Q) \cdot Q - CT(Q) \tag{10.10}$$

Las CPO del problema determinan que el equilibrio se alcanza en el nivel en el que los ingresos marginales se igualan a los costes marginales, $IMg = CMg$, para un nivel de producción que coincida con la intersección de las curvas de demanda efectiva y demanda planeada, $d = D$. Si $IMg \neq CMg$ entonces existirán incentivos para alterar el nivel de producción, provocando reajustes.

[15] Inversamente, si sube el precio y el resto de empresas no la sigue, sus ventas se reducen a lo largo de la curva de demanda planeada, d, más elástica, lo que provoca una mayor reducción de las ventas.

[16] Cabe recordar que la forma de U de los costes deriva de asumir rendimientos mixtos.

Figura 10.3: Demanda Quebrada



Análisis gráfico Gráficamente, en un plano de precios, P, y cantidades, x, como el de la Figura 10.4, se representan las curvas de demanda efectiva, D, planeada, d, y la de coste marginal, CMg:

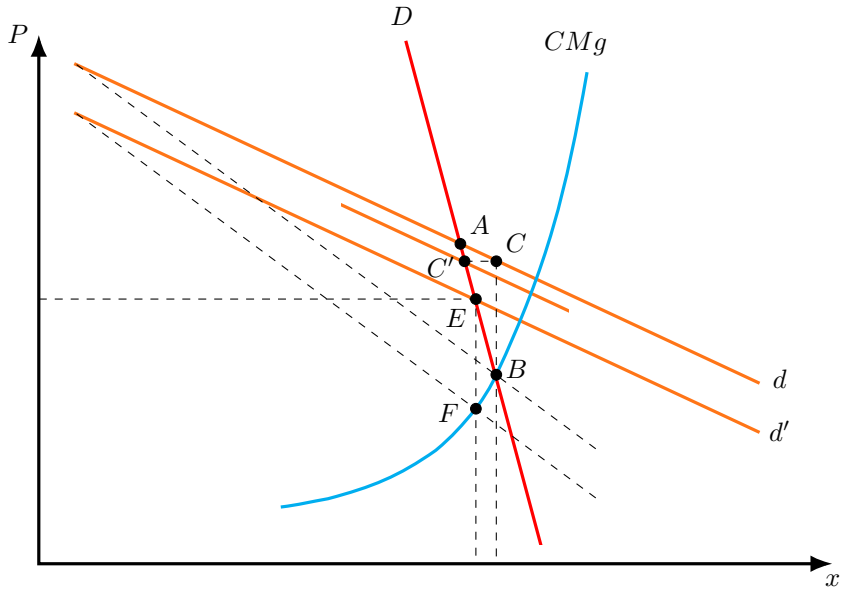
- Partiendo de un equilibrio como A los ingresos marginales (líneas de puntos), asociados a la demanda planeada, d, serán superiores a los costes marginales por lo que habrá incentivos a aumentar la producción hasta C, donde los ingresos y costes marginales coinciden en B.
- Para ello disminuirá los precios teniendo en cuenta su demanda planeada pero los competidores reaccionan, por lo que se alcanzará un nuevo equilibrio en C'. Cuando la empresa constata que los competidores también han reaccionado, revisará su demanda planeada y la desplazará hasta el nuevo equilibrio en C'. No obstante, en este punto los ingresos marginales seguirán siendo superiores a los costes marginales.
- Este movimiento se repetirá sucesivamente hasta alcanzar un equilibrio en E en el que los ingresos marginales coinciden con los costes marginales en F y, además, coincide con la intersección de las dos curvas de demanda. En dicho punto la empresa no tiene incentivos para modificar su producción.
- A su vez, en el corto plazo el equilibrio es compatible con beneficios o pérdidas dependiendo de dónde se sitúe la curva de costes medios, CMe.

Equilibrio en el largo plazo

Sin embargo, en el largo plazo existen dos fuerzas que pueden alterar las condiciones de equilibrio

 © 2026 Manuel Cabado García

Figura 10.4: Modelo de Chamberlin: equilibrio en el corto plazo



1. Las empresas pueden alterar su tamaño de planta y elegir el tamaño óptimo.
2. Existe libertad de entrada y salida dependiendo de la existencia de beneficios.

Esta modificación del número de empresas se verá reflejada en la demanda efectiva, D. Así, el resultado será un equilibrio en el largo plazo en el que desaparecerán los beneficios extraordinarios, determinando un equilibrio estable.

Análisis gráfico Gráficamente, en un plano de precios y cantidades como el de la Figura 10.5:

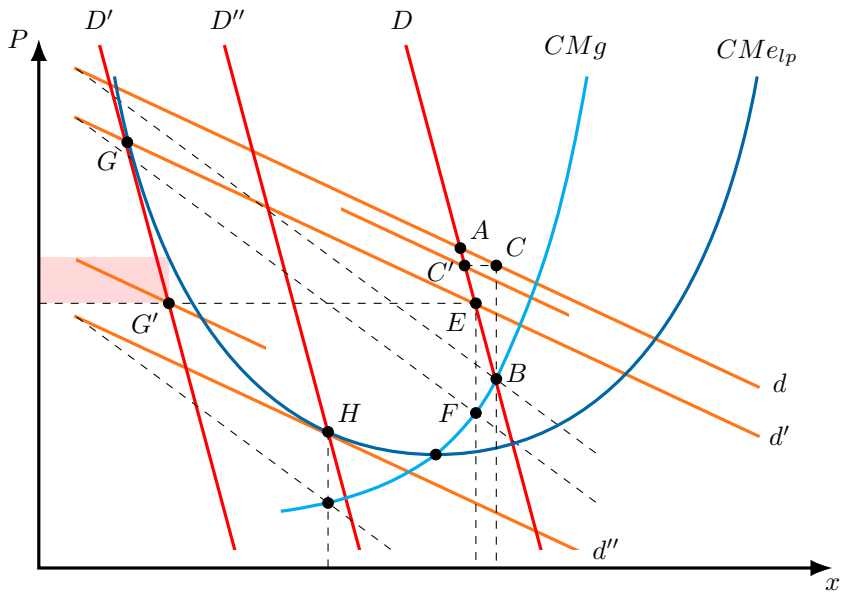
- En el punto E existirán beneficios, pues para este nivel de producción, el precio está por encima de la curva de costes medios a largo plazo, CMe_{lp}
- Ello atraerá a nuevas empresas al mercado, lo que va a producir un desplazamiento de la curva de demanda efectiva hacia la izquierda, D', hasta el punto en el que los beneficios desaparezcan, es decir, un punto situado sobre la curva de costes medios, en G. Como cada empresa tratará de recuperar el nivel de beneficios, disminuirá el nivel de precios hasta el nivel inicial.
- Pero dicho punto G es de desequilibrio ya que el ingreso marginal no se iguala al coste marginal, por lo que disminuirá el precio para recuperar la cantidad en el punto inicial E, considerando su curva d'. No obstante todas las empresas le siguen y solo podrá vender la cantidad correspondiente a G', la cual genera pérdidas dado que está debajo de la CMe, como se observa en el área sombreada.
- De la misma forma que en el corto plazo, esto hará que las empresas reajusten sus demandas planeadas. Además, la situación de pérdidas producirá una salida de

empresas del mercado desplazando la curva de demanda efectiva hacia la derecha. A su vez, como las empresas siguen siendo miopes seguirán disminuyendo los niveles de precios dando lugar a estos dos efectos sucesivamente:

1. En primer lugar el desplazamiento de la demanda planeada hacia abajo, $\Downarrow d$, y
2. En segundo lugar el desplazamiento de la demanda efectiva hacia la derecha, $D \Rightarrow$.

- Esto ocurrirá hasta que en un punto en el que las dos demandas se igualen, concretamente en el punto H con D'' y d'', donde la demanda planeada va a ser tangente a la curva de costes medios a largo plazo, y donde el ingreso marginal es igual al coste marginal. Esto asegura que a largo plazo los beneficios son normales –beneficios nulos–, generando un equilibrio estable donde no entran ni salen empresas.

Figura 10.5: Modelo de Chamberlin: equilibrio en el largo plazo



Teorema de exceso de capacidad

Como se puede observar, en el equilibrio a largo plazo cada empresa produce a un coste medio mayor que el mínimo de la curva de costes medios –punto donde se cortan CMg y CMe_{lp}–, es decir, a un coste mayor que si produjera con su capacidad óptima por lo que las empresas presentan economías de escala no explotadas.

Este hecho es analizado en el teorema de exceso de capacidad (Chamberlin, 1933; Kaldor, 1935)[17], cuya explicación radica en que los consumidores valoran la diferenciación:

[17] Chamberlin (1933) ofreció la primera exposición sistemática de esta doctrina, tal como reconoció el propio Kaldor (1935), cuyo análisis profundizó en sus implicaciones sobre la eficiencia de la libre compe-

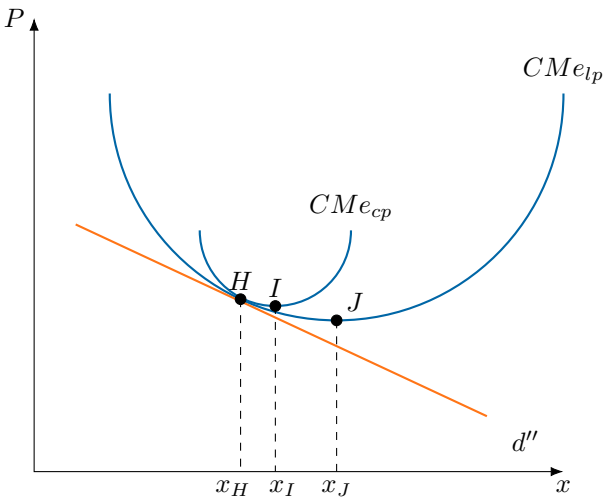
 © 2026 Manuel Cabado García

por lo tanto cuantas más variedades haya, menor será la cantidad producida de cada variedad, lo que dificultará el aprovechamiento de las economías de escala.

Análisis gráfico Gráficamente, en un plano de precios y cantidades producidas como el de la Figura 10.6 se representa la curva de costes medios a corto plazo y la curva de costes medios a largo plazo que es una envolvente de las diferentes curvas de costes medios a corto plazo. Además, representando la curva de demanda planeada, cada monopolista producirá en el punto H una cantidad, X_H, generando dos tipos de exceso de capacidad

1. Exceso de capacidad técnica, es decir, la diferencia entre la cantidad producida en H y la que minimizaría el coste medio de la planta instalada, en I, de modo que la empresa infrautiliza su propia planta y soporta un coste medio mayor del necesario.
2. Exceso de capacidad socioeconómica, determinado por la diferencia entre el mínimo de la curva de costes medios a corto plazo, I, y el mínimo de la curva de costes medios a largo plazo, J. El tamaño de planta no es óptimo desde un punto de vista social, esto se debe a que al competidor monopolista no le interesa construir una planta mayor.

Figura 10.6: Teorema exceso de capacidad



Además, el teorema de exceso de capacidad destaca el hecho de que en competencia monopolística el número de empresas pierde relevancia a la hora de caracterizar el mercado ya que el equilibrio está muy alejado del de competencia perfecta.

tencia –llegó a calificarla como una doctrina "intelectualmente sorprendente" e incluso "revolucionaria", pues invierte el argumento tradicional de que la entrada libre de competidores conduce a un uso eficiente de los recursos–. El propio Kaldor señaló antecedentes previos en Wicksell (1934) y en Cairnes (1874), y la esbozó ya Sraffa (1926) en su célebre artículo sobre los rendimientos bajo competencia. Destacó además el de Shove (1933), "The Imperfection of the Market", como uno de los análisis más penetrantes publicados hasta entonces sobre esta cuestión.

Ineficiencias

Por consiguiente, en el equilibrio en competencia monopolística se genera

- Ineficiencia asignativa, ya que el precio es mayor que el coste marginal, $P > CMg$. Al igual que en el monopolio, existe un rango de producción para el cual la valoración de los consumidores es mayor que el coste de producción.
- Ineficiencia productiva, ya que cada empresa produce con un coste medio superior al mínimo, $CMe >$ Mín CMe, debido al exceso de capacidad analizado anteriormente, y, como a largo plazo $P = CMe$, los consumidores pagan un precio superior a ese mínimo.

Bienestar Esto va a provocar un empeoramiento del bienestar de los consumidores aunque existen dos razones para pensar que la competencia monopolística no es tan negativa socialmente debido a que

1. El poder de mercado en competencia monopolística es reducido, la demanda suele ser muy elástica y las ineficiencias pequeñas.
2. La competencia monopolística ofrece a los consumidores una mayor diversidad de productos.

Por consiguiente, la valoración global de la competencia monopolística dependerá de la comparación de las ineficiencias y ventajas que genere[18].

Críticas al modelo de Chamberlin

Por otro lado, existen una serie de críticas al modelo de Chamberlin:

1. No se define de una forma precisa la industria, es decir, no se define el nivel de elasticidad cruzada que deben tener los bienes sustitutivos para formar parte de la misma industria. De ahí la célebre crítica de Stigler de que la indefinición es tal que, para ex fumadores, el chicle sería un buen sustitutivo del tabaco.
2. No precisa el número de empresas necesarias para distinguir entre una situación de competencia monopolística y de oligopolio, es decir, para que comience a darse interdependencia estratégica entre las empresas.

[18] Para deducir si el número de empresas es inferior o superior al óptimo hay que valorar dos efectos que operan en sentido contrario:

- No apropiación del excedente social (non appropiability of social surplus): las empresas no se pueden apropiar de todo el excedente consumidor asociado a la introducción de una nueva variedad. Esta externalidad positiva conduce a la introducción de menos variedades que el óptimo social.
- Robo de negocio (business stealing): La introducción de una nueva variedad reduce la demanda efectiva de cada empresa, lo que reduce sus beneficios. Esta externalidad negativa conduce a la introducción de demasiados bienes.

 © 2026 Manuel Cabado García

3. Concepto de demanda planeada insuficiente. La curva de demanda planeada solo considera consumidores finales, para los que es importante la diferenciación, pero, como señala Andrews (1949), deja fuera a gran parte de los consumidores intermedios de materias primas, para los que la diferenciación no es tan importante.
4. La diferenciación como concepto no casa con:
   a) El concepto de industria, puesto que impide la agregación de ofertas y de demandas individuales (salvo que el precio sea único, de ahí que Chamberlin se viese obligado a introducir el supuesto heroico).
   b) El supuesto heroico, pues si los productos son diferenciados, ¿por qué cabría esperar que las demandas de todos los productos fueran iguales, y los costes de todas las empresas fueran también iguales?
   c) Miopía, pues la diferenciación implica el estudio del comportamiento de la competencia para distanciarse de ella, por lo que resulta difícil creer que la empresa tenga una variación conjetural nula.
   d) Libre entrada, pues la diferenciación implica costes hundidos (por ejemplo, I+D, prospecciones de mercado, formación del personal, etc.), lo que facilita el surgimiento de barreras de entrada.

En suma, Chamberlin aportó un modelo esencial de la teoría de la competencia monopolística. Sin embargo, aunque con el modelo se representan las ineficiencias productivas y asignativas, no se pueden observar cuáles son las ganancias en bienestar del consumidor: ¿cómo se representa la eficiencia resultante de la preferencia por la variedad? Con el modelo de Dixit y Stiglitz.

Preferencia por la variedad

Por otro lado, los modelos de preferencia por la variedad justificarán la diferenciación y completan o refinan el modelo de Chamberlin. En ese sentido, hasta ahora se ha supuesto que cada consumidor se posicionaba por una variedad preferida. Sin embargo, es posible que el consumidor prefiera diversificar su consumo en diferentes variedades.

Para analizar este enfoque se utilizará el modelo de Dixit y Stiglitz (1977), cuyo rasgo característico es la presencia de un consumidor agente representativo.

Supuestos Parte de los siguientes supuestos:

1. Estructura de mercado de herencia chamberliana, es decir, de diferenciación horizontal del producto con competencia monopolística.
2. Economías de escala internas a la empresa. La tecnología de producción (la misma para todas las empresas) está compuesta por un coste fijo y un coste variable constante: $CT = CF + cq$ (rendimientos crecientes a escala)

3. Existe libre entrada en el mercado, por lo que (E(B^{o})=0)
4. Preferencia por la variedad, es decir, que la utilidad que le reporta al individuo el consumo de n variedades es, ceteris paribus, superior a la que le reporta el consumo de tan solo n–1 variedades. Por lo tanto, los consumidores eligen varios bienes.
5. Supuesto heroico de precio común a la industria.

Problema de maximización del consumidor

En este contexto, la preferencia por la variedad se representa mediante una función de utilidad tipo CES –es decir, con elasticidad de sustitución constante–. Así, la función de utilidad se compone de la suma de x_i, que es cada una de las variedades consumidas del bien, y n representa el número de variedades donde θ determina el grado de sustitución entre los bienes

$$u = \sum_{i=1}^{n} x_i^{\theta}, \quad \theta \in (0,1) \tag{10.11}$$

donde θ es el grado de sustitución de los n bienes, y que será menor que uno y mayor que cero (porque la función de utilidad es cóncava, implica sustituibilidad imperfecta). Por otro lado, esta función verifica la preferencia por la variedad.

La elasticidad de sustitución entre variedades, σ, que mide en qué porcentaje varía el cociente entre las cantidades consumidas de dos variedades cuando su precio relativo varía un 1 %, asociada a esta función de utilidad y a la preferencia por la variedad, PV, será

$$\sigma = \frac{1}{1-\theta} = \frac{1}{PV} \tag{10.12}$$

por lo que cuanto mayor sea la preferencia por la variedad, PV, menor elasticidad de sustitución entre variedades. Es decir, cuanto mayor sea el valor de $1-\theta$ existe mayor preferencia por la variedad, ya que los consumidores consideran menos sustitutivas unas variedades respecto a otras[19].

De esta forma, el consumidor representativo maximizará su utilidad

$$\text{máx} \quad u = \sum_{i=1}^{n} x_i^{\theta} \tag{10.13}$$

sujeto a una restricción presupuestaria

$$\sum_{i=1}^{n} p_i \cdot x_i \leq m \tag{10.14}$$

[19]De esta manera:

- Si $\theta = 1$, entonces los bienes son sustitutivos perfectos, la elasticidad de sustitución es infinita ($1/(1–1) = \infty$) y no existe preferencia por la variedad ($1–1 = 0$).
- Si $\theta = -\infty$ los bienes son perfectamente complementarios.

 © 2026 Manuel Cabado García

donde p_i es el precio del bien x_i y m representa la renta.

La CPO del problema es la siguiente

$$\frac{x_i}{x_j} = \left(\frac{p_j}{p_i}\right)^{\frac{1}{PV}} \tag{10.15}$$

donde, además, la elasticidad de sustitución de los bienes coincide en este modelo con la elasticidad-precio de la demanda en valor absoluto[20].

La interpretación de este resultado es clara: cuanto mayor sea la preferencia por la variedad de un individuo, más rígida será la demanda de cada bien (es decir, menos elástica, más cercana a cero), y menor, por tanto, la sustituibilidad entre dos bienes. Por eso, si la preferencia por la variedad es alta, los precios de las diferentes variedades podrán divergir más sin que ello afecte tanto a las cantidades relativas demandadas.

Problema de maximización del productor

Por el lado de la oferta, se supondrá que cada producto x_i se fabrica solo con trabajo l_i que depende de forma lineal de la cantidad producida

$$l_i = a + b \cdot x_i \tag{10.16}$$

donde a representa el tamaño de las economías de escala, y b la cantidad de trabajo necesaria por cada unidad adicional producida.

Además, dado un salario constante, w, la función de costes de cada variedad, C_i, vendrá determinada por

$$C_i = w \cdot a + w \cdot b \cdot x_i \tag{10.17}$$

donde $w \cdot a$ representan los costes fijos y un coste variable que depende de la cantidad producida, $c \cdot x_i$, donde $c = w \cdot b$ representa el coste marginal. Cabe destacar que la existencia del coste fijo hace que la producción de la empresa tenga economías de escala: su coste medio es decreciente con el nivel de producción de la empresa: $CMe_i = \dfrac{w \cdot a}{x_i} + w \cdot b$

El objetivo de la empresa será maximizar su beneficio

$$\text{máx} \quad \Pi = IT - CT = p_i \cdot x_i - C_i \tag{10.18}$$

y, dado que la empresa tiene poder de mercado sobre la variedad que produce, la condi-

[20] Por tanto, la función de demanda resultante del programa de optimización es una función de demanda isoelástica del tipo $q = a \cdot p^{-b}$, donde la elasticidad-precio de la demanda, ε_{x_i,p_i}, es, como en toda demanda isoelástica, $-b$. Por lo tanto, la elasticidad-precio de la demanda en valor absoluto coincide en este modelo con la elasticidad de sustitución de los bienes: $1/(1-\theta)$.

Así, cuanto mayor es la preferencia del individuo por la variedad ($\uparrow PV$), más se acerca a cero la elasticidad-precio de la demanda pues se hace menos negativa, $-1/(\uparrow PV)$, es decir, se vuelve más rígida y menos sensible, pues, al cambio en los precios.

ción de primer orden establece que los ingresos marginales se deben igualar a los costes marginales, en función de la elasticidad-precio de la demanda, ε, es decir

$$IMg = CMg \Rightarrow p_i \left(1 - \frac{1}{\varepsilon}\right) = w \cdot b \tag{10.19}$$

e introduciendo el supuesto de simetría –todas las empresas son idénticas, tienen la misma función de coste y se enfrentan a demandas idénticas: $p_i = p \quad \forall i$– y sustituyendo el valor de la elasticidad precio de la demanda, ε, se obtiene que el precio será

$$p = \frac{w \cdot b}{\theta} \tag{10.20}$$

De esta forma, las conclusiones son análogas a las obtenidas cuando se analizó el problema del consumidor: el productor podrá lanzar al mercado un precio mayor cuanto menor sea la sustitución de bienes para el individuo ($\downarrow\vartheta$), es decir, cuanto mayor sea su preferencia por la variedad ($\uparrow$PV)[21].

Cantidad producida Por otro lado, bajo la condición de libre entrada y salida del mercado los beneficios deberán ser igual a 0, $\Pi = 0$, por lo que sustituyendo el precio (10.20) en la función de producción (10.18) y despejando en función de x_i

$$x = \frac{a \cdot \theta}{b\,(1 - \theta)} \tag{10.21}$$

por lo tanto, la cantidad producida depende

- Negativamente, de la preferencia por la variedad, $(1 - \theta)$, y no del tamaño de mercado: a mayor preferencia por la variedad, menor será la cantidad producida de cada una de las variedades y menor aprovechamiento de las economías de escala, existiendo exceso de capacidad.
- Negativamente, del parámetro b que determina el coste marginal.
- Positivamente, del parámetro a que determina el coste fijo y por lo tanto el nivel de economías de escala.

Número de empresas Por otro lado, para obtener el número de empresas y, por lo tanto, de variedades, primero se obtendrá la cantidad de trabajo que utiliza cada empresa a partir de la función de producción (10.16)

$$l = \frac{a}{1 - \theta} \tag{10.22}$$

[21] Como se ve, las conclusiones son análogas a las obtenidas en el modelo de Hotelling explicado en el enfoque de direcciones.

 © 2026 Manuel Cabado García

que, al igual que la cantidad producida, depende negativamente de la preferencia por la variedad.

Número de variedades Además, si se supone que la cantidad total de trabajo en la economía está fija y viene dada por el tamaño de la población, L, el número de empresas, y por tanto de variedades, será:

$$n = \frac{L}{l} \Rightarrow n = \frac{L\,(1-\theta)}{a} \tag{10.23}$$

por lo que el número de variedades depende

- Positivamente del tamaño del mercado, L.
- Positivamente de la preferencia por la variedad, $1-\theta$.
- Negativamente de las economías de escala, a través del parámetro a que determina el coste fijo.

Conclusiones del modelo de Dixit y Stiglitz

Por lo tanto, el modelo de Dixit y Stiglitz (1977) implica:

1. La diferenciación aumenta la utilidad de los consumidores porque tienen preferencia por la variedad.
2. Es posible fijar los precios con mark-up (mayores al coste marginal).
3. Suaviza la competencia en precios.
4. Justifica actividades como la publicidad.
5. Beneficio normal a largo plazo ($P = Cme$).
6. El número de empresas dependerá: (-) CF, (+) de la PV, 1–θ, y (+) de las ventas totales de la industria.

Además, su importancia radica en que este modelo de competencia monopolística constituye la base de las nuevas teorías del comercio internacional, del crecimiento económico así como en la Macroeconomía y la Geografía Económica[22].

[22] Concretamente:

- En el modelo de comercio internacional de Krugman la preferencia por la variedad permite explicar el comercio intraindustrial, ya que al eliminar barreras se incrementa el número de variedades y también deriva en ganancias de eficiencia al operar bajo mayores economías de escala y alcanzar un tramo inferior de la curva de costes medios.
- En el modelo canónico de la Nueva Economía Keynesiana desarrollado por Blanchard y Kiyotaki (1987), ya que la preferencia por la variedad va a permitir introducir el poder de mercado de las empresas de competencia monopolística que unido a rigideces nominales y reales supondrá:
  - La no neutralidad del dinero.
  - Ajuste ineficiente a nivel agregado ante un shock.

10.5. Diferenciación Vertical ⏳ 2'

Por último, se va a analizar el concepto de diferenciación vertical, donde las diferencias entre los productos no surgen de diversidad de las preferencias individuales acerca de las características de un bien, sino que se trata de cualidades que todos los consumidores desean en el nivel más alto posible, siendo la calidad el caso clásico[23][24].

La posible diferenciación surge de las divergencias respecto a la disposición a pagar por esa mayor calidad. Esta disposición al pago resume las diferencias que surgen tanto de las intensidades de preferencia por la calidad, como de las que proceden de los diferentes niveles de renta.

Por tanto, la diferenciación vertical halla su encaje dentro del enfoque de direcciones, pero el hecho de que todos los consumidores coincidan en la "dirección" preferida del bien le confiere unas características particulares.

Modelo Para analizar la diferenciación vertical se realizarán unas modificaciones al modelo de Hotelling anteriormente analizado:

- Los consumidores se sitúan en un intervalo $[0, 1]$ según la intensidad de su preferencia por la calidad, x, de modo que un valor más próximo a 1 indica una mayor disposición a pagar por cada unidad de calidad.
- Los consumidores están distribuidos uniformemente a lo largo de dicho intervalo.
- Por simplicidad se analizarán dos empresas, A y B, donde la calidad del bien B es superior a la de A.

La función de utilidad del consumidor situado en x que consume un bien de la empresa i depende positivamente del bien consumido y negativamente del precio:

$$U_x(i) = \begin{cases} a \cdot x - p_A & i = A \\ b \cdot x - p_B & i = B \end{cases} \tag{10.24}$$

donde la utilidad derivada de consumir el bien B es mayor ($b \cdot x > a \cdot x$) por lo que la disposición al pago será también mayor. Además, se supondrá que la utilidad derivada deberá ser igual o superior a 0 para que el consumidor quiera comprar, existiendo así una

- Efectividad real sobre la economía de la política fiscal y monetaria, lo que permitirá suavizar los ciclos económicos.

[23] Este enfoque permite estudiar por qué un mercado puede llegar a segmentarse en función de la calidad y por qué puede ser óptimo para unas empresas elegir como objetivo una calidad inferior a la máxima. Este hecho se observa especialmente en el sector tecnológico, donde, a medida que la demanda de productos ha aumentado, más competidores han entrado, y más se ha segmentado el mercado en función de la calidad. Otro ejemplo serían las marcas de ropa.

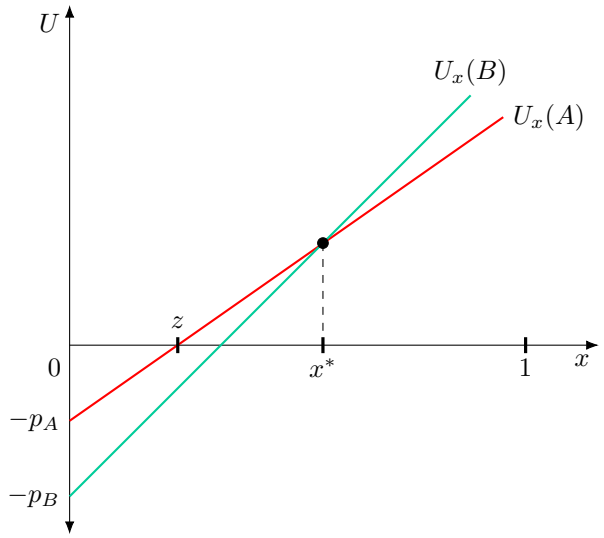
[24] Ejemplos de esto pueden incluir atributos como la potencia del motor en un coche, la capacidad de almacenamiento en un dispositivo electrónico, o la velocidad de procesamiento en un ordenador

 © 2026 Manuel Cabado García

"utilidad de reserva" que deja fuera del mercado a los consumidores con menor valoración, de forma que solo compran los situados en $[z, 1]$, con $z > 0$.

Análisis gráfico Gráficamente, en un plano de utilidad, U, y calidades, x, como el de la Figura 10.7, se representan las preferencias en el espacio 0 y 1, mediante las curvas de indiferencia, U_A y U_B, cuyos interceptos son los precios, p_A y p_B, y que tendrán pendiente positiva dado que el precio es un mal y la calidad un bien. Cuanto mayor sea la valoración de los individuos de las calidades mayor será la pendiente de la curva de indiferencia asociada, en este caso la de B es superior a la de A. El intervalo $[0, z]$ representa la utilidad de reserva, mientras que el consumidor situado en x^* es indiferente entre consumir una variedad u otra.

Figura 10.7: Diferenciación vertical



Los modelos de diferenciación vertical así definidos, como la modificación del modelo de Hotelling para incorporar diferencias de calidad propuesta por Shy (1995), proporcionan algunos resultados interesantes desde el punto de vista de la estrategia empresarial. Así, se considerará una situación donde dos empresas, A y B, producen sendos bienes sin ningún coste de producción, por simplicidad, por lo que se obtendrían funciones de reacción tales que la empresa que produjera un bien con una mayor calidad obtendría un beneficio mayor.

Se plantea un juego de dos etapas donde:

1. En la primera etapa se elige la calidad productiva, con ambas empresas teniendo incentivos a mover primero y elegir la calidad elevada, pues es con la que se obtienen mayores beneficios.
2. En la segunda etapa se elige el precio de venta.

Una vez establecidos los papeles, la empresa que produce una calidad inferior no tendrá incentivo a mejorarla, sino a mantener una diferenciación máxima –es decir, las empresas se situarían en los extremos, repartiéndose A desde z hasta x^* y B desde x^* hasta 1– para evitar una lucha de precios, dando lugar a un equilibrio de Nash[25].

Vertical vs Horizontal Este resultado de estos modelos de diferenciación vertical presenta dos diferencias importantes respecto a los modelos de diferenciación horizontal:

1. El número de variedades no depende tanto del tamaño de mercado, sino de la heterogeneidad de los consumidores en su predisposición a pagar, es decir, la distribución de la renta. De hecho, en un mercado de diferenciación vertical con consumidores muy homogéneos en el que ya existen dos empresas, una tercera que ofreciera una calidad distinta no podría tener cabida, dando lugar a una situación de oligopolio natural.
2. Las ganancias del comercio internacional no se derivan de la creciente gama de productos sino del estímulo –mayor rentabilidad asociada– que genera la mayor competencia del comercio internacional, que fuerza a las empresas a producir una mayor calidad a menores precios, aumentando así el gasto en I+D.

No obstante, es preciso señalar que las diferencias entre los modelos de diferenciación vertical y horizontal se reducen si se rompe con el supuesto de costes variables nulos. Además, en ocasiones la distinción entre diferenciación vertical y horizontal se difumina[26].

10.6. Conclusiones ⏳30"

Como conclusión, a lo largo del tema se han analizado diferentes modelos de competencia monopolística, los cuales mejoran la concepción neoclásica al introducir la heterogeneidad entre bienes. La frontera entre diferenciación vertical y horizontal es difusa ya que, por ejemplo, en el comercio internacional si existe consenso en que la producción nacional es mejor que la extranjera la diferenciación será vertical pero si los consumidores prefieren consumir ambos productos la diferenciación será horizontal. Por lo tanto, la actividad

[25] Implicaciones: En el equilibrio, ambas empresas buscarán mantener la máxima diferenciación posible (pues de este modo la competencia en precios será menor y esto beneficiará a ambas). Aunque la calidad se puede aumentar sin coste y es valorada positivamente por todos los consumidores, la empresa A prefiere producir a una calidad baja para diferenciarse más de su rival.

[26] Por ejemplo, en el ámbito del comercio internacional se aprecian sesgos tanto a favor de los productos nacionales como de los importados. Si hubiera consenso en que respecto al producto i los artículos nacionales son mejores (o peores) que los importados, se estaría ante un caso de diferenciación vertical. Por otro lado, si los consumidores prefieren consumir productos nacionales y parte prefieren consumir productos importados, la situación se acercaría a la diferenciación horizontal. El comercio internacional se vería favorecido en el último caso, de modo que un aumento en la heterogeneidad de preferencias equivale a un aumento de la diferenciación de productos.

 © 2026 Manuel Cabado García

comercial se ve favorecida por la heterogeneidad de las preferencias, aumentando la diferenciación del producto.

Apéndice

10.A.1. El modelo de ciudad circular de Salop

En el modelo de Hotelling se había supuesto que solo hay dos empresas. No obstante, es importante considerar la posible entrada de empresas en el mercado, por lo que en vez de una "ciudad/playa lineal" es más conveniente estudiar una "ciudad circular"[27].

Salop consideró una ciudad circular –por ejemplo, horarios de vuelos de un avión o la programación de televisión–, suponiendo que existen, no solo costes de transporte, sino también costes fijos de entrada, CF, y, por lo tanto, economías de escala.

Así, se define un juego en dos etapas:

1. Primero, las empresas deciden simultáneamente si entran o no. Las empresas que entran se situarán todas equidistantes las unas de las otras –es decir, no hay decisión sobre la localización–, de manera que la diferenciación máxima es exógenamente impuesta.
2. Segundo, deciden simultáneamente sus precios, dada su localización y el número de empresas en el círculo.

El resultado del modelo de ciudad circular es:

- El número de empresas dependerá positivamente de los costes de transporte –es decir, de la heterogeneidad de gustos de los consumidores–, y negativamente de los costes de entrada: $n = f(\underset{+}{t}, \underset{-}{CF})$.
- El precio dependerá positivamente del coste marginal y de los costes de transporte –el aumento de los costes de transporte les da mayor poder de mercado a las empresas dado n y eso les permite poner precios más altos. Los precios más altos permiten la entrada de más empresas–, y negativamente del número de empresas –es decir, de la competencia–: $p = f(\underset{+}{CMg}, \underset{+}{t}, \underset{-}{n})$.

10.A.2. Teoría de la demanda de características de Lancaster

Dentro de los modelos de diferenciación horizontal, el economista australiano Kelvin Lancaster (Lancaster, 1966) propuso una nueva teoría de demanda del consumidor para superar tres cuestiones que la teoría neoclásica no era capaz de explicar:

1. Por qué los consumidores son fieles a las marcas.
2. Por qué se desarrollan nuevos productos.
3. Por qué existe la publicidad.

[27] La ventaja de modelar la "ciudad" o el espacio de productos como un círculo es que ninguna posición es superior a las demás, no hay extremos ni centro. Al contrario de la ciudad lineal donde en el centro se tiene una mayor demanda. Aquí el espacio de los productos es completamente homogéneo.

 © 2026 Manuel Cabado García

Lancaster parte del enfoque de que lo que demandan los consumidores no son bienes en sí mismos, sino las características presentes en ellos, que son las que en realidad les permiten satisfacer sus necesidades. Así, dichas características coexisten en un mismo bien en distintas proporciones.

De esta forma, el consumidor se enfrenta a un problema de maximización de utilidad derivada de la cantidad de características

$$\text{máx} \quad u\left(a_1, \ldots, a_n\right) \tag{10.25}$$

sujeto al conjunto de características a, donde hay n características, y $a_{i,j}$ es la proporción de la característica i que tiene del bien 1, ya que hay l bienes

$$a = a_{1,j} \cdot x_1 + \ldots + a_{i,j} \cdot x_j \tag{10.26}$$

$$m \geq p_1 \cdot x_1 + \ldots + p_j \cdot x_j \tag{10.27}$$

Gráficamente, se supondrá un plano de dos características, a_1 y a_2, como el de la Figura 10.8, y se representarán los vectores de posibilidades de consumo, X, Y y Z, que son las líneas que recogen las proporciones en que cada bien combina las dos características. Las combinaciones X', Y' y Z' corresponden a los casos en que el consumidor gasta toda su renta en dichos bienes, mientras que las líneas que unen estos puntos representan combinaciones intermedias.

En función de las preferencias del individuo, este elegirá una combinación u otra: los individuos que prefieran a_1 tendrán curvas de indiferencia parecidas a u y maximizarán su utilidad escogiendo un punto como el E, donde pueden obtener la combinación de a_1 y a_2 especificada en ese punto consumiendo solo los bienes Y y Z.

Por otro lado, los consumidores cuyas preferencias estén representadas por la curva de indiferencia u' escogerán el punto E' y consumirán únicamente los bienes X e Y.

Así, el modelo de demanda de características predice que las soluciones de esquina en las cuales los individuos consumen una cantidad nula de algunos bienes serán relativamente frecuentes.

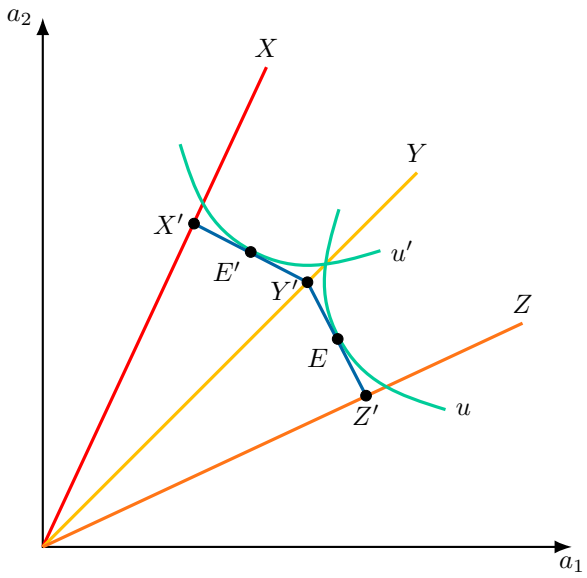
A continuación, se analizará cómo este modelo da respuesta a las tres preguntas formuladas al principio.

Fidelidad a las marcas

La teoría de la demanda de características explica por qué, ante un cambio en el precio de un producto, los consumidores pueden mantenerse fieles a su marca. En efecto, un aumento del precio de un bien daría lugar, por lo general, en un modelo neoclásico tradicional a la sustitución (siquiera parcial) de dicho bien por otro[28].

[28] Esto no ocurre en el modelo neoclásico si las soluciones son de esquina.

Figura 10.8: Modelo de Lancaster



En este modelo, sin embargo, ello no tiene por qué ser así como se observa en la Figura 10.9

1. Se supondrá que, inicialmente, lo máximo que puede consumir un individuo de los bienes X e Y es X^* e Y^*. Dada la curva de indiferencia u_0, preferirá dedicar toda su renta a Y.
2. Si el precio de X baja y la cantidad máxima consumible pasa a ser $X^{*'}$, el consumidor se mantiene fiel a la marca Y.
3. No obstante, si el precio baja aún más y la cantidad máxima consumible pasa a ser $X^{*''}$, el consumidor cambiará de marca y pasará a consumir el bien X, pues su consumo máximo se sitúa ahora en una curva de indiferencia superior, u_1.

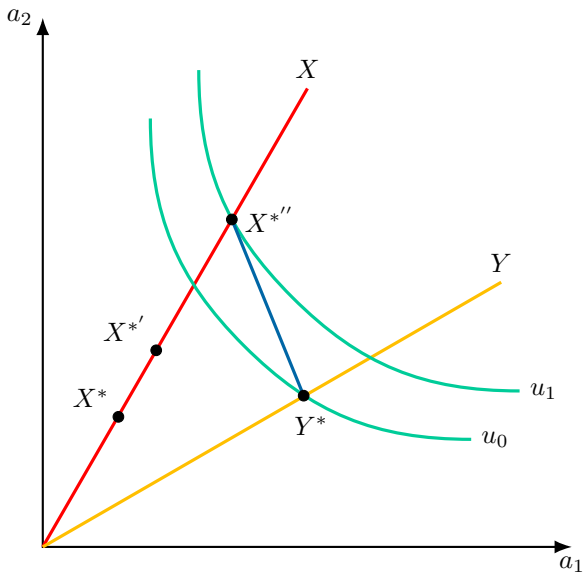
Desarrollo de nuevos productos

Por otro lado, es posible que las preferencias del individuo no se ajusten plenamente a ninguna de las variedades existentes. Como se observa en la Figura 10.10, ello se reflejaría en un equilibrio localizado en medio de algún segmento de la frontera eficiente como E, de forma que el consumo conjunto satisfaría la demanda de características del individuo.

Además, también explicaría la creación de una nueva variedad que contuviera dichas características en esa proporción y desbancara a los otros bienes, lo cual, a su vez, da pie a la teoría de la prospección de mercados –para averiguar exactamente lo que cada cliente demanda– y a la teoría de la segmentación –para ofrecer exactamente a cada cliente lo que demanda–. En el gráfico, el nuevo producto sería Y, y el consumidor se situaría en Y'.

 © 2026 Manuel Cabado García

Figura 10.9: Modelo de Lancaster: fidelidad a las marcas



Publicidad

Teniendo en cuenta todo lo anterior, esta teoría justifica la importancia de dar a conocer las características de un producto mediante la publicidad[29].

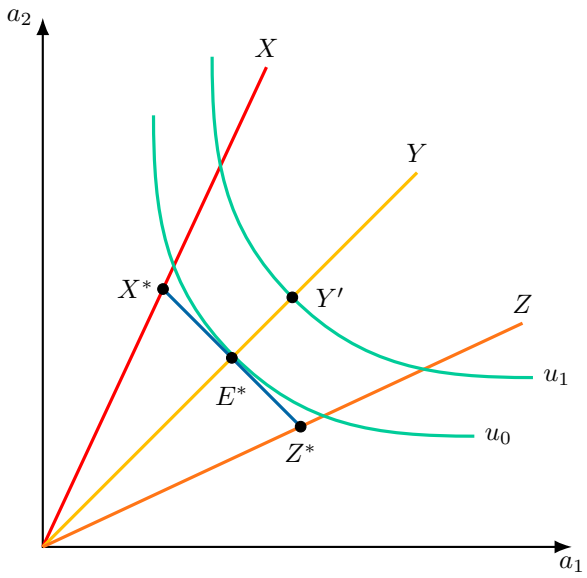
10.A.3. Perspectiva histórica: los enfoques de direcciones y de no direcciones

Para cerrar este apéndice, conviene situar los modelos analizados en su contexto histórico. Como señalan Eaton y Lipsey (1989) en su repaso histórico de la literatura de diferenciación de productos, esta se ha desarrollado a lo largo de dos ramas paralelas: el enfoque de no direcciones, que sitúa las preferencias sobre el espacio de bienes, y el enfoque de direcciones, que las sitúa sobre el espacio de características.

En la rama de no direcciones, Marshall parte del supuesto de bienes homogéneos y competencia perfecta, hasta que Sraffa señala la incoherencia entre las economías de escala no explotadas que se observan en la realidad y ese supuesto marshalliano. Robinson y Chamberlin responden en 1933-1934: el monopolio único de industria de Robinson resulta un callejón sin salida, mientras que el grupo numeroso de Chamberlin, con entrada libre y

[29] El propio modelo de Lancaster se ha aplicado también fuera del ámbito estrictamente económico: Feás (2019) lo utiliza para analizar el comportamiento electoral, argumentando que los votantes no eligen tanto por una posición ideológica en el eje izquierda-derecha como por un conjunto reducido de características concretas –en el caso español, apunta como ejemplo "evitar que gobierne la izquierda o la derecha", "firmeza frente al independentismo" y "regeneración política y lucha contra la corrupción"–, lo que explicaría trasvases de voto aparentemente contradictorios entre partidos ideológicamente distantes. El propio autor cierra su análisis invocando la teoría del segundo óptimo de Lancaster (desarrollada junto con Lipsey) a modo de advertencia: si un votante prioriza una única característica, el resultado no tiene por qué ser la segunda mejor opción posible en el resto de las dimensiones.

Figura 10.10: Modelo de Lancaster: desarrollo de nuevos productos



cada empresa produciendo por debajo de su escala eficiente mínima, triunfa al reconciliar competencia y economías de escala no explotadas –naciendo así el teorema del exceso de capacidad–. Kaldor cuestiona después el supuesto de simetría de Chamberlin y propone un modelo de oligopolios solapados con vecinos cercanos y lejanos, anticipando ya el enfoque de direcciones. El interés por esta rama revive en los años setenta, impulsado por la integración de la Comunidad Económica Europea y el auge del comercio intraindustrial: Spence (1976a) y, después, Dixit y Stiglitz (1977) formalizan el consumidor representativo que se ha empleado en este tema, generalizado posteriormente a consumidores heterogéneos manteniendo el supuesto de simetría.

En la rama de direcciones, el punto de partida es la crítica de Bertrand al duopolio de bien homogéneo de Cournot, de la que parte Hotelling (1929) al introducir la diferenciación –geográfica o de características– para obtener un equilibrio estable con precios por encima del coste marginal, en la línea de los modelos geográficos de localización de Christaller y Lösch. Lerner y Singer amplían después el modelo de Hotelling a más de dos empresas, y Smithies modifica el supuesto sobre la demanda. El interés por esta rama revive también en los años setenta a través de la teoría de la demanda: Lancaster (1966), junto con Baumol y Quandt, desarrollan la teoría de características que se ha analizado en el Apéndice 10.A.2, y el principio de diferenciación mínima –ya visto en el modelo de Hotelling– da paso a los modelos de entrada con previsión estratégica de la localización de los competidores.

Así, los modelos analizados a lo largo de este tema recorren ambas ramas: Chamberlin y Dixit-Stiglitz pertenecen al enfoque de no direcciones, mientras que Hotelling, Salop y Lancaster pertenecen al enfoque de direcciones. El modelo de diferenciación vertical de Shaked y Sutton, por su parte, extiende este último enfoque añadiendo una dimensión

 © 2026 Manuel Cabado García

de calidad sobre la que todos los consumidores coinciden en su valoración, en vez de una característica sobre la que las preferencias difieren.

10.A.4. Otros desarrollos: publicidad y proliferación de marcas

Cerrando ya esta perspectiva histórica, cabe detenerse en dos desarrollos concretos que el enunciado del tema menciona de forma expresa: la publicidad y la proliferación de marcas. Como se ha visto en el Apéndice 10.A.2, el modelo de Lancaster ya justificaba la publicidad como vía para dar a conocer las características de un producto. Dorfman y Steiner (1954) formalizaron esta intuición: la publicidad, al desplazar la demanda de la empresa y hacerla menos elástica, es otro instrumento de diferenciación además de la localización o la calidad.

El modelo de publicidad de Dorfman y Steiner

Se considera una empresa cuya demanda, $q(p, A)$, depende del precio, p, y del gasto en publicidad, A, y cuyo coste de producción, $C(q)$, es independiente de dicho gasto, que se trata como un desembolso adicional. El objetivo de la empresa será maximizar su beneficio

$$\max_{\{p,A\}} \quad \Pi = p \cdot q(p, A) - C(q) - A \tag{10.28}$$

La CPO en precios reproduce el índice de Lerner habitual

$$\frac{p - CMg}{p} = \frac{1}{\varepsilon_p} \tag{10.29}$$

donde ε_p es la elasticidad-precio de la demanda. La CPO en publicidad, por su parte, establece que el beneficio marginal de un euro adicional de publicidad debe igualar su coste marginal, que es la unidad

$$(p - CMg) \cdot \frac{\partial q}{\partial A} = 1 \tag{10.30}$$

Combinando ambas condiciones se obtiene la regla de Dorfman-Steiner

$$\frac{A}{p \cdot q} = \frac{\varepsilon_A}{\varepsilon_p} \tag{10.31}$$

donde ε_A es la elasticidad de la demanda respecto al gasto en publicidad. Es decir, el gasto óptimo en publicidad en relación con los ingresos totales será mayor cuanto más eficaz sea la publicidad desplazando la demanda –ε_A alta– y cuanto menos sensible sea el consumidor al precio –ε_p baja–[30].

[30] La regla de Dorfman-Steiner es óptima desde el punto de vista individual de la empresa, pero no garantiza la eficiencia agregada de la industria: si todas las empresas aumentan simultáneamente su gasto

Proliferación de marcas

Otro desarrollo relacionado es la proliferación de marcas como estrategia de disuasión de la entrada: una empresa que ya controla varios productos dentro de una industria puede lanzar nuevas variedades para ocupar los nichos de mercado más rentables, dificultando así que una empresa entrante encuentre un hueco viable (Spence, 1976b). Schmalensee (1978) formalizó esta lógica para el sector de los cereales de desayuno de Estados Unidos, objeto en los años setenta de una denuncia de la Federal Trade Commission contra sus principales fabricantes –Kellogg's, General Mills, General Foods y Quaker Oats– por lanzar más de ochenta marcas nuevas sin apenas entrada de competidores, pese a no existir patentes que lo impidieran[31].

El test SSNIP y la definición del mercado relevante

Por último, la existencia de productos diferenciados dificulta delimitar el mercado relevante: un mercado definido de forma demasiado estrecha –solo sustitutivos perfectos– dejaría fuera a competidores próximos, mientras que uno demasiado amplio –toda la economía– lo vaciaría de contenido. El criterio más utilizado en la práctica antitrust es el test SSNIP (U.S. Department of Justice, 1982) (*small but significant and non-transitory increase in price*): se pregunta si un hipotético monopolista que controlase todo el conjunto de bienes propuesto como mercado podría subir su precio de forma rentable un margen pequeño pero significativo –habitualmente, un 5 %– de manera sostenida en el tiempo. Si la subida no fuera rentable porque los consumidores se desplazarían hacia bienes de fuera del mercado propuesto, este debe ampliarse hasta que lo sea.

Formalmente, siguiendo a Harris y Simons (1989), se compara la pérdida crítica de ventas, CL, es decir, el porcentaje de ventas que dejaría indiferente al monopolista hipotético ante la subida de precio propuesta, con la pérdida real de ventas que efectivamente se observaría, AL

$$CL = \frac{\delta}{m + \delta} \tag{10.32}$$

donde $m = (p - CMg)/p$ es el índice de Lerner de la empresa y δ es la subida de precio propuesta. Si la pérdida real es menor que la crítica, $AL < CL$, el mercado ya está bien delimitado. Si, por el contrario, $AL > CL$, el mercado debe ampliarse[32].

en publicidad, el resultado puede asemejarse a un dilema del prisionero en el que los gastos se anulan mutuamente sin ganancia neta de cuota de mercado. Solo existe una justificación social clara para este gasto en la medida en que la publicidad aporte información genuina a los consumidores, y no se limite a una redistribución de la demanda entre empresas.

[31] El tribunal concluyó que no había evidencia suficiente de un uso deliberadamente anticompetitivo de esta estrategia, tratándose de una práctica de competencia legítima entre las empresas ya establecidas.

[32] El test presenta una limitación conocida como la "falacia del celofán": si el precio de partida ya incorpora poder de mercado –como ocurrió en el caso antitrust estadounidense del celofán, que da nombre a la falacia–, una subida adicional del 5 % puede parecer no rentable simplemente porque el precio ya es demasiado alto, llevando a definir el mercado de forma artificialmente amplia.

 © 2026 Manuel Cabado García

10.A.5. Aplicaciones del modelo de Dixit y Stiglitz

Cerrado ya el repaso de los desarrollos que menciona expresamente el enunciado, conviene detenerse por último en la relevancia del modelo de Dixit y Stiglitz fuera de la microeconomía pura. Este modelo constituye el instrumental básico de tres desarrollos posteriores de gran relevancia.

Comercio internacional Krugman (1979) mostró que la apertura comercial incrementa el tamaño efectivo del mercado al que se enfrentan las empresas. Esto permite tanto un mayor número de variedades disponibles para los consumidores –lo que aumenta directamente su utilidad, sin necesidad de recurrir a la ventaja comparativa– como un mejor aprovechamiento de las economías de escala internas, al operar cada empresa en un tramo inferior de su curva de costes medios. El comercio intraindustrial genera así ganancias mutuas del comercio incluso entre países idénticos en dotaciones, tecnología y preferencias.

Nueva Economía Keynesiana En el modelo canónico de Blanchard y Kiyotaki (1987), la preferencia por la variedad otorga a las empresas poder de mercado en un contexto de competencia monopolística, lo que, combinado con rigideces nominales y reales, genera tres implicaciones macroeconómicas de primer orden: la no neutralidad del dinero –la inflación distorsiona los precios como señales efectivas de escasez–, el ajuste ineficiente a nivel agregado ante un shock, porque las empresas ajustan sus precios lenta y heterogéneamente por miedo a perder cuota de mercado frente a un ajuste demasiado brusco, y la efectividad de la política monetaria y fiscal para suavizar los ciclos económicos.

Crecimiento endógeno Romer (1990) mostró que la preferencia por la variedad en los bienes de capital o de conocimiento genera rendimientos crecientes que sostienen el crecimiento a largo plazo, endogeneizando el proceso de innovación tecnológica.

© 2026 Manuel Cabado García

Bibliografía

Andrews, P. (1949). *Manufacturing Business*. London: Macmillan.

Blanchard, O. J. y N. Kiyotaki (1987). Monopolistic Competition and the Effects of Aggregate Demand. *The American Economic Review 77*(4), 647–666.

Cairnes, J. E. (1874). *Some Leading Principles of Political Economy Newly Expounded*. London: Macmillan.

Chamberlin, E. H. (1933). *Theory of monopolistic competition: A re-orientation of the theory of value*. Harvard University Press, Cambridge.

d'Aspremont, C., J. J. Gabszewicz y J.-F. Thisse (1979). On Hotelling's "Stability in Competition". *Econometrica 47*(5), 1145–1150.

Dixit, A. K. y J. E. Stiglitz (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review 67*(3), 297–308.

Dorfman, R. y P. O. Steiner (1954). Optimal Advertising and Optimal Quality. *The American Economic Review 44*(5), 826–836.

Eaton, B. C. y R. G. Lipsey (1989). Product differentiation. En R. Schmalensee y R. Willig (eds.), *Handbook of Industrial Organization*, vol. 1 de *Handbook in Economics*, cap. 12, pp. 723–768. North-Holland.

Feás, E. (2019). ¿A quién votaría Kelvin Lancaster? *Vozpópuli*. 24 de abril de 2019.

Harris, B. C. y J. J. Simons (1989). Focusing Market Definition: How Much Substitution is Necessary? *Research in Law and Economics 12*, 207–226.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal 39*(153), 41–57.

Kaldor, N. (1935). Market Imperfection and Excess Capacity. *Economica 2*(5), 33–50.

Krugman, P. R. (1979). Increasing Returns, Monopolistic Competition, and International Trade. *Journal of International Economics 9*(4), 469–479.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy 74*(2), 132–157.

Machado, M. (2008). Vertical Differentiation. Material docente, curso de Organización Industrial, Universidad Carlos III de Madrid.

Neven, D. (1985). Two Stage (Perfect) Equilibrium in Hotelling's Model. *The Journal of Industrial Economics 33*(3), 317.

 © 2026 Manuel Cabado García

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy 98*(5, Part 2), S71–S102.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *The Bell Journal of Economics 10*(1), 141–156.

Schmalensee, R. (1978). Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry. *The Bell Journal of Economics 9*(2), 305–327.

Shove, G. F. (1933). The Imperfection of the Market. *The Economic Journal 43*(169), 113–125.

Shy, O. (1995). *Industrial Organization: Theory and Applications*. The MIT Press.

Spence, M. (1976a). Product Differentiation and Welfare. *The American Economic Review 66*(2), 407–414.

Spence, M. (1976b). Product Selection, Fixed Costs, and Monopolistic Competition. *The Review of Economic Studies 43*(2), 217–235.

Sraffa, P. (1926). The Laws of Returns under Competitive Conditions. *The Economic Journal 36*(144), 535–550.

Sweezy, P. M. (1939). Demand under conditions of oligopoly. *Journal of Political Economy 47*(4), 568–573.

Tirole, J. (1988). *The Theory of Industrial Organization*. The MIT Press.

U.S. Department of Justice (1982). Merger Guidelines.

Wicksell, K. (1934). *Lectures on Political Economy, Volume I: General Theory*. London: George Routledge and Sons.

 © 2026 Manuel Cabado García

Tema 11

Teorías microeconómicas del mercado de trabajo (I). La teoría neoclásica del mercado de trabajo. Análisis intertemporal de la oferta de trabajo. Teoría del capital humano.

11.1. Introducción ⏳ 1'

El mercado de trabajo es un componente fundamental de la economía, pues influye directamente en el bienestar de las familias y en la productividad de un país. Aunque en teoría debería ajustarse de manera fluida, la realidad lleva a plantearse una serie de cuestiones:

- ¿Por qué se trabajan las horas que se trabajan y no más?
- ¿Por qué se trabaja más en unas etapas de la vida que en otras?
- ¿Por qué cobra más quien ha estudiado más?

Responder a estas preguntas exige analizar las decisiones de quienes ofrecen y demandan trabajo. Por consiguiente, el objetivo de la exposición será el análisis del mercado de trabajo desde una óptica microeconómica a través del siguiente esquema:

1. La teoría neoclásica del mercado de trabajo, como punto de partida del análisis, y qué cambia cuando existe competencia imperfecta.
2. El análisis intertemporal de la oferta de trabajo, que explica cómo se reparte el trabajo a lo largo de la vida y del ciclo económico.
3. La teoría del capital humano, que explica por qué los salarios difieren entre trabajadores.

11.2. La teoría neoclásica del mercado de trabajo ⏳ 10'

La teoría neoclásica analiza el mercado de trabajo de forma análoga al mercado de cualquier otro bien, por lo que el nivel de empleo, que es la cantidad, y el salario, que es el precio, quedan determinados por la interacción entre oferta y demanda. Esta visión, que Hicks (1932) sistematizó a partir de la teoría de la productividad marginal, sigue siendo el punto de referencia frente al que se miden las imperfecciones del mercado de trabajo.

Supuestos

La teoría neoclásica parte de los siguientes supuestos:

1. Competencia perfecta en el mercado de bienes y de factores productivos.
2. Plena flexibilidad de precios y salarios.
3. Perfecta movilidad del trabajo.
4. Homogeneidad del factor trabajo.
5. Información perfecta.
6. Ausencia de externalidades.
7. Racionalidad: los individuos maximizan su utilidad y las empresas, su beneficio.

De estos supuestos se deriva, como se verá, un equilibrio único, estable y eficiente, sin desempleo involuntario. Para llegar a él se estudian primero la oferta de trabajo, que nace de las decisiones de los individuos, y después la demanda, que nace de las decisiones de las empresas.

Oferta de trabajo

La oferta de trabajo es el número de horas que un individuo está dispuesto a trabajar a cada salario. Se deriva de la maximización de su utilidad, en la que elige entre consumo y ocio sujeto a una restricción presupuestaria y a otra temporal.[1]

El problema al que se enfrenta el trabajador es la maximización de la utilidad que le reportan el consumo, C, y el ocio, O:

$$\max_{C,O} U(C, O)$$

sujeto a dos restricciones.[2] La presupuestaria exige que el gasto en consumo, valorado al

[1] La intuición es que cada hora de trabajo es una hora menos de ocio, de modo que decidir cuánto trabajar equivale a decidir cuánto ocio se sacrifica a cambio del consumo que paga el salario de esas horas.

[2] La función de utilidad tiene las siguientes propiedades: i) continua, ii) creciente, iii) cuasicóncava, iv) diferenciable y v) ordinal.

 © 2026 Manuel Cabado García

precio del bien, P_C, no supere la suma de la renta no salarial, M, y de la renta salarial, que es el salario nominal, W, multiplicado por las horas trabajadas, L. La temporal reparte la dotación de tiempo total, $\bar{L}$, entre ocio y trabajo:

$$P_C \cdot C \leq M + W\left(\bar{L} - O\right)$$
$$O + L = \bar{L}$$

Como cada hora de ocio supone renunciar al salario que se habría ganado con ella, el salario real, $w = W/P_C$, es el coste de oportunidad del ocio, es decir, su precio.

Las condiciones de primer orden, CPO, del problema establecen que el trabajador ofrecerá trabajo hasta que la relación marginal de sustitución entre ocio y consumo, $RMS_{O,C}$, es decir, el consumo al que está dispuesto a renunciar por una hora más de ocio, se iguale al salario real, w:

$$RMS_{O,C} = \frac{UMg_O}{UMg_C} = \frac{W}{P_C} = w$$

Así, el equilibrio se produce allí donde la tasa a la que el individuo intercambia internamente ocio por consumo se iguala a la tasa objetiva a la que lo hace el mercado.

Para obtener la oferta de trabajo individual, L^S, se realiza un ejercicio de estática comparativa con el salario real, w, en el que hay que distinguir el efecto sustitución del efecto renta.

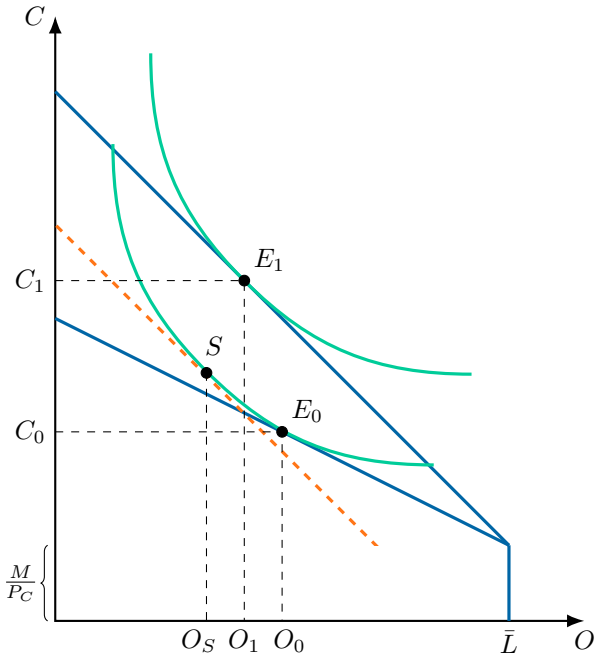
Gráficamente, en un plano de ocio, O, en el eje horizontal y consumo, C, en el vertical, como el de la Figura 11.1, se representan la restricción, que es a la vez presupuestaria y temporal, y las curvas de indiferencia, que recogen las combinaciones de ocio y consumo que generan el mismo nivel de utilidad. Así:

- El equilibrio inicial, E_0, está en el punto de tangencia entre la restricción y la curva de indiferencia más alejada del origen.[3]
- Al subir el salario real, la restricción pivota sobre el punto de dotación, en el que no se trabaja, y el nuevo equilibrio, E_1, se sitúa sobre una curva de indiferencia más alta.
- Para separar los dos efectos se aplica la variación compensatoria en el sentido de Hicks: la nueva restricción se desplaza hasta que es tangente a la curva de indiferencia inicial, en el punto S.

[3]Las preferencias se representan mediante curvas de indiferencia con las siguientes características: i) son continuas, ii) no se cortan, iii) representan más utilidad cuanto más alejadas están del origen, iv) tienen pendiente negativa, igual a la $RMS_{O,C}$ cambiada de signo, y v) son convexas, es decir, ocio y consumo son sustitutivos imperfectos, de forma que, conforme aumenta el ocio, el individuo está dispuesto a sacrificar cada vez menos consumo por una hora adicional.

La pendiente de las curvas de indiferencia refleja la valoración relativa del ocio. Si son muy inclinadas, el individuo está dispuesto a sacrificar mucho consumo por una hora más de ocio. Si son muy planas, valora poco el ocio y trabajará más horas.

Figura 11.1: Determinación de la oferta de trabajo



- Efecto sustitución, de E_0 a S, que siempre reduce el ocio: al aumentar el salario aumenta el coste de oportunidad del ocio, por lo que disminuyen las horas de ocio, $\downarrow O$, y aumenta la oferta de trabajo, $\uparrow L^S$.
- Efecto renta, de S a E_1: al aumentar el salario aumenta la renta salarial, $W\left(\bar{L}-O\right)$, y con ella el consumo de todos los bienes normales, entre ellos el ocio, $\uparrow O$, lo que reduce la oferta de trabajo, $\downarrow L^S$.[4]

Por lo tanto, si el ocio es un bien normal, el efecto renta y el efecto sustitución tienen signos contrarios y la pendiente de la oferta de trabajo depende de la cuantía relativa de ambos.[5]

Ya Robbins (1930) mostró que el efecto de una subida salarial sobre las horas trabajadas es teóricamente ambiguo, porque depende de la elasticidad de la demanda de renta en términos de esfuerzo: si es inferior a la unidad, el individuo responde a un salario más

[4]El efecto renta es doble:

1. Por una parte, el efecto renta ordinario: como sube el precio del ocio, la renta real con la que se compra disminuye, lo que reduce el ocio y aumenta la cantidad de trabajo, $\uparrow L^S$.
2. Por otra parte, el efecto renta dotación: el tiempo del que dispone el individuo vale más, por lo que su renta potencial aumenta y, si el ocio es un bien normal, consume más ocio y trabaja menos, $\downarrow L^S$.

Para quien ofrece trabajo, el efecto renta dotación es mayor que el efecto renta ordinario, por lo que si aumenta w, entonces el efecto renta total hace que aumente el ocio, $\uparrow O$, y, por lo tanto, disminuya la oferta de trabajo, $\downarrow L^S$. La descomposición formal se recoge en el Apéndice 11.A.1.

[5]Si el ocio es un bien normal y el efecto sustitución es mayor en valor absoluto que el efecto renta, $|ES| > |ER|$, la oferta de trabajo tiene pendiente positiva. Si el ocio es un bien inferior, el efecto renta refuerza al efecto sustitución y la oferta de trabajo tiene pendiente positiva en todo caso.

 © 2026 Manuel Cabado García

alto trabajando menos, y la curva de oferta se vuelve hacia atrás.[6] En el plano empírico, Ragnar Frisch (1932) dedujo, a partir de sus estimaciones de la utilidad marginal de la renta, una oferta de trabajo en forma de "S" invertida: decreciente en un primer tramo, creciente en un segundo y, según conjeturaba el propio Frisch, de nuevo decreciente en un tercero.[7]

Por norma general, y para el resto del análisis, se supondrá que el individuo participa en el mercado de trabajo y que su oferta tiene pendiente creciente, lo que exige que:

- El salario de equilibrio sea mayor que el salario de reserva, que es el salario más bajo por el que un trabajador está dispuesto a aceptar un trabajo.[8]
- Si el ocio es un bien normal, como es lo habitual, el efecto sustitución sea mayor que el efecto renta en valor absoluto: $|ES| > |ER|$.

Gráficamente, en un plano de horas trabajadas, L, y salario real, w, como el de la Figura 11.2, se representa la oferta de trabajo individual, L^S.[9] En ella:

- Por debajo del salario de reserva, w_R, el individuo no trabaja, y a partir de él empieza a ofrecer horas.
- Entre w_R y w' domina el efecto sustitución y la oferta es creciente.
- Por encima de w' domina el efecto renta y la curva se vuelve hacia atrás.

En general, se acepta que el tramo relevante de la oferta de trabajo individual es el creciente. La oferta de trabajo de mercado se obtiene por agregación horizontal de las

[6] Con horas extraordinarias el resultado cambia. Mientras que con una subida del salario el efecto es ambiguo, con una prima por horas extraordinarias la oferta de trabajo aumenta inequívocamente. Esta diferencia se debe a que la respuesta a una prima por horas extraordinarias es esencialmente un efecto sustitución puro. La razón es que la prima proporciona una mayor retribución por las horas extraordinarias trabajadas, mientras que una subida del salario proporciona una mayor retribución por todas las horas trabajadas.

[7] Frisch, economista noruego que compartió con Tinbergen el primer Premio Nobel de Economía, en 1969, halló que, para niveles muy bajos de salario, el efecto renta domina al efecto sustitución, por lo que la curva de oferta tendría un tramo inicial con pendiente negativa. Lo asociaba a trabajadores sin más ingresos que su salario, que, cuando este baja, tienen que trabajar más horas para subsistir.

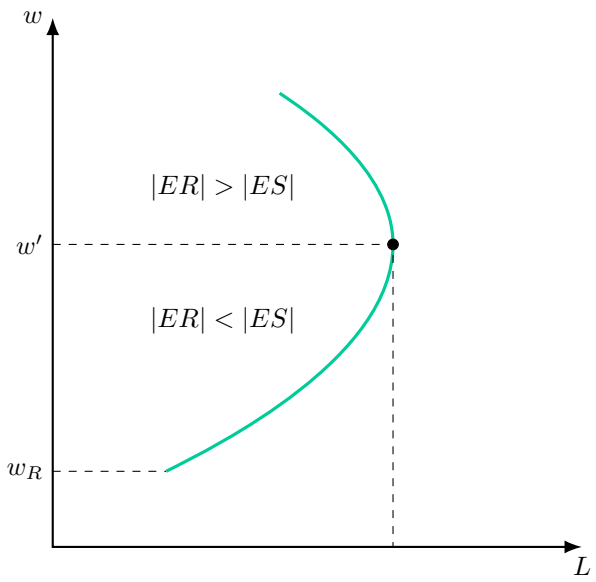
Otra posible explicación, distinta de la de Frisch, es que, con salarios muy bajos, los trabajadores tienen opción a subsidios que perderían si el salario aumentase, por lo que, ante subidas modestas del salario, ofrecerán menos trabajo para que su renta no supere el umbral que les da derecho al subsidio.

Es el problema que intenta corregir el incentivo al empleo del ingreso mínimo vital, regulado por el Real Decreto 789/2022, que permite compatibilizar en parte la prestación con las rentas del trabajo para que trabajar más aumente siempre la renta disponible.

[8] La intuición es que la primera hora de trabajo solo compensa si lo que se gana con ella vale más que la hora de ocio a la que se renuncia, y esa valoración es la pendiente de la curva de indiferencia en el punto en el que no se trabaja. Formalmente, el salario de reserva, w_R, es la $RMS_{O,C}$ evaluada en el punto de dotación, $(O, C) = \left(\bar{L}, M/P_C\right)$ (Borjas, 2016, p. 41). Si el ocio es un bien normal, una renta no salarial más alta eleva el salario de reserva y reduce la probabilidad de participar en el mercado de trabajo. El Tema 13 emplea el mismo concepto en el contexto de la información asimétrica.

[9] La forma de "S" invertida de Frisch añadiría a la curva de la figura un primer tramo decreciente con salarios muy bajos, lo que exige un salario de reserva prácticamente nulo, como el de quien no tiene más ingresos que su salario. La oferta sería entonces decreciente con salarios muy bajos, creciente con salarios intermedios y de nuevo decreciente, hacia atrás, por encima de w'.

Figura 11.2: Curva de oferta de trabajo



ofertas individuales y responde al salario por dos vías: el margen intensivo, es decir, las horas de quienes ya trabajan, y el margen extensivo, es decir, la entrada en el mercado de quienes, al subir el salario, ven superado su salario de reserva.[10]

La incorporación de la mujer al mercado de trabajo ilustra la importancia del margen extensivo. En España, la tasa de actividad de las mujeres pasó del 42,7 % en el segundo trimestre de 2002 al 54,6 % en el de 2026, mientras la de los hombres bajaba del 66,9 % al 64,2 %, de modo que la brecha se redujo de 24 a menos de 10 puntos (INE, 2026). Goldin (2006) la describió como una "revolución silenciosa": desde finales de los años setenta, las mujeres pasaron de ser trabajadoras secundarias, que entraban y salían del mercado según las necesidades del hogar, a planificar carreras profesionales propias, con más formación y mayor apego al empleo. Su efecto sobre el producto a largo plazo es, sin embargo, ambiguo: Evans (2026) sostiene que la incorporación de la mujer al trabajo productivo eleva el PIB, pero que, si los hombres no aceptan relaciones más igualitarias, la caída de la fecundidad puede frenar el crecimiento.[11]

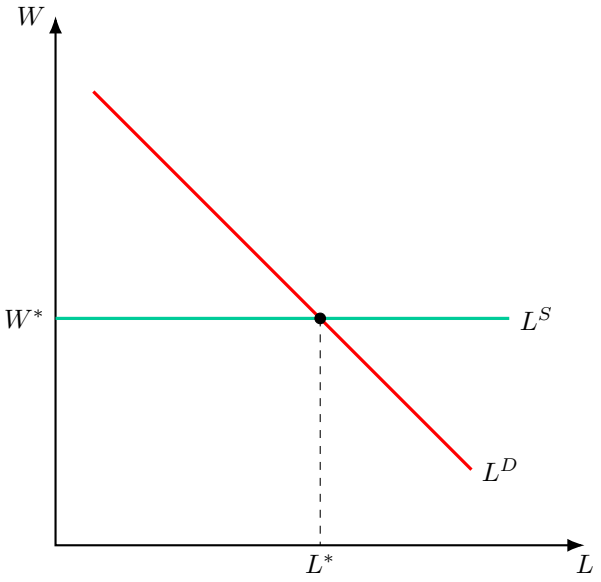
Respecto a la empresa, en un mercado de factores perfectamente competitivo puede contratar tantos trabajadores como desee al salario de mercado, de modo que la oferta de trabajo a la que se enfrenta es perfectamente elástica a ese salario, como se observa en la Figura 11.3.

[10] La distinción importa para la política económica, porque la evidencia muestra que las horas de los varones en edad central apenas responden al salario, mientras que la participación, sobre todo la de las mujeres casadas, es mucho más sensible (Blundell y MaCurdy, 1999) (Cahuc et al., 2014, p. 56). El efecto de los impuestos sobre la oferta de trabajo se estudia en el Tema 21.

[11] Ambas cuestiones se desarrollan en el Apéndice 11.A.8.

 © 2026 Manuel Cabado García

Figura 11.3: Oferta de trabajo a la que se enfrenta la empresa



Demanda de trabajo

La demanda de trabajo es una demanda derivada, es decir, la empresa demanda trabajo porque hay demanda del bien que este ayuda a producir, y varía con ella. Por ello, se obtiene del proceso optimizador de la empresa, que maximiza sus beneficios sujeta a una restricción tecnológica, como se vio en los Temas 5 y 6.[12]

Corto plazo

A corto plazo, la dotación de algún factor de producción es fija. Se supondrá que el factor fijo es el capital, $\bar{K}$, por lo que el productor maximiza solo respecto al trabajo. Además, al haber competencia perfecta en todos los mercados, la empresa es precio-aceptante: el precio del bien, P, y el salario nominal, W, son exógenos.

El programa de optimización será el siguiente:

$$\text{máx} \quad \pi = P \cdot F\left(\bar{K}, L\right) - R \cdot \bar{K} - W \cdot L \tag{11.1}$$

donde K y L son los factores productivos utilizados (imperfectamente sustitutivos), y R y W, sus precios nominales.

De acuerdo con la CPO, la empresa demandará trabajo hasta que su productividad marginal se iguale al salario real o, lo que es lo mismo, hasta que el valor de la produc-

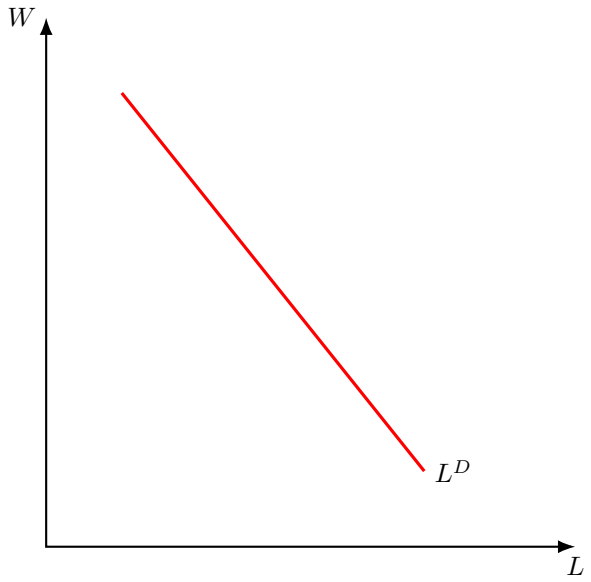
[12]La función de producción tiene las siguientes propiedades: i) productividad marginal positiva pero decreciente, ii) estrictamente cóncava, lo que implica rendimientos a escala decrecientes, iii) continua y iv) dos veces diferenciable, con $f_i' > 0$, $f_j' > 0$ y $f_{ij} > 0$. Esto último implica que la productividad marginal de cada factor aumenta con la cantidad del otro, es decir, que los factores son cooperantes, como en una Cobb-Douglas.

tividad marginal, $P \cdot PMg_L$, es decir, el ingreso que genera un trabajador adicional, se iguale al salario nominal:

$$P \cdot PMg_L = W \;\Rightarrow\; PMg_L = w$$

Gráficamente, en un plano de trabajo, L, y salario nominal, W, como el de la Figura 11.4, se representa la curva del valor de la productividad marginal, $P \cdot PMg_L$. Así:

Figura 11.4: Demanda de trabajo en el corto plazo



- Como la productividad marginal del trabajo es decreciente, la empresa solo contratará más trabajadores si el salario baja.
- La curva $P \cdot PMg_L$ es, por tanto, la demanda de trabajo a corto plazo de la empresa, y tiene pendiente negativa.

Por su parte, la condición de segundo orden, CSO, se cumple porque la productividad marginal del trabajo es decreciente.

Largo plazo

A largo plazo, la dotación de todos los factores es variable, por lo que ahora el productor maximiza respecto al trabajo y al capital.[13]

Las CPO establecen que:

[13]En puridad, el programa de optimización esconde dos subprogramas. El primero consiste en elegir la combinación de factores que minimiza el coste de una producción dada, es decir, el problema dual de costes, que da lugar a la primera condición y, de ella, a las demandas condicionadas de factores, $X_i^D(W, Y)$, donde W es el vector de precios nominales de los factores. El segundo consiste en elegir el nivel de producción que maximiza el beneficio, manteniendo la combinación eficiente de factores, lo que da lugar a la segunda y, de ella, a las demandas incondicionadas, $X_i^D(W, P)$.

 © 2026 Manuel Cabado García

- Minimización de costes: en el largo plazo la empresa minimiza costes para cada nivel de producción, es decir, elige su tamaño de planta óptimo, intercambiando factores hasta que su relación marginal de sustitución técnica, $RMST$, se iguala a la tasa a la que puede hacerlo en el mercado, que es la pendiente de la isocoste, W/R:

$$RMST = \frac{PMg_L}{PMg_K} = \frac{W}{R}$$

- Maximización de beneficios: señala que la empresa contratará factores productivos hasta que sus productividades marginales sean iguales a sus costes reales (o, de igual manera, hasta que los valores de sus productividades marginales sean iguales a sus costes nominales)

$$PMg_L = w \tag{11.2}$$
$$PMg_K = r \tag{11.3}$$

  Estas son las condiciones del óptimo a largo plazo, en las que $w = W/P$ es el salario real y $r = R/P$ el coste real del capital, distinto del tipo de interés real que aparecerá en el análisis intertemporal.[14]

La CSO se cumple si la función de producción es estrictamente cóncava, lo que implica rendimientos a escala decrecientes.[15]

Ahora se observa cómo, en el largo plazo, la demanda de un factor depende no solo de su productividad y de su coste, sino también de la productividad y del coste del otro factor.

Gráficamente, en un plano de trabajo, L, y salario nominal, W, como el de la Figura 11.5, se representan la demanda de trabajo a corto plazo inicial, L^D_{cp}, la que resulta tras ajustarse el capital, $L^{D\prime}_{cp}$, y la demanda a largo plazo, L^D_{lp}. Ante una caída del salario:

- Como se ha visto, la curva del valor de la productividad marginal ($P \cdot PMg_L$) describe la demanda de trabajo cuando el nivel de capital es fijo.
- Si el salario disminuye de W_0 a W_1, se demanda más trabajo para que se cumpla $\downarrow W = P \cdot \downarrow PMg_L$, y se pasa del punto A al B.
- El aumento del trabajo aumenta la PMg_K, por lo que, ahora que el capital es variable, se demanda también más capital para que disminuya la PMg_K y se pueda

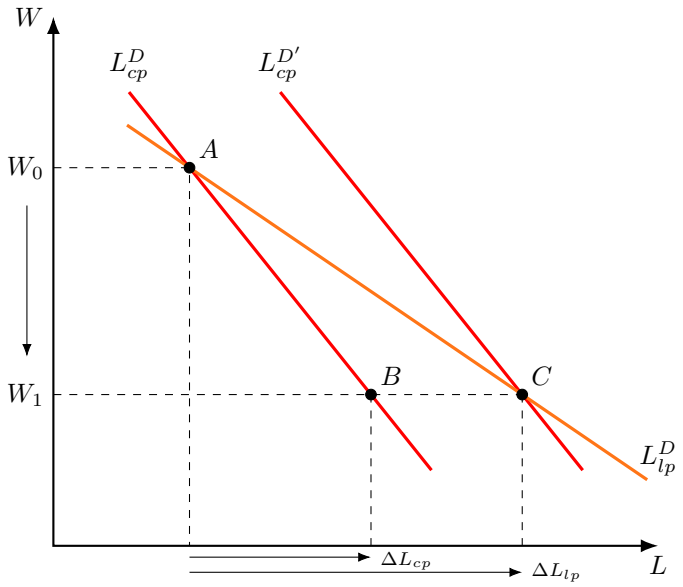
[14] A diferencia de lo que ocurría en el corto plazo, estas ecuaciones no son las demandas de factores, porque en el largo plazo dependen también, como se verá más adelante, de la productividad y del precio del otro factor.

[15] Para que pueda maximizarse el beneficio con todos los factores variables se exige que la función de producción sea estrictamente cóncava, y no basta con que sea estrictamente cuasicóncava, porque la función objetivo, la de beneficios, no está sujeta a ninguna restricción: la tecnológica se inserta en ella. En el problema de maximización de la producción, en cambio, sí hay una restricción, por lo que basta con la cuasiconcavidad estricta y se admite cualquier tipo de rendimientos a escala.

cumplir $R = P \cdot PMg_K$.

- Al demandarse más capital, la PMg_L aumenta y, por tanto, la curva de demanda de trabajo a corto plazo se desplaza a la derecha y el equilibrio pasa al punto C. La demanda a largo plazo, L^D_{lp}, une A y C, y el aumento del empleo es mayor que en el corto plazo, $\Delta L_{lp} > \Delta L_{cp}$.

Figura 11.5: Demanda de trabajo a largo plazo



Del proceso anterior se extraen tres conclusiones:

1. Una disminución del salario aumenta la demanda de trabajo y de capital, $\downarrow W \Rightarrow \uparrow L, \uparrow K$.[16]
2. La demanda individual en el largo plazo ya no coincide con la curva de $P \cdot PMg_L$.
3. La demanda individual a largo plazo es más elástica que la de corto plazo, precisamente por la mayor flexibilidad de la empresa para combinar los factores, lo que se conoce como teorema de Le Chatelier-Samuelson.

La elasticidad de la demanda de trabajo depende, según las reglas de Marshall que formalizó Hicks (1932), de cuatro factores. Es mayor cuanto más fácil es sustituir trabajo por otros factores, cuanto más elástica es la demanda del bien que se produce, cuanto mayor es el peso del trabajo en los costes totales, siempre que la demanda del bien sea más elástica que la sustitución entre factores, y cuanto más elástica es la oferta de los demás factores (Borjas, 2016, cap. 3). En la práctica, Hamermesh (1993), a partir de más

[16] Ocurre siempre que los factores sean cooperantes, es decir, que la productividad marginal del capital aumente con el trabajo, $f_{KL} > 0$, como en las funciones de producción neoclásicas habituales. Para un análisis más general, véase el Apéndice 11.A.2.

 © 2026 Manuel Cabado García

de setenta estudios, sitúa la elasticidad de la demanda de trabajo a producción constante entre 0,15 y 0,75 en valor absoluto, con 0,3 como valor más probable. Si se incluye el efecto escala, se acerca a la unidad (Cahuc et al., 2014, p. 117). De estas reglas depende, por ejemplo, cuánto empleo se pierde cuando se encarece el trabajo con una subida de las cotizaciones sociales.

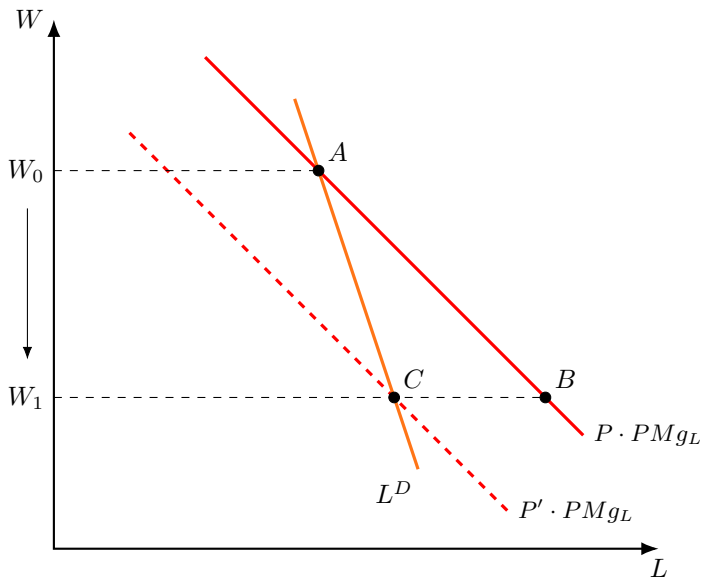
Demanda agregada

El análisis de la demanda agregada se realiza en el corto plazo por simplicidad.

A diferencia de la oferta, la demanda de trabajo agregada no es la suma horizontal de las demandas individuales. Cuando cambia el salario, varía la demanda de trabajo de las empresas y, con ella, la producción agregada y el precio del producto. Cada empresa, al contratar y producir más, hace bajar el precio del bien para todas las demás sin tenerlo en cuenta. Este efecto, que se transmite a través de los precios, se denomina externalidad pecuniaria.

Gráficamente, en un plano de trabajo, L, y salario nominal, W, como el de la Figura 11.6, se representan el valor de la productividad marginal antes y después de la caída del precio del bien, $P \cdot PMg_L$ y $P' \cdot PMg_L$, y la demanda agregada de trabajo, L^D, que une A y C. Así:

Figura 11.6: Demanda agregada de trabajo a corto plazo



- Ante una disminución del salario de W_0 a W_1, las empresas contratarán más trabajadores y producirán más, lo que desplazará a la derecha la curva de oferta de mercado del bien y reducirá su precio de P a P'.

- A causa de dicha reducción, el valor de la productividad marginal se desplaza a la izquierda, de $P \cdot PMg_L$ a $P' \cdot PMg_L$, por lo que se pasa de A a C en lugar de A a B.

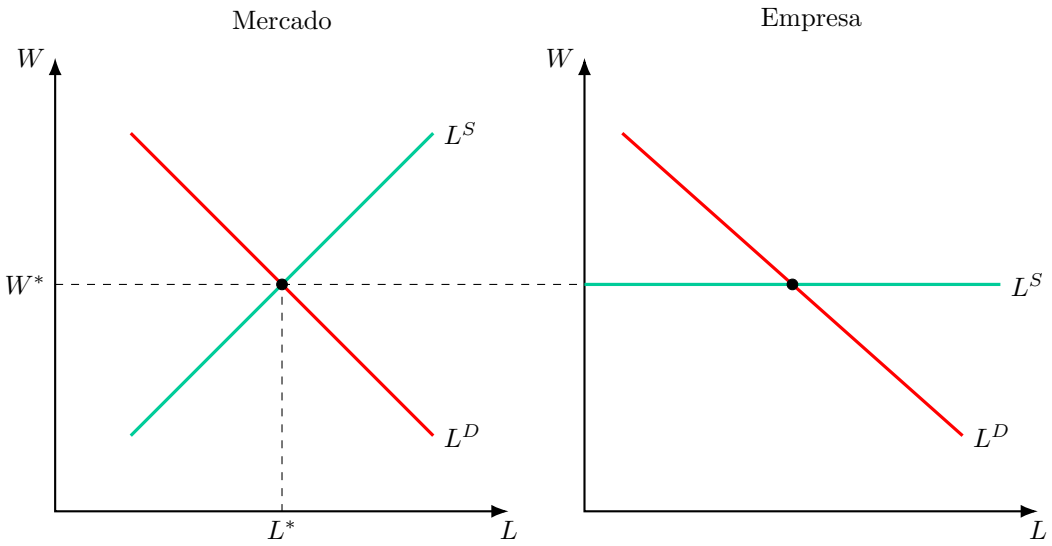
Por tanto, para obtener la demanda de trabajo de una industria deben agregarse las cantidades que demandaría cada empresa teniendo en cuenta el cambio del precio del bien al variar el salario, es decir, las cuasidemandas.

Esto hace que la demanda agregada de trabajo sea más inelástica que la que se obtendría si no variase el precio del bien. Ello se debe a que, al reducirse el salario, también cae el precio del bien, por lo que las empresas tienen menos incentivos a contratar más trabajadores.

Equilibrio

Halladas las curvas de oferta y demanda agregadas de trabajo, el equilibrio neoclásico, es decir, el salario y el nivel de empleo, se produce en la intersección de ambas curvas, como se observa en la Figura 11.7.

Figura 11.7: Equilibrio del mercado de trabajo neoclásico



Se da, pues, un equilibrio único, estable y eficiente, en el que los salarios reales se igualan a la productividad marginal del trabajo y el mercado se vacía de manera instantánea. No hay desempleo involuntario, aunque sí puede haber desempleo voluntario, y la flexibilidad del salario es el mecanismo autocorrector que asegura el ajuste automático. La implicación de política es clara: el mercado no necesita intervención, y cualquier rigidez que impida bajar el salario, como un salario mínimo por encima del de equilibrio, solo puede generar paro.

 © 2026 Manuel Cabado García

Sin embargo, estas conclusiones no se ajustan a la evidencia empírica, en la que no hay ni pleno empleo ni ajuste automático. Por ello, a continuación se relaja el supuesto de competencia perfecta.

Competencia imperfecta en los mercados de bienes y de trabajo

Cuando existe competencia imperfecta, los resultados difieren de los obtenidos en un entorno competitivo. Se estudia primero qué ocurre cuando el poder de mercado lo tiene la empresa, como monopolista en el mercado del bien o como monopsonista en el de trabajo, y después cuando lo tienen los trabajadores organizados en un sindicato.

Poder de mercado de la empresa: monopolio y monopsonio

Un monopsonio es un mercado con un único comprador. En el de trabajo, la empresa monopsonista es la única demandante de trabajo y se enfrenta, por ello, a toda la curva de oferta del mercado, que es creciente, en lugar de tomar el salario como dado, como ya se vio en el Tema 8.[17] Puede ocurrir porque la empresa sea la única de la zona, porque esté muy especializada o porque todas las empresas se hayan agrupado en una patronal que lidera las negociaciones con los trabajadores.

Además, hay que distinguir si la empresa tiene también poder de monopolio en el mercado de bienes, porque la demanda de trabajo es una demanda derivada que depende de su oferta de producto. De ahí que los resultados del monopsonista varíen según sea o no monopolista en el mercado del bien.

Se analiza la demanda de trabajo a corto plazo, con la función de producción $F\left(\bar{K}, L\right)$. La empresa contrata trabajadores hasta que lo que aporta el último se iguala a lo que cuesta:

- Ingreso marginal del factor, $IMgF$: es la variación del ingreso al contratar una unidad adicional de factor, es decir, el ingreso marginal del bien por la productividad marginal del trabajo, lo que en el Tema 8 se denomina ingreso del producto marginal.
- Gasto marginal del factor, $GMgF$: es la variación del coste al contratar una unidad adicional de factor.

En la condición de primer orden intervienen la elasticidad-precio de la demanda del bien, $\varepsilon_{Q^D,P}$, y la elasticidad de la oferta de trabajo respecto al salario, $\varepsilon_{L^S,W}$:

$$P \cdot PMg_L \left(1 - \frac{1}{\left|\varepsilon_{Q^D,P}\right|}\right) = W \left(1 + \frac{1}{\left|\varepsilon_{L^S,W}\right|}\right)$$

[17] La intuición es que el monopsonista sabe que, si contrata menos, puede pagar menos a todos, porque los trabajadores no tienen otra empresa a la que acudir.

El lado izquierdo es el $IMgF$ y el derecho, el $GMgF$. Cada uno cambia según haya o no poder de mercado:

- En el mercado del bien: con competencia perfecta, $\varepsilon_{Q^D,P} = \infty$, y el $IMgF$ es el valor de la productividad marginal, $P \cdot PMg_L$. Con poder de monopolio es menor, porque para vender más la empresa tiene que bajar el precio de todas las unidades:

$$IMgF^* = P \cdot PMg_L \left(1 - \frac{1}{\left|\varepsilon_{Q^D,P}\right|}\right) < P \cdot PMg_L$$

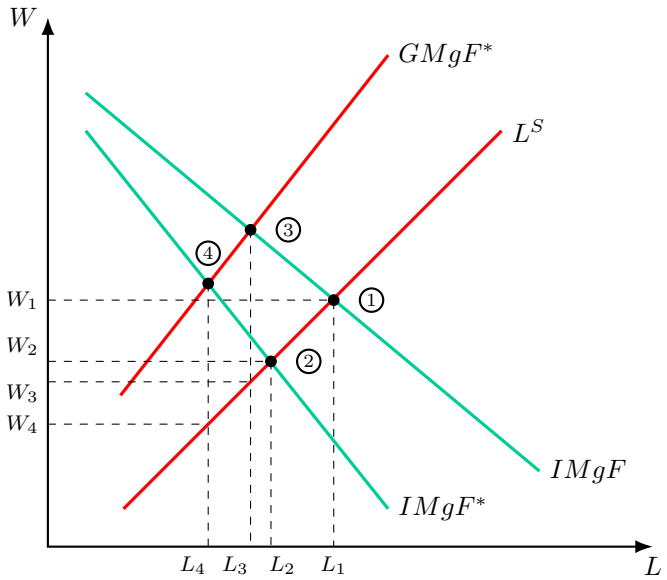
- En el mercado de trabajo: con competencia perfecta, $\varepsilon_{L^S,W} = \infty$, y el $GMgF$ es el salario. Con poder de monopsonio es mayor, porque para contratar a un trabajador más tiene que subir el salario a todos los que ya emplea:

$$GMgF^* = W \left(1 + \frac{1}{\left|\varepsilon_{L^S,W}\right|}\right) > W$$

- La oferta de trabajo, que indica lo mínimo que hay que pagar para atraer a cada número de trabajadores, es la curva de gasto medio, y el salario se lee siempre sobre ella.

Gráficamente, en un plano de trabajo, L, y salario nominal, W, como el de la Figura 11.8, se representan el ingreso marginal del factor en competencia, $IMgF$, igual a $P \cdot PMg_L$, y con poder de monopolio, $IMgF^*$, la oferta de trabajo, L^S, y el gasto marginal del factor con poder de monopsonio, $GMgF^*$. En ella se distinguen cuatro casos:

Figura 11.8: Poder de mercado de la empresa: monopolio y monopsonio



 © 2026 Manuel Cabado García

- Competencia en ambos mercados, punto 1: cada empresa toma como dados el precio y el salario, y el salario y el empleo de mercado, W_1 y L_1, se determinan donde la demanda de trabajo, $IMgF$, corta a la oferta, L^S.
- Monopolio en bienes y competencia en trabajo, punto 2: la demanda de trabajo pasa a ser $IMgF^*$ y corta a la oferta en el punto 2. Como la oferta de mercado es creciente, cae el empleo, a L_2, y también el salario, a W_2 (Koutsoyiannis, 1979, pp. 456-457). Si fuera horizontal, solo caería el empleo.
- Competencia en bienes y monopsonio en trabajo, punto 3: la empresa contrata donde $IMgF$ corta a $GMgF^*$, con el empleo L_3, y paga el salario W_3, que se lee sobre la oferta, por debajo del punto 3.
- Monopolio y monopsonio, punto 4: la empresa contrata donde $IMgF^*$ corta a $GMgF^*$, con el empleo L_4, y paga el salario W_4, que se lee sobre la oferta, por debajo del punto 4.

Joan Robinson llama explotación a la diferencia entre el valor de la productividad marginal del trabajo y el salario.[18] Tiene dos componentes. La explotación de monopolio es la diferencia entre $P \cdot PMg_L$ y el $IMgF^*$, y aparece en los puntos 2 y 4. La explotación de monopsonio es la diferencia entre el ingreso marginal del factor, igual en el óptimo al gasto marginal, y el salario pagado, y aparece en los puntos 3 y 4. Los cuatro casos se resumen en la tabla 11.1.

Cuadro 11.1: Resumen de resultados

	Mercado de producto	**Equilibrio**	**Mercado del factor L**
1	Precio-aceptante	$IMgF = W$	Precio-aceptante
2	Monopolista en el bien	$IMgF^* = W$	Precio-aceptante
3	Precio-aceptante	$IMgF = W\left(1 + \frac{1}{\lvert\varepsilon_{L^S,W}\rvert}\right)$	Monopsonista en L
4	Monopolista en el bien	$IMgF^* = W\left(1 + \frac{1}{\lvert\varepsilon_{L^S,W}\rvert}\right)$	Monopsonista en L

En suma, con competencia perfecta en ambos mercados, el salario y el nivel de empleo son los más altos, en el punto 1, mientras que con la empresa ostentando poder de mercado en ambos mercados, en el punto 4, el salario y el nivel de empleo son los más bajos. No hay,

[18] Robinson (1933, p. 283) considera que un grupo de trabajadores está siendo explotado cuando su salario es menor que "the marginal physical product that they are producing, valued at the price at which it is being sold", y distingue la explotación que se debe al monopolio del bien de la que se debe al monopsonio del trabajo. Parte de la definición de Pigou, a quien cita (Pigou, 1932, p. 549), según la cual hay explotación cuando el salario es menor que el valor del producto marginal neto para la empresa. La considera demasiado restrictiva y la amplía al producto marginal valorado a su precio de venta, lo que permite que también el monopolio del bien genere explotación. Para Robinson, la causa fundamental de la explotación es la falta de elasticidad perfecta de la oferta de trabajo o de la demanda del bien (p. 281).

sin embargo, paro involuntario, porque el salario se fija sobre la curva de oferta y todos los que desean trabajar a ese salario encuentran empleo. El empleo es bajo porque muchos prefieren no trabajar por un salario tan reducido. Cuando la empresa es solo monopolista o solo monopsonista, los resultados se encuentran entre ambos extremos, sin que pueda decirse con carácter general cuál de los dos casos da lugar a más empleo.

La principal implicación de política del monopsonio se refiere al salario mínimo, que, fijado entre el salario de monopsonio, W_3, y el competitivo, W_1, puede elevar a la vez el salario y el empleo.[19] El ejemplo más conocido es el de Card y Krueger (1994), que compararon los restaurantes de comida rápida de Nueva Jersey, que subió su salario mínimo en 1992, con los de la vecina Pensilvania, que no lo hizo, y no hallaron una caída del empleo, resultado difícil de explicar en competencia perfecta pero compatible con el monopsonio (Manning, 2003).[20]

Poder de monopolio de los trabajadores: el sindicato

Cuando son los trabajadores quienes tienen poder en la oferta de trabajo, el resultado es el simétrico del monopsonio. El trabajador es el único oferente de trabajo porque está muy especializado o, sobre todo, porque los trabajadores se agrupan en un sindicato que negocia con las empresas, algo que se viene observando en las economías desarrolladas desde el siglo XIX.

[19] Si se fija un salario mínimo $W_{\text{mín}} < W_3$, este no afecta al equilibrio de mercado, pues ya se estaba pagando un salario mayor y la restricción no es vinculante.

Si se fija un salario mínimo $W_{\text{mín}} \in (W_3, W_1)$, la restricción es vinculante: la empresa paga el salario mínimo y pasa a ser salario-aceptante hasta el empleo en el que este corta la curva de oferta. Al aumentar el salario mínimo en este intervalo, la situación tiende a la de competencia y, lejos de generar más paro, aumenta el empleo.

Si se fija un salario mínimo por encima del competitivo, $W_{\text{mín}} > W_1$, el empleo cae respecto al de competencia y aparece un exceso de oferta de trabajo, que aumenta con el salario mínimo. Mientras no sea demasiado alto, sin embargo, el empleo puede seguir siendo mayor que el del monopsonio sin salario mínimo.

[20] La evidencia posterior es mixta:

- Cengiz et al. (2019) analizan 138 subidas del salario mínimo en estados de EE. UU. entre 1979 y 2016 y no encuentran una pérdida neta de empleos de bajo salario: los que desaparecen por debajo del nuevo mínimo, un 1,8 % del empleo previo, se compensan con los que aparecen justo por encima, un 2,1 %.
- Azar et al. (2022) muestran que los mercados de trabajo locales están muy concentrados y que una mayor concentración se asocia con salarios ofrecidos más bajos, lo que respalda la existencia de monopsonio.

En España, la evidencia apunta a efectos negativos, aunque pequeños:

- Sobre la subida del SMI del 22 % en 2019, Barceló et al. (2021) estiman, como cota máxima, que redujo el empleo de los trabajadores directamente afectados entre 6 y 11 puntos y el empleo total entre 0,6 y 1,1 puntos.
- La AIReF (2025) estima, con datos de afiliación, que esa misma subida supuso entre 40.000 y 65.000 afiliados menos y que la de 2023 supuso entre 55.000 y 85.000, con elasticidades del empleo al salario mínimo de entre −0,01 y −0,02 en 2019 y de entre −0,05 y −0,07 en 2023. El efecto ha sido mayor en la última subida porque el salario mínimo alcanza ya a muchos más trabajadores, sobre todo mujeres, jóvenes y extranjeros.

 © 2026 Manuel Cabado García

La función de utilidad del sindicato depende positivamente del salario y del nivel de empleo, es decir, de la masa salarial, $W(L) \cdot L$, donde $W(L)$ es el salario que la demanda de trabajo permite cobrar con el empleo L, y negativamente de la desutilidad asociada al trabajo, $C(L)$:

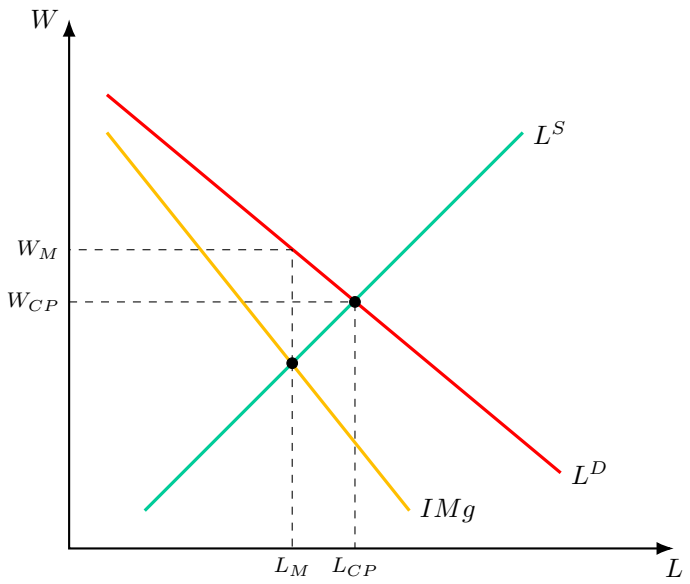
$$U = W(L) \cdot L - C(L)$$

De su CPO se obtiene que el sindicato iguala su ingreso marginal, IMg, al coste marginal del trabajo, CMg_L, en una condición en la que interviene la elasticidad de la demanda de trabajo respecto al salario, $\varepsilon_{L^D,W}$:

$$IMg = W\left(1 - \frac{1}{\left|\varepsilon_{L^D,W}\right|}\right) = CMg_L$$

Gráficamente, en un plano de trabajo, L, y salario nominal, W, como el de la Figura 11.9, se representan la demanda de trabajo, L^D, el ingreso marginal del sindicato, IMg, y la oferta de trabajo, L^S. Así:

Figura 11.9: Poder de monopolio del trabajador



- La curva de ingreso marginal está por debajo de la de demanda porque, al tener el sindicato poder de monopolio, un aumento de su oferta reduce el salario de todas las unidades, $IMg < W$.
- El coste marginal del trabajo, CMg_L, es la propia curva de oferta, L^S, que recoge la desutilidad marginal de trabajar.
- El sindicato elige el empleo L_M, donde IMg corta a L^S, y fija el salario W_M, que se lee sobre la curva de demanda, frente al empleo L_{CP} y el salario W_{CP} del equilibrio competitivo.

En suma, con un sindicato con poder de monopolio en el mercado de trabajo, el nivel de empleo es inferior y el salario mayor que en competencia perfecta, por lo que se produce paro involuntario: a ese salario hay trabajadores que desearían trabajar y no encuentran empleo. El resultado depende, además, del objetivo del sindicato. Si maximiza el empleo, el equilibrio es el competitivo. Si maximiza la masa salarial, fija el salario en el que su ingreso marginal es nulo, siempre que a ese salario haya trabajadores suficientes dispuestos a trabajar. Si maximiza, en cambio, la renta neta de sus afiliados, como en el planteamiento anterior, iguala ingreso y coste marginales y obtiene el salario más alto y el empleo más bajo de los tres casos.

Así, cuanto mayor es el poder de mercado, ya sea de la empresa o del sindicato, menor es el empleo. La diferencia está en el paro: con monopsonio no hay paro involuntario, porque el salario se lee sobre la oferta, y con sindicato sí, porque el salario supera al competitivo.

Cuando un sindicato con poder de monopolio se enfrenta a una empresa monopsonista se tiene un monopolio bilateral, ya presentado en el Tema 8. El resultado queda indeterminado: el salario se sitúa entre el que fijaría el monopsonista y el que pediría el sindicato, según el poder de negociación de cada parte, que formaliza la solución de Nash generalizada que se estudia en el Tema 12. Es el único caso en el que el análisis neoclásico justifica el sindicato, como poder compensador del monopsonio, porque puede acercar el salario y el empleo a los competitivos.[21]

Todo el análisis anterior es estático: el individuo decide cuántas horas trabaja en un único período, y el trabajo se supone homogéneo. Los dos bloques siguientes levantan sucesivamente estos dos supuestos.

11.3. Análisis intertemporal de la oferta de trabajo ⏳ 4’30"

El modelo estático trata cada período como si fuera independiente de los demás, pero las decisiones de trabajo se toman a lo largo de toda la vida. El análisis intertemporal considera que el individuo reparte su consumo y su ocio entre períodos, lo que permite explicar cómo varía la oferta de trabajo a lo largo del ciclo vital y cómo responde a los cambios transitorios del salario, que en los modelos del ciclo real constituyen el mecanismo de propagación de los shocks.

La oferta de trabajo a lo largo del ciclo vital

El salario sigue con la edad un perfil creciente y cóncavo: es bajo al comienzo de la vida laboral, aumenta con la experiencia hasta alcanzar un máximo hacia los cincuenta años

[21] El razonamiento se desarrolla en el Apéndice 11.A.3.

 © 2026 Manuel Cabado García

y después se estanca o cae ligeramente. Las horas trabajadas siguen un perfil parecido (Borjas, 2016, p. 65).

La teoría del ciclo vital explica esta coincidencia. Como el individuo conoce de antemano la senda de su salario, esos cambios previstos no alteran su riqueza a lo largo de la vida, que ya los incorporaba, y solo cambian el precio del ocio en cada momento. Por eso concentrará sus horas de trabajo en los años en que el salario es más alto, que es cuando el ocio es más caro, y suavizará su consumo ahorrando o endeudándose. A diferencia del modelo estático, estos cambios del salario solo generan efecto sustitución.

La jubilación es la última decisión de oferta de trabajo del ciclo vital y también se analiza de forma intertemporal. En el modelo de valor de opción de Stock y Wise (1990), seguir trabajando conserva la opción de jubilarse más adelante en mejores condiciones, y el trabajador se jubila cuando el valor de esa opción deja de ser positivo. Las reglas de las pensiones alteran así la oferta de trabajo de los mayores. En los países de la OCDE, la penalización implícita que los sistemas de pensiones imponen a seguir trabajando explica en torno al 80 % de las diferencias entre países en la inactividad de los trabajadores de 55 a 65 años (Gruber y Wise, 1999, pp. 30-31).

La sustitución intertemporal: Lucas y Rapping (1969)

Lucas y Rapping (1969) fueron los primeros en tratar la oferta de trabajo como una decisión intertemporal, para explicar por qué el empleo fluctúa mucho a lo largo del ciclo mientras el salario real lo hace poco.[22] El modelo se basa en los siguientes supuestos:

1. El individuo vive dos períodos.
2. No dispone de riqueza inicial, es decir, todas sus rentas son salariales.
3. Los mercados de capitales son perfectos, es decir, puede transferir fondos de un período a otro al tipo de interés real, r.
4. La función de utilidad es aditivamente separable en el tiempo y también entre consumo y ocio.
5. La dotación de tiempo de cada período se normaliza a la unidad, $O_t + L_t = \bar{L} = 1$.

Con el consumo de cada período, C_t, las horas trabajadas, L_t, el salario real, w_t, y la tasa subjetiva de descuento, ρ, que define el factor de descuento, β, y con $u(C_t)$ la utilidad del consumo y $v(1 - L_t)$ la del ocio, el individuo resuelve

$$\max_{C_1, C_2, L_1, L_2} U = u(C_1) + v(1 - L_1) + \beta\left[u(C_2) + v(1 - L_2)\right], \qquad \beta = \frac{1}{1 + \rho}$$

[22] El artículo original es un modelo econométrico agregado, estimado con series anuales de Estados Unidos entre 1929 y 1965, que pretendía además dar un fundamento de oferta y demanda a la curva de Phillips, aunque ya parte de un hogar que decide entre dos períodos (pp. 727-728). La versión estilizada que se presenta aquí es la de Romer (2019, pp. 198-199).

sujeto a la restricción presupuestaria intertemporal, según la cual el valor actual del consumo debe igualar al de las rentas salariales:

$$C_1 + \frac{C_2}{1+r} = w_1 L_1 + \frac{w_2 L_2}{1+r}$$

De las CPO se obtienen tres condiciones.[23] Son las siguientes:

- Intratemporal: en cada período, la relación marginal de sustitución entre ocio y consumo se iguala al salario real, $RMS_{O_t,C_t} = v'(1-L_t)/u'(C_t) = w_t$, como en el modelo estático.
- Ecuación de Euler: el consumo se reparte entre períodos hasta que

$$u'(C_1) = \beta(1+r) \cdot u'(C_2)$$

- Intertemporal del ocio, que es la novedad del modelo:

$$\frac{v'(1-L_1)}{v'(1-L_2)} = \frac{1+r}{1+\rho} \cdot \frac{w_1}{w_2} \tag{11.4}$$

  es decir, el individuo trabajará relativamente más en el período en que el salario, corregido por el tipo de interés y la impaciencia, sea más alto.

Por tanto, la oferta de trabajo relativa depende positivamente del salario relativo y del tipo de interés, y negativamente de la tasa de descuento, que mide la impaciencia del individuo:

$$\frac{L_1^S}{L_2^S} = f\left(\underset{(+)}{\frac{w_1}{w_2}}, \underset{(+)}{r}, \underset{(-)}{\rho}\right)$$

De ahí se derivan cuatro implicaciones:

- Una subida permanente del salario, que eleva en la misma proporción el de los dos períodos, no altera la oferta relativa entre períodos. Su efecto sobre el nivel de horas es el del modelo estático: los efectos sustitución y renta tienen signos contrarios y el resultado es ambiguo.
- Una subida transitoria del salario en el primer período aumenta la oferta de trabajo en ese período respecto al segundo. Es la sustitución intertemporal del trabajo: se trabaja más cuando trabajar compensa más.
- Una subida del tipo de interés aumenta la oferta de trabajo presente, porque trabajar y ahorrar hoy resulta más atractivo.
- Una mayor impaciencia la reduce, porque el individuo prefiere disfrutar del ocio cuanto antes.

[23] La derivación, con el lagrangiano del problema, se desarrolla en el Apéndice 11.A.4.

 © 2026 Manuel Cabado García

La implicación de política es que los cambios transitorios tienen efectos mucho mayores sobre el empleo que los permanentes. Una rebaja temporal de los impuestos sobre el trabajo, por ejemplo, desplaza horas hacia el período en que se aplica, mientras que una permanente apenas altera la oferta relativa.

La elasticidad de Frisch y MaCurdy (1981)

El parámetro clave del modelo es la elasticidad de Frisch, que mide la respuesta de las horas trabajadas a un cambio del salario manteniendo constante la utilidad marginal de la riqueza, λ:

$$\varepsilon^F = \left.\frac{\partial \ln L_t}{\partial \ln w_t}\right|_{\lambda \text{ constante}}$$

Recoge, por tanto, un efecto sustitución intertemporal puro, el que se produce cuando el salario cambia de forma prevista a lo largo del ciclo vital o de forma transitoria, sin alterar la riqueza de toda la vida.[24] Por ejemplo, si la utilidad de cada período es $\ln C_t - \varphi\, L_t^{1+1/\eta}/\,(1+1/\eta)$, donde φ mide el peso de la desutilidad del trabajo y η su curvatura, la elasticidad de Frisch es constante e igual a η.

MaCurdy (1981) fue el primero en estimarla de forma coherente con el modelo. Utilizó datos de panel de 513 hombres casados de Estados Unidos entre 1967 y 1976 y aprovechó que el crecimiento previsto del salario a lo largo del ciclo vital no altera la riqueza: al comparar a cada individuo consigo mismo a lo largo del tiempo, el efecto de la riqueza desaparece y queda solo el de sustitución. Concluyó que una subida del 10 % del salario real debida a su crecimiento a lo largo del ciclo vital eleva las horas trabajadas entre un 1 % y un 3 %, es decir, una elasticidad de Frisch de entre 0,1 y 0,3.

Evidencia empírica y papel en el ciclo real

Buena parte de la evidencia posterior, que estima esta elasticidad con efectos fijos individuales, confirma que es positiva pero pequeña, en torno a 0,1 para los varones: las horas apenas cambian en las edades centrales aunque el salario crezca (Borjas, 2016, cap. 2). El caso de los taxistas, que eligen cada día cuántas horas trabajan, ilustra el debate. Los primeros trabajos apuntaban a que dejaban de trabajar al alcanzar un objetivo de ingresos, pero la evidencia más amplia y reciente muestra que trabajan más cuando cada hora rinde más.[25] Chetty et al. (2011), al revisar la evidencia cuasiexperimental, proponen

[24] La intuición es que un cambio pequeño y pasajero del salario apenas modifica lo que el individuo puede gastar a lo largo de su vida, de modo que solo cambia el momento en que le conviene trabajar. Por eso la elasticidad de Frisch es la mayor de las tres que se distinguen en la literatura: supera a la hicksiana, que mantiene constante la utilidad y no la utilidad marginal de la riqueza, y esta, a su vez, supera a la marshalliana, que incluye el efecto renta, siempre que el ocio sea un bien normal: $\varepsilon^F \geq \varepsilon^H \geq \varepsilon^M$ (Cahuc et al., 2014, cap. 1). Debe su nombre a Ragnar Frisch, que estudió las demandas con utilidad marginal de la renta constante.

[25] Camerer et al. (1997) hallaron en Nueva York elasticidades diarias negativas, de entre −0,5 y −1, como si los conductores tuvieran un objetivo de ingresos diario. Farber (2005) atribuyó ese resultado a

valores de 0,5 en el margen intensivo y de 0,25 en el extensivo, lo que da una elasticidad de Frisch de las horas agregadas de alrededor de 0,75.

La elasticidad de Frisch es decisiva para la teoría del ciclo real. Kydland y Prescott (1982) extendieron el razonamiento de Lucas y Rapping a una economía con incertidumbre sometida a shocks reales de productividad. Ante un shock transitorio que eleva el salario real, los trabajadores sustituyen ocio presente por ocio futuro y trabajan más hoy, lo que genera la expansión y la posterior contracción. Un shock permanente, en cambio, no altera la oferta relativa entre períodos: su efecto riqueza compensa en buena parte el de sustitución, de modo que, cuanto más persistente es el shock, menos sustitución intertemporal genera. Su efecto principal se transmite a través de la inversión, por el aumento permanente de la productividad marginal del capital. La implicación de política es que, si las fluctuaciones del empleo son respuestas óptimas a shocks reales, las políticas de estabilización de la demanda no mejoran el bienestar.

Sin embargo, para reproducir las grandes fluctuaciones del empleo con salarios reales poco variables, estos modelos necesitan una elasticidad intertemporal elevada, mayor que la que estiman la mayoría de los trabajos microeconómicos, lo que ha generado un intenso debate (Borjas, 2016, cap. 2). Si la sustitución intertemporal es pequeña, buena parte de las fluctuaciones del empleo no pueden ser voluntarias, y hay que acudir a las rigideces y fricciones que se estudian en el Tema 12 y, desde el punto de vista macroeconómico, en los Temas 11 y 12 de la Parte B.

Tanto el enfoque estático como el intertemporal han supuesto hasta aquí que todos los trabajadores son iguales. La teoría del capital humano levanta este supuesto con un razonamiento también intertemporal: formarse hoy tiene un coste presente a cambio de un salario más alto en el futuro.

11.4. Teoría del capital humano ⏳ 4’

Los trabajadores difieren en su cualificación, que se adquiere en gran medida mediante la educación y la formación, por lo que el salario de equilibrio deja de ser único.

La educación como inversión: Schultz (1961) y Becker (1964)

Schultz (1961), en su discurso como presidente de la Asociación Económica Americana, observó que la producción de Estados Unidos había crecido mucho más que la tierra, las horas de trabajo y el capital físico empleados, y atribuyó buena parte de la diferencia a la inversión de las personas en sí mismas, en educación, en salud o en migraciones. Gary Becker, Premio Nobel en 1992 “por haber extendido el dominio del análisis microeco-

errores de medida del salario diario. Con todos los viajes de taxi de Nueva York entre 2009 y 2013, Farber (2015) estimó elasticidades positivas, de en torno a 0,6, coherentes con la sustitución intertemporal.

 © 2026 Manuel Cabado García

nómico a un amplio abanico de comportamientos e interacciones humanas, incluidos los que tienen lugar fuera del mercado", desarrolló esta idea en la teoría del capital humano (1964), que considera la educación como una inversión que incrementa la productividad del trabajador y, con ella, su salario.[26]

El capital humano es el conjunto de conocimientos y habilidades de un trabajador que elevan su productividad y que, en el modelo de Becker, se adquieren mediante la educación y la formación. El trabajador no puede venderlo, pero sí alquilar sus servicios a las empresas.[27]

Así, la educación deja de ser un fenómeno cultural o social y pasa a ser un fenómeno económico, una decisión de inversión más. Cada individuo compara los costes de formarse, directos, como la matrícula o los materiales, e indirectos, como el salario al que renuncia mientras estudia, con el aumento de sus salarios futuros, y sigue estudiando mientras el valor actual neto de la inversión sea positivo.[28] Este enfoque generó fuertes críticas cuando se formuló, en los años sesenta, porque se consideraba que denigraba a los trabajadores tratándolos como meras máquinas, pero su capacidad para explicar y predecir los fenómenos educativos acabó otorgándole la validez que tiene hoy.

Formación general y formación específica

Becker distinguió entre dos tipos de formación, según dónde pueda aprovecharse:

1. Formación general: incrementa la productividad del individuo en cualquier empresa.
2. Formación específica: incrementa la productividad del individuo solo en una determinada empresa.[29]

[26] Los rasgos de su enfoque se resumen en el Apéndice 11.A.5.

[27] La intuición es que lo aprendido sigue elevando el salario durante toda la vida laboral, pero va unido a la persona: quien lo tiene no puede traspasarlo y solo cobra por usarlo cuando trabaja.

[28] Los agentes deciden formarse en la primera etapa de su vida en función de los rendimientos esperados. Ello permite comprender por qué durante la expansión anterior a 2008 en España la tasa de abandono escolar era tan alta, por la baja prima salarial de la educación, y por qué aumentó la formación cuando estalló la crisis de 2008, con una alta tasa de desempleo, especialmente para los individuos con baja formación. La tasa de abandono temprano de la educación y la formación entre los jóvenes de 18 a 24 años pasó del 31,7 % en 2008 al 12,8 % en 2025, aún por encima del 9,1 % de la media de la UE (Eurostat, 2026).

[29] Por ejemplo, el aprendizaje de los procedimientos internos, de los sistemas informáticos propios o de la organización de una empresa concreta apenas tiene valor en otra. A cambio de esa formación, la empresa suele pagar al principio un salario más bajo que el que el empleado podría ganar fuera, lo que equivale a que este financie parte de su formación, y luego lo compensa con un salario creciente y una carrera interna. Esa dependencia mutua reduce la rotación, porque a ninguna de las dos partes le conviene romper la relación.

Este patrón es habitual en las grandes consultoras y auditoras, como Deloitte, que forman a sus empleados en metodologías y herramientas propias, pero también en la banca, con sus sistemas y productos internos, o en la industria, con la maquinaria y los procesos propios de cada planta. Cuando la formación que paga la empresa es costosa y en parte trasladable a otras empresas, el riesgo de que el trabajador se marche crece. Por eso el Estatuto de los Trabajadores permite pactar su permanencia en la empresa durante un máximo de dos años cuando ha recibido una especialización profesional a cargo del empresario (art. 21.4).

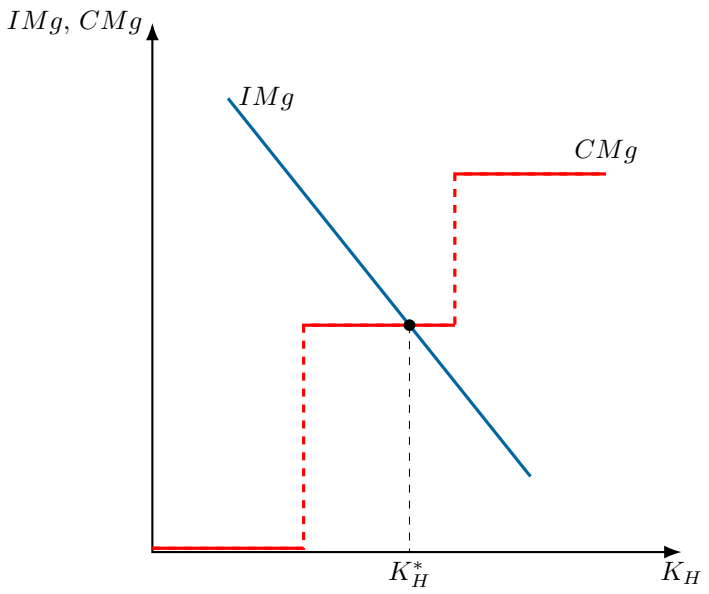
La distinción determina quién paga la formación. La general corresponde al trabajador, pues ninguna empresa tiene incentivos a costear una formación que el trabajador podría aprovechar en otras: si lo hiciera, la competencia de estas le obligaría a subirle el salario hasta su nueva productividad y no recuperaría lo invertido. Por eso el trabajador la paga aceptando un salario más bajo mientras se forma (Borjas, 2016, p. 265). La formación específica, en cambio, se reparte entre la empresa y el trabajador, porque ninguno querría asumirla entera: si la pagara solo la empresa, el trabajador podría marcharse, y si la pagara solo el trabajador, la empresa podría despedirlo, y en ambos casos se perdería la inversión (Borjas, 2016, pp. 266-267).[30] Por eso los trabajadores con formación específica rotan menos y son los últimos en ser despedidos en las recesiones.

En el caso de la formación general, el trabajador invierte en capital humano, K_H, hasta que el aumento del salario que le reporta la última unidad, $\partial W\left(K_H\right)/\partial K_H$, descontado al tipo de interés, r, se iguala a su coste marginal:

$$\frac{\partial W\left(K_H\right)/\partial K_H}{1+r} = CMg_{K_H}$$

Gráficamente, en un plano de capital humano, K_H, y su ingreso y coste marginales, IMg y CMg, como el de la Figura 11.10, se representan ambas curvas. En él:

Figura 11.10: Modelo de capital humano: formación general



- Ingreso marginal: es decreciente porque la inversión en capital humano presenta

[30] Durante la formación, el trabajador cobra menos de lo que ganaría en otra empresa, pero más de lo que produce mientras se forma, de modo que cada parte paga una porción del coste. Una vez completada la formación, la empresa le sube el salario por encima de lo que podría ganar fuera, pero por debajo de su nueva productividad marginal, lo que reparte los rendimientos: al trabajador no le interesa irse y a la empresa no le interesa despedirlo.

 © 2026 Manuel Cabado García

rendimientos decrecientes, por los límites de la memoria y del tiempo disponible.[31]

- Coste marginal: es creciente y discontinuo, pues, según Becker, el individuo acude a fuentes de financiación racionadas, empezando por las más baratas: primero las donaciones familiares y las becas, de coste nulo, después los préstamos en condiciones favorables y los recursos propios y, por último, los préstamos a precios de mercado, cada vez más caros.[32]
- La inversión óptima en capital humano, K_H^*, se alcanza donde se cortan ambas curvas.

El elemento relevante de este problema intertemporal es que puede estudiarse sin tener en cuenta la función de utilidad, gracias al teorema de la separación en la formulación de Acemoglu y Autor (2009, cap. 1): con mercados de capitales perfectos y un ocio que no reporta utilidad, la inversión en capital humano que maximiza la utilidad es la que maximiza el valor actual de los ingresos, con independencia de las decisiones de consumo.

Las decisiones de invertir en educación, que difieren según la capacidad, los costes y el acceso a la financiación de cada individuo, dan lugar a trabajadores heterogéneos. Esto altera los resultados de la teoría neoclásica tradicional, porque segmenta el mercado: ya no existe un único salario de equilibrio, sino distintos salarios según la cualificación, aunque el único desempleo sigue siendo el voluntario. Entre empleos que exigen la misma cualificación también hay diferencias salariales, las diferencias compensatorias.[33]

La ecuación de Mincer (1974)

La aplicación empírica de la teoría es la función de ingresos de Mincer (1974), que ya había modelizado la decisión de invertir en educación (1958). Relaciona el logaritmo del salario, $\ln w$, con los años de escolarización, S, y con los años de experiencia laboral, X, y su cuadrado, con una perturbación aleatoria, ε:

$$\ln w = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \varepsilon$$

El coeficiente β_1 se interpreta como la tasa de rendimiento de un año adicional de educación, de modo que cada año de estudios eleva el salario en torno a un $100 \cdot \beta_1\,\%$.[34] La media mundial de este rendimiento privado, calculada a partir de más de mil estimaciones en 139 países, es de un 9 % por año de escolarización (Psacharopoulos y Patrinos, 2018). En España, los trabajadores con estudios terciarios ganan un 49 % más que los que solo tienen secundaria superior, frente al 54 % de media de la OCDE (OCDE, 2025, nota país

[31] La intuición es que aprender a leer aporta más que el décimo máster.

[32] Un caso muy aplicable a este modelo es el de los préstamos universitarios en Estados Unidos, mientras que en España lo es menos por el peso de las universidades públicas.

[33] Se desarrollan en el Apéndice 11.A.7.

[34] La derivación de la ecuación se recoge en el Apéndice 11.A.6.

de España, p. 3). Los coeficientes de la experiencia, con $\beta_2 > 0$ y $\beta_3 < 0$, recogen la formación en el empleo y dan un perfil de ingresos creciente y cóncavo con la edad, como el que se vio en el análisis intertemporal. Ben-Porath (1967) explica este perfil: como la formación rinde durante menos años a medida que se acerca la jubilación, el valor de invertir en capital humano cae con la edad. El individuo estudia a tiempo completo al principio, invierte cada vez menos en formación después y su salario crece cada vez más despacio hasta alcanzar un máximo entre los 45 y los 60 años (Cahuc et al., 2014, pp. 204-207).[35]

La ecuación de Mincer es el primer pilar del análisis empírico de los rendimientos de la educación y la aplicación más directa de la teoría del capital humano a la distribución de los salarios.

Rendimientos de la educación y problemas de estimación

Aunque la ecuación de Mincer se ajusta bien a los datos, su estimación por mínimos cuadrados ordinarios, MCO, no permite interpretar β_1 como un efecto causal. El principal problema es el sesgo de habilidad: si los individuos más capaces estudian más y, además, ganarían más con cualquier nivel de estudios, parte de lo que se atribuye a la educación se debe en realidad a una capacidad que no se observa, y el rendimiento estimado está sesgado al alza. A ello se suman la influencia del entorno familiar, porque los niveles educativos altos suelen ir unidos a una renta y una educación altas de los padres, y los errores de medida de los años de estudio, que sesgan el rendimiento a la baja.

Por ello, la literatura ha recurrido a experimentos naturales, que buscan variaciones de la educación ajenas a la capacidad:

- Angrist y Krueger (1991) usaron como variable instrumental el trimestre de nacimiento: por las leyes de escolarización obligatoria, quienes nacen al comienzo del año pueden dejar los estudios con menos años de escuela. Sus estimaciones del rendimiento de un año de estudios, de entre el 6 % y el 10 %, apenas difieren de las de MCO, lo que sugiere que el sesgo de habilidad es pequeño (Card, 2001).
- Card (1995) usó como instrumento la cercanía de una universidad al lugar en el que se crece, que abarata los estudios, y obtuvo un rendimiento del 13,2 % por año de estudios, frente al 7,3 % por MCO (Card, 1999, p. 1836).
- Ashenfelter y Krueger (1994) compararon gemelos idénticos, que comparten capacidad y entorno familiar, y estimaron un rendimiento de entre el 12 % y el 16 % por año de estudios.

[35] La crítica más importante es que parece muy simplista suponer que, una vez controlada la experiencia, todas las diferencias salariales se deben únicamente a las diferencias en formación. Por ejemplo, no logra explicar la discriminación salarial entre hombres y mujeres (véase el Apéndice 11.A.8).

 © 2026 Manuel Cabado García

- Duflo (2001) aprovechó el gran programa de construcción de escuelas de Indonesia de los años setenta y obtuvo rendimientos de entre el 6,8 % y el 10,6 %.

En conjunto, las estimaciones con variables instrumentales suelen ser iguales o mayores que las de MCO, lo que Card (1999) atribuye a que los rendimientos son heterogéneos y más altos para quienes estudian más gracias a esas reformas, a menudo personas con dificultades para financiarse.[36] Heckman et al. (2006) advierten, además, de que el coeficiente de Mincer solo coincide con la tasa interna de rendimiento de la educación bajo supuestos muy restrictivos, como la ausencia de costes directos y de impuestos. La implicación de política es que la educación es una inversión muy rentable y que los obstáculos para financiarla pueden explicar buena parte de la inversión insuficiente en ella.

La señalización como alternativa y otras críticas

La teoría del capital humano no es la única explicación de que los más educados ganen más. En el modelo de señalización de Spence (1973), que se estudia en el Tema 13, la educación no aumenta la productividad, sino que, en un contexto de información asimétrica, permite a los trabajadores más productivos distinguirse, porque para ellos estudiar es menos costoso. Las cadenas causales son distintas:

- Becker: ↑ educación (inversión) ⇒ ↑ productividad ⇒ ↑ salario.
- Spence: ↑ productividad ⇒ ↑ educación (señal) ⇒ ↑ salario.

Ambas teorías predicen la misma relación positiva entre educación y salario, pero sus implicaciones de política difieren: si la educación solo señaliza, su rendimiento social es menor que el privado y puede haber sobreeducación (Cahuc et al., 2014, cap. 4). Además, la teoría de Becker está sujeta a otras críticas:

- La teoría de Becker no explica por qué se financia públicamente la educación si el mercado es capaz de proveer el nivel óptimo de formación, porque no tiene en cuenta ni las externalidades del capital humano ni las restricciones al crédito.[37]

Por otra parte, el modelo supone información perfecta sobre costes y rendimientos, olvida que el capital humano se deprecia y que el salario depende también de otros factores, como la antigüedad o la discriminación, y trata como inversión todo gasto en educación, cuando una parte es consumo, como los estudios que se cursan por gusto personal.

[36] David Card recibió la mitad del Premio Nobel de Economía de 2021 por sus aportaciones empíricas a la economía laboral, y Joshua Angrist y Guido Imbens, la otra mitad, por sus aportaciones metodológicas al análisis de las relaciones causales, la base de este enfoque.

[37] Las restricciones de liquidez pueden hacer que individuos que a priori decidirían formarse no lo acaben haciendo. Además, la educación tiene externalidades positivas, por lo que su rendimiento social es superior al privado: los individuos con más educación tienen una menor tasa de criminalidad, incrementan la productividad agregada de las empresas y muestran comportamientos más cívicos. Por ello, la decisión privada de educarse es subóptima en el sentido de Pareto, lo que justifica subvencionarla.

Pese a estas críticas, la teoría del capital humano sigue siendo la explicación de referencia de las diferencias salariales y la base del análisis empírico de los rendimientos de la educación.

11.5. Conclusiones ⏳ 30"

Como conclusión, tanto el empleo como los salarios nacen de decisiones racionales de trabajadores y empresas que miran al futuro, y el poder de mercado y la formación explican buena parte de sus diferencias.

Sin embargo, quedan abiertas cuestiones de gran actualidad:

- Tecnología: la inteligencia artificial obliga a preguntarse qué capital humano se depreciará y cuál ganará valor, y el teletrabajo altera el coste de oportunidad del ocio y la relación entre el lugar de residencia y el empleo.
- Demografía: el envejecimiento de la población plantea cómo prolongar la vida laboral y hacer más flexible la transición a la jubilación, una decisión que el análisis intertemporal permite estudiar.
- Igualdad: aunque la participación de la mujer se ha acercado a la de los hombres, persisten diferencias salariales que se abren con la llegada de los hijos y que la teoría del capital humano no logra explicar por sí sola.
- Desempleo: ¿por qué hay quien quiere trabajar al salario vigente y no encuentra empleo? Responder exige ir más allá de un mercado que se vacía, introduciendo así las fricciones y rigideces del mercado de trabajo.

 © 2026 Manuel Cabado García

Apéndice

11.A.1. Ecuación de Slutsky con efecto renta dotación

En el modelo de oferta de trabajo, el salario aparece en los dos lados de la restricción presupuestaria: es el precio del ocio y, a la vez, determina el valor de la dotación de tiempo, es decir, la renta potencial, $R_0 = M + w\bar{L}$, con el precio del consumo normalizado a la unidad. Por ello, la ecuación de Slutsky incluye un término adicional. Con la demanda hicksiana de ocio, O^h, y la marshalliana, $O\left(w, R_0\right)$, el efecto total de una subida del salario es

$$\frac{dO}{dw} = \underbrace{\left.\frac{\partial O^h}{\partial w}\right|_{\bar{U}}}_{\text{sustitución}} \underbrace{- O \cdot \frac{\partial O}{\partial R_0}}_{\text{renta ordinario}} + \underbrace{\bar{L} \cdot \frac{\partial O}{\partial R_0}}_{\text{renta dotación}} = \frac{\partial O^h}{\partial w} + \left(\bar{L} - O\right) \cdot \frac{\partial O}{\partial R_0}$$

Como $\bar{L} - O = L$ son las horas trabajadas, el efecto renta total aumenta el ocio siempre que el individuo trabaje y el ocio sea un bien normal. Se opone así al efecto sustitución, que siempre lo reduce, y el signo del efecto total, y con él la pendiente de la oferta de trabajo, queda indeterminado. Si el individuo no trabajara, $L = 0$, solo quedaría el efecto sustitución (Cahuc et al., 2014, cap. 1).

11.A.2. Estática comparativa en la demanda de trabajo a largo plazo

A largo plazo, la demanda incondicionada de trabajo puede escribirse como la demanda condicionada, L^c, que minimiza el coste de producir una cantidad Y, evaluada en la producción que maximiza el beneficio, Y^*: $L^D\left(W, R, P\right) = L^c\left(W, R, Y^*\left(W, R, P\right)\right)$. Derivando respecto al salario se separan los dos efectos de su variación:

$$\frac{\partial L^D}{\partial W} = \underbrace{\frac{\partial L^c}{\partial W}}_{\text{efecto sustitución}} + \underbrace{\frac{\partial L^c}{\partial Y} \cdot \frac{\partial Y^*}{\partial W}}_{\text{efecto escala}}$$

El efecto sustitución es siempre negativo, porque la demanda condicionada es decreciente en el precio del propio factor, al ser la función de costes cóncava en los precios de los factores, como se vio en el Tema 6. El efecto escala es también siempre negativo. Si el trabajo es un factor normal, una subida del salario eleva el coste marginal y reduce la producción óptima, $\partial Y^*/\partial W < 0$, y con menos producción se demanda menos trabajo, $\partial L^c/\partial Y > 0$. Si fuera un factor inferior, ambos signos se invertirían y el producto seguiría siendo negativo. Por eso en la teoría de la producción no existen factores Giffen, y la demanda de trabajo a largo plazo tiene pendiente negativa (Cahuc et al., 2014, cap. 2). Es, además, más elástica que la de corto plazo, porque a largo plazo la empresa puede ajustar también el capital, como establece el teorema de Le Chatelier-Samuelson.

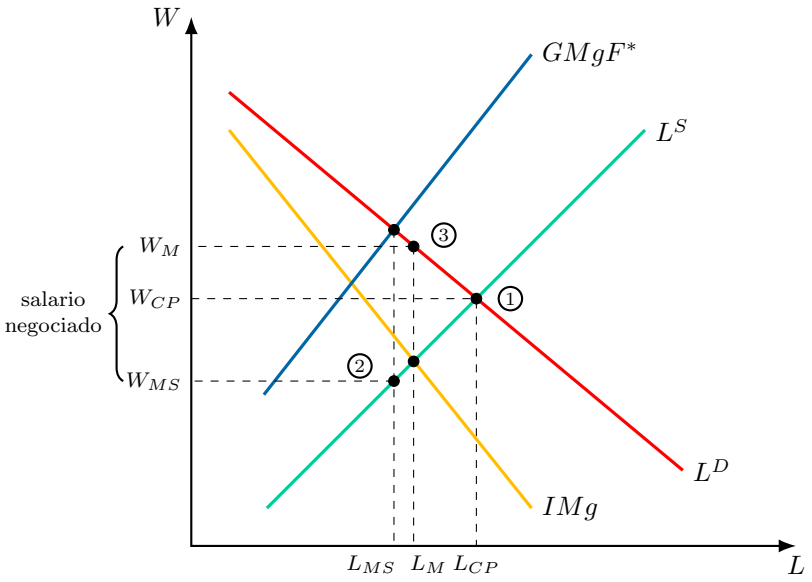
Sobre el capital, en cambio, el efecto de una bajada del salario es ambiguo. El efecto sustitución lo reduce, porque el trabajo se ha abaratado en términos relativos, y el efecto escala lo aumenta si el capital es un factor normal. Si los factores son cooperantes, $f_{KL} > 0$, domina el efecto escala, de modo que una bajada del salario eleva también la demanda de capital, como se vio en el cuerpo del tema.

11.A.3. Monopolio bilateral

Un monopolio bilateral surge cuando un único vendedor se enfrenta a un único comprador. En el mercado de trabajo, es el caso de un sindicato con poder de monopolio en la oferta de trabajo frente a una empresa monopsonista.

Cada parte desea un resultado distinto. Gráficamente, en un plano de trabajo, L, y salario nominal, W, como el de la Figura 11.11, se representan la demanda de trabajo, L^D, que es el ingreso marginal del factor del monopsonista, la oferta, L^S, el gasto marginal del monopsonista, $GMgF^*$, y el ingreso marginal del sindicato, IMg. Así:

Figura 11.11: Monopolio bilateral en el mercado de trabajo



- El monopsonista, que iguala el ingreso marginal del factor a su gasto marginal, querría contratar el empleo L_{MS}, donde L^D corta a $GMgF^*$, y pagar el salario W_{MS}, que se lee sobre la curva de oferta, en el punto 2.
- El sindicato, que iguala su ingreso marginal al coste marginal del trabajo, que es la oferta, querría fijar un salario más alto, W_M, sobre la curva de demanda, con el empleo L_M, en el punto 3.
- Como ninguna de las partes puede imponer su resultado, el salario queda indeterminado entre W_{MS} y W_M, el intervalo de negociación que recoge la figura, y depende

 © 2026 Manuel Cabado García

del poder de negociación de cada una. El salario competitivo, W_{CP}, el del punto 1, queda dentro de ese intervalo.

Desde la óptica neoclásica, este es el único caso en el que el sindicato puede mejorar el resultado del mercado. Si fija un salario entre el de monopsonio y el competitivo, el monopsonista pasa a ser salario-aceptante hasta ese nivel de empleo, igual que con un salario mínimo, por lo que pueden subir a la vez el salario y el empleo. El sindicato actúa así como poder compensador del monopsonio.

11.A.4. El problema intertemporal: lagrangiano y elasticidad de Frisch

El problema de Lucas y Rapping se resuelve con el lagrangiano, en el que el multiplicador, λ, asociado a la restricción presupuestaria intertemporal es la utilidad marginal de la riqueza:

$$\mathcal{L} = u(C_1) + v(1 - L_1) + \beta\left[u(C_2) + v(1 - L_2)\right] + \lambda\left[w_1 L_1 + \frac{w_2 L_2}{1+r} - C_1 - \frac{C_2}{1+r}\right]$$

Las CPO son:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial C_1} &= u'(C_1) - \lambda = 0 \\
\frac{\partial \mathcal{L}}{\partial C_2} &= \beta u'(C_2) - \frac{\lambda}{1+r} = 0 \\
\frac{\partial \mathcal{L}}{\partial L_1} &= -v'(1 - L_1) + \lambda w_1 = 0 \\
\frac{\partial \mathcal{L}}{\partial L_2} &= -\beta v'(1 - L_2) + \frac{\lambda w_2}{1+r} = 0
\end{aligned}$$

Las condiciones de las horas tienen sentido económico: una hora más de trabajo cuesta la utilidad marginal del ocio, v', y aporta w_t unidades de consumo, valoradas a λ. Dividiendo la tercera entre la primera se obtiene la condición intratemporal, $v'(1 - L_1)/u'(C_1) = w_1$. Combinando la primera y la segunda se obtiene la ecuación de Euler, $u'(C_1) = \beta(1 + r)\,u'(C_2)$. Dividiendo, por último, la tercera entre la cuarta:

$$\frac{v'(1 - L_1)}{\beta v'(1 - L_2)} = (1 + r)\frac{w_1}{w_2} \;\Rightarrow\; \frac{v'(1 - L_1)}{v'(1 - L_2)} = \frac{1 + r}{1 + \rho} \cdot \frac{w_1}{w_2}$$

que es la condición 11.4 del cuerpo del tema.

La elasticidad de Frisch se obtiene manteniendo constante la utilidad marginal de la riqueza. Con la utilidad del ejemplo del cuerpo del tema, $\ln C_t - \varphi\, L_t^{1+1/\eta}/(1 + 1/\eta)$, que recoge directamente la desutilidad de las horas en lugar de la utilidad del ocio, y con λ_t la utilidad marginal de la riqueza valorada en el período t, de modo que $\lambda_1 = \lambda$ y

$\lambda_2 = \lambda/\left[\beta\,(1+r)\right]$, la condición de las horas es

$$\varphi L_t^{1/\eta} = \lambda_t w_t \;\Rightarrow\; L_t = \left(\frac{\lambda_t w_t}{\varphi}\right)^{\eta} \;\Rightarrow\; \ln L_t = \eta \ln w_t + \eta \ln \lambda_t - \eta \ln \varphi$$

de modo que, con λ_t constante, $\varepsilon^F = \partial \ln L_t / \partial \ln w_t = \eta$. Un cambio transitorio del salario apenas altera λ_t, y su efecto sobre las horas lo mide η. Un cambio permanente, en cambio, eleva la riqueza y reduce λ_t, lo que compensa en parte el efecto anterior. En esto se basa la estrategia de MaCurdy (1981), que trata $\ln \lambda$ como un efecto fijo de cada individuo y lo elimina al trabajar con las variaciones de las horas y del salario a lo largo del tiempo.

11.A.5. Gary Becker

Gary Becker, Premio Nobel en 1992, es conocido por aplicar el instrumental económico a problemas que trascienden el ámbito tradicional de la economía, como la discriminación, el crimen, la familia o la propia educación. En su conferencia Nobel resumió así los rasgos de su enfoque (Becker, 1993):

- Los individuos maximizan su bienestar "tal y como ellos lo conciben", sean egoístas, altruistas, leales, rencorosos o masoquistas. El enfoque no supone que el comportamiento sea egoísta, sino coherente con las propias preferencias.
- Su comportamiento mira hacia el futuro y es coherente a lo largo del tiempo.
- Sus decisiones están limitadas por la renta, el tiempo, la memoria y la capacidad de cálculo imperfectas y las oportunidades que ofrece la economía. La restricción más importante es el tiempo limitado.

Este enfoque conecta con el tema por dos vías. La teoría del capital humano es su aplicación más directa: la educación se decide como cualquier inversión, mirando al futuro. Además, el tiempo como recurso escaso enlaza con la elección entre ocio y consumo de la oferta de trabajo y con la asignación del tiempo dentro del hogar que se vio en el Tema 3.

11.A.6. Derivación de la ecuación de Mincer

La primera parte de la ecuación procede del modelo de escolarización. Supóngase que el único coste de estudiar es el salario que se deja de ganar mientras se estudia, sin costes directos, que la vida laboral es muy larga y que los ingresos se descuentan al tipo de interés, r, en tiempo continuo. El valor actual de los ingresos de quien estudia S años y después cobra el salario w_S es

$$V(S) = \int_S^{\infty} w_S\, e^{-rt} dt = \frac{w_S}{r}\, e^{-rS}$$

 © 2026 Manuel Cabado García

En equilibrio, los individuos deben ser indiferentes entre estudiar más o menos años, $V(S) = V(0)$, de donde $w_S = w_0 e^{rS}$ y

$$\ln w_S = \ln w_0 + rS$$

El coeficiente de la escolarización es, por tanto, el tipo de interés, que en equilibrio coincide con la tasa de rendimiento de la educación (Heckman et al., 2006).

La segunda parte procede de la formación en el empleo. Si, tras dejar los estudios, el trabajador dedica a formarse una fracción de su capacidad de ingresos, k_t, que decrece linealmente con la experiencia, t, desde k_0 hasta anularse al cabo de T años, $k_t = k_0(1 - t/T)$, y esa inversión rinde la tasa r, la capacidad de ingresos, E_t, que al terminar los estudios es E_S, crece de forma aproximadamente cuadrática con la experiencia (Mincer, 1974, cap. 5):

$$\ln E_t \approx \ln E_S + rk_0\, t - \frac{rk_0}{2T}\, t^2$$

Como el trabajador dedica a formarse la fracción k_t de su capacidad de ingresos, el salario observado es $Y_t = E_t(1 - k_t)$, es decir, $\ln Y_t = \ln E_t + \ln(1 - k_t)$. Con la aproximación $\ln(1 - k_t) \approx -k_t$, este término reduce la constante en k_0 y añade k_0/T al coeficiente de la experiencia (Mincer, 1974, p. 86). Sumando ambas partes se obtiene, de forma aproximada, la forma de la ecuación del cuerpo del tema, en la que $\beta_1 = r$, $\beta_2 = k_0(r + 1/T) > 0$ y $\beta_3 = -rk_0/(2T) < 0$. Los signos, y con ellos el perfil creciente y cóncavo, no cambian.[38]

Estos supuestos explican las críticas de Heckman et al. (2006): con costes directos, impuestos o una vida laboral finita, el coeficiente β_1 deja de coincidir con la tasa interna de rendimiento de la educación.

11.A.7. Diferencias salariales compensatorias

Las diferencias salariales compensatorias son las diferencias de salario entre empleos que exigen la misma cualificación y que compensan a los trabajadores por las características no salariales de cada empleo, como el riesgo, el esfuerzo o el horario. La idea se remonta a Adam Smith, que en *La riqueza de las naciones* (1776) ya sostuvo que los salarios varían para compensar lo agradable o desagradable de cada ocupación (Borjas, 2016, p. 196).[39]

Su formalización moderna es la función salarial hedónica, que desarrolló Sherwin Rosen en los años setenta (Rosen, 1986). Cada trabajador elige la combinación de salario y riesgo que le sitúa en la curva de indiferencia más alta, y cada empresa ofrece las combinaciones que le permiten obtener beneficios nulos en competencia. El mercado empareja así a los trabajadores que más rechazan el riesgo con las empresas a las que menos cuesta

[38] En la práctica, la ecuación suele estimarse sin el término $\ln(1 - k_t)$, y entonces $\beta_2 = rk_0$ (Cahuc et al., 2014, p. 217).

[39] La intuición es que, si dos empleos pagaran lo mismo y uno fuera más peligroso, nadie elegiría el peligroso, de modo que la empresa que lo ofrece tiene que pagar más para atraer trabajadores.

ofrecer seguridad, y la curva que une los equilibrios observados, la función hedónica, tiene pendiente positiva: los empleos más arriesgados pagan más (Borjas, 2016, pp. 206-207).

Su aplicación más conocida es el cálculo del valor estadístico de la vida. Si los trabajadores aceptan un salario algo mayor a cambio de un pequeño aumento de la probabilidad de sufrir un accidente mortal, de ese precio se deduce cuánto están dispuestos a pagar conjuntamente por evitar una muerte, un dato que se utiliza para evaluar la regulación de la seguridad laboral. La evidencia, sin embargo, solo es concluyente para el riesgo de muerte: para otras características, la estimación se ve sesgada porque los trabajadores más capaces, cuya capacidad no se observa, tienen a la vez mejores salarios y mejores condiciones de trabajo (Borjas, 2016, pp. 214-215).

Ejemplos habituales son los trabajos peligrosos, como la minería o la construcción en altura, los turnos de noche o de fin de semana y los trabajos en condiciones desagradables, como la limpieza de aguas residuales. En todos ellos la diferencia salarial no se debe a una mayor productividad, sino a un pago por los atributos negativos del empleo, lo que muestra que el mercado de trabajo valora tanto las condiciones del empleo como la productividad del trabajador.

11.A.8. La mujer en el mercado de trabajo: Goldin, la penalización por hijos y Evans

Claudia Goldin recibió en 2023 el Premio Nobel de Economía "por haber hecho avanzar nuestra comprensión de los resultados de las mujeres en el mercado de trabajo" (Real Academia Sueca de Ciencias, 2023). Su obra, y la literatura que ha seguido sobre la penalización por hijos, explican cómo se incorporó la mujer al mercado de trabajo, por qué persiste la brecha salarial y qué políticas pueden reducirla.

La curva en U y la revolución silenciosa A partir de más de doscientos años de datos de Estados Unidos, Goldin mostró que la participación de la mujer casada no ha crecido de forma continua, sino que dibuja una U. Cayó con el paso de una sociedad agraria a una industrial en el siglo XIX, cuando el trabajo se trasladó del hogar a la fábrica, y empezó a recuperarse con el crecimiento del sector servicios en el siglo XX. Detrás de esa evolución están el cambio estructural y las normas sociales sobre el papel de la mujer en el hogar.

Goldin explicó también por qué la incorporación fue lenta. Las jóvenes forman sus expectativas a partir de la experiencia de las generaciones anteriores, como la de sus madres, que no volvían a trabajar hasta que los hijos eran mayores, lo que frenaba su inversión en formación. Desde finales de los años setenta se produjo lo que Goldin (2006) llamó una "revolución silenciosa": las mujeres empezaron a esperar carreras largas y continuas, a invertir más en formación y a retrasar el matrimonio y la maternidad. La píldora an-

 © 2026 Manuel Cabado García

ticonceptiva aceleró el cambio al permitirles planificar su carrera profesional, y hoy su nivel educativo supera al de los hombres en la mayoría de los países de renta alta (Real Academia Sueca de Ciencias, 2023).

Cómo se mide la brecha: la descomposición de Oaxaca-Blinder Pese a ello, persiste una brecha salarial entre hombres y mujeres. Para saber qué parte se debe a que difieren sus características productivas y qué parte a que el mercado las retribuye de forma distinta, Oaxaca (1973) y Blinder (1973) propusieron descomponer la diferencia entre los salarios medios en logaritmos de hombres y mujeres, $\overline{\ln w}_H - \overline{\ln w}_M$, a partir de sus ecuaciones de Mincer. Sean $\bar{X}_H$ y $\bar{X}_M$ las características medias de cada grupo, como la educación o la experiencia, y $\hat{\beta}_H$ y $\hat{\beta}_M$ los rendimientos estimados de esas características:

$$\overline{\ln w}_H - \overline{\ln w}_M = \left(\bar{X}_H - \bar{X}_M\right)\hat{\beta}_H + \bar{X}_M\left(\hat{\beta}_H - \hat{\beta}_M\right)$$

El primer término es la parte explicada por las diferencias en dotaciones. El segundo es la parte no explicada, que recoge las diferencias en la retribución de unas mismas características.

En España, Murillo Huertas et al. (2017) aplican estas técnicas a la Encuesta de Estructura Salarial. En 2010 la brecha bruta era de 0,201 puntos logarítmicos, unos 20 puntos porcentuales. Con solo las características individuales, la brecha es casi toda no explicada, porque las mujeres tienen de media más formación. Al incorporar las características del puesto y del establecimiento, 0,125 puntos se explican por las diferencias de dotaciones, sobre todo por la segregación de las mujeres en establecimientos peor pagados, y 0,077 puntos siguen sin explicar.[40] Esa parte no explicada no equivale por sí sola a discriminación, porque recoge también todo lo que los datos no observan.

La penalización por hijos Una parte importante de la brecha salarial actual se da entre hombres y mujeres de la misma ocupación y surge, sobre todo, con el nacimiento del primer hijo (Real Academia Sueca de Ciencias, 2023). Es lo que la literatura denomina penalización por hijos, la caída de los ingresos laborales que sufren las madres tras tener un hijo y que no sufren los padres.[41]

Kleven et al. (2019) la midieron con datos administrativos daneses mediante un estudio de eventos: compararon los ingresos de cada persona en los años anteriores y posteriores al nacimiento de su primer hijo, de modo que el salto que se produce justo en ese momento se atribuye a la llegada del hijo. A largo plazo, el primer hijo abre una brecha de ingresos

[40] Estas cifras proceden de la descomposición de Juhn et al. (1993), una extensión de la de Oaxaca-Blinder que permite incluir efectos fijos del establecimiento (Murillo Huertas et al., 2017, tabla 4).

[41] La intuición es que, mientras los padres siguen trabajando igual que antes, las madres reducen sus horas, se pasan a empleos más compatibles con el cuidado o salen del mercado, y esa diferencia ya no se recupera. El término no implica por sí mismo discriminación: recoge la caída de ingresos sea cual sea su causa.

entre madres y padres de en torno al 20 %, que se reparte a partes casi iguales entre participación, horas trabajadas y salario por hora. Además, la parte de la desigualdad de género que se debe a esta penalización pasó de en torno al 40 % en 1980 al 80 % en 2013, a medida que otras causas de la brecha, como las diferencias educativas, desaparecían. La penalización se transmite también de padres a hijas, lo que apunta a que el entorno familiar moldea las preferencias de las mujeres sobre familia y carrera.

En España, De Quinto, Hospido y Sanz (2021), con registros de la Seguridad Social, estiman que los ingresos de las madres caen un 11 % el año siguiente al primer hijo y un 28 % a los diez años, mientras que los de los padres no cambian. El ajuste se produce con menos días trabajados, más empleo a tiempo parcial y más contratos temporales.

¿Qué causa la penalización? La evidencia de las parejas del mismo sexo Identificar la causa es difícil, porque en las parejas heterosexuales todas las explicaciones posibles predicen lo mismo, una caída de los ingresos de la madre. Andresen y Nix (2022) resuelven este problema con un modelo de oferta de trabajo del hogar y datos administrativos noruegos que incluyen parejas del mismo sexo. El modelo distingue cuatro posibles causas, y cada una predice un patrón distinto en las parejas de dos mujeres o de dos hombres:

- Especialización por ventaja comparativa: si el miembro con mayor salario se especializa en el mercado y el otro en el hogar, la penalización debería aparecer también en las parejas del mismo sexo, en el miembro con menor salario.
- Coste biológico del parto: debería penalizar solo a la mujer que da a luz, también en las parejas de dos mujeres, y solo en los primeros años.
- Preferencias de las mujeres por el cuidado de los hijos: deberían penalizar a las dos mujeres de una pareja femenina y a ninguno de los hombres de una pareja masculina.
- Normas de género sobre el cuidado: solo deberían afectar a las parejas heterosexuales, sin penalización relativa en las del mismo sexo.

En las parejas heterosexuales, las madres pierden en torno al 22 % de sus ingresos tras el primer hijo, una caída que persiste en el tiempo, mientras que los padres no pierden nada. En las parejas de dos mujeres, la que da a luz pierde al principio un 13 % y su pareja un 5 %. A los dos años, sin embargo, ambas se igualan, y a los cuatro la penalización ha desaparecido. Los ingresos del hogar heterosexual caen entre un 6 % y un 8 % de forma persistente, mientras que en las parejas de dos mujeres la caída desaparece a los cinco años (Andresen y Nix, 2019).

La versión publicada del trabajo, que añade la comparación con las parejas heterosexuales que adoptan, descarta en buena medida el parto y la ventaja salarial del padre como explicaciones, y deja como causas principales las preferencias, las normas de género y la discriminación (Andresen y Nix, 2022). De ahí se deduce que, si la penalización no

 © 2026 Manuel Cabado García

es una respuesta eficiente del hogar, sino el resultado de normas sociales, una parte de la caída de ingresos del hogar es evitable.

Implicaciones de política Los mismos autores evaluaron dos políticas en Noruega. Las ampliaciones de la cuota de permiso reservada al padre, analizadas con un diseño de regresión en discontinuidad que compara a los padres de niños nacidos justo antes y justo después de cada reforma, aumentaron los permisos que disfrutan los padres, pero no redujeron la penalización de las madres. La ampliación de las plazas de guardería subvencionada a partir de 2002, en cambio, redujo la penalización en torno a un 25 % por cada año de uso, aunque solo mientras el niño acudía a la guardería (Andresen y Nix, 2019). Para el debate sobre los permisos iguales e intransferibles, la lección es que igualar los permisos no basta por sí solo y que el cuidado infantil asequible puede ser más eficaz.

Esta evidencia enlaza con el cuerpo del tema por dos vías. Por el lado de la oferta de trabajo, la llegada de un hijo actúa sobre el margen extensivo y sobre las horas, y su efecto depende de cómo se reparte el cuidado dentro del hogar, algo que permite analizar el modelo colectivo de Chiappori (1992), en el que cada miembro tiene sus propias preferencias y las decisiones del hogar son eficientes, a diferencia del modelo unitario, que trata al hogar como si fuera un solo individuo. Por el lado del capital humano, quien espera interrumpir su carrera tiene menos incentivos a invertir en formación, sobre todo en la específica de una empresa, lo que amplía la brecha.

Emancipación femenina y crecimiento económico Evans (2026) plantea la otra cara de esta emancipación: su efecto sobre el PIB. Parte de una analogía con la abolición de la esclavitud en Estados Unidos, que elevó el producto no solo porque dejó de extraerse el fruto del trabajo de los esclavos, sino porque estos pudieron elegir en qué sectores y lugares trabajar según su propio bienestar. Del mismo modo, el patriarcado extraía de las mujeres trabajo y fecundidad, lo que la autora llama "rentas patriarcales", y su desaparición tiene dos efectos de signo contrario:

- La industrialización trajo un crecimiento a la vez extensivo e intensivo. Fue extensivo porque mujeres antes recluidas en el hogar entraron en el mercado de trabajo y aumentó el número de trabajadores. Fue intensivo porque pasaron de la producción doméstica no remunerada y de la agricultura de baja productividad a fábricas, oficinas, escuelas, hospitales y servicios, lo que elevó la productividad media sin necesidad de cambio técnico. A la inversa, la reclusión de la mujer reduce mecánicamente la fuerza de trabajo y desperdicia talento: según la autora, la India tiene tantas mujeres en edad de trabajar como China, pero trabajan muchas menos, aunque ya superan a los hombres en educación superior.
- Con la urbanización, las mujeres ganaron autonomía y elevaron lo que la autora llama su "salario de reserva" para casarse, es decir, las condiciones mínimas que

> exigen para asumir el conjunto de cuidados y tareas que tradicionalmente acompañaba al matrimonio. Es el mismo concepto que decide la entrada en el mercado de trabajo. Si muchos hombres siguen prefiriendo una relación de deferencia y tienen alternativas al matrimonio, aumentan la soltería y la caída de la fecundidad.

Una fecundidad baja envejece la población, eleva la tasa de dependencia y el ahorro y debilita el consumo interno. Japón ilustra su efecto sobre el PIB. Entre 1991 y 2019 su PIB creció un 0,83 % anual, frente al 2,53 % de Estados Unidos, pero por adulto en edad de trabajar creció un 1,39 %, frente al 1,65 % (Fernández-Villaverde et al., 2023, p. 1). El estancamiento japonés se explica, sobre todo, por la caída de su población en edad de trabajar, más que por un menor crecimiento por trabajador.

La conclusión de Evans es que la emancipación femenina mejora el bienestar de las mujeres, pero que su efecto sobre el PIB a largo plazo depende de cómo reaccionen los hombres. La evidencia de Andresen y Nix sugiere una vía para conciliar ambos objetivos: si el cuidado de los hijos se reparte de forma más igualitaria, como ocurre en las parejas del mismo sexo, tener hijos deja de suponer una pérdida permanente de ingresos para las madres.

 © 2026 Manuel Cabado García

Bibliografía

Acemoglu, D. y D. Autor (2009). Lectures in Labor Economics. Manuscrito, Massachusetts Institute of Technology (versión de 2009).

AIReF (2025, septiembre). Análisis microeconómico del impacto del salario mínimo interprofesional en el empleo. Documento Técnico 9/25, Autoridad Independiente de Responsabilidad Fiscal, Madrid.

Andresen, M. E. y E. Nix (2019). What Causes the Child Penalty? Evidence from Same Sex Couples and Policy Reforms. Discussion Papers 902, Statistics Norway, Oslo.

Andresen, M. E. y E. Nix (2022). What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples. *Journal of Labor Economics 40*(4), 971–1004.

Angrist, J. D. y A. B. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics 106*(4), 979–1014.

Ashenfelter, O. y A. B. Krueger (1994). Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review 84*(5), 1157–1173.

Azar, J., I. Marinescu y M. Steinbaum (2022). Labor Market Concentration. *Journal of Human Resources 57*(S), S167–S199.

Barceló, C., M. Izquierdo, A. Lacuesta, S. Puente, A. Regil y E. Villanueva (2021). Los efectos del salario mínimo interprofesional en el empleo: nueva evidencia para España. Documentos Ocasionales 2113, Banco de España, Madrid.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research / Columbia University Press.

Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy 101*(3), 385–409.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy 75*(4), 352–365.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources 8*(4), 436–455.

Blundell, R. y T. MaCurdy (1999). Labor Supply: A Review of Alternative Approaches. En O. Ashenfelter y D. Card (eds.), *Handbook of Labor Economics*, vol. 3A, cap. 27, pp. 1559–1695. Amsterdam: Elsevier.

Borjas, G. J. (2016). *Labor Economics* (7 ed.). New York: McGraw-Hill Education.

Cahuc, P., S. Carcillo y A. Zylberberg (2014). *Labor Economics* (2 ed.). Cambridge, MA: MIT Press.

Camerer, C., L. Babcock, G. Loewenstein y R. Thaler (1997). Labor Supply of New York City Cabdrivers: One Day at a Time. *Quarterly Journal of Economics 112*(2), 407–441.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. En L. N. Christofides, E. K. Grant y R. Swidinsky (eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pp. 201–222. Toronto: University of Toronto Press.

Card, D. (1999). The Causal Effect of Education on Earnings. En O. Ashenfelter y D. Card (eds.), *Handbook of Labor Economics*, vol. 3A, cap. 30, pp. 1801–1863. Amsterdam: Elsevier.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica 69*(5), 1127–1160.

Card, D. y A. B. Krueger (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *American Economic Review 84*(4), 772–793.

Cengiz, D., A. Dube, A. Lindner y B. Zipperer (2019). The Effect of Minimum Wages on Low-Wage Jobs. *Quarterly Journal of Economics 134*(3), 1405–1454.

Chetty, R., A. Guren, D. Manoli y A. Weber (2011). Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins. *American Economic Review 101*(3), 471–475.

Chiappori, P.-A. (1992). Collective Labor Supply and Welfare. *Journal of Political Economy 100*(3), 437–467.

De Quinto, A., L. Hospido y C. Sanz (2021). The Child Penalty: Evidence from Spain. *SERIEs 12*(4), 585–606.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review 91*(4), 795–813.

Eurostat (2026). Early Leavers from Education and Training by Sex and Labour Status (edat_lfse_14). Base de datos de Eurostat, datos de 2025.

 © 2026 Manuel Cabado García

Evans, A. (2026). How does Female Emancipation affect GDP? The Great Gender Divergence (Substack), 31 de mayo.

Farber, H. S. (2005). Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers. *Journal of Political Economy 113*(1), 46–82.

Farber, H. S. (2015). Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers. *Quarterly Journal of Economics 130*(4), 1975–2026.

Fernández-Villaverde, J., G. Ventura y W. Yao (2023). The Wealth of Working Nations. NBER Working Paper 31914, National Bureau of Economic Research, Cambridge, MA. Revisado en agosto de 2024. Publicado en *European Economic Review* 173 (2025).

Frisch, R. (1932). *New Methods of Measuring Marginal Utility*. núm. 3 de Beiträge zur ökonomischen Theorie. Tübingen: J. C. B. Mohr (Paul Siebeck).

Goldin, C. (2006). The Quiet Revolution That Transformed Women's Employment, Education, and Family. *American Economic Review 96*(2), 1–21.

Gruber, J. y D. A. Wise (eds.) (1999). *Social Security and Retirement around the World*. Chicago: University of Chicago Press.

Hamermesh, D. S. (1993). *Labor Demand*. Princeton: Princeton University Press.

Heckman, J. J., L. J. Lochner y P. E. Todd (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. En E. A. Hanushek y F. Welch (eds.), *Handbook of the Economics of Education*, vol. 1, cap. 7, pp. 307–458. Amsterdam: Elsevier.

Hicks, J. R. (1932). *The Theory of Wages*. London: Macmillan.

INE (2026). Encuesta de Población Activa. Tasas de actividad por sexo y grupo de edad (tabla 65081). Instituto Nacional de Estadística, datos del segundo trimestre de 2026.

Juhn, C., K. M. Murphy y B. Pierce (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy 101*(3), 410–442.

Kleven, H., C. Landais y J. E. Søgaard (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

Koutsoyiannis, A. (1979). *Modern Microeconomics* (2 ed.). London: Macmillan.

Kydland, F. E. y E. C. Prescott (1982). Time to Build and Aggregate Fluctuations. *Econometrica 50*(6), 1345–1370.

Lucas, Jr., R. E. y L. A. Rapping (1969). Real Wages, Employment, and Inflation. *Journal of Political Economy 77*(5), 721–754.

MaCurdy, T. E. (1981). An Empirical Model of Labor Supply in a Life-Cycle Setting. *Journal of Political Economy 89*(6), 1059–1085.

Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton: Princeton University Press.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy 66*(4), 281–302.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research / Columbia University Press.

Murillo Huertas, I. P., R. Ramos y H. Simón (2017). Regional Differences in the Gender Wage Gap in Spain. *Social Indicators Research 134*(3), 981–1008.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review 14*(3), 693–709.

OCDE (2025). *Education at a Glance 2025: OECD Indicators*. París: OECD Publishing.

Pigou, A. C. (1932). *The Economics of Welfare* (4 ed.). London: Macmillan.

Psacharopoulos, G. y H. A. Patrinos (2018). Returns to Investment in Education: A Decennial Review of the Global Literature. *Education Economics 26*(5), 445–458.

Real Academia Sueca de Ciencias (2023). The Prize in Economic Sciences 2023: She uncovered key drivers of gender differences in the labour market. Nota de prensa, 9 de octubre.

Robbins, L. (1930). On the Elasticity of Demand for Income in Terms of Effort. *Economica* (29), 123–129.

Robinson, J. (1933). *The Economics of Imperfect Competition*. London: Macmillan.

Romer, D. (2019). *Advanced Macroeconomics* (5 ed.). New York: McGraw-Hill Education.

Rosen, S. (1986). The Theory of Equalizing Differences. En O. Ashenfelter y R. Layard (eds.), *Handbook of Labor Economics*, vol. 1, cap. 12, pp. 641–692. Amsterdam: North-Holland.

Schultz, T. W. (1961). Investment in Human Capital. *American Economic Review 51*(1), 1–17.

 © 2026 Manuel Cabado García

Spence, M. (1973). Job Market Signaling. *Quarterly Journal of Economics 87*(3), 355–374.

Stock, J. H. y D. A. Wise (1990). Pensions, the Option Value of Work, and Retirement. *Econometrica 58*(5), 1151–1180.

 © 2026 Manuel Cabado García

Tema 12

Teorías microeconómicas del mercado de trabajo (II). La curva de Beveridge, el modelo de búsqueda y emparejamiento y costes de ajuste. Mención a los modelos de negociación, salarios de eficiencia y contratos implícitos.

12.1. Introducción ⏳ 1'

En el primer trimestre de 2026, la tasa de paro en España era del 10,2 %, y sin embargo había unas 158.000 vacantes, 0,63 por cada cien activos, que las empresas no conseguían cubrir.[1] Esa coexistencia, que se repite en todas las economías, es imposible en el modelo neoclásico, en el que el salario se ajusta hasta vaciar el mercado y el único desempleo es voluntario.[2]

Las teorías microeconómicas que se estudiarán nacen de abandonar tres supuestos de ese modelo: la información perfecta, el ajuste de la plantilla sin coste y el salario como mecanismo de vaciado del mercado.

Por consiguiente, el objetivo de la presente exposición es analizar por qué el mercado laboral no se ajusta a las predicciones del modelo competitivo a través del siguiente esquema:

[1] Datos corregidos de variaciones estacionales y efecto calendario (Felgueroso et al., 2026, pp. 3-4).

[2] El modelo neoclásico de oferta y demanda de trabajo, que aquí se da por conocido, se desarrolla en el Tema 11 de la Parte A.

1. La curva de Beveridge y el modelo de búsqueda y emparejamiento, que explican el desempleo friccional.
2. Los costes de ajuste, que convierten la demanda de trabajo en un problema dinámico.
3. Una mención a las teorías de la rigidez salarial: modelos de negociación, salarios de eficiencia y contratos implícitos.

12.2. Desempleo friccional: la curva de Beveridge y el modelo de búsqueda y emparejamiento ⏳ 10'

Si la información fuera perfecta y los intercambios instantáneos, ningún puesto quedaría vacante mientras hubiera un parado dispuesto a ocuparlo. Al abandonar ese supuesto aparece el desempleo friccional.

Desempleo friccional El desempleo friccional es el que se debe al tiempo que tardan trabajadores y empresas en encontrarse y en formalizar una relación laboral, aunque haya puestos disponibles para todos ellos.[3]

La curva de Beveridge

La relación entre paro y vacantes lleva el nombre de William Beveridge, que ya en 1944 advertía que el pleno empleo no podía significar un paro nulo, porque siempre habría personas cambiando de un trabajo a otro, y lo definía como la situación en la que hay "más empleos vacantes que hombres parados, y no un poco menos" (Beveridge, 1944, p. 18). Sin embargo, quien primero la representó no fue él, sino Dow y Dicks-Mireaux (1958), con datos británicos de 1946 a 1956 (Elsby et al., 2015, p. 571).

Curva de Beveridge La curva de Beveridge es la relación decreciente y convexa que se observa a lo largo del ciclo entre la tasa de paro, u, y la tasa de vacantes, v, ambas medidas en proporción a la población activa.[4] El cociente entre ambas, la tensión del mercado de trabajo, $\theta = v/u$, indica cuántas vacantes hay por cada parado.

Gráficamente, en un plano de tasa de paro, u, y tasa de vacantes, v, como el de la Figura 12.1, se representan dos curvas de Beveridge, BC_0 y BC_1, y la bisectriz, $u = v$. Así:

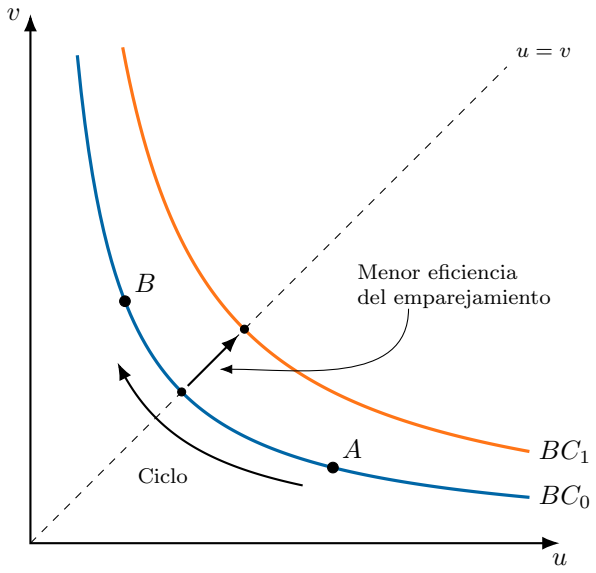
[3] La intuición es que buscar empleo, o buscar a quien contratar, lleva tiempo y cuesta recursos, porque la información sobre qué puestos existen, qué exigen y qué pagan está dispersa y no llega a todos a la vez. Mientras dura esa búsqueda, el trabajador figura como parado y el puesto como vacante.

[4] La intuición es que, cuando la actividad se acelera, las empresas abren más puestos y una parte de ellos se cubre con parados, de modo que las vacantes suben mientras el paro baja. La convexidad refleja que, en un mercado ya muy tenso, cada vacante adicional encuentra menos parados con los que emparejarse y reduce el paro cada vez menos.

 © 2026 Manuel Cabado García

- Los cambios en la actividad desplazan la economía a lo largo de una misma curva: en las recesiones las empresas abren pocas vacantes y el paro es alto, como en el punto A, y en las expansiones ocurre lo contrario, como en el punto B.
- Los desplazamientos de la curva reflejan cambios en la eficiencia con la que se emparejan parados y vacantes o en el ritmo al que se destruyen empleos. Cuanto más alejada del origen está, como BC_1 frente a BC_0, más paro coexiste con cada nivel de vacantes, lo que suele interpretarse como un deterioro estructural del emparejamiento por desajustes de cualificación, sectoriales o geográficos (Cahuc et al., 2014, pp. 574-576).[5]
- Dow y Dicks-Mireaux leían la curva en términos de exceso de demanda de trabajo, y en esa lectura su corte con la bisectriz se toma a menudo como aproximación al paro friccional, porque en él hay una vacante por cada parado. El pleno empleo de Beveridge exigía situarse por encima de ella.

Figura 12.1: Curva de Beveridge



Implicaciones de política económica La distinción tiene una implicación de política directa. Si la economía se mueve a lo largo de la curva, el paro es cíclico y responde a las políticas de demanda. Si la curva se desplaza hacia fuera, el problema está en el emparejamiento y lo que procede son políticas activas de empleo, como la intermediación de los servicios públicos de empleo, la formación de los parados o el fomento de la movilidad geográfica.

[5] No todo desplazamiento aparente es estructural. Tras una recesión, la economía suele recorrer un bucle en sentido antihorario alrededor de la curva, porque entre la apertura de una vacante y la contratación pasa tiempo. Por eso Lazear y Spletzer consideraban prematuro atribuir a desajustes el desplazamiento observado en Estados Unidos tras 2009 (Cahuc et al., 2014, p. 576).

Evidencia empírica española La experiencia española ilustra ambos movimientos. Con una serie homogénea de vacantes desde 1980, Boscá et al. (2017) muestran que la curva se desplazó hacia fuera en 2008 y 2009 por una pérdida de eficiencia del emparejamiento, y que desde 2014 ha ido recuperando gradualmente su posición anterior a la crisis. Aun así, el mercado español sigue lejos del pleno empleo de Beveridge. Desde mediados de 2022 la tasa de vacantes oscila en torno a 0,6 por cada cien activos con una tasa de paro superior al 10 %, es decir, unos dieciséis parados por cada vacante, y los parados de larga duración, que son los que peor se emparejan, siguen siendo 900.000, el 38 % del total, aunque su número casi se ha reducido a la mitad desde 2021 (Felgueroso et al., 2026, pp. 3-4).

Ahora bien, la curva solo describe la relación entre paro y vacantes, no la explica. ¿Por qué un parado no acepta la primera vacante que encuentra? La respuesta está en los modelos de búsqueda.

Los modelos de búsqueda

La respuesta empezó a construirse con George Stigler, que denunció que la teoría económica apenas prestaba atención a la información (Stigler, 1961, p. 213). Su punto de partida es una observación sencilla: un mismo trabajo se paga de forma distinta en empresas distintas, porque nadie conoce todas las ofertas, y encontrar la mejor exige buscar, lo que cuesta tiempo y dinero. En el mercado de trabajo, el parado decide de antemano cuántas ofertas examinar, tantas como sea necesario para que lo que espera ganar con una oferta más iguale lo que le cuesta conseguirla (Stigler, 1962).[6]

Una década después, McCall (1970) y Mortensen (1970) sustituyeron esa regla por otra más realista, la búsqueda secuencial. El parado recibe ofertas de una en una, a una tasa a, con salarios, w, extraídos de una distribución conocida, $H(w)$, y ante cada oferta decide si la acepta o sigue buscando. La estrategia óptima consiste en aceptar cualquier oferta que supere un umbral, el salario de reserva, y rechazar el resto (Cahuc et al., 2014, p. 262).

Salario de reserva El salario de reserva, w_R, es el salario mínimo con el que a un parado le compensa aceptar un empleo en lugar de seguir buscando.[7] Si se llama z a la renta neta del parado, que incluye la prestación y descuenta los costes de buscar, r al tipo

[6] Esa estrategia no secuencial no es óptima, porque puede llevar a seguir buscando después de haber recibido una oferta superior al salario de reserva. McCall demostró la optimalidad de la regla de parada secuencial (Cahuc et al., 2014, p. 262).

[7] La intuición es que aceptar una oferta tiene un coste de oportunidad, que es renunciar a encontrar otra mejor. El salario de reserva es el punto en el que ese coste se iguala con lo que se gana al dejar de buscar.

 © 2026 Manuel Cabado García

de interés y λ a la tasa a la que se destruyen los empleos, el salario de reserva cumple

$$w_R = z + \frac{a}{r+\lambda} \cdot \int_{w_R}^{\infty} (w - w_R)\, dH(w) \tag{12.1}$$

donde el último término es el valor esperado y descontado de lo que una nueva búsqueda puede aportar por encima del salario de reserva (Cahuc et al., 2014, p. 264). Como el parado sale del paro cuando recibe una oferta y esta supera su umbral, la duración media del paro, $T_u = 1/\{a \cdot [1 - H(w_R)]\}$, es mayor cuanto más alto es el salario de reserva.[8]

Implicaciones de política económica Una prestación más generosa eleva z y, con él, el salario de reserva y la duración del paro, de modo que el seguro de desempleo genera un problema de riesgo moral.[9] Por eso su diseño combina la protección de la renta con incentivos a aceptar empleo, como una cuantía decreciente con la duración o la exigencia de búsqueda activa. En España, la prestación contributiva se paga al 70 % de la base reguladora durante los primeros 180 días y al 60 % a partir del día 181.[10]

Paradoja de Diamond La paradoja de Diamond es el resultado según el cual, si las empresas fijan los salarios y los trabajadores tienen costes de búsqueda, todas las empresas pagan el salario de reserva y desaparece la dispersión salarial que justificaba buscar (Diamond, 1971).[11] El problema de fondo es que la distribución de salarios, $H(w)$, se toma como dada, y deja de ser coherente en cuanto se modeliza el comportamiento de las empresas (Cahuc et al., 2014, pp. 304-306).

Las respuestas siguieron dos caminos: introducir heterogeneidad o búsqueda desde el empleo (Burdett y Mortensen, 1998) en los modelos en los que la empresa anuncia el salario, o suponer que el salario se negocia una vez que trabajador y empresa se han encontrado.[12] Este segundo camino, en el que las empresas también buscan y el salario se negocia, es el de los modelos de emparejamiento, que culminan en el modelo DMP.

[8] La derivación completa del salario de reserva y de la duración media del paro figura en el Apéndice 12.A.1.

[9] El riesgo moral se estudia en el Tema 13 de la Parte A. La predicción es teóricamente robusta, pero su magnitud es empírica: muchos parados no cobran prestación, y para ellos una prestación más generosa puede reducir el salario de reserva, porque aceptar un empleo da derecho a cobrarla en el futuro (Cahuc et al., 2014, p. 265).

[10] Artículo 270.2 del texto refundido de la Ley General de la Seguridad Social, en la redacción dada por la Ley 31/2022, de 23 de diciembre, de Presupuestos Generales del Estado para 2023, en vigor desde el 1 de enero de 2023.

[11] La intuición es que ninguna empresa gana nada pagando más que el salario de reserva, porque cualquier oferta que lo iguale ya se acepta. Si todas pagan lo mismo, no hay ninguna oferta mejor que buscar, y el salario cae hasta la renta del parado, z.

[12] En el modelo de Burdett y Mortensen los trabajadores siguen buscando cuando ya tienen empleo, de modo que las empresas compiten por atraerlos y retenerlos. En equilibrio surge una distribución de salarios no degenerada, en la que las empresas que pagan más son también las más grandes (Cahuc et al., 2014, p. 306).

El modelo de búsqueda y emparejamiento de Diamond, Mortensen y Pissarides

Diamond, Mortensen y Pissarides, que recibieron el Premio Nobel de Economía de 2010 por su análisis de los mercados con fricciones de búsqueda, llevaron la búsqueda al equilibrio del mercado de trabajo: buscan a la vez los parados y las empresas con vacantes, y el salario se negocia una vez que se encuentran (Diamond, 1982; Pissarides, 1985). En la versión básica de Pissarides (2000, cap. 1), trabajadores y empresas son idénticos, cada empresa tiene un solo puesto, que produce p cuando está ocupado y cuesta $p \cdot c$ por unidad de tiempo mientras está vacante, y los empleos se destruyen a una tasa exógena, λ.[13]

Función de emparejamiento La función de emparejamiento es la relación que da el número de contratos que se firman por unidad de tiempo en función del número de parados y de vacantes, $m = m(u, v)$.[14] Se supone creciente en ambos argumentos, cóncava y con rendimientos constantes a escala, y la evidencia la respalda: Petrongolo y Pissarides (2001, p. 393) sitúan su elasticidad respecto al paro, η, entre 0,5 y 0,7. Con rendimientos constantes, todo depende de la tensión, θ. Una vacante se cubre a la tasa $q(\theta) = m/v$, decreciente en θ, y un parado encuentra empleo a la tasa $\theta \cdot q(\theta) = m/u$, creciente en θ. Cada nueva vacante facilita así que los parados encuentren empleo, pero dificulta que se cubran las demás vacantes. Son las externalidades de congestión del modelo.

El equilibrio queda determinado por tres relaciones.

Curva de Beveridge En estado estacionario, los empleos que se destruyen, $\lambda \cdot (1 - u)$, igualan a los parados que encuentran empleo, $\theta \cdot q(\theta) \cdot u$, de donde

$$u = \frac{\lambda}{\lambda + \theta \cdot q(\theta)} \tag{12.2}$$

que es la curva del apartado anterior, ahora derivada de los flujos: el paro es mayor cuanto más alta es la tasa de destrucción y cuanto menos eficiente es el emparejamiento.

Curva de creación de empleo Las empresas abren vacantes mientras les resulte rentable, y con libre entrada el valor de una vacante se anula. Entonces, el coste esperado de cubrir un puesto, $p \cdot c/q(\theta)$, que es el coste de la vacante por su duración media, debe igualar el valor de un puesto ocupado, que es el beneficio, $p - w$, descontado al tipo de

[13] Mortensen y Pissarides (1994) endogeneizaron después la destrucción de empleo, que pasa a depender de perturbaciones de productividad específicas de cada puesto, de modo que la creación y la destrucción de empleo se mueven en sentido opuesto a lo largo del ciclo.

[14] La intuición es que funciona como una función de producción de contratos cuyos factores son los parados y las vacantes. Resume en una sola relación todas las fricciones del mercado, como la información incompleta o los desajustes de cualificación y de localización, sin modelizarlas una a una.

 © 2026 Manuel Cabado García

interés más la tasa de destrucción:

$$\frac{p \cdot c}{q(\theta)} = \frac{p - w}{r + \lambda} \tag{12.3}$$

La relación entre salario y tensión es decreciente, porque un salario más alto hace menos rentable cada empleo y se abren menos vacantes. Desempeña el papel de la demanda de trabajo.

Curva salarial Empresa y trabajador se reparten el excedente que genera el emparejamiento mediante la solución de Nash, en la que el poder de negociación del trabajador es β, y el resultado es

$$w = (1 - \beta) \cdot z + \beta \cdot p \cdot (1 + c \cdot \theta) \tag{12.4}$$

El salario crece con la tensión porque, cuanto más fácil es encontrar otro empleo, mejor es la posición negociadora del trabajador.[15] Con un poder de negociación nulo, el salario se iguala a la renta del parado, z, como en la paradoja de Diamond.[16]

Gráficamente, en dos planos, uno de tensión, θ, y salario, w, y otro de tasa de paro, u, y tasa de vacantes, v, como los de la Figura 12.2, se representan la curva de creación de empleo, CCE, y la curva salarial, CS, en el primero, y la curva de Beveridge, BC, y el rayo $v = \theta^* \cdot u$ en el segundo. Así:

- El corte de la curva de creación de empleo con la curva salarial determina la tensión, θ^*, y el salario, w^*, de equilibrio.
- El rayo $v = \theta^* \cdot u$ corta la curva de Beveridge en las tasas de paro y de vacantes de equilibrio, u^* y v^*.
- El paro de equilibrio es positivo aunque el salario sea plenamente flexible, porque es desempleo friccional en sentido estricto.[17]

Implicaciones de política económica Una prestación más generosa, que eleva z, desplaza hacia arriba la curva salarial, reduce la tensión y aumenta el paro, y lo mismo ocurre si aumenta el poder de negociación de los trabajadores (Cahuc et al., 2014, pp. 597-599). Un aumento de la productividad eleva el salario, pero eleva más la rentabilidad de crear empleo, y el paro baja.[18] Un emparejamiento menos eficiente o una destrucción

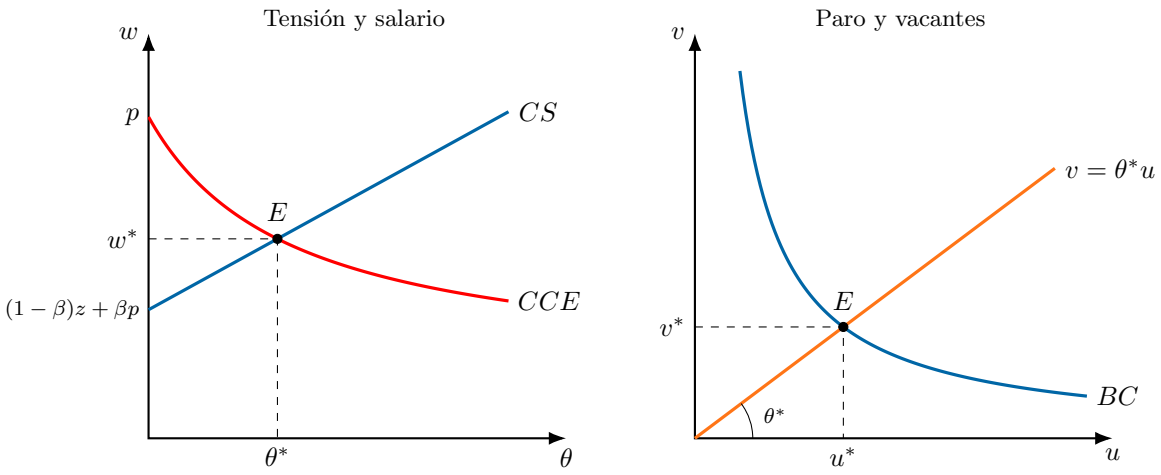
[15] La intuición es que, en un mercado tenso, el trabajador que rompe la negociación encuentra pronto otro empleo, mientras que la empresa tendría que volver a buscar. El término $p \cdot c \cdot \theta$ es precisamente el coste medio de contratación por cada parado, y el trabajador se queda con una parte de ese ahorro proporcional a su poder de negociación.

[16] La derivación de las tres relaciones y del equilibrio figura en el Apéndice 12.A.2.

[17] El paro de equilibrio del modelo DMP es uno de los fundamentos de la tasa natural de paro, que se estudia en el Tema 11 de la Parte B.

[18] El resultado depende de que la renta del parado, z, no crezca con la productividad. Si z y el coste de la vacante crecieran en proporción a los salarios, el nivel de productividad no afectaría al paro a largo plazo (Cahuc et al., 2014, p. 599).

Figura 12.2: Equilibrio del modelo DMP



de empleo más rápida alejan la curva de Beveridge del origen y elevan el paro, que es lo que tratan de evitar las políticas activas de empleo.

Condición de Hosios La condición de Hosios es la que hace eficiente el equilibrio del modelo: el poder de negociación de los trabajadores debe coincidir con la elasticidad de la función de emparejamiento respecto al paro, $\beta = \eta$ (Hosios, 1990).[19] Si los trabajadores tienen más poder del que indica esa elasticidad, las empresas abren pocas vacantes y el paro es mayor que el socialmente óptimo, y si tienen menos, ocurre lo contrario (Cahuc et al., 2014, p. 602). Nada garantiza, por tanto, que el paro de equilibrio sea el eficiente, lo que abre la puerta a la intervención pública.[20]

Evidencia empírica El modelo DMP es hoy el marco de referencia del análisis del mercado de trabajo, y su contraste empírico más conocido es también su crítica más seria. Shimer (2005) mostró que en Estados Unidos la desviación típica de la tensión, v/u, es casi veinte veces la de la productividad, mientras que el modelo con salarios negociados predice volatilidades parecidas.[21] Es el puzle de la volatilidad del paro, y una

[19] La intuición es que cada vacante y cada parado que entran en el mercado congestionan a los de su mismo lado y facilitan la búsqueda a los del otro, y nadie tiene en cuenta esos efectos al decidir. El reparto del salario solo los compensa cuando lo que obtiene cada parte coincide con lo que aporta al emparejamiento, que es lo que mide la elasticidad.

[20] Cahuc et al. (2014, pp. 600-603) la llaman condición de Hosios-Pissarides. La ineficiencia es propia de la búsqueda aleatoria con salarios negociados: si los trabajadores pueden dirigir su búsqueda hacia los salarios que anuncian las empresas, como en el modelo de Moen, la competencia por atraerlos restaura la eficiencia del equilibrio. La condición supone, además, que la renta del parado solo recoge el valor del ocio. Si incluye una prestación, que para la sociedad es una mera transferencia, el equilibrio genera demasiado paro aunque se cumpla $\beta = \eta$.

[21] La desviación típica de la tensión es de 0,38 en los datos y de 0,035 en el modelo calibrado. Además, el modelo predice una elasticidad del salario a la productividad prácticamente unitaria, frente a unos 0,45 en las estimaciones de Hagedorn y Manovskii. Pissarides objeta, sin embargo, que los salarios de los nuevos contratos sí son flexibles (Cahuc et al., 2014, pp. 613-616).

 © 2026 Manuel Cabado García

de las respuestas más extendidas, desde Hall (2005), ha sido introducir rigidez salarial en el modelo (Cahuc et al., 2014, pp. 613-616). En España, el modelo ayuda a entender el peso de la tasa de destrucción, λ: las finalizaciones de contrato explican dos tercios de las salidas del empleo hacia el paro o la inactividad, y junto con los despidos, en torno al 90 % (Felgueroso et al., 2026, p. 4).

Ahora bien, el modelo supone que, una vez encontrado el trabajador, el empleo se crea o se destruye al instante y sin coste. En la realidad, contratar y despedir cuesta, y esos costes cambian por completo la forma en que se ajusta el empleo.

12.3. Costes de ajuste y dinámica de la demanda de trabajo ⏳ 4'

El modelo neoclásico supone que la empresa ajusta su plantilla de forma instantánea, de modo que el empleo responde de inmediato a cualquier cambio del salario o de la productividad. Con costes de ajuste, la demanda de trabajo se convierte en un problema dinámico: la empresa no solo decide cuánto empleo quiere, sino a qué ritmo llegar a él.

Costes de ajuste Los costes de ajuste son los costes en que incurre la empresa al variar su plantilla, como los de selección, contratación y formación de los nuevos trabajadores o las indemnizaciones por despido.[22] Se distinguen los costes internos, que son pérdidas de eficiencia mientras la empresa se reorganiza, de los externos, como los honorarios de una agencia o las indemnizaciones, y los costes brutos, ligados a cada entrada y salida de trabajadores, de los netos, ligados a la variación neta del empleo (Cahuc et al., 2014, pp. 119-120).[23]

La idea se remonta a Oi (1962), que trató el trabajo como un factor cuasifijo. Si contratar y formar a un trabajador exige un coste fijo, la empresa lo amortiza a lo largo de la relación laboral y mantiene al trabajador mientras su productividad marginal cubra el salario, aunque no cubra el coste total. Los trabajadores más cualificados, en cuya formación la empresa ha invertido más, sufren por eso menos despidos en las recesiones.

Formalmente, la empresa elige la senda del empleo, L_t, que maximiza el valor descontado de sus beneficios, que son la producción, $F(L_t)$, menos los salarios, $w \cdot L_t$, y menos

[22] La intuición es que incorporar a un trabajador no termina al firmar el contrato: hay que buscarlo, seleccionarlo y formarlo, y desprenderse de él también cuesta. Cambiar de plantilla tiene así un precio propio, que se suma al salario.

[23] Su cuantía varía mucho entre países. En Estados Unidos, reponer a un trabajador cuesta entre el 25 % del salario anual, en los menos cualificados, y más del 100 %, en los más cualificados, y contratar es más caro que despedir. En Francia ocurre lo contrario: según Abowd y Kramarz, una separación cuesta en promedio el 56 % del coste laboral anual y una contratación el 3,3 % (Cahuc et al., 2014, p. 120).

el coste de ajuste, $C(\dot{L}_t)$, que depende de la variación del empleo, $\dot{L}_t$:

$$\max_{\{L_t\}} \int_0^\infty e^{-rt} \cdot \left[F(L_t) - w \cdot L_t - C(\dot{L}_t)\right] dt$$

La dinámica del empleo depende de la forma de C. Con costes cuadráticos, la forma que introdujeron Holt et al. (1960) y que suele usarse en su versión simétrica, $C = (b/2) \cdot \dot{L}_t^2$, donde b mide la intensidad del coste (Cahuc et al., 2014, p. 121), ajustar deprisa es más caro que ajustar poco a poco, y la empresa reparte el ajuste en el tiempo: el empleo converge gradualmente a su nivel óptimo. Este caso explica que el empleo reaccione con retraso a la producción y que la productividad aparente del trabajo sea procíclica, porque en las recesiones las empresas retienen mano de obra que no necesitan.

Los costes cuadráticos implican, sin embargo, ajustes pequeños y continuos, mientras que los datos muestran ajustes bruscos alternados con periodos sin cambios. Por eso la literatura posterior prefiere costes lineales y asimétricos, con un coste por cada contratación, c_h, y otro por cada despido, c_f (Nickell, 1986). La empresa contrata hasta que la productividad marginal iguala el salario más el coste de contratación anualizado, y despide hasta que la iguala el salario menos el coste de despido anualizado:

$$F'(L_h) = w + r \cdot c_h \qquad F'(L_f) = w - r \cdot c_f$$

Gráficamente, en un plano de empleo, L, y productividad marginal y salario, $F'(L)$ y w, como el de la Figura 12.3, se representan la productividad marginal del trabajo, $F'(L)$, el salario, w, con el que se contrataría sin costes hasta L^*, y los umbrales $w + r \cdot c_h$ y $w - r \cdot c_f$.[24] En él:

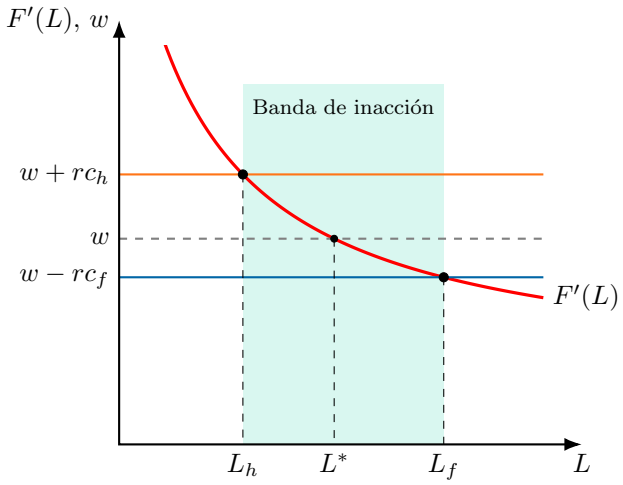
- Si el empleo de partida es menor que L_h, la productividad marginal supera el umbral de contratación y la empresa contrata hasta L_h.
- Si es mayor que L_f, la productividad marginal queda por debajo del umbral de despido y la empresa despide hasta L_f.
- Entre L_h y L_f se abre una banda de inacción: a la empresa no le compensa modificar el empleo, que se queda donde estaba (Cahuc et al., 2014, pp. 126-128).

Implicaciones de política económica Los costes de despido estabilizan el empleo, pero no tienen un efecto claro sobre su nivel medio. En un entorno incierto, al encarecer el despido la empresa despide menos en las recesiones, pero también contrata menos en las expansiones, porque anticipa que más adelante tendrá que despedir. Bentolila y Bertola (1990) mostraron que el primer efecto domina ligeramente, de modo que los costes de despido pueden incluso elevar algo el empleo medio a largo plazo, aunque reducen la rotación de trabajadores (Cahuc et al., 2014, pp. 133-135).

[24] La derivación de estos umbrales y de la demanda de trabajo resultante figura en el Apéndice 12.A.3.

 © 2026 Manuel Cabado García

Figura 12.3: Banda de inacción con costes de ajuste lineales y asimétricos



Evidencia empírica española El caso español muestra que lo decisivo puede ser la diferencia de costes entre contratos. En España, el despido improcedente de un trabajador indefinido se indemniza con 33 días de salario por año trabajado, con un máximo de 24 mensualidades, mientras que la extinción de un contrato temporal da derecho a 12 días por año.[25] Con esa dualidad, el ajuste recae sobre los temporales. Bentolila et al. (2012) estiman, con un modelo de búsqueda calibrado, que España habría evitado alrededor del 45 % del aumento del paro durante la Gran Recesión si hubiera tenido la legislación francesa de protección del empleo: el paro, del 8 % en ambos países antes de la crisis, subió al 10 % en Francia y al 23 % en España. La reforma laboral de 2021 atacó esa dualidad limitando el contrato temporal a las circunstancias de la producción y a la sustitución de trabajadores, y fomentando el contrato fijo discontinuo.[26] La tasa de temporalidad de los asalariados ha bajado del 25,4 % a finales de 2021 al 15,1 % en el segundo trimestre de 2026.[27] Desde finales de 2019, el empleo indefinido ha crecido un 31,4 % y el temporal ha caído un 33,2 % (Felgueroso et al., 2026, p. 2).

Hasta aquí, el salario se ha tomado como dado o como resultado de una negociación sin fricciones. Queda por explicar por qué, cuando hay paro, no baja lo suficiente para vaciar el mercado.

[25] Artículos 56.1 y 49.1.c del texto refundido de la Ley del Estatuto de los Trabajadores. La cuantía del despido improcedente procede de la reforma de 2012 (Real Decreto-ley 3/2012 y Ley 3/2012). El despido objetivo procedente se indemniza con 20 días por año y un máximo de 12 mensualidades (artículo 53.1.b).

[26] Real Decreto-ley 32/2021, de 28 de diciembre, de medidas urgentes para la reforma laboral, la garantía de la estabilidad en el empleo y la transformación del mercado de trabajo.

[27] INE, Encuesta de Población Activa del cuarto trimestre de 2021, y Ministerio de Trabajo y Economía Social, a partir de la Encuesta de Población Activa del segundo trimestre de 2026. El Banco de España advierte de que ese descenso no puede atribuirse sin más a la reforma, y de que parte de la rotación se ha trasladado a los fijos discontinuos, cuya tasa de bajas casi se ha duplicado desde 2022 (Banco de España, *Informe Anual 2023*, capítulo 3).

12.4. Mención a las teorías de la rigidez salarial ⏳ 4'30"

Las respuestas a esa pregunta comparten una idea: el salario deja de ser un precio que se ajusta libremente. Lo negocian empresa y sindicato, lo fija la empresa por encima del de mercado para que sus trabajadores rindan más, o se pacta de forma implícita como un seguro. Varias de ellas generan, además, paro involuntario: trabajadores dispuestos a trabajar al salario vigente, o incluso a uno menor, que no encuentran empleo.

Modelos de negociación

La teoría de la negociación salarial parte de Hicks (1932), que explicaba el salario pactado por la capacidad de cada parte para aguantar una huelga. La curva de concesión de la empresa, el salario que está dispuesta a aceptar, crece con la duración de la huelga, y la de resistencia del sindicato, el salario que exige, decrece con ella. El salario se fija en su cruce y, si ambas partes lo prevén, la huelga no llega a producirse (Cahuc et al., 2014, pp. 414-415).

Nash (1950) dio después a la negociación su solución axiomática, que el juego de ofertas alternativas de Rubinstein (1982) permitió fundamentar como resultado de un comportamiento estratégico (Cahuc et al., 2014, pp. 416-422).[28]

En su versión generalizada, aplicada a un sindicato y a una empresa, la solución maximiza el producto de lo que gana cada parte respecto a su situación sin acuerdo, ponderado por su poder de negociación.[29] Llamando $V(w, L)$ a la utilidad del sindicato, $\pi(w, L)$ al beneficio de la empresa, $\bar{V}$ y $\bar{\pi}$ a sus valores sin acuerdo y β al poder de negociación del sindicato:

$$\text{máx} \left[V(w, L) - \bar{V}\right]^{\beta} \cdot \left[\pi(w, L) - \bar{\pi}\right]^{1-\beta}$$

Los modelos de negociación colectiva aplican esta solución y se distinguen por lo que se negocia.

[28] Cuando el intervalo entre ofertas tiende a cero, el equilibrio del juego de Rubinstein converge a la solución de Nash generalizada, en la que el poder de negociación de cada parte es mayor cuanto más paciente es, como mostraron Binmore, Rubinstein y Wolinsky. El punto de desacuerdo es entonces lo que cada parte obtiene mientras dura el conflicto, y no sus opciones externas (Cahuc et al., 2014, pp. 421-422).

[29] Nash exigió a la solución cuatro propiedades. La eficiencia en el sentido de Pareto obliga a que las partes aprovechen todas las ganancias mutuas. La invariancia ante transformaciones afines crecientes de las utilidades hace que el resultado no dependa de cómo se representen las preferencias. La independencia de las alternativas irrelevantes garantiza que, si se descartan acuerdos que de todos modos no se habrían elegido, el resultado no cambia. Por último, la simetría supone que dos partes intercambiables obtienen lo mismo, es decir, que tienen igual poder de negociación. La única solución que cumple las cuatro maximiza el producto de las ganancias netas de las partes. Al suprimir la simetría se obtiene la solución generalizada, en la que el exponente β mide el poder de negociación (Cahuc et al., 2014, pp. 415-416).

 © 2026 Manuel Cabado García

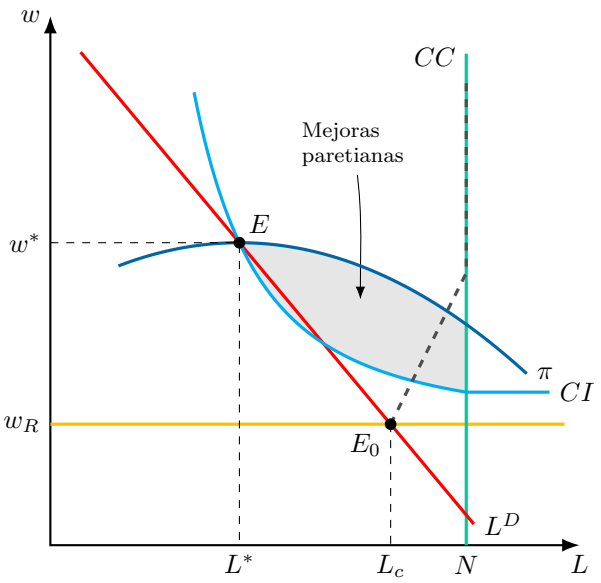
Modelo de derecho a gestionar En el modelo de derecho a gestionar de Nickell y Andrews (1983), el sindicato y la empresa negocian solo el salario, y la empresa elige después el empleo sobre su curva de demanda. Con todo el poder de negociación en manos del sindicato, el modelo se reduce al sindicato monopolista de Dunlop (1944), y sin poder alguno, al resultado competitivo.

Gráficamente, en un plano de empleo, L, y salario, w, como el de la Figura 12.4, se representan la demanda de trabajo, L^D, el salario de reserva, w_R, el número de afiliados, N, una isobeneficio de la empresa, π, una curva de indiferencia del sindicato, CI, y la curva de contrato, CC. Así:

- Sin poder sindical, el acuerdo es el competitivo, E_0, sobre la demanda y al salario de reserva.
- Con poder sindical, el acuerdo de derecho a gestionar, E, tiene un salario mayor y menos empleo, y es ineficiente: la curva de indiferencia y la isobeneficio se cortan, y entre ellas queda una zona de mejoras paretianas sin explotar (Cahuc et al., 2014, pp. 433-435).

Si la función de ingresos de la empresa es isoelástica, la elasticidad de la demanda de trabajo es constante, el salario negociado no responde a las perturbaciones y todo el ajuste recae en el empleo.[30]

Figura 12.4: Modelos de derecho a gestionar y de contrato eficiente



[30] La intuición es que el sindicato fija el salario como un margen sobre el salario de reserva que depende de la elasticidad de la demanda de trabajo. Si la perturbación no altera esa elasticidad, tampoco altera el salario.

Modelo de contrato eficiente McDonald y Solow (1981) mostraron que esa ineficiencia desaparece si se negocian a la vez el salario y el empleo.[31] El acuerdo se sitúa entonces sobre la curva de contrato, CC, de la Figura 12.4, donde se tocan las curvas de indiferencia y las isobeneficio, a la derecha de la demanda de trabajo, y con trabajadores aversos al riesgo hay más empleo que en la situación competitiva.[32]

Implicaciones de política económica En el modelo de contrato eficiente, la moderación salarial no crea empleo, y un mayor poder sindical puede incluso aumentarlo. En la práctica, sin embargo, el salario suele negociarse antes que el empleo y de forma más centralizada, por lo que ambos modelos se interpretan como casos límite de un mismo marco (Cahuc et al., 2014, pp. 440-441).[33]

Modelo de insiders-outsiders El modelo de insiders-outsiders de Lindbeck y Snower (1988) parte de que la negociación no representa por igual a todos los trabajadores. Los insiders, que ya tienen empleo, negocian con la empresa, mientras que los outsiders, parados o con contratos más precarios, quedan fuera. El poder de los insiders procede de los costes de rotación, como los de contratar, formar o despedir: sustituirlos por outsiders dispuestos a trabajar por menos le cuesta a la empresa, y mientras ese coste supere la diferencia salarial no le compensa hacerlo (Lindbeck y Snower, 2001). En su versión más sencilla, si contratar a un outsider cuesta c_h y despedir a un insider c_f, el salario de los insiders, w_I, puede superar al de los outsiders, w_O, hasta en la suma de ambos costes:

$$w_I = w_O + c_h + c_f$$

El modelo explica así que haya paro involuntario, porque los outsiders trabajarían por menos pero la empresa no los contrata, y también que el paro persista. Tras una caída de la demanda, los despedidos pasan a ser outsiders y dejan de contar en la negociación, y los insiders que quedan fijan salarios que solo protegen sus propios puestos (Romer, 2019, p. 549).[34]

Su implicación de política es reducir el poder de los insiders y facilitar la entrada de

[31] La idea se remonta a Leontief, que ya advirtió que negociar a la vez salario y empleo permite alcanzar acuerdos eficientes (Cahuc et al., 2014, p. 435).

[32] Si además se negocian las prestaciones de los parados o las indemnizaciones por despido, el sindicato puede asegurar a sus afiliados frente al riesgo de perder el empleo, y el empleo vuelve al nivel competitivo. Es el contrato fuertemente eficiente, que reconcilia la eficiencia productiva con la eficiencia en el sentido de Pareto (Cahuc et al., 2014, pp. 437-439).

[33] También importa el nivel en que se negocia. Calmfors y Driffill (1988) encontraron una relación en forma de U invertida entre la centralización de la negociación y el paro: los resultados son mejores con una negociación muy centralizada, que tiene en cuenta sus efectos sobre el conjunto de la economía, o muy descentralizada, disciplinada por la competencia, y peores en los casos intermedios, como la negociación por sectores.

[34] Este modelo es una de las explicaciones de la persistencia del paro y de la histéresis, que se estudian en el Tema 11 de la Parte B.

 © 2026 Manuel Cabado García

los outsiders: abaratar los costes de rotación, permitir que las empresas se descuelguen de los convenios cuando lo necesiten y mejorar la formación de los parados. En España, la dualidad entre indefinidos y temporales es la versión más clara de esta división, porque los temporales soportan la mayor parte del ajuste.[35]

Salarios de eficiencia

Salario de eficiencia El salario de eficiencia es el salario que la empresa fija por encima del que vaciaría el mercado porque la productividad de sus trabajadores depende de lo que les paga.[36] Las razones pueden ser la reducción de la rotación (Salop, 1979), la selección de mejores candidatos (Weiss, 1980), la reciprocidad de quien se siente bien tratado, que Akerlof (1982) llamó intercambio de regalos, o la disuasión del escaqueo (Shapiro y Stiglitz, 1984).

Solow (1979) mostró el resultado común a todas ellas. Si el esfuerzo de los trabajadores, $e(w)$, crece con el salario, la empresa elige el salario que minimiza el coste de cada unidad de esfuerzo, $w/e(w)$, y lo encuentra donde la elasticidad del esfuerzo respecto al salario es unitaria (Romer, 2019, p. 525):

$$\frac{e'(w) \cdot w}{e(w)} = 1$$

Gráficamente, en dos planos, uno de salario, w, y esfuerzo, e, y otro de salario y coste por unidad de esfuerzo, w/e, como los de la Figura 12.5, se representan la función de esfuerzo, $e(w)$, y el coste por unidad de esfuerzo, $w/e(w)$. En ellos:

- El salario de eficiencia, w^*, es aquel en el que un rayo desde el origen es tangente a la función de esfuerzo, donde la elasticidad es unitaria.
- En ese mismo salario, el coste por unidad de esfuerzo alcanza su mínimo, $(w/e)^*$.
- Ese salario no depende de la demanda de trabajo, de modo que las perturbaciones se trasladan al empleo y no al salario real.[37]

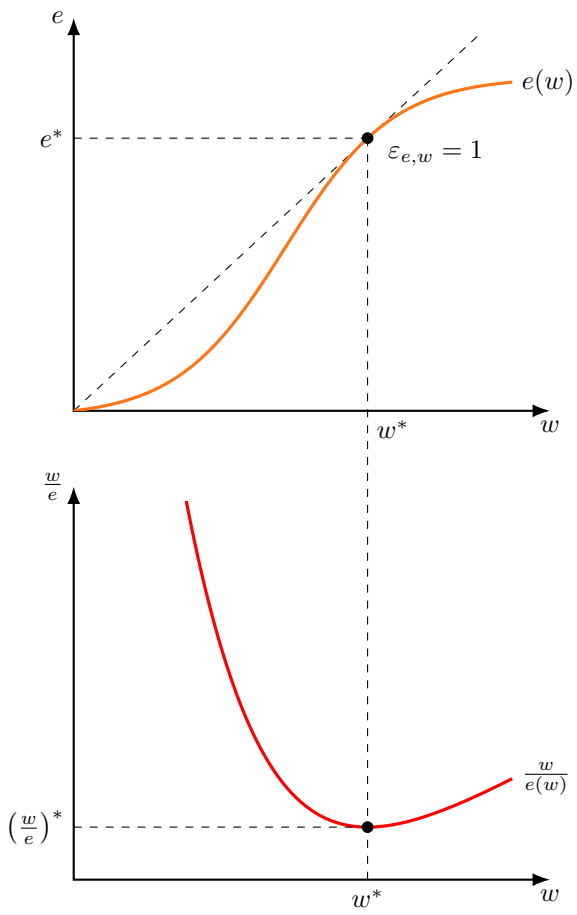
Shapiro y Stiglitz (1984) dieron a la idea un fundamento de riesgo moral. La empresa no observa el esfuerzo y solo puede despedir a quien sorprende escaqueándose, pero con pleno empleo el despedido encontraría otro trabajo enseguida y el despido no disciplinaría a nadie. La curva de no escaqueo, CNE, da para cada nivel de empleo el salario mínimo

[35] Si la empresa puede pagar salarios distintos a insiders y entrantes, el poder de los insiders genera más discriminación que paro: los entrantes cobran menos, pero el empleo total no cambia. El paro solo aparece cuando la empresa debe pagar a todos el mismo salario (Cahuc et al., 2014, pp. 443-445).

[36] La intuición es que un salario más alto puede hacer que el trabajador se esfuerce más, que no se marche a otra empresa o que se presenten mejores candidatos. Si esas ganancias compensan el mayor coste, a la empresa le conviene pagar más de lo necesario para contratar.

[37] El paro de equilibrio que resulta cuando todas las empresas pagan salarios de eficiencia se obtiene en el Apéndice 12.A.7.

Figura 12.5: Condición de Solow



que hace que no compense escaquearse, y es creciente porque, cuanto menor es el paro, menos cuesta perder el empleo.

Gráficamente, en un plano de empleo, L, y salario, w, como el de la Figura 12.6, se representan la oferta de trabajo con información simétrica, L^S, horizontal en $\bar{e}$ hasta el pleno empleo, $\bar{L}$, la curva de no escaqueo, CNE, y la demanda de trabajo, L^D. Así:

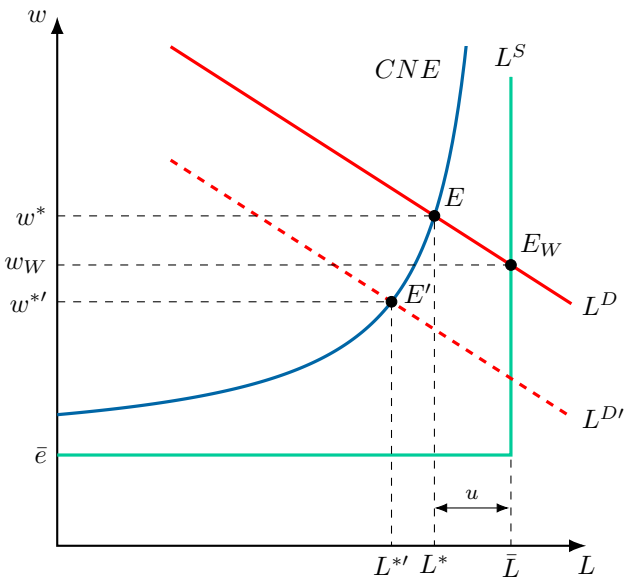
- El equilibrio, E, está en el corte de la curva de no escaqueo con la demanda, con un salario, w^*, superior a $\bar{e}$ y un empleo, L^*, inferior al pleno empleo.
- La diferencia entre $\bar{L}$ y L^* es paro involuntario: los parados trabajarían al salario vigente, pero no pueden comprometerse de forma creíble a no escaquearse. El paro actúa así como mecanismo de disciplina.[38]
- Una caída de la demanda de trabajo, de L^D a $L^{D\prime}$, desplaza el equilibrio de E a E' a lo largo de la curva de no escaqueo: bajan el salario y el empleo y aumenta el paro (Romer, 2019, p. 540).

[38] La teoría tiene un punto débil reconocido: si los trabajadores prefieren el empleo al paro, la empresa podría exigirles una fianza que perderían si se les sorprende escaqueándose, o cobrarles por el puesto, y el racionamiento desaparecería. Que esas prácticas apenas se observen es un enigma para el modelo (Romer, 2019, p. 542).

 © 2026 Manuel Cabado García

Implicaciones de política económica El equilibrio es además ineficiente. Con pleno empleo, la productividad marginal del trabajo supera lo que a los trabajadores les cuesta esforzarse, de modo que lo óptimo sería que todos trabajaran y se esforzaran. El Estado no puede lograrlo obligando a las empresas a contratar hasta el pleno empleo: sin paro que discipline, los trabajadores se escaquearían y la producción se hundiría. Lo que sí mejora el bienestar es un subsidio al empleo financiado con impuestos de suma fija o sobre los beneficios. El subsidio desplaza hacia arriba la demanda de trabajo, y la economía sube a lo largo de la curva de no escaqueo, con más salario y más empleo. Como el valor de la producción adicional supera su coste de oportunidad, el bienestar aumenta. El reparto de esa ganancia entre trabajadores y empresas depende de cómo se financie el subsidio (Romer, 2019, p. 541).

Figura 12.6: Modelo de Shapiro y Stiglitz



Evidencia empírica El ejemplo clásico es el de Henry Ford, que en 1914 duplicó el salario de sus trabajadores hasta cinco dólares diarios. Según Raff y Summers (1987), la rotación anual en la fábrica pasó del 370 % en 1913 al 16 % en 1915 y el absentismo diario cayó del 10 % al 2,5 %, con aumentos notables de la productividad.[39]

Contratos implícitos

La teoría de los contratos implícitos, desarrollada por Baily (1974), Azariadis (1975) y Gordon (1974), explica la rigidez salarial como un seguro. Los trabajadores son aversos al riesgo y no pueden asegurar su renta en los mercados financieros, mientras que la empresa,

[39] Los propios autores advierten de que la recesión de 1914 en Detroit pudo explicar parte de la caída de la rotación.

que sí tiene acceso a ellos, se comporta como neutral al riesgo.[40] El contrato óptimo, que no está escrito y se sostiene en la reputación de la empresa, garantiza al trabajador una renta estable a lo largo del ciclo a cambio de un salario medio algo menor, y deja que el ajuste recaiga en las horas o en el empleo. La teoría explica así la rigidez del salario real, pero no el paro involuntario. En el contrato óptimo, la empresa no elige el empleo tomando el salario como dado. Lo fija el propio contrato, en el nivel en que la productividad marginal del trabajo iguala a la desutilidad marginal, es decir, a lo que al trabajador le cuesta trabajar una hora más, y nadie queda sin trabajar contra su voluntad. Lo que sí puede haber es paro aparente: en las recesiones se trabaja menos por la misma renta, el salario por hora resulta alto y los trabajadores querrían trabajar más a ese salario, aunque el contrato sea óptimo (Romer, 2019, pp. 545-547).

Evidencia empírica española Los ERTE de la pandemia aplicaron el mismo principio desde lo público: en el peor momento llegaron a cubrir a unos 3,6 millones de trabajadores, cuya renta quedó asegurada mientras el ajuste recaía en las horas y no en el empleo.[41]

12.5. Conclusiones ⏳ 30"

Como conclusión, encontrar empleo lleva tiempo, cambiar de plantilla cuesta y el salario no siempre es un precio que vacía el mercado. Esta lección está hoy más viva que nunca:

- Tecnología: las plataformas digitales de empleo y la inteligencia artificial pueden mejorar la eficiencia del emparejamiento y acercar la curva de Beveridge al origen, pero también destruir empleos y exigir cualificaciones nuevas que la alejen. ¿Qué efecto dominará?
- Desajuste: los 900.000 parados de larga duración son el reflejo más claro de un emparejamiento que falla. Si ese desajuste es estructural, no se corrige bajando salarios, sino con formación, intermediación y movilidad.
- Vivienda: si los parados no pueden trasladarse a donde están las vacantes porque la vivienda es cara o escasa, el emparejamiento empeora. El mercado de trabajo y el de la vivienda están más conectados de lo que parece.
- Transición ecológica: el cierre de sectores intensivos en carbono y el auge de otros es una perturbación de reasignación de manual, que puede alejar la curva de Beveridge del origen durante años. ¿Está preparada la política de empleo para recolocar a los trabajadores afectados?

[40] El supuesto es razonable ante perturbaciones específicas de la empresa, que sus propietarios pueden diversificar, pero no tanto ante las agregadas, que son precisamente las que la teoría quiere explicar (Romer, 2019, p. 544).

[41] Según la Seguridad Social, el máximo, de 3.608.148 trabajadores, se alcanzó en la primavera de 2020 (comparecencia del Secretario de Estado de la Seguridad Social en el Senado, mayo de 2021).

 © 2026 Manuel Cabado García

Apéndice

12.A.1. El salario de reserva en el modelo de búsqueda secuencial

Se trabaja en tiempo continuo con la notación del cuerpo del tema (Cahuc et al., 2014, pp. 262-264). Un trabajador empleado con salario w pierde su empleo a la tasa λ, de modo que el valor de estar empleado, $V_e(w)$, cumple

$$r \cdot V_e(w) = w + \lambda \cdot [V_u - V_e(w)] \qquad \Longrightarrow \qquad V_e(w) = \frac{w + \lambda \cdot V_u}{r + \lambda}$$

donde V_u es el valor de estar parado. El parado obtiene la renta z y recibe ofertas a la tasa a, que acepta si le dan un valor mayor que seguir buscando:

$$r \cdot V_u = z + a \cdot \int_0^{\infty} \text{máx}\left\{V_e(w) - V_u, 0\right\} dH(w)$$

El salario de reserva es el que deja al parado indiferente, $V_e(w_R) = V_u$. Sustituyendo en la expresión de V_e se obtiene $w_R = r \cdot V_u$, y por tanto $V_e(w) - V_u = (w - w_R)/(r + \lambda)$, que es positivo solo para las ofertas que superan el salario de reserva. Llevando estos dos resultados a la ecuación de V_u se llega a la ecuación (12.1) del cuerpo:

$$w_R = z + \frac{a}{r + \lambda} \cdot \int_{w_R}^{\infty} (w - w_R)\, dH(w)$$

Como el parado sale del paro cuando recibe una oferta y esta supera su umbral, la tasa de salida del paro es $a \cdot [1 - H(w_R)]$ y la duración media del paro, su inversa. Diferenciando la ecuación anterior respecto a z se obtiene el efecto de la prestación:

$$\frac{\partial w_R}{\partial z} = \frac{1}{1 + \dfrac{a}{r + \lambda} \cdot [1 - H(w_R)]} \in (0, 1)$$

de modo que una prestación más alta eleva el salario de reserva, aunque menos que proporcionalmente, y alarga la duración media del paro.

12.A.2. El equilibrio del modelo de Diamond, Mortensen y Pissarides

Se sigue la versión básica de Pissarides (2000, cap. 1). Para la empresa, el valor de una vacante, V, y el de un puesto ocupado, J, cumplen

$$r \cdot V = -p \cdot c + q(\theta) \cdot (J - V) \qquad\qquad r \cdot J = p - w - \lambda \cdot J$$

La vacante cuesta $p \cdot c$ por unidad de tiempo y se cubre a la tasa $q(\theta)$, y el puesto ocupado produce p, paga w y se destruye a la tasa λ. Con libre entrada, $V = 0$, de modo que

$J = p \cdot c/q(\theta)$, y como $J = (p - w)/(r + \lambda)$, se obtiene la curva de creación de empleo, (12.3).

Para el trabajador, el valor de estar parado, U, y el de estar empleado, W, cumplen

$$r \cdot U = z + \theta \cdot q(\theta) \cdot (W - U) \qquad r \cdot W = w + \lambda \cdot (U - W)$$

El salario resuelve el problema de Nash, $\max_w (W - U)^{\beta} \cdot (J - V)^{1-\beta}$, cuya condición de primer orden es $(1 - \beta) \cdot (W - U) = \beta \cdot (J - V)$: el trabajador se queda con una fracción β del excedente total del emparejamiento. Como $(r + \lambda) \cdot (W - U) = w - r \cdot U$ y $(r + \lambda) \cdot J = p - w$, la regla de reparto da

$$w = (1 - \beta) \cdot r \cdot U + \beta \cdot p$$

y, sustituyendo $W - U = \beta \cdot J/(1 - \beta)$ y $J = p \cdot c/q(\theta)$ en la ecuación de U, se obtiene $r \cdot U = z + \beta \cdot p \cdot c \cdot \theta/(1 - \beta)$. Llevando este valor a la expresión del salario resulta la curva salarial, (12.4).

Eliminando el salario entre la curva de creación de empleo y la curva salarial se obtiene la condición que determina la tensión de equilibrio:

$$(1 - \beta) \cdot (p - z) = [r + \lambda + \beta \cdot \theta \cdot q(\theta)] \cdot \frac{p \cdot c}{q(\theta)} \tag{12.5}$$

El primer miembro no depende de θ y el segundo es creciente en θ, porque $q(\theta)$ es decreciente y $\theta \cdot q(\theta)$ creciente, así que el equilibrio, θ^*, es único. Con θ^*, la curva de Beveridge, (12.2), da la tasa de paro, u^*, y las vacantes son $v^* = \theta^* \cdot u^*$. La estática comparativa del cuerpo se lee directamente en esta ecuación: un aumento de z o de β reduce el primer miembro de la ecuación o eleva el segundo, y con ello reduce la tensión de equilibrio, mientras que un aumento de p la eleva, porque el coste de la vacante, $p \cdot c$, crece en la misma proporción que p pero z no.

12.A.3. La demanda de trabajo con costes de ajuste

Con costes cuadráticos, $C(\dot{L}_t) = (b/2) \cdot \dot{L}_t^2$, el problema de la empresa se resuelve con el hamiltoniano en valor presente, en el que la variable de control es $x_t = \dot{L}_t$ y μ_t es la variable de coestado:

$$H = e^{-rt} \cdot \left[F(L_t) - w \cdot L_t - \frac{b}{2} \cdot x_t^2 \right] + \mu_t \cdot x_t$$

La condición $\partial H/\partial x_t = 0$ da $\mu_t = e^{-rt} \cdot b \cdot x_t$, y la ecuación de la coestado, $\dot{\mu}_t = -\partial H/\partial L_t = -e^{-rt} [F'(L_t) - w]$. Derivando la primera respecto al tiempo e igualando con la segunda

 © 2026 Manuel Cabado García

se obtiene la ecuación de Euler (Cahuc et al., 2014, pp. 123-124):

$$F'(L_t) = w + r \cdot b \cdot \dot{L}_t - b \cdot \ddot{L}_t$$

En el estado estacionario, $\dot{L} = \ddot{L} = 0$ y se recupera la condición estática, $F'(L^*) = w$. Linealizando alrededor de L^*, la senda estable es $L_t = L^* + (L_0 - L^*) \cdot e^{\nu t}$, con $\nu = \left[r - \sqrt{r^2 + 4\kappa}\right]/2 < 0$ y $\kappa = -F''(L^*)/b > 0$. Cuanto mayor es el coste de ajuste, b, menor es κ, menor es el valor absoluto de ν y más lento es el ajuste.

Con costes lineales y asimétricos, contratar a un trabajador cuesta c_h una sola vez y despedirlo, c_f. Contratar un trabajador más compensa si el valor actual de lo que aporta, $[F'(L) - w]/r$, supera el coste de contratarlo, y despedirlo, si el ahorro, $[w - F'(L)]/r$, supera el coste del despido. Los umbrales son, por tanto,

$$F'(L_h) = w + r \cdot c_h \qquad\qquad F'(L_f) = w - r \cdot c_f$$

y la demanda de trabajo, a partir de un empleo inicial L_0, es $L = L_h$ si $L_0 < L_h$, $L = L_0$ si $L_h \leq L_0 \leq L_f$ y $L = L_f$ si $L_0 > L_f$ (Cahuc et al., 2014, pp. 126-127). En este caso determinista, un aumento de c_f desplaza L_f hacia la derecha y amplía la banda de inacción sin afectar al umbral de contratación. El caso con incertidumbre se estudia en el apartado siguiente.

12.A.4. Los costes de ajuste con incertidumbre

Con incertidumbre, la empresa decide la plantilla de hoy pensando en las perturbaciones futuras, y sus expectativas pasan a determinar la senda del empleo. Si los costes son cuadráticos y la empresa forma expectativas racionales, la demanda de trabajo en tiempo discreto cumple

$$L_t = \varphi \cdot L_{t-1} + \mu_0 \cdot \sum_{i=0}^{\infty} (a_0 \cdot \mu_0)^i \cdot E_t\, a_{t+i}$$

En ella, a_t mide la rentabilidad del empleo en cada periodo, que crece con la perturbación de productividad y cae con el salario, y E_t son las expectativas formadas en t. Los parámetros φ, μ_0 y a_0 dependen de la tecnología, del tipo de interés y del coste de ajuste, b. El coeficiente φ, entre cero y uno, crece con b, de modo que, cuanto más cuesta ajustar, más pesa el empleo pasado (Cahuc et al., 2014, pp. 128-131). El empleo depende así del pasado, a través de los costes de ajuste, y del futuro esperado, a través de las expectativas. Una perturbación transitoria apenas altera lo que se espera de los periodos siguientes y mueve poco el empleo. Una permanente, en cambio, desplaza todas las expectativas y provoca un ajuste mayor, aunque gradual.

Con costes lineales, la productividad alterna entre un estado bueno, A_G, y uno malo, A_B. Se pasa del bueno al malo con probabilidad instantánea q_G y del malo al bueno con

q_B. La empresa mantiene un empleo L_G en el estado bueno y L_B en el malo, que cumplen

$$F_L(A_G, L_G) = w_G + q_G \cdot c_f + (r + q_G) \cdot c_h \qquad F_L(A_B, L_B) = w_B - q_B \cdot c_h - (r + q_B) \cdot c_f$$

En el estado bueno, la productividad marginal debe cubrir el salario, el coste de contratar y también el coste esperado de despedir cuando el ciclo cambie. Por eso L_G cae con los dos costes. Simétricamente, L_B sube con ambos: un despido caro frena los despidos en las recesiones, y una contratación cara también, porque la empresa sabe que tendrá que volver a contratar en la recuperación. Los costes de ajuste estabilizan así el empleo, y la rotación, τ, que es la suma de empleos creados y destruidos,

$$\tau = \frac{2 \cdot q_G \cdot q_B}{q_G + q_B} \cdot (L_G - L_B)$$

cae cuando el mercado de trabajo se vuelve más rígido (Cahuc et al., 2014, pp. 132-134).

12.A.5. La condición de Hosios

El planificador elige la tensión para maximizar el valor descontado de la producción neta, que es lo que producen los ocupados, más la renta de los parados, menos los costes de las vacantes, sujeto a la dinámica del paro:

$$\max_{\{\theta_t\}} \int_0^\infty e^{-rt} \cdot [p \cdot (1 - u_t) + z \cdot u_t - p \cdot c \cdot \theta_t \cdot u_t]\, dt \qquad \text{s.a.} \qquad \dot{u}_t = \lambda \cdot (1 - u_t) - \theta_t \cdot q(\theta_t) \cdot u_t$$

En este problema, z recoge solo el valor del ocio o de la producción doméstica, y no transferencias como la prestación por desempleo, que únicamente redistribuyen (Cahuc et al., 2014, p. 601). Llamando $\eta = -\theta \cdot q'(\theta)/q(\theta)$ a la elasticidad de la función de emparejamiento respecto al paro, las condiciones de optimalidad en el estado estacionario se reducen a

$$(1 - \eta) \cdot (p - z) = [r + \lambda + \eta \cdot \theta \cdot q(\theta)] \cdot \frac{p \cdot c}{q(\theta)}$$

que tiene la misma forma que la condición de equilibrio, (12.5), con η en el lugar de β (Hosios, 1990). Las dos coinciden, y el equilibrio es eficiente, si y solo si $\beta = \eta$, siempre que la renta del parado del equilibrio no incluya tampoco la prestación por desempleo. Con una prestación positiva, el primer miembro de la condición de equilibrio es menor que en el óptimo y, aunque se cumpla $\beta = \eta$, la tensión es demasiado baja y el paro, demasiado alto (Cahuc et al., 2014, p. 601). Si $\beta > \eta$, el primer miembro de la condición de equilibrio es menor y el segundo mayor que en el óptimo, la tensión es demasiado baja y el paro, demasiado alto, y si $\beta < \eta$, ocurre lo contrario (Cahuc et al., 2014, pp. 600-602).

 © 2026 Manuel Cabado García

12.A.6. La solución de Nash en la negociación colectiva

La empresa obtiene unos ingresos $R(L)$, cóncavos, y su beneficio es $\pi = R(L) - w \cdot L$, con una demanda de trabajo dada por $R'(L) = w$, cuya elasticidad respecto al salario, en valor absoluto, es ε_L. El sindicato tiene N afiliados y es utilitarista: cada empleado obtiene $u(w)$ y cada parado $u(w_R)$, con u cóncava. Sin acuerdo, la empresa no produce y los afiliados obtienen $u(w_R)$, de modo que las ganancias de cada parte son $L \cdot [u(w) - u(w_R)]$ y π (Cahuc et al., 2014, pp. 431-437).

En el modelo de derecho a gestionar se negocia solo el salario, y el empleo lo fija la empresa sobre su demanda, $L = L^D(w)$. El problema es $\max_w \beta \cdot \ln \left\{ L^D(w) \cdot [u(w) - u(w_R)] \right\} + (1 - \beta) \cdot \ln \pi(w)$. Como $d\pi/dw = -L$, la condición de primer orden, multiplicada por w, da

$$\frac{u(w) - u(w_R)}{w \cdot u'(w)} = \frac{\beta}{\beta \cdot \varepsilon_L + (1 - \beta) \cdot \dfrac{w \cdot L}{\pi}}$$

Con $\beta = 1$ se obtiene el sindicato monopolista de Dunlop, que con neutralidad al riesgo fija $w = w_R / (1 - 1/\varepsilon_L)$, un margen sobre el salario de reserva. Si los ingresos son isoelásticos, $R(L) = A \cdot L^\alpha$, se cumple $\varepsilon_L = 1/(1 - \alpha)$ y $w \cdot L/\pi = \alpha/(1 - \alpha)$, ambos constantes: el salario negociado es proporcional al de reserva y no responde a las perturbaciones de A, que recaen por entero en el empleo.

En el modelo de contrato eficiente se negocian el salario y el empleo, y el acuerdo está sobre la curva de contrato, donde la relación marginal de sustitución del sindicato iguala la de la empresa:

$$R'(L) = w - \frac{u(w) - u(w_R)}{u'(w)}$$

Como $u(w) > u(w_R)$, se cumple $R'(L) < w$ y el acuerdo está a la derecha de la demanda de trabajo. Con neutralidad al riesgo, $u(w) = w$, la condición se reduce a $R'(L) = w_R$: la curva de contrato es vertical en el empleo competitivo y la negociación solo reparte la renta. Con aversión al riesgo, la concavidad de u implica $u(w) - u(w_R) > u'(w) \cdot (w - w_R)$, de modo que $R'(L) < w_R$ y hay sobreempleo respecto al nivel competitivo.

12.A.7. El paro de equilibrio con salarios de eficiencia

Se sigue a Romer (2019, pp. 528-530), que formaliza una idea de Summers (1988). La condición de Solow fija el salario de cada empresa, pero no dice cuánto paro hay cuando todas pagan por encima del salario que vaciaría el mercado. Para responder, el esfuerzo debe depender también del salario que pagan las demás empresas, w_a, y de la tasa de paro, u: $e = e(w, w_a, u)$. El esfuerzo crece con el salario propio y con el paro, porque perder el empleo cuesta más, y cae con el salario de las demás empresas, porque el despido es menos grave si fuera se paga bien. Cada empresa es pequeña y toma w_a y u como dados, de modo que sigue eligiendo el salario en el que la elasticidad del esfuerzo es unitaria.

Con la forma funcional de Summers, el esfuerzo depende de la prima que el salario propio paga sobre las oportunidades del trabajador fuera de la empresa, x:

$$e = \left(\frac{w-x}{x}\right)^{\gamma} \qquad x = (1 - b \cdot u) \cdot w_a$$

donde γ, entre cero y uno, es la elasticidad del esfuerzo respecto a esa prima, y b mide cuánto pesa el paro en la valoración del trabajador. La condición de Solow da

$$w = \frac{x}{1-\gamma} = \frac{1 - b \cdot u}{1-\gamma} \cdot w_a$$

de modo que cada empresa paga una prima aproximadamente igual a γ sobre las oportunidades externas. En equilibrio todas pagan lo mismo, $w = w_a$, lo que solo es posible si $1 - \gamma = 1 - b \cdot u$, es decir, si

$$u = \frac{\gamma}{b}$$

Si el paro fuera menor, cada empresa querría pagar más que las demás, el salario subiría y el empleo caería hasta restablecerlo. El paro de equilibrio depende solo de la función de esfuerzo y no de la de producción, de modo que el progreso técnico no lo reduce a largo plazo. Además, bastan elasticidades pequeñas para generar un paro considerable: con $\gamma = 0{,}06$ y $b = 1$, el paro de equilibrio es del 6 % (Romer, 2019, p. 530).

12.A.8. La condición de no escaqueo de Shapiro y Stiglitz

Se sigue la formulación de Romer (2019, pp. 532-539). Un trabajador obtiene utilidad $w - \bar{e}$ si se esfuerza y w si se escaquea, y un parado obtiene cero. Los empleos se destruyen a la tasa λ, el escaqueo se detecta a la tasa q y los parados encuentran empleo a la tasa a. Con un tipo de descuento ρ, los valores de estar empleado y esforzarse, V_E, de estar empleado y escaquearse, V_S, y de estar parado, V_U, cumplen

$$\rho \cdot V_E = (w - \bar{e}) - \lambda \cdot (V_E - V_U) \qquad \rho \cdot V_S = w - (\lambda + q) \cdot (V_S - V_U) \qquad \rho \cdot V_U = a \cdot (V_E - V_U)$$

La empresa paga el salario mínimo que evita el escaqueo, $V_E = V_S$. Igualando las dos primeras ecuaciones se obtiene $V_E - V_U = \bar{e}/q$: el empleo tiene que valer más que el paro exactamente lo que cuesta el esfuerzo, corregido por la probabilidad de ser descubierto. Restando la tercera ecuación de la primera, $\rho \cdot (V_E - V_U) = w - \bar{e} - (a + \lambda) \cdot (V_E - V_U)$, y sustituyendo, el salario necesario es

$$w = \bar{e} + (\rho + a + \lambda) \cdot \frac{\bar{e}}{q}$$

En estado estacionario, con $\bar{L}$ trabajadores y L empleados, las salidas del empleo igualan a las entradas, $\lambda \cdot L = a \cdot (\bar{L} - L)$, de modo que $a + \lambda = \lambda \cdot \bar{L}/(\bar{L} - L)$. Sustituyendo

 © 2026 Manuel Cabado García

se obtiene la curva de no escaqueo:

$$w = \bar{e} + \left(\rho + \frac{\lambda \cdot \bar{L}}{\bar{L} - L}\right) \cdot \frac{\bar{e}}{q}$$

que es creciente en el empleo y tiende a infinito cuando L se acerca a $\bar{L}$: con pleno empleo ningún salario evita el escaqueo. Si la detección es perfecta, $q \to \infty$, el salario se reduce a $\bar{e}$, el de información simétrica.

12.A.9. El contrato implícito óptimo

Se sigue a Romer (2019, pp. 543-547). La productividad toma el valor A_i con probabilidad p_i, $i = 1, \ldots, n$. La empresa, neutral al riesgo, paga al trabajador C_i y le pide L_i horas en cada estado, y el trabajador obtiene $U(C_i) - V(L_i)$, con U cóncava y V convexa. El contrato maximiza el beneficio esperado sujeto a que el trabajador alcance al menos la utilidad esperada u_0, con multiplicador μ:

$$\mathcal{L} = \sum_{i=1}^{n} p_i \cdot [A_i \cdot F(L_i) - C_i] + \mu \cdot \left\{ \sum_{i=1}^{n} p_i \cdot [U(C_i) - V(L_i)] - u_0 \right\}$$

Las condiciones de primer orden respecto a C_i y a L_i son

$$U'(C_i) = \frac{1}{\mu} \qquad\qquad A_i \cdot F'(L_i) = \frac{V'(L_i)}{U'(C)}$$

La primera implica que la renta del trabajador es la misma en todos los estados, $C_i = C$: la empresa le asegura por completo. La segunda iguala en cada estado la productividad marginal del trabajo con la relación marginal de sustitución entre ocio y consumo, de modo que el empleo es eficiente. Como el empleo es mayor cuando A_i es alto y la renta es constante, el salario por hora, C/L_i, es contracíclico, y no hay paro involuntario.

Bibliografía

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics 97*(4), 543–569.

Azariadis, C. (1975). Implicit contracts and underemployment equilibria. *Journal of Political Economy 83*(6), 1183–1202.

Baily, M. N. (1974). Wages and employment under uncertain demand. *Review of Economic Studies 41*(1), 37–50.

Bentolila, S. y G. Bertola (1990). Firing costs and labour demand: How bad is Eurosclerosis? *Review of Economic Studies 57*(3), 381–402.

Bentolila, S., P. Cahuc, J. J. Dolado y T. Le Barbanchon (2012). Two-tier labour markets in the Great Recession: France versus Spain. *Economic Journal 122*(562), F155–F187.

Beveridge, W. H. (1944). *Full Employment in a Free Society.* London: George Allen & Unwin.

Boscá, J. E., R. Doménech, J. Ferri y J. R. García (2017). Shifts in the Beveridge curve in Spain and their macroeconomic effects. *Revista de Economía Aplicada 25*(75), 5–27.

Burdett, K. y D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review 39*(2), 257–273.

Cahuc, P., S. Carcillo y A. Zylberberg (2014). *Labor Economics* (Segunda ed.). Cambridge, MA: MIT Press.

Calmfors, L. y J. Driffill (1988). Bargaining structure, corporatism and macroeconomic performance. *Economic Policy 3*(6), 13–61.

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory 3*(2), 156–168.

Diamond, P. A. (1982). Wage determination and efficiency in search equilibrium. *Review of Economic Studies 49*(2), 217–227.

Dow, J. C. R. y L. A. Dicks-Mireaux (1958). The excess demand for labour: A study of conditions in Great Britain, 1946–56. *Oxford Economic Papers 10*(1), 1–33.

Dunlop, J. T. (1944). *Wage Determination under Trade Unions.* New York: Macmillan.

Elsby, M. W. L., R. Michaels y D. Ratner (2015). The Beveridge curve: A survey. *Journal of Economic Literature 53*(3), 571–630.

 © 2026 Manuel Cabado García

Felgueroso, F., R. Doménech, J. R. García, M. Jansen y A. Viola (2026). Observatorio trimestral del mercado de trabajo. boletín n.º 17. Apuntes 2026/14, FEDEA y BBVA Research, Madrid. 27 de mayo de 2026.

Gordon, D. F. (1974). A neo-classical theory of Keynesian unemployment. *Economic Inquiry 12*(4), 431–459.

Hall, R. E. (2005). Employment fluctuations with equilibrium wage stickiness. *American Economic Review 95*(1), 50–65.

Hicks, J. R. (1932). *The Theory of Wages*. London: Macmillan.

Holt, C. C., F. Modigliani, J. F. Muth y H. A. Simon (1960). *Planning Production, Inventories, and Work Force*. Englewood Cliffs, NJ: Prentice-Hall.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies 57*(2), 279–298.

Lindbeck, A. y D. J. Snower (1988). *The Insider-Outsider Theory of Employment and Unemployment*. Cambridge, MA: MIT Press.

Lindbeck, A. y D. J. Snower (2001). Insiders versus outsiders. *Journal of Economic Perspectives 15*(1), 165–188.

McCall, J. J. (1970). Economics of information and job search. *Quarterly Journal of Economics 84*(1), 113–126.

McDonald, I. M. y R. M. Solow (1981). Wage bargaining and employment. *American Economic Review 71*(5), 896–908.

Mortensen, D. T. (1970). Job search, the duration of unemployment, and the Phillips curve. *American Economic Review 60*(5), 847–862.

Mortensen, D. T. y C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies 61*(3), 397–415.

Nash, J. F. (1950). The bargaining problem. *Econometrica 18*(2), 155–162.

Nickell, S. J. (1986). Dynamic models of labour demand. En O. Ashenfelter y R. Layard (eds.), *Handbook of Labor Economics*, vol. 1, cap. 9, pp. 473–522. Amsterdam: North-Holland.

Nickell, S. J. y M. Andrews (1983). Unions, real wages and employment in Britain 1951–79. *Oxford Economic Papers 35*(Supplement), 183–206.

Oi, W. Y. (1962). Labor as a quasi-fixed factor. *Journal of Political Economy 70*(6), 538–555.

Petrongolo, B. y C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature 39*(2), 390–431.

Pissarides, C. A. (1985). Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *American Economic Review 75*(4), 676–690.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory* (Segunda ed.). Cambridge, MA: MIT Press.

Raff, D. M. G. y L. H. Summers (1987). Did Henry Ford pay efficiency wages? *Journal of Labor Economics 5*(4, Part 2), S57–S86.

Romer, D. (2019). *Advanced Macroeconomics* (Quinta ed.). New York: McGraw-Hill Education.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica 50*(1), 97–109.

Salop, S. C. (1979). A model of the natural rate of unemployment. *American Economic Review 69*(1), 117–125.

Shapiro, C. y J. E. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review 74*(3), 433–444.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review 95*(1), 25–49.

Solow, R. M. (1979). Another possible source of wage stickiness. *Journal of Macroeconomics 1*(1), 79–82.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy 69*(3), 213–225.

Stigler, G. J. (1962). Information in the labor market. *Journal of Political Economy 70*(5, Part 2), 94–105.

Summers, L. H. (1988). Relative wages, efficiency wages, and keynesian unemployment. *American Economic Review 78*(2), 383–388.

Weiss, A. (1980). Job queues and layoffs in labor markets with flexible wages. *Journal of Political Economy 88*(3), 526–538.

 © 2026 Manuel Cabado García

Tema 13

Economía de la información y teoría de la agencia: selección adversa y riesgo moral.

13.1. Introducción ⏳ 1'

Tradicionalmente, la teoría microeconómica parte del supuesto clave de la existencia de información perfecta. No obstante, en la realidad se observa como la información imperfecta es el caso más frecuente como, por ejemplo, en el hecho de la necesidad de garantías para comprar un electrodoméstico, las entrevistas de trabajo, la supervisión bancaria o en la diligencia debida (due diligence) antes de una fusión empresarial.

Así, la información imperfecta puede ser:

- Incompleta pero simétrica, es decir, los agentes tienen la misma información pero no toda[1].
- Incompleta pero asimétrica, es decir, una de las partes de una relación contractual dispone de información relevante que la otra no tiene. Cuando, además, sus intereses están enfrentados, la parte informada puede aprovechar esa ventaja en perjuicio de la otra[2].

[1]Dentro de la información incompleta, cabe hallarse en una situación de riesgo (el individuo no sabe qué estado de la naturaleza va a ocurrir, pero puede asignar probabilidades) o de incertidumbre (si el individuo conoce todos o algunos de los estados de la naturaleza pero no puede asignar probabilidades).

Un mercado contingente es aquel en el que se compra hoy un bien que solo se entrega si se produce un determinado estado de la naturaleza, como un seguro de hogar, que solo indemniza si se produce el estado "incendio". Según el resultado de Arrow y Debreu, recogido en el manual de Arrow y Hahn (1971), si existiera un mercado contingente para cada estado posible, la asignación con información incompleta pero simétrica sería eficiente. El problema es que crear todos esos mercados suele tener un coste prohibitivo. Además, con selección adversa o riesgo moral estos mercados no resuelven el problema, porque los contratos solo pueden condicionarse a lo que ambas partes observan.

[2]La intuición es que quien vende su coche usado sabe si ha tenido averías y el comprador no. Si a ambos les interesara lo mismo, el vendedor se lo contaría sin más. Por eso los mecanismos que estudia la

Así, el funcionamiento del mercado y las relaciones contractuales bajo información asimétrica son estudiados por la economía de la información, cuyo marco analítico es la teoría de la agencia, es decir, el estudio de las relaciones en las que una parte, el principal, contrata con otra, el agente, que dispone de información que el principal no observa[3]. Este campo ha ganado importancia en las últimas décadas, con los diferentes premios Nobel otorgados en 2001 a Akerlof, Spence y Stiglitz, en 2007 a Myerson, Hurwicz y Maskin, en 2016 a Hart y Holmström[4] y en 2020 a Paul Milgrom y Robert Wilson.

Por lo tanto, el objetivo de la presente exposición será el estudio de este problema y de las soluciones propuestas por la teoría económica, que solo pueden alcanzar un óptimo de second best, es decir, la mejor asignación posible dadas las restricciones de información, mediante el siguiente esquema:

1. Selección adversa.
2. Riesgo moral.
3. Selección adversa y riesgo moral a la vez.

13.2. Selección Adversa ⏳ 9'30"

La selección adversa es un problema de información asimétrica que surge con anterioridad a la firma del contrato (ex-ante), y se debe a información oculta por una de las partes (ventaja informativa)[5]. Para que se dé una situación de selección adversa, debe darse:

1. Asimetría informativa: información oculta, es decir, características de una de las partes que no son perfectamente observables por la otra parte.
2. Heterogeneidad de la contraparte, ya que si todos los agentes fueran iguales, el tratamiento podría ser homogéneo sin pérdida de eficiencia.
3. Conflicto de intereses, ya que de lo contrario la contraparte no ocultaría información.

Dada la imposibilidad del principal de distinguir el tipo de contraparte en el contrato (es decir, si se trata de un conductor prudente o de uno temerario, si se trata de un

teoría de la agencia deben lograr que decir la verdad sea la mejor estrategia del agente, algo que, según el principio de revelación de Myerson (1979), puede exigirse sin pérdida de generalidad.

[3]La intuición es la del propietario que encarga la venta de su piso a una inmobiliaria: él es el principal, y la inmobiliaria, que sabe cuánto se esfuerza realmente en venderlo, es el agente.

[4]No confundir con Elinor Ostrom (también Premio Nobel en 2009, famosa por el análisis institucional de la gestión comunitaria de recursos de propiedad común –commons–, como pastos o pesquerías) ni con Mancur Olson (famoso por su teoría de la acción colectiva, que explica la organización de individuos para perseguir objetivos comunes). Hart estudió, junto con Grossman y Moore, los contratos incompletos y los derechos de propiedad. Holmström estudió el diseño de incentivos en contratos que especifican el pago correspondiente a cada resultado observable, como en los problemas de riesgo moral.

[5]Para un tratamiento unificado de la selección adversa y la señalización de mercado, véase (Kreps, 1990), capítulo 17.

 © 2026 Manuel Cabado García

trabajador responsable o de uno vago, etc.), el principal acabará diseñando un contrato que sea adecuado al promedio de la contraparte, es decir, tratará de manera uniforme a todos los agentes, lo que eventualmente conducirá a la expulsión de los agentes buenos (es decir, los conductores prudentes, los trabajadores responsables). Se habrá operado una selección adversa.

Mercado de los limones El tratamiento teórico del problema comenzó con Akerlof en su famoso paper "The market for lemons: Quality uncertainty and the market mechanism" (Akerlof, 1970).

En él, Akerlof explicó las implicaciones de las distintas calidades y de la información asimétrica en el mercado de coches usados. El autor se preguntó por qué un coche apenas usado puede llegar a tener un precio tan bajo en el mercado de segunda mano. El Premio Nobel descubrió que no se debe a que la gente no lo valore, sino a que existe asimetría informativa.

Akerlof analizó el mercado de coches de segunda mano en el que existen dos tipos:

- Coches de buena calidad, en proporción $1-\lambda$.
- Coches de mala calidad –o cacharros en el paper de Akerlof–, en proporción λ.

Este mercado se caracteriza por la existencia de información asimétrica, de manera que el vendedor conoce la calidad del coche usado, pero el comprador la desconoce. Si los compradores pudiesen distinguir la calidad de los coches, pagarían según su disposición a pagar por las distintas calidades: será mayor, P_B, para los coches buenos que para los coches malos, P_M.

Sin embargo, si no pueden distinguir entre los tipos de coche, entonces estarán dispuestos a pagar según la calidad media –al ser neutrales al riesgo, les basta con el valor esperado, sin que las probabilidades que asignan a cada tipo tengan por qué ser del 50 %: vienen dadas por la proporción real de cada tipo en el mercado, λ y $1-\lambda$–, probabilidades que se van ajustando a la nueva información:

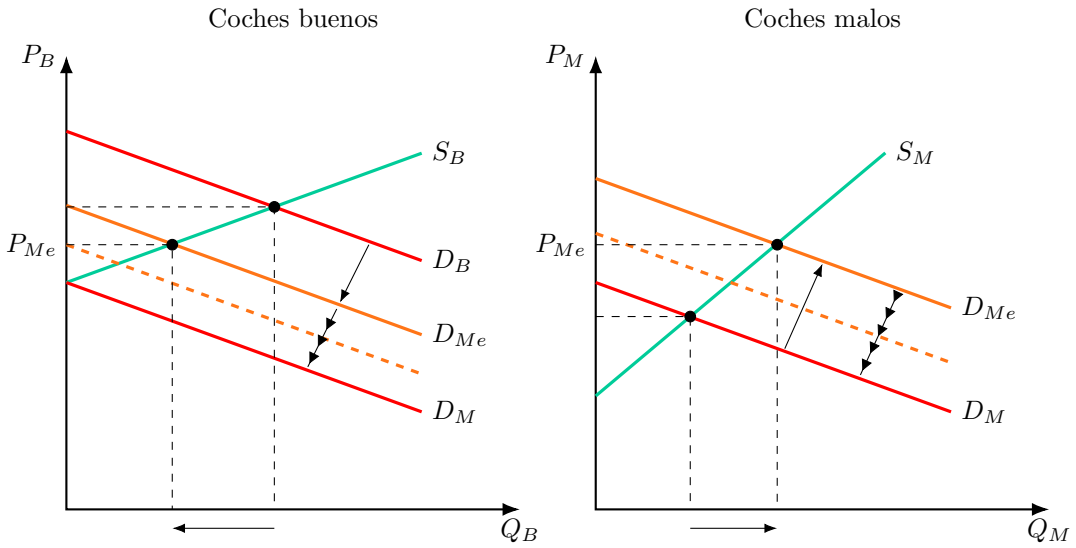
$$P_{Me} = (1-\lambda)\, P_B + \lambda \cdot P_M \tag{13.1}$$

En estas circunstancias, los vendedores de coches malos saldrían ganando, porque cobran más de lo que vale su coche, pero los de coches buenos saldrían perdiendo si el precio medio cae por debajo de la valoración mínima que exigen por su coche, en cuyo caso abandonarán el mercado[6].

[6]La intuición es que, si nadie puede comprobar el estado de un coche usado, quien tiene uno en perfecto estado prefiere quedárselo antes que venderlo al precio de un coche cualquiera. Al caer la calidad media de los coches en el mercado, los compradores van ajustando las probabilidades y de forma progresiva van bajando el precio que están dispuestos a pagar, lo que supone expulsar a más coches. Al final la calidad buena es expulsada del mercado y solo quedarán cacharros. El mercado de coches buenos desaparece por la información asimétrica, de manera que transacciones mutuamente beneficiosas no pueden llevarse a

Es posible analizar el problema gráficamente en dos planos de coches de calidad buena y mala con sus respectivos precios y cantidades demandadas, como se observa en la Figura 13.1

Figura 13.1: Modelo del mercado de limones de Akerlof



En el mercado de coches de segunda mano, los vendedores exigen un precio mayor por un coche bueno, recogido en la oferta S_B, que por un coche malo, S_M, y los compradores están dispuestos a pagar más por un coche bueno, recogido en la demanda D_B , que por un coche malo, D_M. En ausencia de información asimétrica se venderá cada coche en el equilibrio entre oferta y demanda.

Si existe información asimétrica y los compradores no pueden distinguir entre un coche de buena calidad y uno de mala calidad, su disposición a pagar vendrá dada por la demanda media, D_{Me}, con un precio medio, P_{Me}.

Dada la demanda media, se reducirá la cantidad de coches buenos ofrecidos en el mercado. Al haber menos coches buenos, los compradores revisarán su disposición a pagar, aumentando la cantidad de coches malos vendidos. El proceso se repetirá hasta que en el mercado no haya coches de buena calidad y solo se oferten coches de baja calidad. Este colapso total no es, sin embargo, un resultado automático: si la proporción de coches malos, λ, es reducida, el precio medio puede mantenerse por encima de la valoración mínima de los vendedores de coches buenos, y el mercado seguiría funcionando de forma

cabo. La clave para que los agentes buenos sean expulsados del mercado es que ambos pertenecen a un mismo mercado, llevando a un equilibrio agrupador (*pooling*). Podría ser eficiente intentar separar los mercados.

Este problema de información asimétrica también podría verse desde el punto de vista de las externalidades: el problema se halla en que los vendedores de cacharros generan una externalidad negativa, pues cuando deciden poner a la venta sus cacharros, afectan a la percepción que tienen los compradores de la calidad del coche medio del mercado, lo cual reduce el precio que están dispuestos a pagar por el coche medio, lo que perjudica a los vendedores de coches buenos.

 © 2026 Manuel Cabado García

mezclada (pooling), sin llegar a colapsar.

Los problemas de selección adversa no solo están presentes en el mercado de coches de segunda mano, sino también en muchos otros mercados.

Aplicaciones

Existen diversas aplicaciones de la teoría de la selección adversa como:

- Mercado de seguros: el asegurador no puede distinguir entre personas con un riesgo elevado y un riesgo reducido –salud, accidentes de circulación, etc.– y, de este modo establecerá un precio o prima media. Si este precio es elevado para los que tienen menores riesgos entonces no contratarán seguros, aumentando la proporción de individuos de riesgo elevado. Esto llevará a una subida de prima de modo que solo se asegurarían las personas de mayor riesgo, llevando a la desaparición del mercado ya que no sería rentable vender seguros. Esta situación justificaría la intervención estatal, estableciendo el seguro obligatorio.
- Mercado de capitales: Stiglitz y Weiss (1981) analizaron los problemas de selección adversa en los mercados de crédito: el prestamista –principal– concede préstamos y créditos a prestatarios cuya probabilidad de devolver la deuda desconoce. Si cobra un tipo de interés medio, este resulta demasiado alto para los prestatarios de menor riesgo y barato para los de mayor riesgo, de modo que cada subida del tipo expulsa a los mejores prestatarios y empeora la calidad media de su cartera. Por eso el prestamista puede preferir no subir el tipo de interés y racionar el crédito.
- Mercado de trabajo: el empleador no puede, a priori, distinguir entre los trabajadores más eficientes y productivos de los menos eficientes y productivos. Esto llevaría a que los trabajadores más productivos recibieran un salario inferior a su productividad y los menos productivos un salario superior a esta.

Soluciones: señalización y autoselección

Existen dos formas de resolver la selección adversa:

1. Señalización por parte del agente, que trata de señalizar información sobre sus características no observables a través de acciones sí observables. La iniciativa surge, pues, de la parte informada.
2. Autoselección o filtro (“Screening”) por parte del principal, que desarrolla mecanismos para distinguir a los tipos de agentes, de forma que el agente se identifica al elegir (autoselección). La iniciativa surge, pues, de la parte no informada.

Para el análisis, se toma el mercado de trabajo, donde se supone que hay dos tipos de trabajadores, diferenciados por su productividad: los trabajadores tipo A, más productivos,

y los B, menos productivos, de forma que $PMgL_A > PMgL_B$. La mitad de la población es del tipo A y la otra mitad del tipo B, pero no se puede determinar ex ante de qué tipo es cada empleado, pues la productividad no es observable a priori, ya que se trata de una situación de información asimétrica.

Modelo de Señalización de Spence

El modelo de (Spence, 1973) muestra cómo los trabajadores más productivos pueden señalizarse a través de la educación.

En efecto, aunque no se puede establecer a priori a qué tipo corresponde cada trabajador, sí puede analizarse una característica observable: la educación, E (medida en un índice de años de estudio, titulaciones, etc.). La clave radica en que la educación tiene un coste (financiero, de oportunidad, desutilidad, etc.), $C(E)$, y que ese coste es mayor para los individuos poco productivos que para los más productivos, $C_B(E) > C_A(E)$.

La empresa decidirá qué salario abona a los trabajadores en función de su nivel de educación. Para ello fija un nivel de educación umbral, $\hat{E}$, a partir del cual cree que el trabajador es productivo. De modo que:

- Si $E \geq \hat{E}$, el empresario pagará w_A, pues considera que el trabajador es del tipo productivo (es decir, tipo A).
- Si $E < \hat{E}$, el empresario pagará w_B, siendo menor que w_A.

Es decir, la empresa clasifica a los trabajadores como productivos o poco productivos según alcancen o no el nivel de educación umbral[7].

Condiciones de señalización Para que la educación actúe como una señal los salarios deben cumplir tres condiciones:

1. Restricción de participación: $w_A - C_A\left(\hat{E}\right) > w_R$, donde w_R es el salario de reserva, para que a los trabajadores A les compense educarse y trabajar antes que quedarse con su salario de reserva.
2. Restricción de incentivos 1: $w_A - C_A\left(\hat{E}\right) > w_B$, entonces a los trabajadores A les interesará obtener el nivel de educación $\hat{E}$.
3. Restricción de incentivos 2: $w_A - C_B\left(\hat{E}\right) < w_B$, entonces a los trabajadores B no les compensa obtener el nivel de educación $\hat{E}$.

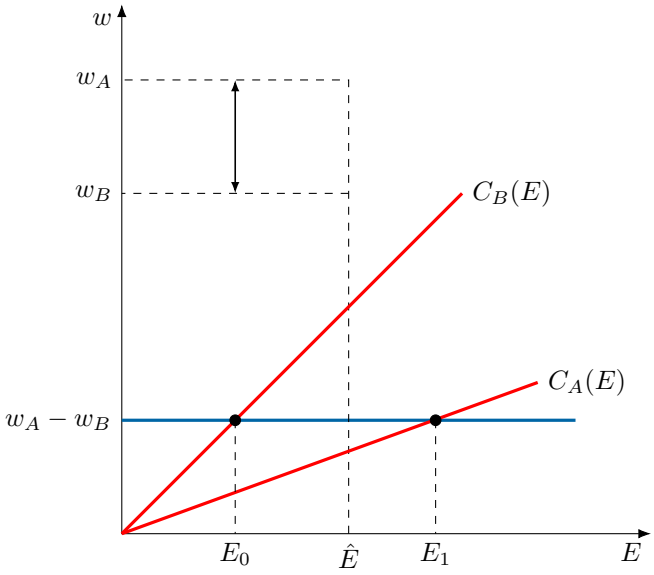
Gráficamente, en un plano de salarios, w, y nivel de educación, E, como el de la Figura 13.2, se representan las curvas de costes de la educación para los trabajadores A y B y para el nivel de educación señal, $\hat{E}$, se va a calcular la diferencia entre w_A y w_B,

[7] La intuición es la de un aspirante con un título exigente: a la empresa no le importa tanto lo que aprendió como el hecho de que haya sido capaz de obtenerlo.

 © 2026 Manuel Cabado García

y trasladar dicha diferencia hacia el eje de abscisas. Esto determinará cómo el punto de la diferencia quedará por encima de la curva de costes de los trabajadores productivos, $C_A(E)$, es decir, les compensará acceder al nivel de educación señal $\hat{E}$, pero quedará por debajo de la curva de costes de los trabajadores poco productivos, $C_B(E)$, por lo que no les compensará acceder a dicho nivel de educación.

Figura 13.2: Modelo de señalización de Spence



Asimismo, cabe destacar que los valores de equilibrio de E son múltiples: todos aquellos para los que el coste de incurrir en ese nivel de educación es lo suficientemente alto para el tipo B como para que no le compense incurrir en él para acceder al mayor salario (es decir, por encima de E_0), pero lo suficientemente bajo como para que al tipo A sí le interese educarse (es decir, por debajo de E_1). Esta estructura de salarios dará lugar a un equilibrio separador (separating equilibrium), es decir, los individuos de diferente tipo toman decisiones diferentes respecto de la señalización[8].

Otras señales serían las garantías en el mercado de coches usados, marcas, estándares de calidad, licencias/permisos/homologaciones (profesionales liberales, estrellas de un hotel), auditorías, calificaciones de agencias de rating (para inversiones), etc.

No obstante, es preciso subrayar que las señales son efectivas porque son costosas en su justa medida: si fuesen poco costosas, todo el mundo se señalizaría[9]. Si fuesen muy

[8] Si, para esos costes de educarse, el nivel de educación exigido por las empresas fuera menor que E_0, todos los individuos se educarían, pues para ambos la ganancia salarial sería mayor que el coste, por lo que no habría señal ninguna y el equilibrio sería agrupador (*pooling equilibrium*). Si, por el contrario, el nivel de educación exigido por las empresas fuera mayor que E_1, entonces ningún individuo se educaría, por lo que, de nuevo, no habría señalización, y se generaría también un equilibrio agrupador.

[9] Algunos autores han utilizado el modelo de señalización de Spence para explicar las consecuencias de la implantación de los planes de Bolonia en España. En la medida en que, dicen, los actuales planes de Bolonia son más fáciles que las antiguas licenciaturas (por el mero hecho, siquiera, de que se da menos contenido al ser las carreras más cortas), los costes del esfuerzo son menores que antes. De esta manera,

costosas, nadie se señalaría. Así, se alcanzaría un equilibrio agrupador.

De esta manera, porque hay que incurrir en un coste adicional, se trata de una solución de second best:

- Conforme aumenta el nivel de educación exigido, $\hat{E}$, los trabajadores A ven reducido su bienestar, porque pagan un coste mayor por el mismo salario. Los trabajadores B, que no se educan, cobran w_B en lugar del salario medio que obtendrían si nadie pudiera señalizarse.
- La señalización tiene un coste que disminuye el bienestar global. En la actualidad los estudios de grado se han normalizado, por lo que los agentes más productivos tienen que incurrir en costes mayores (máster).
- Crítica desde la teoría del capital humano de Becker (1964): el modelo no considera que la educación pueda aumentar la productividad, al concebirla únicamente como un filtro de habilidades preexistentes. En la práctica, la educación cumple probablemente ambas funciones a la vez[10].

Señal vs Capital Humano Como curiosidad, el modelo de Spence contrasta con la teoría del capital humano de Becker (1964), en la que la educación es útil porque aumenta la productividad (algo que no asume el modelo de Spence):

- Spence: ↑ productividad ⇒ ↑ educación (señal) ⇒ ↑ salario.
- Becker: ↑ educación (inversión) ⇒ ↑ productividad ⇒ ↑ salario.

Autoselección: el principio de revelación de Myerson

Estos modelos se basan en el principio de revelación de Myerson (1979)[11]. En ellos el principal diseña un menú de contratos de manera que cada uno de ellos va dirigido a un tipo específico de agentes. El diseño de este menú ha de ser tal que cada individuo esté

al nivel de educación exigido anteriormente por las empresas (por ejemplo, una carrera, $\hat{E}$) se produce un equilibrio agrupador, pues al ser las carreras más fáciles (es decir, menos costosas), las curvas de coste se desplazan hacia abajo, y más agentes tienen incentivos a obtener ese nivel de educación. De ahí el fenómeno que se ha observado en los últimos años de que las empresas están exigiendo, además de una carrera, un máster, pues, ante unos menores costes de educarse, es necesario exigir mayor nivel de educación para que se genere un equilibrio separador.

Kelly Bedard (Universidad de California, Santa Barbara) encontró precisamente este hecho al estudiar el efecto del acceso a la universidad sobre el abandono escolar en EEUU: en los mercados laborales locales donde había una universidad cercana aumentó el número de universitarios, pero también se incrementó la tasa de abandono escolar en secundaria. La razón es que, al caer la productividad media del pool de individuos con un diploma de estudios secundarios, cayó el salario esperado, y para los agentes menos productivos el coste del esfuerzo de acabar la educación no compensaba el incremento salarial que obtendrían al recibir el diploma.

[10] Arrow (1973) llega a una conclusión parecida a la de Spence al concebir la enseñanza superior como un filtro que identifica a los más capaces, aunque no aumente su productividad.

[11] El principio de revelación también fue planteado por otros autores como Gibbard (1973) y Green y Laffont (1977).

 © 2026 Manuel Cabado García

mejor con el contrato pensado para él (restricción de autoselección o de compatibilidad de incentivos), de forma que, al elegir, revele su tipo. El principio de revelación garantiza que limitarse a este tipo de menús no supone ninguna pérdida: cualquier resultado que el principal pudiera alcanzar con un mecanismo más complicado puede lograrse también con uno en el que a cada agente le conviene revelar la verdad[12].

Algunos ejemplos de menús de contratos son:

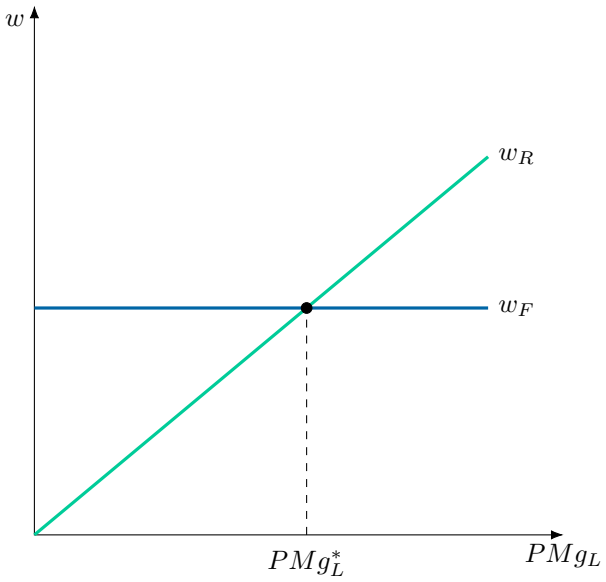
- Modelo de grilletes dorados de Salop y Salop (1976): la empresa, que invierte en formar a sus trabajadores, ofrece un contrato con salario bajo los primeros años, que irá aumentando con la antigüedad (por ejemplo, un contrato de prácticas). El trabajador que prevé marcharse pronto sabe que no llegará a cobrar los salarios más elevados, por lo que no le interesará el contrato, y solo lo aceptarán quienes piensan quedarse[13].
- Modelo de autoselección de Milgrom y Roberts (1992): la empresa ofrece dos tipos de salarios: uno fijo y otro variable. Gráficamente, en un plano de salario, w, y productividad marginal del trabajo, $PMgL$, como el de la Figura 13.3, se pueden representar los salarios fijos, w_F, y los salarios variables, w_R, que serán crecientes con el rendimiento. Si se trata de un trabajador de productividad baja, es decir, si su $PMgL \in [0, PMgL^*]$, elegirá el salario fijo, mientras que si tiene productividad alta, elegirá el salario variable.

[12] La intuición es que una tienda que quiere distinguir a los clientes con prisa de los que no la tienen no necesita interrogarlos: le basta con ofrecer un envío urgente caro y uno normal barato y dejar que cada cliente elija. Estos menús de contratos, basados en el principio de revelación de Myerson, son los que se ven por ejemplo en:

- El mercado de seguros de Rothschild y Stiglitz (1976): En este modelo, las compañías de seguros (principales) se enfrentan a individuos (agentes) con diferentes niveles de riesgo de accidente (por ejemplo, alto y bajo riesgo), pero no pueden observarlo directamente. Para inducir a los agentes a revelar su tipo, las aseguradoras ofrecen un menú de dos contratos: uno con alta cobertura y alta prima (atractivo para los de alto riesgo) y otro con baja cobertura y baja prima (atractivo para los de bajo riesgo). La clave es que los contratos están diseñados de manera que los individuos de bajo riesgo no quieran elegir el contrato de alta cobertura (debido a su alta prima para un riesgo que no tienen), y los de alto riesgo no quieran elegir el de baja cobertura (porque no les brinda suficiente protección), logrando así la auto-selección.
- El modelo de Laffont y Tirole (1986), que es parte de una vasta literatura sobre la regulación de monopolios naturales bajo asimetría de información, también emplea menús de contratos, pero con un enfoque que puede incluir tanto selección adversa como riesgo moral. En este contexto, el regulador (principal) busca que una empresa monopolista (agente) produzca de manera eficiente, pero desconoce tanto la verdadera estructura de costes de la empresa (selección adversa, por ejemplo, si es de bajo o alto coste) como el esfuerzo real que la empresa dedica para reducir costes (riesgo moral). El regulador ofrece un menú de contratos que relaciona la tarifa/precio con el beneficio o los costes declarados por la empresa. Los contratos se diseñan para incentivar a la empresa a revelar su tipo de coste y, a la vez, a esforzarse por reducir costes. Así, el regulador ofrece a las empresas de bajo coste un contrato de incentivos fuertes, próximo al de precio fijo, en el que se quedan con todo el ahorro, y a las de alto coste uno de incentivos débiles, más cercano al de reembolso de costes, respetando las restricciones de compatibilidad de incentivos y de participación.

[13] La intuición es la de los planes de pensiones o las acciones de la empresa que solo se consolidan tras varios años: atan al trabajador y, por eso, atraen sobre todo a quien ya pensaba quedarse.

Figura 13.3: Modelo de autoselección



En suma, cuando las empresas compiten ofreciendo menús de contratos, como en el mercado de seguros de Rothschild y Stiglitz (1976), puede no existir equilibrio, y si existe ha de ser separador. Además, en ese equilibrio los agentes de peor tipo obtienen la misma utilidad que con información simétrica, mientras que los de mejor tipo obtienen menos, lo que refleja el coste asociado a la existencia de información asimétrica.

13.3. Riesgo Moral ⏳ 7'

El riesgo moral es un problema de información asimétrica que surge con posterioridad a la firma del contrato, y se debe al riesgo de acciones ocultas por una de las partes que modifican el valor de la transacción para la otra parte.

Para que exista una situación de riesgo moral, debe darse:

1. Información incompleta: se pueden dar distintos estados de la naturaleza, y los agentes son capaces de asignar probabilidades a cada uno de ellos.
2. Información asimétrica: acciones ocultas, es decir, acciones de una de las partes que no son perfectamente observables por la otra.
3. Conflicto de intereses entre el principal y el agente. Mientras que el objetivo del agente es la maximización de su utilidad, el del principal es la maximización de su beneficio. El problema es que cuanto mayor sea el esfuerzo del agente, mayor será el beneficio del principal, pero el esfuerzo genera desutilidad para el agente.
4. Debe haber un problema de reparto eficiente de riesgos que entre en conflicto con el diseño de incentivos.

 © 2026 Manuel Cabado García

El modelo paradigmático de riesgo moral es el del mercado de trabajo, que es el que se presenta a continuación:

- El problema puede modelizarse a partir de la relación entre un empresario (principal) que ofrece un contrato a un trabajador (agente), cuyo nivel de esfuerzo no puede observar.
- Los resultados no son un buen indicador del esfuerzo, ya que la relación entre ellos no es unívoca: cuanto más se esfuerce el agente, más probabilidades habrá de que los resultados sean altos, pero un esfuerzo alto no genera necesariamente un beneficio alto. De ahí que la relación entre esfuerzo y resultado se defina a partir de una función de distribución probabilidad.
- La solución al problema pasa por el diseño por parte del principal de un contrato que incentive al agente a desempeñar el nivel de esfuerzo óptimo.

Para analizar cómo sería el contrato óptimo, se parte (para centrar el tema) de una situación en la que no hay asimetrías de información, para ver posteriormente cómo varía la solución con asimetría.

Contrato óptimo con esfuerzo observable (no existe riesgo moral)

Con el fin de resolver el problema se parte de los siguientes supuestos:

- Información simétrica pero incompleta: los ingresos son una variable aleatoria cuyos posibles valores y probabilidades conocen ambas partes, es decir, hay riesgo.
- Relación bilateral entre principal y agente, entre los que tiene lugar un juego en dos etapas:
    - El principal (empresario) propone al agente un contrato exigiéndole el nivel de esfuerzo que maximiza sus beneficios, e*.
    - El agente (trabajador) decide si acepta o no el contrato con el nivel de esfuerzo exigido por el principal. Las preferencias sobre el riesgo vienen expresadas por una función tipo von Neumann-Morgenstern (vNM).
- Los resultados esperados (ingresos, X) del principal dependen de:
    - El estado de naturaleza que se dé, i, de todos los posibles, n, teniendo cada uno una probabilidad pi de materializarse. De ahí que se hable de resultados esperados. O lo que es lo mismo: la transacción está sujeta a un riesgo, lo que enriquece el modelo, ya que el esquema de pagos óptimo dependerá también de la actitud del principal y del agente ante el riesgo.
    - El esfuerzo del agente puede ser más alto o más bajo, y afecta a las probabilidades de que se dé un estado de la naturaleza u otro . El esfuerzo, en este

primer modelo, es verificable (es decir, información simétrica –no hay riesgo moral–). Concretamente, la distribución de probabilidad de un esfuerzo alto domina estocásticamente a la de un esfuerzo bajo, es decir, que cuanto mayor sea el esfuerzo más probabilidades habrá de que se dé un buen resultado.

Por lo tanto, la función de resultados (ingresos) esperados es:

$$EX = \sum_{i=1}^{n} p_i(e) \cdot x_i \tag{13.2}$$

donde x_i es el resultado si ocurre el estado de la naturaleza i. Además, como se apuntaba antes, cuanto más alto sea e, más probabilidades habrá de que x sea alto.

Por su parte, la utilidad esperada del principal, que depende del beneficio que le queda tras pagar el salario, será

$$BE = \sum_{i=1}^{n} p_i(e^*) \cdot B[x_i - w(x_i)] \tag{13.3}$$

donde el salario del agente dependerá, en el planteamiento inicial, del resultado finalmente obtenido.

El principal se enfrenta a un trade-off: cuanto mayor sea el esfuerzo del agente, mayor será el ingreso esperado, pero mayor será también el salario que tiene que pagarle. Así, el nivel de esfuerzo que maximiza su función objetivo será e^*. Nótese que el problema del principal no termina aquí: su problema más importante, de hecho, es determinar el salario asociado al esfuerzo que maximiza sus beneficios, e^*, en función tanto de su actitud frente al riesgo como de la del agente.

La pregunta es: ¿aceptará el trabajador el contrato? La utilidad esperada del agente será:

$$UE = \sum_{i=1}^{n} [p_i(e^*) \cdot u(w(x_i))] - v(e^*) \tag{13.4}$$

donde v es la desutilidad del agente generada por el esfuerzo exigido por el principal, que será mayor cuanto mayor sea el esfuerzo.

De esta forma, el agente aceptará el contrato si la utilidad esperada de trabajar con ese nivel de esfuerzo es al menos igual a su utilidad de reserva, es decir, a la que obtendría con su mejor alternativa (por ejemplo, el salario en otra empresa o el subsidio por desempleo). Esta es la llamada restricción de participación[14].

Así, en el programa de optimización el problema del principal será determinar el esquema de pagos óptimo, es decir, el que maximiza sus beneficios y, al mismo tiempo,

[14] La intuición es que nadie acepta un empleo que le deje peor que el que ya tiene o que el subsidio por desempleo.

 © 2026 Manuel Cabado García

hace que el trabajador prefiera aceptarlo a quedarse con su utilidad de reserva:

$$\text{máx} \quad BE = \sum_{i=1}^{n} p_i\left(e^*\right) \cdot B\left[x_i - w\left(x_i\right)\right] \tag{13.5}$$

sujeto a la restricción de participación

$$\sum_{i=1}^{n} \left[p_i\left(e^*\right) \cdot u\left(w\left(x_i\right)\right)\right] - v\left(e^*\right) \geq U_R \tag{13.6}$$

donde la CPO

$$\frac{B'\left[x_i - w\left(x_i\right)\right]}{u'\left(w\left(x_i\right)\right)} = \lambda \tag{13.7}$$

La restricción de participación se va a cumplir siempre en igualdad, es decir, el empresario no va a pagar más de lo que sea necesario, obteniendo el trabajador la Utilidad de reserva (utilidad mínima por la que estaría dispuesto a trabajar), U_R. Por otro lado, λ mide el aumento marginal de la FO cuando la utilidad aumenta marginalmente. La restricción se satura (con B', $U' > 0$ por lo que $\lambda > 0$).

Solución: actitudes frente al riesgo La solución al problema dependerá de la actitud frente al riesgo del principal y del agente:

1. Principal neutral y agente averso: si el principal es neutral, entonces B' (utilidad marginal del beneficio) será constante, porque su función de utilidad sobre el beneficio, B, es lineal ($B' > 0$; $B'' = 0$; $U' > 0$, $U'' < 0$) (si el principal es neutral será indiferente a incurrir en un mayor riesgo, solo le importa el beneficio esperado). Como el esquema óptimo implica que $B'/u' = \lambda$, y λ también es constante, entonces u' (utilidad marginal) deberá ser necesariamente constante para satisfacer dicha condición. Pero si el agente es averso al riesgo, se sabe que u' no será constante, sino decreciente. Por tanto, la única forma de que u' sea constante es que el salario sea siempre el mismo con independencia del resultado empresarial. Conclusión: si el principal es neutral y el agente es averso, el reparto óptimo del riesgo consiste en que el principal asuma todo el riesgo de los resultados, de forma que el principal ofrecerá un contrato al agente en el que aquél deberá asumir todo el riesgo de la operación y pagar al agente un salario constante[15], de manera que el agente estará completamente asegurado. El contrato ofrecido por el principal sería del tipo: "Tendrías que realizar un esfuerzo e*, y tu salario será fijo". Y si el trabajador acepta el contrato, solo puede realizar e*, y no menos, ya que éste es observable, y recibirá un salario fijo (se verá con más detalle más adelante).

[15] Como el agente es averso al riesgo, valora un salario variable por debajo de su esperanza matemática (su equivalente cierto es menor). Por eso, para garantizarle su utilidad de reserva, al principal le resulta más barato pagarle una cantidad fija que un salario variable, que tendría que incluir una prima de riesgo.

2. Principal averso y agente neutral: si el principal es averso, entonces B' será decreciente (porque ahora $B'' < 0$ y $U'' = 0$). Como el esquema óptimo implica que $B'/u' = \lambda$, y λ también es constante, entonces B' deberá ser necesariamente constante para satisfacer dicha condición. Pero si el principal es averso al riesgo, se sabe que B' no será constante, sino decreciente. Por tanto, la única forma de que B' sea constante es que el beneficio del principal sea siempre el mismo con independencia del resultado empresarial. Conclusión: si el principal es averso y el agente es neutral, el reparto óptimo del riesgo consiste en que el agente asuma todo el riesgo de los resultados, de forma que el principal ofrecerá un contrato al agente en el que éste deberá asumir todo el riesgo de la operación y pagar al principal un canon constante (es decir, contrato de franquicia), de manera que el principal estará completamente asegurado.
   ¿Cuál de los dos escenarios es más común? El primero, ya que: i) el principal puede diversificar y ii) el empresario puede hacer difusión del riesgo entre los accionistas.
3. Principal y agente aversos: Será igual que en los casos anteriores, pero la distribución dependerá de la aversión al riesgo de cada uno. Por lo tanto, si el principal es averso al riesgo y el agente es averso, la solución pasará por que ni el beneficio ni el salario sean constantes. Concretamente será mayor en la situación de éxito que en la de fracaso. Así, el reparto óptimo del riesgo consiste en que ambos asuman parte del riesgo. El agente más averso estará más protegido.
4. Principal neutral y agente neutral: Cualquier reparto es óptimo (suponiendo que las pendientes de las curvas de indiferencia son iguales).

En suma, bajo información simétrica se produce un reparto óptimo de riesgos siendo además la solución óptima de Pareto en términos de eficiencia. La maximización del beneficio esperado sujeto a la restricción de participación garantiza que el agente realice el esfuerzo deseado por el principal. La solución en cambio variará en función de los perfiles con respecto al riesgo de ambos.

Contrato óptimo con esfuerzo no observable (existe riesgo moral)

En el caso de que el esfuerzo no sea observable (existe riesgo moral), se mantienen los mismos supuestos que antes, con la diferencia de que ahora el esfuerzo del agente no es observable por parte del principal, por lo que habrá información asimétrica. El juego ahora es de tres etapas:

1. El principal (empresario) propone al agente un contrato exigiéndole el nivel de esfuerzo que maximiza sus beneficios, e^*.
2. El agente (trabajador) decide si acepta o no el contrato.

 © 2026 Manuel Cabado García

3. Tras aceptarlo, el agente decide si realiza el esfuerzo exigido por el principal, e^*, o uno más bajo, e^b[16].

Dado que ahora el nivel de esfuerzo del agente no es observable, si el principal ofrece un salario constante como antes, el agente tendrá incentivos a realizar un nivel de esfuerzo menor al óptimo, $e^b < e^*$. De ahí que, en presencia de riesgo moral, el principal tenga que rediseñar el esquema de pagos.

Esquema de pagos óptimo Ahora, el esquema de pagos óptimo deberá contemplar, además de la maximización del beneficio y de la restricción de participación, una restricción adicional[17]: la de incentivos[18]. En efecto, tendrá que ofrecer un esquema de pagos tal que la utilidad del agente sea mayor (o igual) si realiza el esfuerzo óptimo exigido por el principal, e^*, que si realiza uno menor, e^b:

$$\sum_{i=1}^{n} \left[p_i\left(e^*\right) \cdot u\left(w\left(x_i\right)\right)\right] - v\left(e^*\right) \geq \sum_{i=1}^{n} \left[p_i\left(e^b\right) \cdot u\left(w\left(x_i\right)\right)\right] - v\left(e^b\right) \tag{13.8}$$

ahora ambas restricciones también se saturan (la restricción de participación y la de incentivos), dando lugar a la nueva CPO:

$$\frac{B'\left[x_i - w\left(x_i\right)\right]}{u'\left(w\left(x_i\right)\right)} = \lambda + \mu\left[1 - \frac{p_i\left(e^b\right)}{p_i\left(e^*\right)}\right], \quad \forall i \tag{13.9}$$

donde ahora aparecen dos multiplicadores, λ y μ, al haber sujetado la función objetivo a dos restricciones.

La ratio de verosimilitud es la probabilidad de obtener el resultado x_i si el esfuerzo ha sido menor al óptimo, $p_i\left(e^b\right)$, dividida por la probabilidad de obtener el resultado x_i si el esfuerzo ha sido el óptimo, $p_i\left(e^*\right)$. μ es el multiplicador de la restricción de incentivos y mide cuánto le cuesta al principal, en el margen, tener que incentivar al agente. Si μ fuese igual a 0, la restricción no sería activa y el contrato coincidiría con el de información simétrica. Con riesgo moral la restricción es activa, de modo que μ es positivo y la asimetría de información tiene un coste.

De la CPO se derivan tres implicaciones encadenadas:

1. Relación negativa entre los ingresos y la ratio de verosimilitud: por la propiedad de la ratio de verosimilitud monótona, cuanto mayor es el resultado obtenido, x_i,

[16] Nótese que para que se dé el problema de riesgo moral es necesario que el nivel de esfuerzo que maximiza los beneficios del principal, e*, no sea el mínimo que puede realizar el agente, pues en ese caso no sería necesario incentivarle (pues ya realizaría por sí mismo este esfuerzo óptimo).

[17] Basta con añadirla al programa de optimización anterior, sin repetir sus ecuaciones.

[18] Es decir, la idea de la restricción de incentivos es que el agente prefiera realizar el esfuerzo alto a pesar de que le suponga un mayor esfuerzo (y por tanto mayor desutilidad) porque aumenta sus probabilidades de tener un output mayor y por tanto mayor utilidad esperada que si hiciese el esfuerzo bajo, que aunque supusiese una menor desutilidad del esfuerzo, en términos totales de utilidad esperada sería menor que si se realizase el nivel alto de esfuerzo.

menor es la ratio, ya que un resultado alto es relativamente más probable bajo el esfuerzo óptimo que bajo un esfuerzo bajo[19].

2. Relación negativa entre la ratio de verosimilitud y el salario: por la CPO arriba vista, y asumiendo que el principal es neutral (por lo que B' es constante) y el agente averso, cuanto mayor es la ratio, menor es el término entre corchetes, $(1 - \text{ratio})$, y por tanto menor es B'/u', lo que implica una u' mayor. Y como u' es decreciente en el salario por la aversión al riesgo del agente[20], una u' mayor exige un salario menor.
3. Combinando ambas relaciones (negativa con negativa), existe una relación positiva entre los ingresos y el salario. Por tanto, el esquema de pagos óptimo es un salario variable que aumenta conforme lo hacen los ingresos (los beneficios para el principal son menores que con información simétrica pero mayores que sin incentivos, porque ahora lo óptimo para el principal es que el agente asuma parte del riesgo de la relación, pero como es averso al riesgo habrá que compensarle dándole un salario esperado mayor para mantener su utilidad en el nivel de la utilidad de reserva). Es decir, se traslada parte del riesgo al agente (subóptimo).

En suma, en este caso existe un coste de agencia, es decir, una pérdida de eficiencia respecto a la situación con información simétrica, que da lugar a un second best (equilibrio Pareto inferior). La razón de esta ineficiencia es que se traslada parte del riesgo al agente averso, cuando lo óptimo es que lo soporte íntegramente el neutral.

Implicaciones en política económica

A la hora de diseñar un contrato para cualquier política pública, en el caso de que exista un problema de Riesgo Moral, será necesario diseñar un esquema de pagos dependiente de los resultados para asegurar que el esfuerzo del agente sea óptimo.

Debido a que el Riesgo Moral es un fallo de mercado que conduce a una situación subóptima en términos de bienestar, el diseño de políticas que eliminen la información asimétrica puede ser muy beneficioso (eliminar este fallo de mercado conducirá a una situación Pareto Superior con total certeza cuando no existan otros fallos de mercado, de acuerdo con la Teoría del Second Best)[21].

[19] Cuanto mayor sea $p_i\left(e^b\right)/p_i\left(e^*\right)$, menor probabilidad de que x_i sea alto.

[20] Recuérdese que $\uparrow u'$ si $\downarrow$w porque la función de utilidad, u, es cóncava, por lo que la utilidad marginal, u', es decreciente a medida que aumenta el salario (o, lo que es lo mismo, u' es creciente a medida que disminuye el salario).

[21] La clave estará en encontrar el método que elimine la información asimétrica con el menor coste. Esto es especialmente importante en los países en desarrollo, donde muchas veces no se concede crédito por el riesgo de impago. Una solución estudiada en la literatura es la responsabilidad conjunta o "peer monitoring" (Stiglitz, 1990): al condicionar las deudas de un grupo de agentes al pago de cada uno de los miembros, se traslada el coste de supervisión de la entidad financiera al grupo de deudores, permitiendo ofrecer crédito a costes menores, a cambio de trasladar parte del riesgo a los propios prestatarios.

 © 2026 Manuel Cabado García

Aplicaciones

A continuación se presentan diferentes aplicaciones de la teoría de la agencia:

- Contrato de seguros: El campo de los seguros es el caso típico donde se pueden generar problemas de riesgo moral : el asegurado no tendrá incentivos a ser diligente y evitar el riesgo. Una posible solución es la cobertura parcial de los costes del siniestro, trasladando parte del riesgo al agente.
- Separación entre propiedad y gestión: En empresas grandes, los directivos suelen ser agentes distintos a los propietarios, y sus incentivos pueden estar desalineados (pudiendo llevar a cabo acciones que conlleven un mayor beneficio a corto plazo a cambio de un mayor riesgo). El diseño de un contrato correcto será crucial.
- Paquetes de rescate macroeconómicos: Una vez recibida la ayuda, los gobiernos pueden no tener incentivos a reconducir la política económica que les había abocado a la situación de emergencia. La solución pasará por imponer condicionalidad a las ayudas.
- Mercado de crédito: una entidad financiera puede preferir racionar el crédito antes que subir los tipos de interés, ya que la subida empujaría a los deudores hacia proyectos más arriesgados, con mayor rendimiento si salen bien pero mayor probabilidad de impago, lo que aumentaría la morosidad. Es el modelo de racionamiento del crédito de Stiglitz y Weiss (1981).

13.4. Selección adversa y riesgo moral a la vez ⏳ 2'

Los problemas de selección adversa y riesgo moral no son mutuamente excluyentes. En efecto, pueden encontrarse situaciones en las que ambos problemas se dan simultáneamente.

En el marco de la Nueva Economía Keynesiana, Stiglitz y Weiss (1981) mostraron que el mercado de crédito puede estar caracterizado en equilibrio por un racionamiento del crédito. La existencia de información asimétrica en los mercados financieros podría provocar que ante un exceso de demanda de crédito para unos tipos de interés dados, el banco prefiera racionar la oferta y mantener los tipos de interés y el exceso de demanda en vez de aumentar los tipos de interés debido a los problemas de selección adversa y riesgo moral que podrían darse.

La principal razón para la no subida de los tipos es la posible aparición de problemas de dos tipos:

- Selección Adversa: Los agentes con mejor expediente crediticio no estarán dispuestos a pagar dicho tipo de interés dado que es demasiado elevado para las operaciones de inversión que desean hacer. No obstante aquellos agentes con peor expediente

crediticio y con proyectos de inversión más arriesgados sí estarán dispuestos a pedir prestado. Los prestatarios con buen perfil de crédito se verán expulsados del mercado. Los restantes inversores pondrán en peligro la estabilidad financiera.

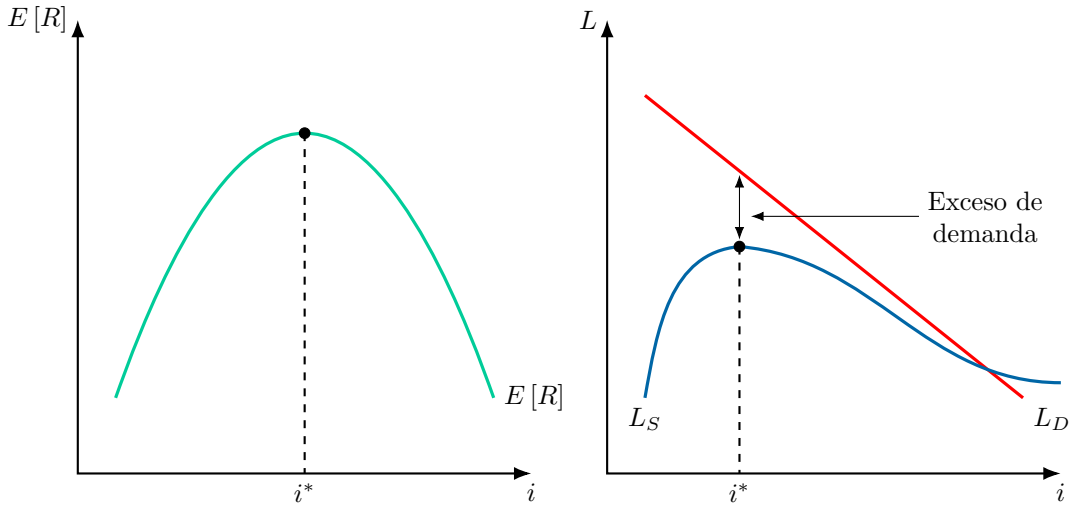
- Riesgo Moral: los mayores costes de financiación harán que los individuos se vean obligados a llevar a cabo inversiones más arriesgadas con el dinero prestado surgiendo así problemas de riesgo moral que de nuevo pondrán en peligro la estabilidad de las entidades de crédito.

Gráficamente, como se observa en la Figura 13.4, se puede representar el problema:

- El gráfico de la izquierda muestra el retorno esperado del banco, $E[R]$, el cual se calcula como la probabilidad de repago multiplicada por el tipo de interés, i. Al principio la curva es creciente hasta el punto i^*, a partir del cual la probabilidad de repago es tan reducida que las posibles ganancias de un tipo mayor son superadas por la baja probabilidad de repago.
- El gráfico de la derecha muestra una curva de demanda de préstamos, L_D, decreciente y una curva de oferta, L_S, cóncava con un máximo correspondiente al tipo de interés máximo, i^*, mencionado anteriormente.

Para que se vacíe el mercado de crédito, los bancos deberían fijar el tipo de interés correspondiente al cruce entre ambas curvas. Sin embargo, por encima de i^* su rendimiento esperado disminuye, por lo que prefieren mantener i^*. No obstante los problemas de riesgo moral y selección adversa llevan a un racionamiento de la provisión de crédito dado por la diferencia entre oferta y demanda.

Figura 13.4: Modelo de racionamiento del crédito de Stiglitz y Weiss



 © 2026 Manuel Cabado García

Implicaciones

En definitiva, la información asimétrica en el mercado de crédito provoca un racionamiento, es decir, una pérdida de transacciones potencialmente beneficiosas. En la práctica existen mecanismos para desvelar más información, como por ejemplo el empleo por parte del demandante de crédito de la existencia de un colateral o de un historial de crédito[22].

Este modelo tiene importantes implicaciones en crecimiento económico, ya que el racionamiento del crédito puede llevar a la falta de implementación de proyectos de inversión potencialmente rentables y con ello, reducir el crecimiento económico[23].

Este modelo, además, arroja implicaciones de política monetaria. En una situación de racionamiento como la anterior, una política monetaria expansiva del Banco Central podría aumentar la oferta de préstamos hasta hacer desaparecer el racionamiento de crédito. Por lo tanto, dicha política tendría efectos reales. Un buen ejemplo serían los planes de Quantitative Easing llevados a cabo por el Banco Central Europeo, la Reserva Federal en EEUU y otros Bancos Centrales.

En la Figura 13.4, esto provocaría un desplazamiento hacia arriba de la curva de oferta de préstamos, de modo que esta llegaría a cortar a la curva de demanda justo en su máximo, haciendo desaparecer el racionamiento de crédito, es decir, el exceso de demanda sería cero.

Finalmente, una forma de medir la información asimétrica en el mercado de crédito es la constatación por parte de Holmström y Tirole (2011) de que las grandes empresas del S&P500 en Estados Unidos acumulan una cifra cercana a 1,4 billones de dólares de efectivo. Este comportamiento se justifica precisamente por el racionamiento en el mercado de crédito derivado de la información asimétrica. Las empresas acumulan efectivo ya que la autofinanciación es una forma alternativa de financiarse, de acuerdo con la teoría de las jerarquías financieras.

13.5. Conclusiones ⏳ 30"

En conclusión, la información asimétrica genera fallos de mercado que impiden alcanzar el óptimo de Pareto, ya sea por selección adversa –cuando la parte no informada no puede distinguir la calidad de su contraparte antes de contratar, como en el mercado de limones de Akerlof– o por riesgo moral –cuando no puede observar su comportamiento tras la firma del contrato–, pudiendo darse ambos problemas a la vez, como en el racionamiento del crédito de Stiglitz y Weiss (1981). Las soluciones estudiadas –señalización, criba, menús

[22] Por ejemplo, es práctica habitual para alquilar un inmueble que se requieran nóminas o declaraciones de la renta que puedan garantizar al arrendador que el arrendatario podrá hacer frente al pago del alquiler.

[23] Una posible solución que ha tenido cierto éxito en los países en desarrollo son los microcréditos. Los miembros de una comunidad tienen más información que el banco de forma que acaban excluyendo a los propios morosos mediante un sistema tal que cuando se produce un impago se le deniega el crédito en el futuro.

de contratos incentivo-compatibles– nunca alcanzan el óptimo de información perfecta, sino una solución de second best.

Esta distorsión no solo afecta a los mercados tradicionales, sino que se manifiesta de forma aún más compleja en la era digital y en las cadenas de agencia o modelos multicapa que involucran a varios agentes y principales interconectados, desde la fiabilidad de las plataformas online hasta la gestión pública delegada. Las soluciones, si bien no siempre alcanzan el óptimo teórico, deben ser innovadoras, aprovechando herramientas digitales para la señalización o la supervisión, como las tecnologías blockchain y el big data. Asimismo, la regulación y el diseño institucional son esenciales para solucionar estas asimetrías, como muestran las recientes Leyes europeas de Mercados Digitales, de Servicios Digitales, y de IA.

 © 2026 Manuel Cabado García

Apéndice

13.A.1. El papel de la CESCE ante la selección adversa

CESCE (Compañía Española de Seguros de Crédito a la Exportación) es un excelente ejemplo de cómo una política económica, o en este caso una empresa pública (aunque con actividad comercial), puede actuar para mitigar la selección adversa en un mercado específico: el seguro de crédito y de caución para operaciones internacionales. CESCE es una sociedad anónima mayoritariamente participada por el Estado que actúa como la agencia oficial de crédito a la exportación española, gestionando por cuenta del Estado el seguro de los riesgos comerciales, políticos y extraordinarios asociados a la internacionalización de las empresas españolas[24].

- Contexto de la selección adversa:
  - Principales: Las empresas exportadoras españolas que venden bienes o servicios en el extranjero, o las entidades financieras que les otorgan créditos o avales para estas operaciones.
  - Agentes: Los compradores extranjeros (importadores) o los países donde se realizan las operaciones.
  - Asimetría de información: Los compradores extranjeros (o los entornos económicos y políticos de sus países) tienen información privada sobre su solvencia real, su probabilidad de impago, o los riesgos políticos y económicos que pueden afectar el cumplimiento de sus obligaciones. Es mucho más difícil para una empresa española evaluar la calidad crediticia de un cliente en un país lejano que la de un cliente local.
  - Problema de selección adversa: sin una intervención, las aseguradoras privadas de crédito comerciales podrían estar poco dispuestas a ofrecer seguros de crédito a la exportación a ciertos mercados o clientes, o lo harían a precios muy elevados. Esto se debe a que son los exportadores con clientes más arriesgados (o los que operan en mercados más inestables) los que tienen más incentivos para buscar este seguro. Si la aseguradora no puede distinguir entre buenos y malos riesgos, podría acabar asegurando solo los malos, lo que llevaría a un mercado de seguros ineficiente o inexistente.
- El papel de CESCE como instrumento de política pública:
  - Proveedor de información y evaluación de riesgos: CESCE se especializa en el análisis de riesgo país y de crédito a nivel internacional. Recopila y procesa

[24]Fuente: CESCE, "Quiénes somos" (cesce.es/corporativo/quienes-somos) y Ministerio de Economía, Comercio y Empresa, "CESCE" (comercio.gob.es), consultados en 2026. Un intento de privatización parcial quedó aparcado en 2015 y CESCE sigue operando con participación mayoritaria del Estado.

una vasta cantidad de información sobre la solvencia de empresas extranjeras y la estabilidad de diferentes mercados. Esta información es crucial y difícil de obtener para una empresa individual. Al hacerlo, reduce la asimetría de información que sufren los exportadores y las entidades financieras.

- Proveedor de seguros de crédito a la exportación: al ofrecer seguros de crédito y de caución (aval) a las empresas españolas (o a sus bancos), CESCE asume el riesgo de impago por parte de los compradores extranjeros o los riesgos políticos/comerciales inherentes a la operación internacional.
- Función pública y estratégica: CESCE actúa como un instrumento de política económica para fomentar el comercio exterior de España. Interviene en mercados donde las aseguradoras privadas pueden no estar dispuestas a operar debido a la alta asimetría de información y el potencial de selección adversa, o donde los riesgos son demasiado grandes para ser asumidos solo por el sector privado. Al cubrir estos riesgos, facilita que las empresas españolas se aventuren en mercados internacionales que de otra manera serían inaccesibles o prohibitivamente arriesgados.

En suma, CESCE funciona como un mecanismo para abordar la selección adversa. Al centralizar y especializar el análisis de riesgo en mercados complejos y al ofrecer seguros donde la información es asimétrica y el sector privado podría fallar, CESCE reduce la incertidumbre para los exportadores y sus bancos, incentivando la participación en operaciones internacionales y mitigando el problema de que solo las "malas" operaciones busquen cobertura. Es una forma en que el Estado, a través de una entidad especializada, interviene para corregir un fallo de mercado causado por la asimetría de información.

13.A.2. El papel de las Webs de Rating en la Teoría de la Agencia

Las webs de rating operan principalmente como mecanismos de señalización o filtrado (screening), aunque su impacto exacto puede variar según el diseño de la plataforma y el tipo de mercado. La literatura empírica confirma que estos sistemas de reputación son predictivos del comportamiento futuro de los agentes (Resnick y Zeckhauser, 2002) y que replican, a una escala mucho mayor, los mecanismos tradicionales de boca a boca (Dellarocas, 2003).

- Frente a la selección adversa: antes de que se realice una transacción o se establezca una relación, el principal (consumidor, comprador, empleador, etc.) tiene información limitada sobre las características ocultas del agente (vendedor, proveedor de servicios, producto, etc.). Las webs de rating permiten a los usuarios, que actúan como principales anteriores, compartir sus experiencias y evaluaciones sobre la calidad o fiabilidad del agente: al acumular un gran número de reseñas y calificaciones,

 © 2026 Manuel Cabado García

estas plataformas crean una reputación pública que actúa como señal de la calidad del agente, de forma que los de alta calidad tienen un incentivo para mantenerla y los de baja calidad quedan expuestos. El principal puede así discriminar entre agentes de diferente calidad, eligiendo aquellos con mejores calificaciones, lo que reduce la probabilidad de incurrir en selección adversa.

- Frente al riesgo moral: una vez establecida la relación, el principal no puede observar directamente el esfuerzo o comportamiento del agente. La expectativa de ser calificado ex post crea un incentivo para que el agente actúe de forma que maximice la satisfacción del principal: si sabe que un mal servicio resultará en una baja calificación que dañará su reputación futura (afectando su capacidad para atraer nuevos clientes), tendrá un mayor incentivo para esforzarse. El temor a una mala reputación, o la búsqueda de una buena, mitiga así el riesgo moral al alinear los incentivos del agente con los del principal incluso en ausencia de supervisión directa.

Estos dos mecanismos se observan con claridad en distintas plataformas:

- Comercio electrónico (Amazon, eBay): el agente es el vendedor y el principal el comprador. Los compradores califican a los vendedores y dejan comentarios sobre la calidad del producto, la velocidad de envío y el servicio al cliente. Frente a la selección adversa, un comprador puede consultar las calificaciones antes de comprar, de modo que los vendedores con malas calificaciones son menos propensos a atraer compradores, lo que incentiva a los vendedores de baja calidad a mejorar o a salir del mercado. Frente al riesgo moral, el vendedor sabe que su desempeño (tiempo de envío, precisión de la descripción del producto, manejo de problemas) será calificado, lo que le incentiva a mantener un alto nivel de servicio para proteger su reputación y sus ventas futuras.
- Transporte y alojamiento (Uber, Airbnb, BlaBlaCar): el agente es el conductor o anfitrión y el principal el pasajero o huésped, que se califican mutuamente tras cada interacción. Frente a la selección adversa, un usuario puede elegir a un conductor o anfitrión basándose en sus calificaciones y reseñas pasadas, evitando a aquellos con reputaciones negativas. Frente al riesgo moral, el conductor sabe que si maneja de forma imprudente o es descortés recibirá una baja calificación que podría llevar a su expulsión de la plataforma o a una reducción en la asignación de viajes, y los anfitriones tienen incentivos para mantener sus propiedades limpias y brindar buena hospitalidad para obtener buenas reseñas.
- Reseñas de restaurantes y negocios locales (Yelp, TripAdvisor, Google Reviews): el agente es el negocio (restaurante, hotel, tienda) y el principal el cliente potencial, que comparte sus experiencias y calificaciones. Frente a la selección adversa, un nuevo cliente puede consultar las reseñas para decidir dónde comer o alojarse, evitando lugares con malas críticas sobre higiene o servicio. Frente al riesgo moral, los pro-

pietarios tienen un incentivo continuo para mantener la calidad de sus productos y servicios, ya que saben que una mala experiencia puede ser rápidamente compartida y dañar su reputación y, por ende, sus ingresos futuros.

- Plataformas de freelancers y servicios profesionales (Upwork, Fiverr, LinkedIn): el agente es el freelancer o consultor y el principal el cliente que busca el servicio. Los clientes califican el trabajo y la comunicación, y los freelancers construyen así un portafolio de reputación. Frente a la selección adversa, un cliente puede elegir un freelancer basándose en sus calificaciones, reseñas y trabajos anteriores, lo que le permite identificar talentos de alta calidad. Frente al riesgo moral, el freelancer sabe que debe entregar un trabajo de calidad y cumplir con los plazos para obtener buenas calificaciones, ya que estas son cruciales para conseguir futuros proyectos.

Aunque las webs de rating son herramientas poderosas, no son perfectas: pueden existir incentivos para dejar reseñas falsas, positivas o negativas, o para manipular las calificaciones, lo que introduce una nueva asimetría de información y puede socavar la confianza (Dellarocas, 2003). Además, los usuarios pueden tener diferentes estándares o sesgos al calificar. No todos los principales dejan reseñas, lo que puede llevar a una muestra sesgada de información. Por último, en los sistemas bidireccionales, el miedo a una calificación negativa recíproca puede disuadir a los usuarios de dejar reseñas honestas.

A pesar de estas limitaciones, el papel general de las webs de rating es fundamental para aumentar la transparencia del mercado al reducir la asimetría de información, lo que lleva a una asignación más eficiente de recursos y a la mejora de la calidad de los bienes y servicios.

13.A.3. Modelos multicapa

En el contexto de la economía de la información y la teoría de la agencia, existe un tipo de modelos denominados modelos multicapa, es decir, situaciones donde la asimetría de información o la relación principal-agente no se limita a una única interacción, sino que se extiende a múltiples niveles o etapas, o involucra a varios agentes o principales interconectados. Este tipo de estructuras jerárquicas de agencia, con delegación sucesiva en varios niveles, fueron formalizadas por (Tirole, 1986), quien mostró que la posibilidad de colusión entre los niveles intermedios de una organización condiciona el diseño óptimo de los contratos e incentivos en toda la cadena.

Un modelo multicapa puede darse, en primer lugar, cuando hay múltiples agentes o principales: un principal puede delegar tareas a varios agentes que interactúan entre sí o cuyos comportamientos le afectan por igual (por ejemplo, una empresa y sus diferentes departamentos o empleados), un agente puede rendir cuentas o tener contratos con varios principales (por ejemplo, un consultor que trabaja para varias empresas), o pueden darse redes complejas de relaciones contractuales, con delegación y asimetría de información en

 © 2026 Manuel Cabado García

diferentes nodos (por ejemplo, cadenas de suministro con subcontratistas).

En segundo lugar, puede darse una delegación en cadena o cascada de agencias: un principal delega una tarea en un agente, y ese agente, a su vez, se convierte en principal para otro agente, y así sucesivamente. Es lo que ocurre, por ejemplo, en la cadena Gobierno (principal) → Ministerio (agente/principal) → Departamento (agente/principal) → Funcionario (agente final), o en la cadena Accionistas (principal) → Consejo de Administración (agente/principal) → Alta Dirección (agente/principal) → Gerentes de nivel medio (agente final). En cada nivel de la cadena surge el problema de la asimetría de información, ya sea selección adversa o riesgo moral, entre el principal de ese nivel y su agente, lo que amplifica los problemas y hace que su gestión sea mucho más compleja, ya que los incentivos y las acciones de un nivel afectan a los demás.

Por último, aunque no son estrictamente multicapa en el sentido de una jerarquía, las interacciones dinámicas o repetidas también añaden una capa de complejidad: en contratos de largo plazo, el comportamiento actual de un agente puede influir en su reputación y en futuros contratos, lo que añade otra capa de incentivos.

Los modelos multicapa son relevantes porque la mayoría de las organizaciones y sistemas económicos del mundo real son inherentemente complejos y jerárquicos, y estos modelos intentan capturar esa complejidad para analizar cómo se propagan los problemas de un nivel a otro (el riesgo moral o la selección adversa en un nivel pueden exacerbar problemas en otros niveles), cómo diseñar contratos e incentivos que funcionen eficazmente a través de múltiples niveles de delegación evitando una alineación de intereses defectuosa, cómo la estructura de una organización puede mitigar o agravar estos problemas, y cómo de efectivas pueden ser las regulaciones cuando hay múltiples capas de asimetría de información.

Así, los modelos multicapa son un concepto importante en este campo y se refieren a la extensión de la teoría de la agencia a escenarios más realistas y complejos con múltiples niveles de interacción y asimetría de información.

A continuación se presentan una serie de ejemplos que ilustran cómo las relaciones de agencia con múltiples capas son ubicuas y cómo los desafíos de la asimetría de información se propagan y se hacen más complejos en sistemas con delegación en cadena. El diseño de contratos, sistemas de supervisión e incentivos se vuelve crucial para mitigar estos problemas:

- Grandes corporaciones multinacionales: la cadena de delegación típica es accionistas → consejo de administración → CEO (alta dirección) → gerentes de división → gerentes de departamento → empleados, y en cada eslabón surge un problema de agencia distinto. Entre accionistas (principal) y consejo/CEO (agente), los accionistas quieren maximizar el valor de la empresa, pero el consejo o el CEO pueden tener otros intereses, como maximizar su propia remuneración, construir un imperio corporativo o evitar riesgos, dando lugar tanto a riesgo moral (falta de esfuerzo,

decisiones subóptimas) como a selección adversa (¿se contrató al CEO más competente?). Entre el CEO (principal) y los gerentes de división (agente), estos últimos pueden tener incentivos para inflar sus resultados, ocultar problemas o evitar proyectos arriesgados que beneficien a la empresa pero les expongan personalmente. Y entre los gerentes de departamento (principal) y los empleados (agente), estos pueden incurrir en riesgo moral (holgazanería, bajo esfuerzo) o el gerente puede enfrentar selección adversa al contratar (¿es el candidato tan productivo como dice?). Los problemas de agencia en un nivel pueden además propagarse a los demás: si el CEO no está bien incentivado, es menos probable que supervise eficazmente a los gerentes de división, lo que a su vez puede llevar a un peor rendimiento de los empleados, por lo que el diseño de contratos e incentivos debe considerar esta cascada.

- Contratos en cadenas de suministro: la estructura habitual es empresa final (principal) → proveedor principal (agente/principal) → subproveedores (agente/principal) → productores de materias primas (agente final). Entre la empresa final y el proveedor principal, esta quiere que los componentes sean de alta calidad y se entreguen a tiempo, mientras que el proveedor (agente) podría reducir la calidad para ahorrar costes (riesgo moral) o exagerar sus capacidades de producción (selección adversa). Entre el proveedor principal y los subproveedores, estos pueden tener incentivos similares para recortar esquinas o no revelar completamente sus capacidades. Así, si una empresa automotriz (principal) encarga el diseño de un componente a un proveedor de primer nivel (agente 1) y este, a su vez, subcontrata la fabricación de una parte específica a un segundo proveedor (agente 2), la calidad final del automóvil depende de la alineación de incentivos en toda la cadena y de cómo se gestionan las asimetrías de información en cada eslabón.
- Servicios de salud: la cadena es paciente (principal) → médico (agente/principal) → laboratorio o farmacéutica (agente). Entre el paciente y el médico, el paciente tiene asimetría de información sobre su propio cuerpo, mientras que el médico tiene más conocimiento, pudiendo surgir riesgo moral si el médico recomienda tratamientos innecesarios para aumentar ingresos, aunque hay un fuerte componente ético que lo mitiga. Entre el médico y el laboratorio o la farmacéutica, el primero solicita pruebas o prescribe medicamentos, y el segundo podría realizar pruebas innecesarias u ofrecer incentivos al médico para prescribir un medicamento específico sobre otro. La confianza y la información fluyen aquí en múltiples direcciones, y los incentivos en cada nivel pueden afectar la calidad y el coste de la atención médica.
- Franquicias: la estructura básica es franquiciador (principal) → franquiciado (agente), y si el franquiciado tiene a su vez empleados se añade otra capa. El franquiciador quiere mantener la calidad y la reputación de la marca, pero el franquiciado, como

 © 2026 Manuel Cabado García

agente, podría tener incentivos para reducir costes de forma que afecte a la calidad del servicio o producto, o para no esforzarse lo suficiente en marketing local, por lo que el franquiciador debe diseñar contratos que alineen los incentivos del franquiciado con la marca global.

- Contratación pública: la cadena es ciudadanos o contribuyentes (principales últimos) → gobierno o agencia pública (agente/principal) → empresa privada contratada (agente). Entre los ciudadanos y el gobierno, los primeros delegan la provisión de servicios públicos, pudiendo darse riesgo moral si los políticos o burócratas actúan en su propio interés (corrupción, proyectos de prestigio innecesarios). Entre el gobierno y la empresa contratada, la agencia pública (principal) contrata a una empresa privada (agente) para construir una infraestructura o proveer un servicio, pudiendo darse tanto selección adversa (empresas que exageran sus capacidades) como riesgo moral (sobrecostes, retrasos, baja calidad una vez firmado el contrato).

13.A.4. Soluciones alternativas en una situación de riesgo moral

Hasta aquí se ha supuesto que la única variable de la que podía hacerse depender la remuneración eran los ingresos. Pero si la correlación entre el esfuerzo y los ingresos es muy imperfecta[25], la ineficiencia será muy grande en un contexto de información asimétrica.

De ahí que surjan fórmulas alternativas para incentivar al agente a que realice el esfuerzo óptimo:

- Criterios comparativos, como otorgar incentivos según los resultados relativos entre miembros de la misma empresa o entre distintas empresas –evaluación relativa del desempeño–. Se justifica por el principio de informatividad de Holmström (1979): toda señal que aporte información sobre el esfuerzo del agente, además de la que ya contiene el resultado, como el resultado de sus comparables, debe incorporarse al esquema de pagos óptimo[26].
- Competencia, pues los comportamientos oportunistas derivados del riesgo moral disminuyen los resultados de la empresa, pudiendo ser desplazada del mercado.
- Juegos dinámicos, pues con períodos infinitos, el propósito de labrarse una buena reputación (por ejemplo, de "trabajador responsable", de "conductor prudente", o de "prestatario serio") desincentiva al agente de ser un oportunista.
- Supervisión: La solución anterior se basa en el supuesto de que existe una correlación entre la acción oculta (el esfuerzo) y el objetivo observable del principal (los beneficios).

[25] Como se ha visto, los resultados no dependen solo del esfuerzo, sino también de si se da un estado de la naturaleza favorable y se podrían incluir otras variables.

[26] La intuición es que, si en una crisis todas las sucursales de un banco venden menos, comparar a cada director con los demás permite descontar la mala suerte común y premiar solo su esfuerzo.

  - En ciertas situaciones, puede ocurrir que sea difícil encontrar una variable, que represente el objetivo, correlacionada con la acción oculta, o que la correlación sea muy imperfecta. En estos casos, puede ser más aconsejable recurrir a otras soluciones. Por ejemplo, una opción es supervisar al agente.
  - En algunos casos, aumentar la supervisión del agente también puede incentivar una conducta del mismo que sea acorde a los intereses del principal, aunque esto tiene un coste.
  - El control y la supervisión serán rentables para el principal mientras que el coste marginal no exceda el beneficio marginal.

- Condicionalidad: El riesgo moral también afecta, a menudo, a los gobiernos. Por ejemplo, el hecho de que existan instituciones financieras multilaterales que puedan actuar como prestamistas de última instancia en caso de serias dificultades económicas puede reducir los incentivos de un gobierno para aplicar políticas económicas prudentes. Un problema análogo surge dentro de un país cuando el gobierno confía en que el banco central actuará como prestamista de última instancia, lo que puede conducir a situaciones de dominancia fiscal. En estos casos, la solución tradicional al problema de riesgo moral es establecer una fuerte condicionalidad, que haga que el recurso a este tipo de financiación no sea atractivo excepto en situaciones bastante extremas y, de este modo, reduzca la incidencia del riesgo moral.

13.A.5. Oliver Hart y Bengt Holmström (Nobel 2016)

Ambos destacan en el campo de la teoría de los contratos: cómo diseñar contratos más eficientes que resuelvan los problemas de agencia y permitan alcanzar situaciones en las que ganen todas las partes.

- Hart señaló la imposibilidad de especificar en un contrato todo lo que podría ocurrir –son, por definición, contratos incompletos–, por lo que conviene preestablecer las reglas y derechos ante incumplimientos o fallos del contrato, por ejemplo en caso de bancarrota (Hart y Moore, 1988). Junto con Grossman, desarrolló la teoría de los derechos de propiedad. Los derechos residuales de control son el derecho a decidir sobre todos los usos de un activo que el contrato no ha previsto, y su titularidad es lo que define la propiedad. Cuantos más derechos residuales de control se tengan sobre un activo, más incentivos habrá para realizar inversiones específicas en la relación contractual (Grossman y Hart, 1986)[27]. Entre sus aplicaciones destaca el análisis de la gestión pública frente a la privada: que un hospital o una prisión deban gestionarse de forma pública o privada dependerá de las condiciones concretas de cada servicio,

[27] La intuición es que el dueño de una furgoneta decide qué se hace con ella en todo lo que el contrato de alquiler no dice, y por eso es quien más se preocupa de cuidarla y mejorarla.

 © 2026 Manuel Cabado García

como la verificabilidad de la calidad, y no del tipo de servicio en sí. De hecho, en Estados Unidos se dejó de privatizar prisiones apoyándose en los estudios de Hart sobre los incentivos a recortar calidad que genera la gestión privada cuando esta no es verificable.

- Holmström, por su parte, estableció pautas sobre cómo diseñar la remuneración y los incentivos dentro de los contratos: no conviene vincular la remuneración únicamente al resultado observado, porque este depende también de factores aleatorios ajenos al esfuerzo del agente, y conviene aprovechar cualquier otra señal informativa sobre ese esfuerzo –el principio de informatividad ya visto en el apartado de soluciones alternativas al riesgo moral (Holmström, 1979)–. Así, los contratos en empresas de alto riesgo deben tener sueldos fijos, para ser cooperativos y generar confianza, mientras que para un operario cuyo rendimiento pueda medirse con precisión conviene un pago en proporción directa a su rendimiento. Del mismo modo, los profesores deben tener salario fijo, para no menoscabar la enseñanza de habilidades que no se pueden medir con facilidad. Holmström también estudió cómo fijar la retribución de los directivos de una empresa para beneficiar a los accionistas y conseguir un mayor valor a largo plazo, y por qué los trabajadores jóvenes se esfuerzan más que los de mayor edad: al principio de su carrera tienen más en juego en términos de reputación y expectativas profesionales futuras. Entre sus aplicaciones destaca, de nuevo, el sector público: profesores o funcionarios de prisiones deben tener retribuciones fijas o variables en función de si su desempeño es o no verificable.

Bibliografía

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics 2*(3), 193–216.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research and Columbia University Press.

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science 49*(10), 1407–1424.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica 41*(4), 587–601.

Green, J. y J.-J. Laffont (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica 45*(2), 427–438.

Grossman, S. J. y O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy 94*(4), 691–719.

Hart, O. y J. Moore (1988). Incomplete contracts and renegotiation. *Econometrica 56*(4), 755–785.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics 10*(1), 74–91.

Holmström, B. y J. Tirole (2011). *Inside and Outside Liquidity*. MIT Press.

Kreps, D. M. (1990). *A Course in Microeconomic Theory*, cap. 17 Adverse selection and market signaling. Princeton University Press.

Laffont, J.-J. y J. Tirole (1986). Using cost observation to regulate firms. *Journal of Political Economy 94*(3), 614–641.

Milgrom, P. R. y J. Roberts (1992). *Economics, Organization, and Management*, cap. 5, pp. 157–158. Prentice-Hall International editions. Prentice-Hall.

Myerson, R. B. (1979). Incentive compatibility and the bargaining problem. *Econometrica 47*(1), 61–73.

 © 2026 Manuel Cabado García

Resnick, P. y R. Zeckhauser (2002). Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. En M. R. Baye (ed.), *The Economics of the Internet and E-Commerce*, vol. 11 de *Advances in Applied Microeconomics*, pp. 127–157. Elsevier Science.

Rothschild, M. y J. Stiglitz (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics 90*(4), 629–649.

Salop, J. y S. Salop (1976). Self-selection and turnover in the labor market. *The Quarterly Journal of Economics 90*(4), 619–627.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Stiglitz, J. E. (1990). Peer monitoring and credit markets. *The World Bank Economic Review 4*(3), 351–366.

Stiglitz, J. E. y A. Weiss (1981). Credit Rationing in Markets with Imperfect Information. *The American Economic Review 71*(3), 393–410.

Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics, and Organization 2*(2), 181–214.

 © 2026 Manuel Cabado García

Tema 14

Economía del bienestar (I). Óptimo económico y equilibrio general competitivo. Los teoremas fundamentales del bienestar. Teoría del "second best".

14.1. Introducción ⏳ 1'

La Economía del bienestar es la rama de la teoría económica que intenta valorar, desde un punto de vista colectivo, las distintas situaciones en las que puede encontrarse un sistema económico. Estudia la deseabilidad social de los estados económicos alternativos a través de distintos criterios de valoración.

Las funciones de la Economía del Bienestar son[1]:

- Función positiva o descriptiva, proporciona instrumentos para valorar la deseabilidad social de los estados alternativos, caracterizados por una asignación de recursos y una distribución de la renta. A priori, valorar estos estados implica hacer juicios de valor.
- Función normativa o prescriptiva, proporciona criterios de decisión para maximizar el bienestar social (política económica).

El criterio fundamental que ha surgido en el campo de la Economía del Bienestar para valorar el bienestar social es el de eficiencia económica, entendida como el mejor uso posible

[1]Por esto, será una rama de la economía con un notable carácter normativo, propio a la acción de establecer un orden de preferencia entre distintas situaciones en función de la distribución del consumo, producción o desigualdad de la renta. Los criterios de decisión, en la medida en que se utilizan para maximizar el bienestar social, deben incorporar necesariamente algún juicio de valor (que incorporan inherentemente un conjunto de valores particulares o creencias personales).

de los recursos limitados: un sistema económico es eficiente si no desperdicia recursos, es decir, si no es posible aumentar el bienestar de un individuo sin reducir el de otro.

Existen varios enfoques para valorar la eficiencia económica, siendo el más aceptado el de Vilfredo Pareto, cuya importancia radica en que:

1. Se basa en pocos juicios de valor.
2. Dichos juicios de valor son débiles, por lo que son generalmente aceptados.
3. Permite valorar el sistema competitivo como mecanismo de asignación de los recursos de la economía.

No obstante, dadas las limitaciones de este enfoque surgieron nuevas aportaciones posteriores como

- Nueva Economía del Bienestar, desarrollada en los años 30 y 40 de la mano de Kaldor, Hicks, Scitovsky y Little, que trataron de superar la indeterminación paretiana mediante los criterios de compensación (vía neoparetiana) y el criterio de la función de bienestar social (vía neoutilitarista).
- Novísima Economía del Bienestar, en los años 50 y 60 por parte de Arrow, Nash, Lipsey y Lancaster, con la Teoría del Second Best o la Teoría de elección colectiva.

Por consiguiente, el objetivo del tema será analizar los criterios de ordenación más importantes desde el punto de vista de la eficiencia a través del siguiente esquema[2]:

1. Óptimo económico.
2. Equilibrio competitivo y teoremas del bienestar.
3. Los criterios de compensación.
4. Teoría del Second Best.

14.2. Óptimo Económico ⏳ 10'

En su obra *Manual de economía política* (1906) Pareto estableció un criterio de ordenación basado en la "superioridad". Para ello incorpora una serie de juicios de valor que suelen ser generalmente aceptados[3]:

[2]La exposición se va a centrar en gran medida en cómo se obtiene el óptimo económico u óptimo de Pareto. Ya al final se intentarán explicar mecanismos que superen las dos limitaciones más importantes del criterio de Pareto:

- Puede generar una indeterminación, al arrojar infinitos óptimos no comparables entre sí. Se verá cómo los criterios de compensación tratan de resolver esta indeterminación.
- Se encuadra en un marco de competencia perfecta, que es muy restrictivo. Se verá cómo la teoría del second best trata de dar respuesta cuando existe alguna imperfección en el mercado.

[3]El criterio de Pareto surge ante la insatisfacción que producían las teorías acerca de la optimalidad existentes hasta el momento:

 © 2026 Manuel Cabado García

1. Individualismo, las variables que afectan al bienestar social son las utilidades individuales. El bienestar social depende del bienestar individual, considerando que el bienestar puede aumentar con el consumo
2. Unanimidad, establece el principio de superioridad. El bienestar no es comparable. Solo cambios unánimemente aceptados mejoran el bienestar.
3. Independencia del proceso, no importan los mecanismos utilizados para asignar los recursos y distribuir la renta. Concretamente, no importa si una situación es fruto de la acción libre o de decisiones forzadas.
   Discutible, la democracia es preferible a una dictadura.
4. No paternalismo, las funciones de utilidad de referencia son las individuales, pues se considera que el individuo es el mejor juez de su propio bienestar.[4]
   Discutible, por ejemplo ante el consumo de bienes de consumo no socialmente aceptados como pueden ser las drogas o el porno.
5. Benevolencia, en el sentido de que un incremento en la utilidad de cualquier individuo, ceteris paribus, se considera una mejora en el bienestar. Esto implica que el criterio de Pareto no emite juicios de valor sobre las desigualdades.
   Discutible, no tiene en cuenta una potencial distribución equitativa de la renta, por ende cuestiones como la pobreza o la justicia social.

A partir de estos juicios de valor, Pareto estableció dos conceptos cruciales:

- Criterio de bienestar social: una asignación x es preferida a otra y si, en el paso de y a x, algún individuo mejora y ninguno empeora. En este caso, x es Pareto superior a y[5]. No obstante, que una asignación sea Pareto superior a otra no significa que la primera sea un óptimo de Pareto.
- Pareto óptimo: o criterio de óptimo económico. Una asignación es Pareto óptimo si no es posible reasignar recursos de manera que algún individuo mejore sin que otro empeore. Es decir, no existe ninguna asignación Pareto superior al mismo.

Con el fin de estudiar cómo llega el mercado al óptimo de Pareto y a las condiciones de

- Marginalistas de primera generación (Jevons, Walras): consideraban que los 'bienestares individuales' se podían medir, pues concebían una función de utilidad cardinal.
- Utilitaristas (Bentham): consideraban que el bienestar social era igual a la suma de los 'bienestares individuales'.

Partiendo de esas ideas, estos autores hallaban el bienestar social sumando las utilidades cardinales, valorando los efectos sobre el bienestar de la política económica en términos de si la suma total de utilidades aumentaba o disminuía. El problema es que estos enfoques se basaban en unos supuestos muy fuertes y restrictivos (utilidad cardinal y comparable entre individuos).

[4] El paternalismo libertario de Thaler y Sunstein cuestiona este supuesto: si lo que se elige depende de cómo se presentan las opciones, la elección observada no siempre revela el bienestar del individuo. Véase el apartado 4.A.8 del Apéndice del Tema 4.

[5] La intuición es la de un reparto de tarta en el que alguien recibe un trozo más sin que a nadie se le quite nada: el nuevo reparto es mejor sin discusión, aunque quizá aún sobren migas que repartir.

óptimo económico, se utilizará un modelo estático 2x2x2 con los siguientes siete supuestos:

1. Existen dos factores de producción, capital K y trabajo L, cuya oferta total está dada (exógena) y que son perfectamente divisibles y homogéneos.
2. Existen dos bienes, X e Y, con características de bienes privados[6], que se producen a partir de K y L mediante una función de producción $F(K, L)$ neoclásica y rendimientos no crecientes a escala[7].
3. Existen dos consumidores A y B, cuyas preferencias vienen representadas por las funciones de utilidad $U_A = U_A(X_A, Y_A)$ y $U_B = U_B(X_B, Y_B)$. Las preferencias son convexas, es decir, las funciones de utilidad son cuasicóncavas, de modo que los consumidores muestran gusto por la variedad[8].
4. Inexistencia de efectos externos tanto en la producción como en el consumo[9].
5. Inexistencia de bienes públicos[10].
6. Los consumidores y los productores son racionales, es decir, que desean maximizar sus funciones objetivo –utilidad y beneficios respectivamente–.
7. Modelo cerrado, es decir, la renta se agota y hay plena ocupación.

Una vez planteados los supuestos, para que una asignación de recursos sea un óptimo de Pareto es necesario que la asignación sea simultáneamente óptima en la producción, óptima en el consumo –o intercambio– y óptima globalmente.

Óptimo en la producción

Una asignación es óptima/eficiente en la producción si, mediante la reasignación de factores de producción entre industrias, no es posible producir una mayor cantidad de un bien sin reducir la producción del otro. Es decir, implica una asignación óptima de los factores entre las empresas.

[6] Si algún bien fuera público, entonces los precios dejarían de ser señales válidas (al no reflejar íntegramente la utilidad del consumo ni los costes de la producción), y el equilibrio general competitivo no alcanzaría una solución eficiente de Pareto.

[7] Bajo rendimientos decrecientes, los pagos de los factores no agotarían el producto y los beneficios de las empresas serían positivos –y, por lo tanto, en la dotación de los consumidores/propietarios se incluirían dichos beneficios–, pero se preservaría la concavidad requerida en la Frontera de Posibilidades de Producción, por lo que se mantendrían las conclusiones. El problema vendría con rendimientos crecientes, ya que en ese caso la FPP podría ser convexa y, por tanto, no asegurar un máximo de bienestar.

[8] Las funciones de utilidad tienen las propiedades habituales: i) utilidad marginal positiva pero decreciente, ii) continuas y iii) diferenciables. Además, son independientes (es decir, no hay externalidades en el consumo –como efecto imitación o efecto snob–).

[9] Si hubiera externalidades en la producción o en el consumo, entonces los precios dejarían de ser señales válidas y el equilibrio general competitivo no alcanzaría una solución eficiente de Pareto.

[10] Si algún bien fuera público, entonces los precios dejarían de ser señales válidas –al no reflejar íntegramente la utilidad del consumo ni los costes de la producción–, y el equilibrio general competitivo no alcanzaría una solución eficiente de Pareto.

 © 2026 Manuel Cabado García

Para obtener un óptimo en producción se plantea un problema de maximización de la cantidad del bien X, que utilizará una determinada cantidad de factor capital, K_X, y trabajo, L_X

$$\max_{K_X, L_X} X\left(K_X, L_X\right) \tag{14.1}$$

sujeto a una cantidad dada del bien Y, $\bar{Y}$, y a las cantidades totales de factores, $\bar{K}$ y $\bar{L}$, utilizadas en la producción de ambos bienes

$$\bar{Y} = Y\left(K_Y, L_Y\right) \tag{14.2}$$

$$\bar{K} = K_X + K_Y \tag{14.3}$$

$$\bar{L} = L_X + L_Y \tag{14.4}$$

La condición de primer orden, CPO, establece que la relación marginal de sustitución técnica, $RMST$, entre dos factores de producción debe igualarse entre industrias

$$RMST_{K,L}^{X} = RMST_{K,L}^{Y} \tag{14.5}$$

es decir, deberán igualarse las tasas a las que se intercambian factores manteniendo constante el nivel de producción, definidas como el cociente entre las productividades marginales de los factores

$$RMST_{K,L}^{X} = -\left.\frac{dL_X}{dK_X}\right|_{\bar{X}} = \frac{PMg_K^X}{PMg_L^X} \tag{14.6}$$

Gráficamente, el óptimo de Pareto en la producción se puede representar mediante la caja de Edgeworth[11] cuyas dimensiones vienen determinadas por las dotaciones de capital y trabajo como se observa en la Figura 14.1.

- A partir de las curvas isocuantas, que muestran las combinaciones de factores productivos que permiten alcanzar el mismo nivel de producción y cuya pendiente es la $RMST$, se puede hallar la producción óptima que se dará en los puntos de tangencia de las mismas, donde las $RMST$ de ambos bienes se igualan.
- De esta forma, se obtiene la curva de contrato, definida como el lugar geométrico de las combinaciones de capital y trabajo que determinan un óptimo en la producción[12].
- El conjunto de asignaciones de la curva de contrato no serán comparables entre sí en el sentido de eficiencia de Pareto, dado que todas son óptimas. Además, no todas

[11] La caja de Edgeworth debe su nombre a Francis Ysidro Edgeworth (1845-1926), economista irlandés que en 1881 representó por primera vez juntas las curvas de indiferencia de dos agentes. La forma de caja cerrada se debe a Pareto y su difusión a Bowley, de ahí que también se la llame caja de Edgeworth-Bowley.

[12] Los dos vértices de la caja de Edgeworth, O_X y O_Y, en los que una empresa usa todos los factores productivos y la otra nada de ninguno también forman parte de la curva de contrato

las tangencias se producen necesariamente en la misma pendiente, es decir, la curva de contrato no es necesariamente una isoclina.

Así:

- Si se parte de un punto como 1, sería posible reasignar los factores productivos y aumentar la producción de un bien manteniendo constante la del otro en los puntos 2 y 3, o bien aumentar la producción de ambos bienes en cualquiera de los puntos dentro de la lente –la región formada por las isocuantas que pasan por la dotación inicial 1– sobre la curva de contrato.
- De esta forma, por analogía con el intercambio entre consumidores, se puede definir el núcleo como el conjunto de asignaciones eficientes que ningún agente rechazaría frente a la asignación inicial 1, es decir, el tramo de la curva de contrato situado dentro de la lente. Por lo tanto, si una asignación pertenece al núcleo, se han agotado todas las ganancias del intercambio y nadie queda peor que en su situación de partida[13].

Por lo tanto, para que una asignación pueda considerarse núcleo de la economía es necesario que cumpla:

- Estar dentro de la caja, es decir, que sea alcanzable.
- Estar dentro de la lente, es decir, que ningún agente quede peor que en la asignación inicial (en la caja de producción, que ninguna industria reduzca su producción).
- Estar sobre la curva de contrato, es decir, que sea Pareto eficiente.

La Frontera de Posibilidades de Producción, FPP, es la curva que muestra la máxima cantidad de Y que puede producirse para cada cantidad de X, dados los factores disponibles y la tecnología. Se obtiene trasladando la curva de contrato al plano de bienes, X e Y, como en la Figura 14.2, pues cada asignación eficiente de K y L da lugar a una combinación de producciones[14]. En cualquier punto de la curva todos los factores están empleados de forma eficiente.

Posee las siguientes propiedades:

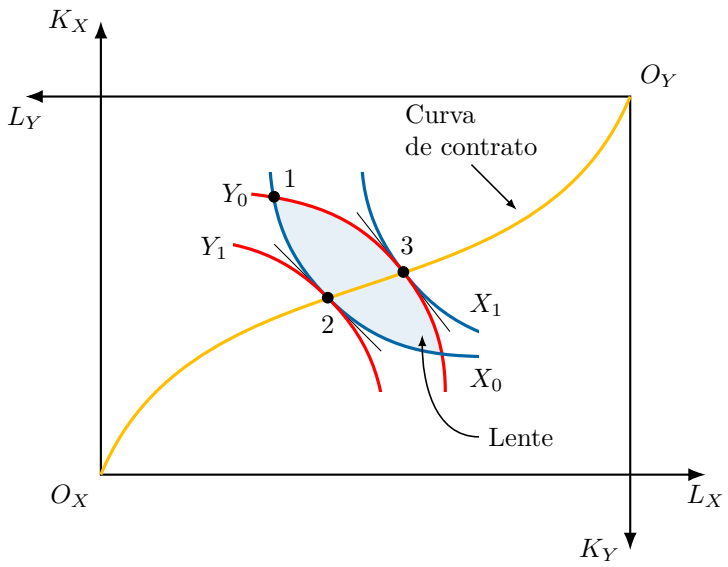
1. Pendiente negativa, porque las funciones de producción son crecientes.

[13] La intuición es la de dos niños que intercambian cromos: ninguno acepta un trato peor que quedarse con los suyos y no paran mientras quede un cambio que convenga a ambos. Técnicamente, el núcleo es el conjunto de asignaciones que ninguna coalición de agentes puede mejorar por su cuenta, y con dos agentes coincide con ese tramo de la curva de contrato.

[14] La FPP es, dicho de otra manera, el equilibrio de las empresas en el espacio de bienes. Esta traslación es necesaria porque, en el óptimo, las cantidades producidas deben ser iguales a las cantidades consumidas, y los consumidores toman sus decisiones de compra en el espacio de bienes. De ahí que la FPP resulte necesaria para coordinar los planes de producción con los de consumo.

 © 2026 Manuel Cabado García

Figura 14.1: Caja de Edgeworth de producción



2. Concavidad, porque, a medida que se trasladan factores de un bien a otro, cada unidad adicional de un bien exige renunciar a cantidades crecientes del otro, de modo que la RMT aumenta a lo largo de la frontera[15].
3. Su pendiente es la relación marginal de transformación, RMT, que mide la cantidad de Y a la que hay que renunciar para producir una unidad más de X, es decir, el coste de oportunidad de X en términos de Y[16]

$$RMT_{X,Y} = -\frac{dY}{dX} \tag{14.7}$$

4. Todos los puntos de la FPP son eficientes en la producción, por lo que:
   - Existen infinitos óptimos en la producción.
   - Los puntos interiores a la FPP son ineficientes, y los puntos situados fuera son inalcanzables, a menos que se descubran recursos adicionales y/o una nueva tecnología.

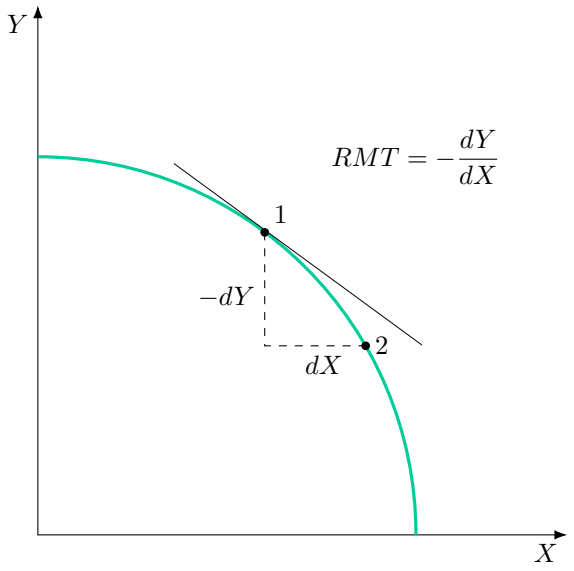
Óptimo en el consumo

Una vez alcanzado el óptimo en producción y ante la existencia de infinitos óptimos, se analizarán los óptimos en el consumo, es decir, aquellas asignaciones en las que no sea

[15]Esto se debe a las productividades marginales decrecientes y, sobre todo, a que la intensidad factorial de las dos industrias es distinta (de hecho, esta es la clave de la forma de la FPP). Si ambas usaran los factores en la misma proporción y tuvieran rendimientos constantes, la FPP sería una recta.

[16]La intuición es que, si un taller que fabrica bicicletas y patinetes debe dejar de hacer dos patinetes para hacer una bicicleta más, la RMT es dos.

Figura 14.2: Frontera de posibilidades de producción



posible a través de una redistribución de los bienes aumentar la utilidad de un consumidor sin disminuir la del otro, por lo que serán eficientes. Por lo tanto, implica una distribución óptima de los bienes distribuidos entre los consumidores.

Para obtener un óptimo en consumo se plantea un problema de maximización de la utilidad del individuo A, que le reporta el consumo del bien X y del bien Y

$$\max_{X_A, Y_A} u_A\left(X_A, Y_A\right) \tag{14.8}$$

sujeto a una utilidad dada del individuo B , $\bar{u}_B$, y a las cantidades totales consumidas de ambos bienes por ambos individuos

$$\bar{u}_B = u_B\left(X_B, Y_B\right) \tag{14.9}$$

$$\bar{X} = X_A + X_B \tag{14.10}$$

$$\bar{Y} = Y_A + Y_B \tag{14.11}$$

La condición de primer orden, CPO, establece que la relación marginal de sustitución de A deberá igualarse a la del individuo B

$$RMS^A_{X,Y} = RMS^B_{X,Y} \tag{14.12}$$

es decir, la cantidad de Y a la que cada individuo está dispuesto a renunciar a cambio de una unidad más de X, manteniendo constante su utilidad, debe ser la misma para

 © 2026 Manuel Cabado García

ambos, y cada una es igual al cociente entre las utilidades marginales de los bienes

$$RMS^A_{X,Y} = -\frac{dY}{dX}\bigg|_{\bar{u}_A} = \frac{UMg^A_X}{UMg^A_Y} \tag{14.13}$$

Gráficamente, el problema se puede representar a través de una caja de Edgeworth como la de la Figura 14.3 cuyas dimensiones ahora vendrán determinadas por las cantidades consumidas por los individuos A y B tanto del bien X como del bien Y.

- A su vez, se podrán representar las curvas de indiferencia, que indican las combinaciones de bienes X e Y que proporcionan el mismo nivel de utilidad y cuya pendiente, en valor absoluto, es la Relación Marginal de Sustitución[17], RMS. Los puntos óptimos de consumo[18] vendrán representados por los puntos de tangencia de las curvas de indiferencia para el consumo del individuo A y el individuo B. A partir de la unión de todos los puntos óptimos en el consumo se podrá representar la curva de contrato.
- A partir de las curvas de indiferencia, que indican las combinaciones de bienes X e Y que representan el mismo nivel de utilidad y cuya pendiente es la RMS, se puede hallar el consumo óptimo que se dará en los puntos de tangencia de las mismas, donde las RMS de ambos consumidores se igualan. De esta forma, se obtiene la curva de contrato, definida como el lugar geométrico de las combinaciones de bien X y bien Y que determinan un óptimo en el consumo.
- De la misma forma que en el análisis de producción, el conjunto de asignaciones de la curva de contrato no serán comparables entre sí en el sentido de eficiencia de Pareto, dado que todas son óptimas. Además, no todas las tangencias se producen necesariamente en la misma pendiente, es decir, la curva de contrato no es necesariamente una isoclina.

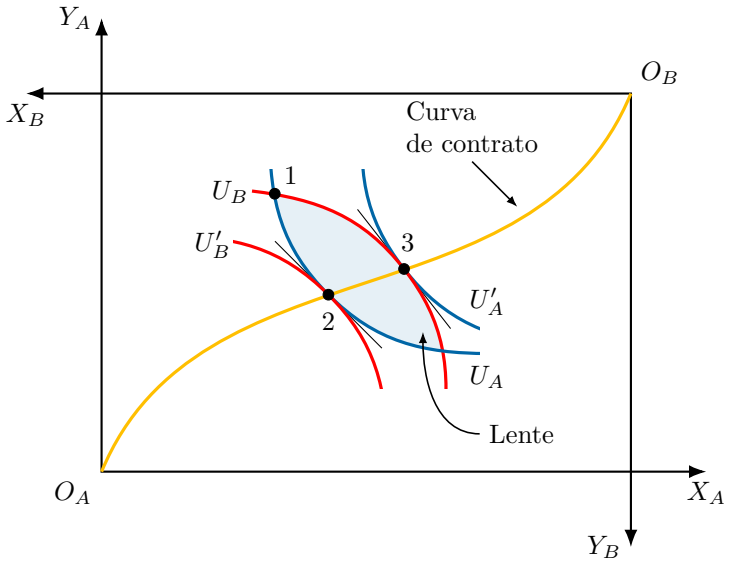
Además, si se representa la curva de contrato en el espacio de las utilidades de los agentes como en la Figura 14.4, se podrá obtener la Frontera de Posibilidades de Utilidad, FPU, que recoge los puntos de utilidad óptimos, es decir, aquellos en los que no es posible aumentar la utilidad de B sin disminuir la utilidad de A. Además, posee las siguientes propiedades

1. Pendiente negativa, ya que las funciones de utilidad son crecientes y porque no existen externalidades derivadas del consumo.

[17] Mide la cantidad de un bien a la que el consumidor está dispuesto a renunciar por una unidad más del otro sin que varíe su utilidad.

[18] Solución: la eficiencia en el consumo ocurrirá allí donde las RMS de consumir ambos bienes se igualen (es decir, allí donde las curvas de indiferencia sean tangentes). Si no se cumple esta igualdad, podría aumentarse la utilidad de un individuo sin disminuir la del otro, o aumentar la de ambos a la vez.

Figura 14.3: Caja de Edgeworth de consumo



2. No va a tener una forma concreta ya que se basa en una ordenación ordinal, las funciones de utilidad no tienen necesariamente una forma cóncava porque las funciones de utilidad no son estrictamente cóncavas, sino cuasicóncavas.
3. Todos sus puntos son eficientes en el consumo, por lo que
   - Existen infinitos óptimos en el consumo
   - Los puntos interiores de la FPU son ineficientes mientras que los situados fuera son inalcanzables
4. Para cada óptimo en producción va a existir una FPU, por lo que, al existir infinitos óptimos de producción, existirán infinitas FPUs.

Óptimo global

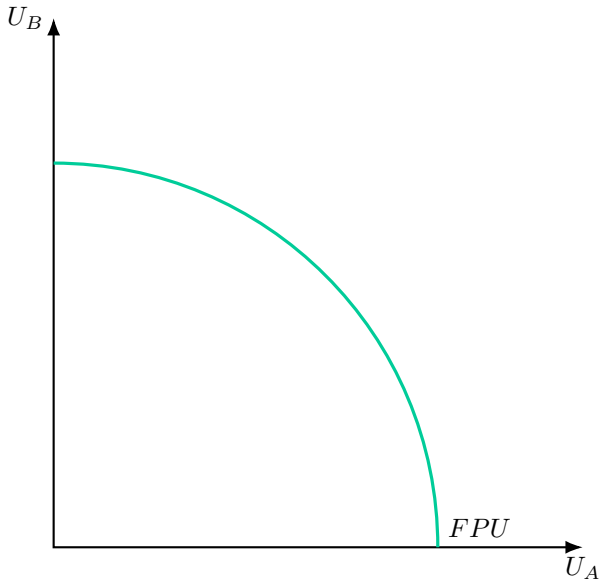
Por último, para que una asignación sea un óptimo económico debe ser un óptimo global, es decir, una asignación eficiente en la producción y en el consumo en la que, además, la valoración relativa de los bienes por los consumidores coincide con su coste de oportunidad en la producción. De esta forma:

- Para ello, se añadirá al problema de maximización de consumo (14.8) una restricción adicional, que implica que Y es una función de X, de modo que la producción esté sobre la FPP, es decir, se garantiza que se alcanza un óptimo en la producción.

$$\bar{Y} = f(X) \tag{14.14}$$

 © 2026 Manuel Cabado García

Figura 14.4: Frontera de posibilidades de utilidad



- Las CPO del problema establecerán que la relación marginal de sustitución de ambos individuos debe coincidir con la relación marginal de transformación, RMT, en la producción de los bienes[19]

$$RMS^A = RMS^B = RMT \tag{14.15}$$

  es decir, la valoración relativa de los bienes en el consumo debe coincidir con la valoración relativa en producción[20].

- Por lo tanto, la eficiencia en la producción y en el consumo es una condición necesaria pero no suficiente para alcanzar un óptimo global, pues además debe cumplirse la igualdad entre las RMS y la RMT[21].

Gráficamente, se introducirá una caja de Edgeworth en una FPP, como se observa en la Figura 14.5. Sobre la FPP, el punto E supone una producción de X e Y, X^*e Y^*, que

[19]Puesto que la RMT muestra la tasa a la cual un bien puede transformarse en otro, desde el punto de vista de la producción, y la RMS muestra la tasa a la cual los consumidores están dispuestos a cambiar un bien por el otro, el sistema no está en equilibrio a menos que todas las razones sean iguales.

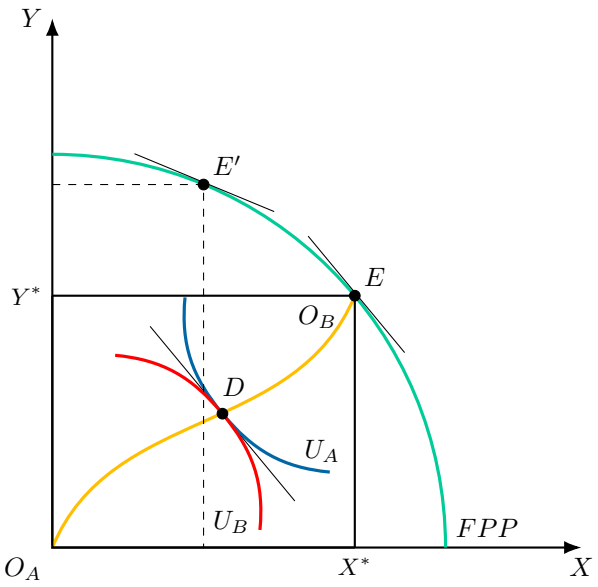
[20]Esta misma condición se obtiene resolviendo directamente el problema del dictador benevolente o planificador central, que maximiza el bienestar social sujeto únicamente a la restricción de factibilidad de la economía, sin que intervengan precios de mercado. Al no existir ninguna distorsión, ambas soluciones —la descentralizada, a través del mercado, y la centralizada, a través del planificador— coinciden exactamente. Solo cambia el mecanismo mediante el cual se alcanza el óptimo. Para una demostración completa de esta equivalencia, resolviendo ambos problemas paso a paso en un modelo dinámico, véase Torres (2010), sección 2.4. Para su versión más general y rigurosa –el Segundo Teorema del Bienestar en un contexto de crecimiento– véase Acemoglu (2009), capítulo 5.

[21]Pueden existir asignaciones eficientes en la producción o en el consumo, pero que no sean óptimos globales porque no son eficientes en ambas dimensiones simultáneamente. Incluso pueden existir asignaciones eficientes en la producción y en el consumo pero no ser óptimos globales.

determina las dimensiones de la caja de Edgeworth. Por tanto:

- Todos los puntos de las líneas de contrato son eficientes en el sentido de Pareto y, dado que dichas dotaciones de bienes se han producido también eficientemente –es decir, producción en la FPP–, se puede decir que se trata de asignaciones todas ellas eficientes tanto en consumo como en producción.
- Sin embargo, no todas son eficientes desde un punto de vista global. Será un óptimo de Pareto global aquel reparto de los bienes producidos donde se igualen las RMS de ambos consumidores entre sí e iguales a la RMT. Así, no todas las tangencias de las curvas de indiferencia situadas en la línea de contrato se producen con la misma pendiente, por lo que no todas ellas deben coincidir con la de la FPP. Ese punto –o puntos– será el óptimo global para esa producción, siendo en el gráfico el punto D, donde, para la producción del punto E, se cumple $RMS_A = RMS_B = RMT$.
- De esta forma, para otras producciones eficientes, como podría ser el punto E', se podría obtener otra caja de Edgeworth para el consumo, con otro óptimo global –o con varios o con ninguno–. Por lo tanto, puede haber infinitos óptimos globales que no serían comparables, por lo que no sería posible determinar cuál es el estado económico en el que la economía debe situarse, dándose una indeterminación.

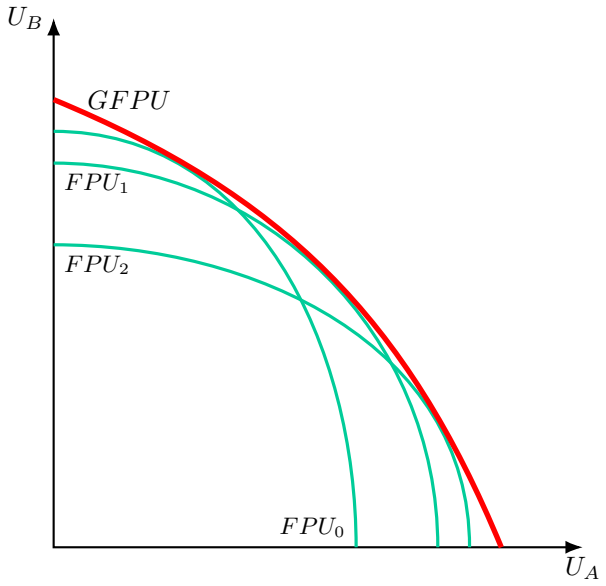
Figura 14.5: Óptimo global



Por otro lado, los diferentes óptimos globales se pueden recoger en la Gran Frontera de Posibilidades de Utilidad, $GFPU$, que gráficamente, en un plano de utilidad de los individuos A y B como el de la Figura 14.6 se define como la envolvente de los puntos de las fronteras de posibilidades de utilidad que se corresponden con óptimos globales.

 © 2026 Manuel Cabado García

Figura 14.6: Gran frontera de posibilidades de utilidad



No obstante, al existir infinitos óptimos no comparables, el criterio de Pareto puede no ser suficiente para determinar qué asignación maximiza el bienestar social, que es el objetivo último de la Economía del Bienestar. Así, la eficiencia u optimalidad paretiana global es una condición necesaria pero no suficiente para garantizar la maximización del bienestar social.

14.3. Equilibrio competitivo y Teoremas del Bienestar ⏳ 3'

Para determinar qué óptimo de Pareto maximiza el bienestar social (el óptimo optimorum), es necesario algún criterio adicional. En este sentido, surgió durante los años 30 y 40 una nueva línea de pensamiento, la Nueva Economía del Bienestar, que trató de superar esta indefinición a través de dos vías:

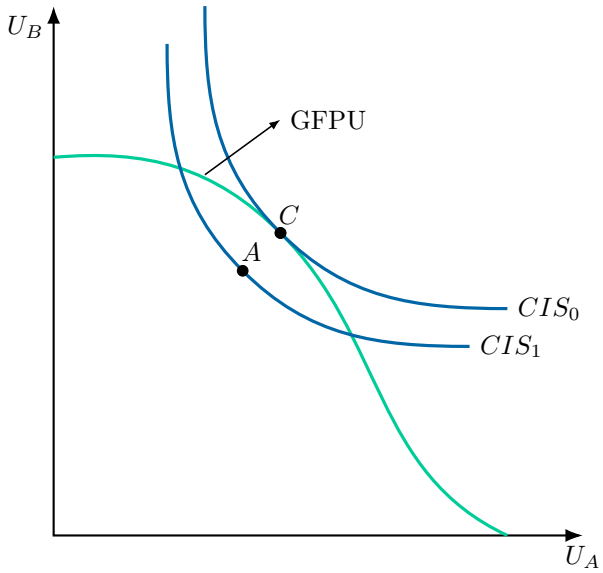
1. Línea neoutilitarista, que introduce el criterio de la función de bienestar social. Bergson desarrolló una función capaz de evaluar sin ambigüedad los estados alternativos de la economía. Ésta permitiría encontrar el punto de máxima felicidad. Es decir, aquel punto tangente a la GFPU que maximiza el bienestar social (ser óptimo global será condición necesaria pero no suficiente). Este punto será meramente tecnocrático: es independiente de los precios o del sistema económico con el que se alcance.
2. Línea neoparetiana, que introduce los criterios de compensación basados en el principio de "compensación potencial": una situación económica será preferible siempre

que los individuos que mejoran puedan indemnizar a los perjudicados.

Respecto a la primera vía, Bergson (1938) elaboró una función de bienestar social capaz de valorar sin ambigüedad los estados alternativos de la economía. Gráficamente, en un plano de utilidades de A y B como el de la Figura 14.7, se representa la $GFPU$ junto con dos curvas de indiferencia social, CIS_0 y CIS_1, cada una de las cuales une las combinaciones de utilidades de A y B a las que la función de bienestar social asigna el mismo nivel de bienestar[22].

Por ejemplo, un cambio que desplazara a la sociedad desde un punto A hacia un punto C aumentaría el bienestar de la sociedad ya que se encontraría sobre una curva de indiferencia social más alejada del origen. El punto máximo de bienestar social se alcanzará en el punto de tangencia entre la $GFPU$ y la curva de indiferencia social más alejada del origen. Este punto C se denomina "punto de felicidad". El problema de la función de bienestar social es que no cuenta con un método para construirla, por lo que se presume su existencia de forma axiomática[23].

Figura 14.7: Criterio de la función de bienestar social



No obstante, dicho punto de felicidad es puramente tecnocrático, es decir, se alcanza por las formas funcionales supuestas, que permiten la igualación de las valoraciones de distintos agentes –$RMS = RMT$– con la igualación social –pendiente de la curva de indiferencia social–, con independencia de los precios.

[22] La intuición es que la sociedad aceptaría que un individuo pierda algo de utilidad si otro gana lo suficiente, igual que un consumidor acepta menos de un bien a cambio de más de otro.

[23] La teoría de la elección social estudia precisamente este problema, pero su resultado central, el teorema de imposibilidad de Arrow, muestra que no existe ninguna regla que obtenga un orden social a partir de las preferencias individuales y respete a la vez unas condiciones mínimas de racionalidad y de no dictadura

 © 2026 Manuel Cabado García

¿Existe algún sistema económico con el que se pueda alcanzar el óptimo global? Sí, la competencia perfecta (Equilibrio General Competitivo). Se puede demostrar que la competencia perfecta conduce a una situación de equilibrio general que es óptima de Pareto global (aunque no necesariamente una situación que maximice el bienestar social[24]). Se basa en los siguientes supuestos:

1. Los mercados son completos, es decir, todos los bienes se intercambian en el mercado a precios públicamente conocidos.
2. Información perfecta, no existen fallos de información.
3. Agentes precio-aceptantes.
4. Todas las acciones quedan determinadas por los precios.

Bajo estos supuestos, un equilibrio competitivo se define como un vector de precios y de asignaciones de bienes y factores tal que los agentes maximizan su utilidad –por el lado del consumo– y los beneficios –por el lado de la producción– y todos los mercados se vacían.

A partir de esta noción de equilibrio competitivo se formulan los dos teoremas fundamentales de la Economía del Bienestar:

- Primer Teorema de la Economía del Bienestar: en ausencia de fallos de mercado, todo equilibrio general competitivo es un óptimo de Pareto[25].
- Segundo Teorema de la Economía del Bienestar, cualquier asignación eficiente en sentido de Pareto puede alcanzarse como un equilibrio competitivo a través de la redistribución de las dotaciones iniciales. Este resultado exige que las preferencias de los consumidores y los conjuntos de producción de las empresas sean convexos –condiciones ya recogidas en los supuestos del modelo–, y que la redistribución se realice mediante transferencias de suma fija que no distorsionen las decisiones de los agentes.

Por un lado, una consecuencia directa del segundo teorema del bienestar es que el sector público podría llevar a cabo transferencias de suma fija y dejar posteriormente trabajar a los mecanismos de mercado para alcanzar un óptimo de Pareto global que maximiza el

[24] Que el punto de equilibrio general competitivo (que es un óptimo global) maximice, a su vez, el bienestar social, dependerá de si dicho equilibrio coincide con el punto de máxima felicidad, esto es, con el punto donde la Gran Frontera de Posibilidades de Utilidad es tangente a la curva de indiferencia derivada de la función de bienestar social. Precisamente esta posible divergencia entre el punto eficiente de equilibrio y el punto eficiente de maximización del bienestar social es lo que da pie al segundo teorema del bienestar, que, como se verá, defiende que, a través de una redistribución previa vía transferencias de suma fija, y dejando después actuar a los mecanismos del mercado, se puede alcanzar otra solución eficiente pero que, además, maximice el bienestar social.

[25] La intuición es que, si los precios recogen todos los costes y beneficios, cada consumidor iguala su RMS al precio relativo y cada empresa iguala su RMT a ese mismo precio, de modo que no queda ningún intercambio ni cambio en la producción que mejore a alguien sin perjudicar a otro.

bienestar social[26].

A su vez, el segundo teorema implica que es posible separar las dimensiones de equidad y eficiencia, por lo que los defensores de la equidad utilizarán este resultado para defender que, aun en el caso de que no haya imperfecciones, el Estado debe intervenir para mejorar la equidad, sin que por ello se tenga que resentir la eficiencia.

Por otro lado, el primer teorema del bienestar sirve en defensa del libre mercado, limitando así el papel del Estado a la corrección de imperfecciones.

Valoración del criterio paretiano

En definitiva, el criterio de Pareto ofrece ciertas ventajas:

1. Proporciona un criterio de eficiencia económica.
2. Emplea un análisis generalmente aceptado, especialmente por el marginalismo.
3. Evita consideraciones distributivas y comparaciones interpersonales de utilidad.
4. Introduce juicios débiles y generalmente aceptados.
5. Formaliza las ventajas de la competencia perfecta como sistema de asignaciones.

Aun así, presenta también una serie de problemas desde el punto de vista colectivo debido a

1. Existencia de infinitas soluciones.
2. No se pueden realizar comparaciones.
3. Independencia del proceso. Para el criterio de Pareto es igualmente válida, ceteris paribus, una decisión que haya sido adoptada democráticamente que otra que ha sido impuesta de manera dictatorial.
4. No tiene en cuenta la distribución, siendo la solución de esquina –un individuo consume todo y el otro nada– un óptimo en el consumo –pues para que el que no tiene nada mejore, el que tiene todo debe empeorar–[27].
5. Tiranía de las minorías, como analizó Baumol (1952). El criterio de Pareto puede resultar inoperante para la política económica, ya que una nueva asignación puede

[26] El contra-argumento que se ha esgrimido es que es prácticamente imposible introducir un sistema de impuestos y transferencias óptimo (es decir, que no distorsione las decisiones y el comportamiento de los agentes económicos), por lo que pueden surgir pérdidas de eficiencia y aparecer con fuerza el trade-off equidad vs. eficiencia.

Pero, ¿cuál es la diferencia entre igualdad y equidad?: Igualdad, consiste en darles a todos lo mismo, por ende el resultado dependerá de las características del individuo. La equidad consiste en el reparto de bienes en función de las necesidades de cada individuo buscando que todos obtengan el mismo resultado.

[27] En ese sentido, se ha propuesto el concepto de "asignaciones justas en el sentido de Foley" (Foley, 1967), que serían aquellas que, además de Pareto eficientes, son equitativas. Así, se entendería como "equitativas" aquellas asignaciones libres de envidia, es decir, en las que ningún agente prefiere la asignación que tiene el otro.

 © 2026 Manuel Cabado García

ser vetada si tan solo un individuo la rechaza, lo que implica la unanimidad en las decisiones.

6. Supuestos muy restrictivos: los fallos de mercado son comunes.
7. No es capaz de ofrecer una ordenación completa entre los óptimos de Pareto ni determinar cuál es el que maximiza el bienestar.

Por este último motivo, se plantean a continuación dos aportaciones de la Nueva y la Novísima Economía del Bienestar:

1. Los criterios de compensación, que intentan romper con la indeterminación del criterio de Pareto y con la regla de la unanimidad –algo que también intentaba hacer la función de bienestar social analizada anteriormente–.
2. La teoría del second best, que trata de dar una respuesta a los casos en los que se producen imperfecciones en el mercado.

14.4. Los criterios de compensación ⏳2'30"

Para superar tanto la indeterminación como la restrictividad del criterio de Pareto surgen los criterios de compensación (Boadway y Bruce, 1984). Concretamente, se busca un criterio que permita pronunciarse sobre la deseabilidad de cualquier combinación de situaciones económicas, permitiendo una ordenación completa de los estados económicos alternativos para encontrar aquella que maximice el bienestar social.

Para poder analizarlos, gráficamente en un plano de utilidad de A y B como el de la Figura 14.8, se representan dos FPU con diferentes puntos: 1, 2, 3 y 4. A partir de los mismos se presentan los criterios de compensación:

1. Criterio de Kaldor (1939): una asignación será preferible a otra si los ganadores pueden compensar a los perdedores para que haya cambio y, aun así, seguir beneficiándose del cambio. En la Figura 14.8a, el paso de 1 a 2 no se puede determinar por el criterio de Pareto porque ambos no son comparables. Sin embargo, según el criterio de Kaldor deberá llevarse a cabo porque el ganador –consumidor A– puede compensar al consumidor B situándose en 3 y, aun así, estar mejor que en la situación inicial.
2. Criterio de Hicks (1939): una asignación será preferible a otra si los perdedores no pueden compensar a los ganadores para que no haya cambio y, aun así, beneficiarse de que no haya cambio. Dicho de otro modo: los perdedores podrían intentar "sobornar" a los ganadores, ofreciéndoles algo a cambio de que renuncien al cambio. El criterio de Hicks se cumple si ese soborno no les compensaría, es decir, si aun pagándolo y evitando el cambio, los perdedores seguirían estando peor que si el cambio se hubiera producido. A diferencia de Kaldor, cuyo ejemplo se ilustraba con

el paso de 1 a 2, el criterio de Hicks se ilustra aquí con una transición distinta que resulta concluyente por sí sola –aplicar Hicks directamente a la transición de 1 a 2 ya adelantaría la inconsistencia entre criterios que se desarrollará más abajo con la paradoja de Scitovsky–. En la misma Figura 14.8a, el paso de 3 a 4 deberá llevarse a cabo porque el perdedor –consumidor B– no puede compensar al consumidor A situándose en 2 y estar mejor que si hubiese cambio.

3. Criterio de Scitovszky (1941): Scitovsky demostró que existirán casos donde el criterio de Kaldor o el de Hicks podrían entrar en contradicción con ellos mismos, dando lugar a la paradoja de Scitovsky. En la Figura 14.8a, según Kaldor es óptimo pasar de 1 a 2, pero una vez en 2 el criterio de Kaldor indica que es óptimo pasar de 2 a 1. Por lo tanto, bajo este criterio la economía estaría sometida a cambios cíclicos entre ambos estados.
   Esta paradoja se produce porque entre las FPU que se comparan existe un punto de corte, es decir, ninguna domina completamente a la otra.
   Para solucionarlo, Scitovsky propuso un doble criterio, es decir, que se cumplan a la vez el criterio de Kaldor y el de Hicks. Así, un cambio será potencialmente Pareto superior si los ganadores pueden compensar potencialmente a los perdedores y aun así alcanzar un nivel de utilidad superior al nivel inicial y, a su vez, los perdedores no pueden compensar a los ganadores a cambio de que no se produzca el cambio y, a su vez, alcanzar un nivel de utilidad superior al percibido tras el cambio.
   En la Figura 14.8b, la transición de 1 a 2 cumple dicho criterio ya que el ganador –consumidor B– puede compensar a A en 3. A su vez, A no puede compensar a B en 4.

4. Criterio de Samuelson (1950): Samuelson solucionó la posible intransitividad que genera el criterio de Scitovsky. Como se observa en la Figura 14.8c, según Samuelson, una asignación 2 será preferible a una asignación 1 si la Frontera de Posibilidades de Utilidad correspondiente a 2 es completamente exterior a 1.
   Se trata de un criterio muy exigente, porque solo permite pronunciarse cuando una frontera domina por completo a la otra, pero no es redundante con el de Pareto: los puntos concretos 1 y 2 pueden no ser comparables según Pareto y, aun así, la situación 2 es preferible, porque con una redistribución adecuada todos podrían estar mejor que en cualquier reparto posible de la situación 1[28].

5. Criterio de Little (1950): Little se apartó del criterio paretiano y consideró que hay que tener en cuenta aspectos redistributivos aunque ello suponga consideraciones de equidad. Este autor propuso que se cumplan no solo los criterios de Kaldor y de Hicks, sino que, además, se produzca una mejoría en la distribución de la renta.

[28] La intuición es que se juzga el tamaño de la tarta y no el reparto concreto de cada trozo: si cualquier forma de repartir la tarta antigua puede mejorarse para todos con la nueva, la nueva es preferible aunque, tal como se reparte hoy, alguien salga perdiendo.

 © 2026 Manuel Cabado García

Como se observa en la Figura 14.8d, la aplicación simultánea de los criterios de Kaldor y de Hicks determina que tanto el punto 2 como el punto 3 sean preferibles al punto 1. En cambio, si se añade el requisito adicional de Little de que la distribución de la renta mejore, entonces el punto 2 seguirá siendo preferible a 1, pues se cumplen simultáneamente los criterios de Kaldor y de Hicks y, además, la distribución de la renta mejora pues está más cerca de la bisectriz –que reflejaría aquellos casos en que la distribución es totalmente igualitaria–. Sin embargo, el punto 3 ya no es preferible al 1 debido a que la distribución de la renta empeora.

Sin embargo, los criterios de compensación presentan un problema[29]: que suponen compensaciones potenciales, no efectivas, lo que conlleva implícita una comparación interpersonal de la utilidad (la ganancia de utilidad de los que mejoran es más deseable que la disminución de la utilidad de los que empeoran), algo que una función de utilidad ordinal[30] como tal no permite[31].

14.5. Teoría del Second Best ⏳ 3'

Por último, ya dentro de la Novísima Economía del Bienestar, Lipsey y Lancaster (1956) formalizaron la teoría del second best, según la cual, si una de las condiciones de óptimo de Pareto no puede cumplirse, cumplir las demás deja de ser necesariamente deseable.

Teoría tradicional Cuando alguna de las condiciones de óptimo de Pareto no se cumplía, se creía deseable alcanzar el máximo número posible de condiciones de óptimo, pues con ello se creía que la economía se acercaría más al óptimo global. En términos de política económica esto tenía unas implicaciones muy claras: había que llevar a cabo medidas dirigidas a restaurar el máximo número posible de condiciones de óptimo.

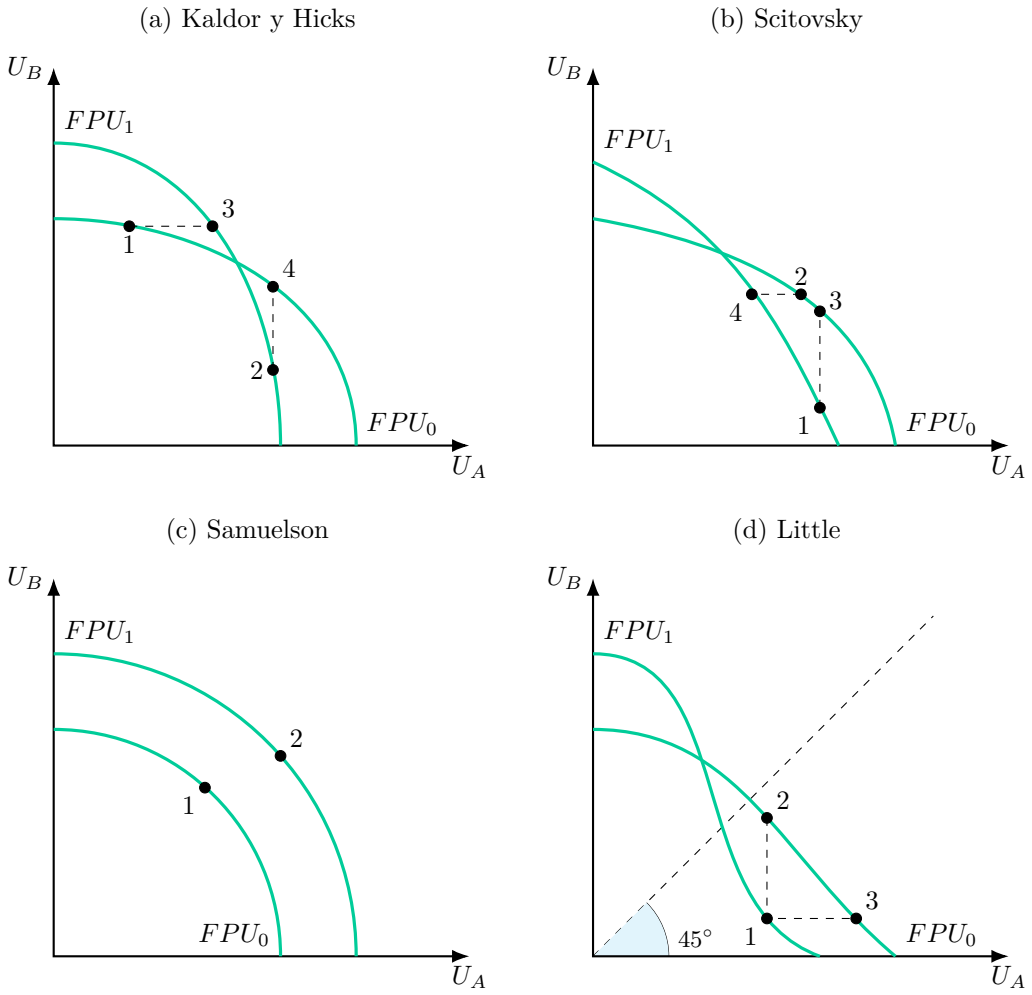
[29] Además, los criterios de compensación son tan parecidos que cuando se aplican a situaciones no comparables según el criterio de Pareto conducen a resultados inconsistentes, y cuando se aplican a situaciones comparables según Pareto no son necesarios.

[30] No se puede comparar la utilidad entre personas porque la utilidad es un concepto ordinal, no cardinal. En otras palabras, la utilidad permite ordenar las preferencias de una persona (por ejemplo, me gusta más el chocolate que la vainilla), pero no da un valor numérico exacto para la satisfacción. No hay una unidad de medida universal para la felicidad o el bienestar. Dado que la satisfacción es una experiencia subjetiva y personal, es imposible saber si la satisfacción que siente una persona al obtener algo es mayor o menor que la que siente otra persona. Por lo tanto, no se pueden hacer comparaciones interpersonales de utilidad.

De esta forma, tal y como hace la Microeconomía moderna (dado que es un enfoque más realista y objetivo), es mejor usar la utilidad ordinal porque la utilidad cardinal es imposible de medir de forma objetiva. La utilidad ordinal, al simplemente ordenar las preferencias, es un modelo más realista y menos exigente. Permite analizar el comportamiento del consumidor sin necesidad de asignarle un valor numérico a la satisfacción, algo que la ciencia no puede hacer.

[31] Además, existen una serie de puntos complicados como por ejemplo: i) ¿es ético que los ricos sean más ricos si pueden compensar a los pobres? ii) ¿existe algún mecanismo institucional que permita transformar las compensaciones potenciales en efectivas? Sería muy complicado iii) la obtención de información también es un proceso complicado y complejo iv) inexistencia de mecanismos que hagan cumplir los acuerdos.

Figura 14.8: Criterios de compensación



Teoría del second best Cuando no se cumple alguna de las condiciones de óptimo, entonces las otras condiciones de Pareto, aunque se puedan cumplir o alcanzar, no son necesariamente deseables. Por lo tanto, las condiciones de Pareto son deseables sólo si se cumplen todas de manera simultánea. Así:

1. El número de condiciones de óptimo que cumplan distintas situaciones en las que existen distorsiones no aporta información. Es decir, una situación que cumpla un mayor número de condiciones de óptimo no es necesariamente superior a otra que satisfaga un menor número de condiciones.
2. Cuando existan varias distorsiones, eliminar una de ellas puede afectar al bienestar de forma positiva, negativa o neutral.
   En términos de política económica, esto significa que, en presencia de al menos una distorsión no evitable, el policymaker no estaría entonces obligado a eliminar el resto de distorsiones (y podría, incluso añadir alguna más).

Gráficamente, en un plano de utilidades como el de la Figura 14.9, se representa la $GFPU$,

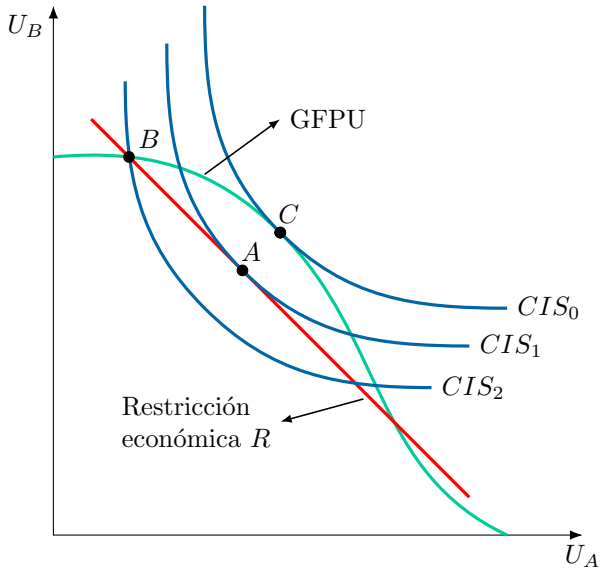
 © 2026 Manuel Cabado García

donde todos los puntos son óptimos de Pareto[32] (pues se cumplen las condiciones $RMS^A = RMS^B = RMT$), y diferentes curvas de indiferencia social con diferentes puntos, siendo el punto C el punto de felicidad –máximo bienestar social–.

Si se introduce una restricción, R, en la economía tal que los puntos por encima de esa recta ya no son alcanzables –es decir, el punto C no es alcanzable–, cabe preguntarse qué punto es ahora mejor, si A o B:

- El punto B es un óptimo paretiano
- El punto A, por su parte, no es un óptimo, pero da lugar a un mayor nivel de bienestar social, por lo que esta asignación A será preferible socialmente a B.

Figura 14.9: GFPU con restricción en la economía



Esto demuestra que, ante la existencia de una restricción, será mejor, en este ejemplo, abandonar las condiciones de optimalidad que sí se cumplen, pues así se mejora el bienestar social.

Ejemplo: monopolio contaminante. Se supondrá que existe un monopolio contaminante, por lo que existirán dos distorsiones: poder de mercado y externalidad.

- El policymaker puede pensar que eliminando el monopolio –es decir, eliminando una distorsión– se puede estar más cerca del óptimo paretiano. Sin embargo, al existir más de una ineficiencia –además del poder de mercado, la externalidad de la contaminación–, la teoría del second best dice que esto no tiene por qué ser necesariamente así. De hecho, si se desmantela el monopolio y se introduce competencia

[32] Se asume que existe una restricción adicional al problema de Pareto tradicional

en el mercado, se producirá más y, por tanto, la externalidad negativa será mayor, pudiendo disminuir el bienestar social.

- Del mismo modo, el policymaker puede pensar que introducir una distorsión adicional –por ejemplo, un impuesto a la producción– podría alejar más a la economía del óptimo de Pareto. Pero lo cierto es que la introducción de un impuesto podría disminuir la producción y, con ésta, la contaminación.

Davis y Whinston (1965) trataron de rescatar lo fundamental del criterio de Pareto, sosteniendo que si la restricción es muy independiente del resto de la economía, entonces en el sector de la distorsión se debería aplicar una solución de segundo óptimo, mientras que en el resto de sectores se debería rescatar la asignación paretiana.

Otros ejemplos de la Teoría del Second Best son:

1. Teoría de la integración económica: Las uniones aduaneras. El efecto final depende de la desviación de comercio y de la creación de comercio.
2. Modelo de Brecher y Díaz Alejandro (1977): Liberalizar la Cuenta Financiera sin haber liberalizado la Cuenta Corriente puede llevar a una reducción del bienestar económico[33].
3. Modelos de Riesgo Moral y de Selección Adversa.

14.6. Conclusiones ⏳30"

Como conclusión, la Economía del Bienestar ofrece herramientas para evaluar la eficiencia de las políticas económicas, combinando análisis técnico con juicios de valor. El criterio de Pareto, plasmado en los dos teoremas del bienestar, sostiene que el libre mercado puede alcanzar asignaciones eficientes, y que la intervención estatal se justifica para corregir fallos o redistribuir equitativamente la riqueza.

No obstante, este criterio tiene limitaciones: no permite comparar situaciones donde unos ganan y otros pierden. Por ello, se han desarrollado enfoques complementarios como los criterios de compensación, que admiten mejoras potenciales aunque no sean Pareto óptimas. La teoría del "Second Best" advierte que corregir una sola imperfección en un sistema con múltiples fallos puede empeorar el resultado.

Además, el análisis del bienestar se ha ampliado para abordar contextos más realistas, como mercados incompletos, externalidades o poder de mercado, adaptando sus herramientas a la complejidad del mundo actual.

[33] Así, si un país mantiene un arancel que protege a su sector importador-competidor intensivo en capital –una distorsión en la Cuenta Corriente– y, sin corregirla, liberaliza la Cuenta Financiera, la entrada de capital extranjero expande aún más ese sector ya protegido, alejando a la economía de la situación de libre comercio. Como el capital extranjero recibe el valor íntegro –sin gravar– de su producto marginal, el resultado es una pérdida neta de bienestar para el país receptor: liberalizar una sola de las dos cuentas, sin corregir la distorsión de la otra, empeora el resultado.

 © 2026 Manuel Cabado García

Apéndice

14.A.1. Planteamiento analítico de la teoría del second best

A continuación se formaliza la intuición ya presentada en el cuerpo del tema: partiendo de una situación en la que se cumplen todas las condiciones de optimalidad, se rompe una de ellas para comprobar qué ocurre con las demás. Por simplicidad, se modela a la sociedad como si fuera un único consumidor representativo, cuya restricción $T(x_1, x_2, \ldots, x_n) = 0$ recoge, de forma agregada, las posibilidades de transformación de la economía –el equivalente, para n bienes, a la FPP del cuerpo del tema–.

Hay un consumidor (la sociedad) y n bienes. Sin restricciones, el problema del consumidor es

$$\text{máx} \quad U(x_1, x_2, \ldots, x_n)$$

sujeto a

$$T(x_1, x_2, \ldots, x_n) = 0$$

Se construye el lagrangiano

$$L = U(x_1, x_2, \ldots, x_n) - \lambda \cdot T(x_1, x_2, \ldots, x_n)$$

Se obtienen las CPO

$$\begin{aligned}
&\frac{\partial L}{\partial x_i} = 0 \rightarrow \frac{\partial U}{\partial x_i} - \lambda \cdot \frac{\partial T}{\partial x_i} = 0 \\
&\quad \vdots \\
&\frac{\partial L}{\partial x_n} = 0 \rightarrow \frac{\partial U}{\partial x_n} - \lambda \cdot \frac{\partial T}{\partial x_n} = 0 \\
&\frac{\partial L}{\partial x_n} = 0 \rightarrow T(x_1, x_2, \ldots, x_n) = 0
\end{aligned}$$

por lo que

$$RMS_i^j = RMT_i^j$$

Supóngase ahora que no se cumple una de las condiciones (por ejemplo, la primera):

$$\frac{\partial U}{\partial x_1} - \lambda \cdot \frac{\partial T}{\partial x_1} \neq 0 \rightarrow \frac{\partial U}{\partial x_1} - \beta \cdot \frac{\partial T}{\partial x_1} = 0, \quad \beta \neq \lambda, \quad \beta > 0$$

Ahora el problema será:

$$\text{máx} \quad U(x_1, x_2, \ldots, x_n)$$

sujeto a

$$T(x_1, x_2, \ldots, x_n) = 0$$
$$\frac{\partial U}{\partial x_1} - \beta \cdot \frac{\partial T}{\partial x_1} = 0$$

Se construye el lagrangiano, donde el nuevo multiplicador μ recoge el coste marginal, en términos de bienestar, de mantener rota la condición del bien 1 –es decir, cuánto "cuesta" no poder ajustar libremente x_1–:

$$L = U(x_1, x_2, \ldots, x_n) - \lambda \cdot T(x_1, x_2, \ldots, x_n) - \mu \cdot (u_1' - \beta t_1')$$

Se obtienen las CPO

$$\frac{\partial L}{\partial x_i} = 0 \to \frac{\partial U}{\partial x_i} - \lambda \cdot \frac{\partial T}{\partial x_i} - \mu (u_{1i}'' - \beta t_{1i}'') = 0, \quad \forall i = 1, \ldots, n$$
$$\frac{\partial L}{\partial \lambda} = 0 \to T(x_1, x_2, \ldots, x_n) = 0$$
$$\frac{\partial L}{\partial \mu} = 0 \to u_1' - \beta t_1' = 0$$

por lo que

$$RMS_i^j = \frac{\lambda t_i' + \mu (u_{1i}'' - \beta t_{1i}'')}{\lambda t_j' + \mu \left(u_{1j}'' - \beta t_{1j}''\right)}$$

Y, en principio, no puede decirse nada sobre los signos de las derivadas incluidas en el segundo miembro, por lo que, aunque la solución siga estando sobre la FPP, las demás condiciones $RMS_i^j = RMT_i^j$ dejan, en general, de ser óptimas. Esta es, precisamente, la idea central de la teoría del second best: si una condición de óptimo no puede cumplirse, cumplir todas las demás no garantiza el máximo bienestar alcanzable, y puede convenir apartarse también de ellas[34].

14.A.2. Ejemplos de la teoría del Second Best

Aquí se dan seis ejemplos alternativos al monopolio contaminante visto anteriormente:

1. Subsidios en mercados con externalidades negativas. En un mercado ideal, se corrigen las externalidades negativas (como la contaminación) a través de un impuesto, que hace que el coste social se refleje en el precio. Sin embargo, si por motivos políticos o de viabilidad no se puede aplicar un impuesto, un subsidio a una actividad complementaria puede ser una alternativa. Por ejemplo, en lugar de gravar el uso de coches contaminantes, se podría subsidiar el transporte público o la bicicleta.

[34] Salvo si i) la función de utilidad es aditiva y separable y ii) las tecnologías son lineales, en cuyo caso $u_{1i}'' = 0$ y $t_{1i}'' = 0$, por lo que para la maximización del bienestar seguiría siendo necesario cumplir las condiciones de óptimo.

 © 2026 Manuel Cabado García

Esta medida es un "second best" porque, aunque no ataca directamente la fuente de la contaminación, puede incentivar un cambio de comportamiento que reduzca su impacto, logrando un resultado más eficiente que no hacer nada.

2. Controles de precios en mercados con información asimétrica. En un mercado ideal, la información es perfecta. Cuando hay información asimétrica (una de las partes sabe más que la otra), como ocurre en el mercado de seguros (los clientes saben más de su riesgo que las aseguradoras), pueden surgir fallos (Akerlof, 1970). La teoría del "first best" sugiere que se debe solucionar la asimetría de información. Sin embargo, en la práctica, esto es muy difícil. La teoría del "second best" sugiere que una regulación, como un control de precios, puede ser una solución viable. Aunque el control de precios no es una herramienta ideal, puede ayudar a mitigar los efectos de la selección adversa y aumentar la eficiencia general del mercado.
3. Subsidios a la vivienda en mercados inmobiliarios ineficientes. En teoría, el mercado de vivienda es más eficiente sin intervención. Sin embargo, si existe una ineficiencia previa, como una regulación de zonificación que limita la oferta de terrenos para construcción, el precio de la vivienda puede dispararse (Glaeser y Gyourko, 2003). En este escenario, la mejor solución ideal sería eliminar la regulación de zonificación. Si eso no es posible, una política "second best" podría ser un subsidio a la vivienda para las familias de bajos ingresos. Si bien el subsidio es una distorsión en sí mismo, puede compensar la distorsión del mercado y mejorar el bienestar de la población afectada.
4. Políticas de salario mínimo. En un mercado laboral de competencia perfecta, un salario mínimo por encima del salario de equilibrio puede generar desempleo. Sin embargo, en mercados con monopsonio (un solo comprador de trabajo, como una gran empresa en un pueblo pequeño), la empresa puede fijar salarios más bajos (Stigler, 1946). En este caso, la política "first best" sería eliminar el poder de monopsonio, lo cual es complejo. La teoría del "second best" sugiere que un salario mínimo moderado puede en realidad mejorar la eficiencia y el bienestar social, ya que contrarresta el poder de mercado del monopsonio y podría aumentar el empleo.
5. Regulaciones de seguridad laboral. La solución "first best" para la seguridad en el lugar de trabajo sería que las empresas y los trabajadores, con información completa, negociaran libremente los salarios y el nivel de riesgo. No obstante, a menudo existe información asimétrica (los trabajadores no saben todos los riesgos que existen en su trabajo), lo que lleva a un nivel de seguridad subóptimo. La teoría del "second best" indica que, en lugar de esperar a que la información sea perfecta (algo casi imposible), es mejor implementar regulaciones de seguridad laboral obligatorias. Aunque estas regulaciones son una distorsión del mercado, son un "segundo mejor" porque mejoran la seguridad y el bienestar de los trabajadores, compensando el fallo

de mercado original.

6. Fijación de precios de Ramsey en monopolios regulados. En un mercado ideal, el precio de un monopolio natural (agua, electricidad) se fijaría igual al coste marginal. Sin embargo, si el coste marginal es inferior al coste medio, esa regla genera pérdidas que deben cubrirse con subvenciones no siempre viables. La teoría del "second best" propone la regla de Ramsey (Ramsey, 1927; Corlett y Hague, 1953): fijar en cada mercado un margen relativo sobre el precio, $(p_i - CMg_i)/p_i$, inversamente proporcional a la elasticidad-precio de su demanda, de forma que se minimice la pérdida de eficiencia total con la condición de que la empresa cubra sus costes[35].

14.A.3. Críticas, limitaciones y extensiones del criterio de Pareto

Antes de presentar las extensiones más recientes, conviene recapitular brevemente las críticas ya avanzadas al criterio de Pareto:

1. Tiranía de las minorías (Baumol, 1952). El criterio de Pareto puede resultar inoperante para la política económica, ya que una nueva asignación puede ser vetada si tan solo un individuo la rechaza, lo que implica la necesidad de unanimidad en las decisiones. Por ejemplo, la regla del consenso en la OMC.
2. Independencia del proceso: para el criterio de Pareto es igualmente válida, ceteris paribus, una decisión que ha sido adoptada democráticamente que otra que ha sido impuesta de manera dictatorial.
3. No tiene en cuenta aspectos redistributivos. De hecho, la solución de esquina (un individuo consume todo y el otro nada) siempre es un óptimo en el consumo (pues para que el que no tiene nada mejore, el que tiene todo tiene que empeorar)[36].
4. Indeterminación: puede no ofrecer una ordenación completa de los estados económicos alternativos, pues pueden existir infinitos óptimos de Pareto que no serían comparables entre sí. Así como tampoco son comparables puntos óptimos y no óptimos.
5. Supuestos muy restrictivos: los fallos de mercado son comunes.

Para responder a estas limitaciones, con el paso del tiempo han ido surgiendo nuevas aportaciones que pretenden superar algunas de sus principales deficiencias. Así:

- Los criterios de compensación permiten valorar y comparar una serie de situaciones

[35] La intuición es que el agua o la electricidad se siguen consumiendo casi igual aunque se encarezcan, de modo que cargar en ellas la mayor parte del coste fijo apenas distorsiona las decisiones, mientras que hacerlo en un servicio con muchos sustitutos espantaría a los usuarios.

[36] En respuesta a esto, se ha propuesto el concepto de "asignaciones justas en el sentido de Foley", que serían aquellas que, además de Pareto eficientes, son equitativas (entendiendo "equitativas" como aquellas asignaciones libres de envidia, es decir, en las que ningún agente prefiere la asignación que tiene el otro).

 © 2026 Manuel Cabado García

no comparables bajo el criterio de Pareto, y hacer dicho criterio más útil en su aplicación política al romper la regla de unanimidad.

- La teoría del second best muestra que las condiciones de optimalidad son deseables solo si se cumplen todas de manera simultánea. En efecto, en presencia de alguna distorsión, la conducta optimizadora puede llevar a abandonar otras condiciones alcanzables, lo que presenta importantes lecciones para la política económica.

Por último, el criterio de Pareto queda sujeto a las condiciones de competencia perfecta. Sin embargo, el análisis del bienestar buscará evaluar también aquellas situaciones en las que existan fallos de mercado. Por ello, se han introducido extensiones posteriores:

- Ruptura de los supuestos de mercados completos e información perfecta: Hart (1975) mostró que, con mercados incompletos, el equilibrio competitivo puede no ser eficiente ni siquiera respecto a los mercados existentes, lo que dio lugar al análisis de la "eficiencia paretiana restringida".
- Presencia de externalidades y bienes públicos: El sistema de precios ya no recogerá todos los costes y beneficios. Como solución se ha desarrollado el análisis coste-beneficio (Prest y Turvey, 1965), que permite tener en cuenta los costes y beneficios tanto privados como sociales.
- Existencia de poder de mercado: En el marco de la Teoría de la Organización Industrial, Baumol (1982) desarrolló la Teoría de los mercados contestables, argumentando que bajo ciertas condiciones, incluso en competencia imperfecta los agentes se pueden comportar competitivamente.

14.A.4. La liberalización del sector del taxi

La liberalización del sector del taxi es un tema complejo, sin una respuesta única, que genera fuertes debates entre los defensores de la competencia y los partidarios de la regulación. Desde el punto de vista económico, una opinión equilibrada reconoce tanto los beneficios potenciales de la liberalización como los riesgos asociados.

Criterios de eficiencia para valorar la liberalización: Para valorar la liberalización del sector del taxi, es esencial usar varios criterios económicos centrados en la eficiencia:

- Eficiencia productiva: Se refiere a si los servicios se producen al menor coste posible. La liberalización, al permitir la entrada de nuevos competidores (como los VTC), podría fomentar una mayor competencia y, por lo tanto, forzar a las empresas a ser más eficientes, reduciendo sus costes y mejorando la calidad de sus servicios para no perder clientes. De hecho, la evidencia empírica confirma una mayor tasa de utilización de la capacidad –tiempo y kilómetros con pasajero– en las plataformas de VTC frente al taxi tradicional (Cramer y Krueger, 2016). Por otro lado, la saturación del mercado podría llevar a una subutilización de los vehículos, reduciendo la eficiencia.

- Eficiencia asignativa: existe cuando cada servicio se presta hasta el punto en que lo que los usuarios están dispuestos a pagar por la última unidad coincide con su coste marginal. Un mercado liberalizado podría conducir a una mayor variedad de servicios (taxis de lujo, económicos, etc.) y a una mejor cobertura geográfica, satisfaciendo la demanda donde el servicio tradicional de taxi no llega de forma óptima. Esto podría reducir los tiempos de espera y mejorar la accesibilidad, beneficiando a los consumidores. Además, las aplicaciones de movilidad mejoran la transparencia de la información para los consumidores –precio y ruta antes de iniciar el viaje–, lo que reduce la incertidumbre y evita los precios abusivos en horas pico o para turistas, permitiendo así una asignación más ajustada a las preferencias reales del consumidor. En un sistema de taxis regulado, las tarifas son fijas, pero la falta de transparencia sobre la disponibilidad puede ser un problema. Sin embargo, un exceso de oferta podría generar congestión vial, lo que perjudicaría a toda la población.
- Eficiencia dinámica: Se centra en la capacidad del mercado para innovar con el tiempo. La liberalización, especialmente a través de la introducción de plataformas de movilidad (Uber, Cabify), ha impulsado la innovación tecnológica (aplicaciones, pagos electrónicos, geolocalización) que mejora la experiencia del usuario. La competencia fomenta la mejora continua de los servicios y la adaptación a las nuevas necesidades del mercado.

Así, la liberalización ilustra bien el tipo de disyuntiva que ocupa a la economía del bienestar: las ganancias de eficiencia productiva, asignativa y dinámica que trae la competencia solo se traducen en una mejora social neta si el diseño regulatorio internaliza las externalidades negativas –congestión, contaminación– que la propia liberalización puede generar.

14.A.5. La importancia de los precios relativos en Economía

Los precios relativos son la herramienta fundamental que coordina la asignación de recursos en una economía de mercado. A diferencia del precio nominal de un bien, que es su valor en términos monetarios, un precio relativo es el precio de un bien comparado con el precio de otro.

Su importancia radica en que guían las decisiones de productores y consumidores. Si el precio relativo de un bien aumenta, los productores tienen un incentivo para desviar más recursos hacia su producción, ya que se ha vuelto más rentable. Por otro lado, los consumidores buscan sustitutos más baratos, lo que genera una nueva demanda. Este cambio en la asignación de recursos es la base de la eficiencia en el mercado. Los precios relativos informan a los agentes económicos sobre la escasez y el valor de los bienes, permitiendo que la economía se ajuste dinámicamente.

En un mercado de competencia perfecta, los precios relativos reflejan los costes marginales de producción y las preferencias de los consumidores, asegurando que los recursos

 © 2026 Manuel Cabado García

se destinen a los usos más valorados. Sin ellos, el mercado sería incapaz de operar eficientemente.

Este es, de hecho, el papel que cumplen la RMS y la RMT en el modelo desarrollado en este tema: los precios relativos de equilibrio son precisamente los que igualan ambas relaciones marginales, coordinando las decisiones de consumo y producción sin necesidad de un planificador central.

14.A.6. Mercados completos e incompletos

La distinción entre mercados completos e incompletos se basa en la capacidad de los agentes económicos para asegurar la gestión de riesgos futuros y la realización de transacciones en todas las contingencias posibles. Esta diferencia es fundamental para el estudio de la eficiencia de los mercados.

Mercados Completos Un mercado completo es un ideal teórico[37] en el que existe un mercado para cada bien y servicio en cada estado futuro de la naturaleza. Esto significa que los agentes pueden realizar contratos hoy que especifican un resultado para cualquier evento futuro imaginable. Por ejemplo, se puede comprar un seguro contra cualquier tipo de riesgo, desde la pérdida de un trabajo hasta la caída de un cultivo por una sequía. En un mercado completo, no hay sorpresas, ya que cualquier riesgo puede ser cubierto con un instrumento financiero. Bajo ciertas condiciones, los mercados completos conducen a una asignación de recursos eficiente en el sentido de Pareto: es la versión del Primer Teorema Fundamental del Bienestar para una economía con tiempo e incertidumbre, en la que los bienes se definen como contingentes al estado de la naturaleza. El Segundo Teorema también presupone mercados completos, pero añade la convexidad de las preferencias y de los conjuntos de producción y la redistribución de las dotaciones iniciales mediante transferencias de suma fija.

[37] No existen en la práctica, pero el concepto se utiliza para ilustrar cómo el mercado podría funcionar de manera eficiente si no hubiera fallos:

- Los "valores de Arrow-Debreu" (Arrow, 1964; Debreu, 1959): Este es un concepto teórico fundamental en economía. Un valor de Arrow-Debreu es un instrumento financiero hipotético que paga una unidad de riqueza si un evento futuro específico ocurre y cero en cualquier otro caso. En un mercado completo, existiría un valor de este tipo para cada evento concebible (por ejemplo, "llueve mañana", "el precio del trigo sube", etc.). La existencia de estos valores permitiría a los agentes económicos asegurarse contra todos los riesgos posibles y alcanzar una asignación de recursos óptima.
- Un mercado de futuros ideal: En un mercado de futuros real, los agricultores pueden vender su cosecha a un precio acordado hoy para una entrega futura, lo que les permite asegurar el precio. En un mercado completo, este tipo de contrato existiría para todos los bienes y servicios, en todas las fechas futuras y para todas las posibles condiciones de mercado.

Mercados Incompletos En el mundo real, los mercados son casi siempre incompletos[38]. La incompletitud surge cuando no existen mercados para ciertos bienes, servicios o riesgos futuros. Esto ocurre por varias razones:

1. Costes de transacción: Crear un mercado para cada evento futuro es prohibitivamente costoso.
2. Asimetrías de información: Es muy difícil asegurar riesgos si una de las partes tiene más información que la otra (por ejemplo, el comprador de un seguro de vida sabe más sobre su salud que la compañía de seguros).
3. Riesgos no asegurables: Algunos eventos, como una crisis financiera global, son tan amplios que no pueden ser asegurados en un mercado privado.

En un mercado incompleto, los agentes no pueden cubrir todos los riesgos futuros. Esto puede llevar a ineficiencias, ya que las personas no pueden diversificar completamente sus carteras y pueden tomar decisiones que no son óptimas. Por ejemplo, un agricultor que no puede asegurarse contra una sequía puede decidir no invertir en nuevas tecnologías por miedo a la quiebra.

La distancia entre ambos casos –el ideal de mercados completos y la incompletitud que domina la práctica– es, en definitiva, la que justifica buena parte de la intervención pública en riesgo: redes de seguridad social frente al desempleo o los desastres naturales, y regulación de los mercados financieros allí donde los costes de transacción o la información asimétrica impiden que el mercado los complete por sí mismo.

14.A.7. Otros enfoques para valorar la eficiencia económica

En la literatura económica han surgido varios enfoques para valorar la eficiencia económica. Por ejemplo

[38] Los mercados incompletos son la norma en la vida real. Aquí se dan algunos ejemplos tangibles de por qué son incompletos y qué implicaciones tiene.

- Mercado de seguros de desempleo: No puedes comprar un seguro privado para el riesgo de perder tu trabajo. Las aseguradoras no ofrecen este tipo de pólizas debido a la asimetría de la información (el solicitante sabe más sobre su probabilidad de ser despedido que la aseguradora) y al riesgo moral (la persona asegurada podría tener menos incentivos para buscar trabajo). Como resultado, este mercado es incompleto y el gobierno interviene a través de programas de seguridad social.
- Mercados de seguros de desastres naturales a gran escala: Es muy difícil o imposible para las personas asegurarse completamente contra riesgos como un gran terremoto, una inundación masiva o los efectos del cambio climático. El mercado privado no puede asumir estos riesgos sistémicos tan grandes. Como consecuencia, el mercado es incompleto y los gobiernos a menudo tienen que intervenir con programas de ayuda o seguros respaldados por el estado.
- Servicios de salud en regiones rurales: En áreas de baja densidad de población, es posible que el mercado privado no ofrezca ciertos servicios médicos (un especialista, por ejemplo) porque no sería rentable. Los costes de establecer una clínica son altos y la demanda es baja. Este mercado es incompleto, lo que justifica la intervención del gobierno para asegurar el acceso a la atención médica en estas áreas.

 © 2026 Manuel Cabado García

- Los mercantilistas asociaban el bienestar a la acumulación de metales preciosos,
- Adam Smith (1776) implícitamente relacionó el bienestar social con el volumen total de producción.
- Un precedente más cercano es Jeremy Bentham (1789), que asociaba el bienestar con utilidad, calculó el bienestar mediante la agregación de funciones de utilidad cardinales.
- No obstante, uno de los precursores más relevantes fue Alfred Marshall (1890), quien propuso una concepción del bienestar susceptible de ser medido en términos monetarios a través del excedente del consumidor, lo que permite hacer comparaciones interpersonales.
- Sin embargo, una de las figuras más destacables en el campo de la economía del bienestar es la de Vilfredo Pareto (1848-1923), quien utilizó el razonamiento de la Teoría del Equilibrio General para examinar las implicaciones de distintas medidas económicas para el bienestar (tratando de llevar la economía walrasiana a la política económica). Pareto abordó cómo evaluar la eficiencia de la asignación de recursos en una economía o en una determinada estructura de mercado. Rechazó la concepción cardinalista de la utilidad (negando que el bienestar sea mesurable y que por tanto se puedan hacer comparaciones interpersonales) y propuso una nueva teoría del bienestar basada en el individualismo y la unanimidad.

Por tanto, el criterio más aceptado ha sido el aportado por el economista italiano Vilfredo Pareto (1848-1923), según el cual un estado es eficiente si ningún agente puede mejorar sin que empeore otro. Otros enfoques:

- Excedente del consumidor de Marshall (valora el bienestar económico en términos monetarios siendo posible realizar comparaciones interpersonales)
- Pigou (1920): El dividendo nacional (producción) tiene efectos sobre el bienestar económico en función de su volumen y distribución: El bienestar social es una función con dos argumentos fundamentales: la eficiencia y la equidad, por lo tanto se puede elevar el bienestar social produciendo más o logrando una distribución más equitativa de la renta (Pigou (1877-1959) es posterior a Pareto). Pigou comparó grupos sociales pero no individuos.

14.A.8. Origen histórico del instrumental analítico: la caja, la lente y el núcleo

El instrumental gráfico empleado a lo largo de este tema –la caja, la curva de contrato, la lente y el núcleo– no es una invención simultánea, sino la acumulación de casi un siglo de desarrollo teórico, en el que cada pieza fue aportada por un autor distinto.

Fue Edgeworth (1881) quien, en su obra *Mathematical Psychics*, introdujo por primera vez la superposición de curvas de indiferencia de dos agentes y el concepto embrionario de curva de contrato, aunque todavía no con la forma de "caja" cerrada que hoy resulta familiar. Fue Pareto (1906), en su *Manuale di economia politica*, quien reelaboró la representación de Edgeworth hasta darle la forma de caja con dos orígenes enfrentados que se ha empleado en este tema. La difusión de esta herramienta entre los economistas de habla inglesa se debe a Bowley (1924), en *The Mathematical Groundwork of Economics* –de ahí que en la literatura anglosajona se hable a menudo de la "caja de Edgeworth-Bowley"–.

El concepto de núcleo, por su parte, es bastante más tardío y procede de un ámbito distinto: la teoría de juegos cooperativos. Lo introdujo Gillies (1959) y fue Shubik (1959) quien lo relacionó con la curva de contrato de Edgeworth. Debreu y Scarf (1963), en su artículo "A Limit Theorem on the Core of an Economy", demostraron con generalidad el célebre teorema del límite del núcleo, ya intuido por Edgeworth: a medida que una economía se replica –aumentando el número de consumidores de cada tipo manteniendo las proporciones–, el núcleo se va reduciendo progresivamente hasta converger exactamente con el conjunto de asignaciones de equilibrio competitivo. Este resultado proporciona un fundamento de teoría de juegos –independiente del mecanismo de precios– a la intuición de que los mercados competitivos descentralizados alcanzan asignaciones eficientes y estables, sin necesidad de suponer de antemano la existencia de un sistema de precios walrasiano[39].

Así, la relación caja $\supset$ curva de contrato, caja $\supset$ lente, núcleo $=$ lente $\cap$ curva de contrato, que se ha aplicado en este tema a la caja de consumo y, por analogía, a la de producción, no es un mero recurso gráfico, sino la traza visible de dos programas de investigación distintos –el análisis marginalista de Edgeworth y Pareto, y la teoría de juegos cooperativos de Scarf y Debreu– que confluyen en un mismo instrumento.

14.A.9. El debate del cálculo económico en la era de la inteligencia artificial

La naturaleza tecnocrática del punto de felicidad –alcanzable, en principio, sin necesidad de que exista mercado alguno, con solo conocer las formas funcionales de las preferencias y la tecnología– no es una curiosidad aislada de este tema: es la puerta de entrada a uno de los debates más antiguos y persistentes de la ciencia económica, el del cálculo económico bajo el socialismo. Si el planificador pudiera conocer directamente las funciones de utilidad y de producción de toda la economía, ¿para qué se necesitaría el mercado?

Esta es precisamente la pregunta que Hayek (1945) respondió negativamente en uno de los artículos más influyentes del siglo XX. Su argumento no es que el problema sea de mero volumen de cálculo, sino de naturaleza del conocimiento: la información relevante

[39] La intuición es que, cuantos más compradores y vendedores hay, más difícil es que alguien imponga un reparto ventajoso, porque siempre aparece otro grupo dispuesto a ofrecer mejores condiciones, hasta que solo sobreviven los repartos a precios de competencia.

 © 2026 Manuel Cabado García

para la asignación de recursos –las circunstancias particulares de tiempo y lugar, las preferencias subjetivas y cambiantes de millones de agentes– está dispersa, es en buena parte tácita (no articulable en forma de datos) y se genera y destruye continuamente en el propio proceso de intercambio. El sistema de precios, sostiene Hayek, no es solo un mecanismo de asignación, sino sobre todo un mecanismo de transmisión y agregación de esa información dispersa, que ningún organismo central, por potente que sea, podría replicar recopilándola de antemano.

Antes de ese artículo, Oskar Lange (1936) había dado a la crítica de Mises una respuesta influyente, el llamado socialismo de mercado, a la que el propio Hayek replicaba. Según Lange, un planificador central podría, en principio, resolver el problema fijando precios sombra para los bienes y factores, y ajustándolos por tanteo –a la manera del *tâtonnement* walrasiano– hasta que la oferta y la demanda se igualaran en cada mercado, sin necesidad de propiedad privada de los medios de producción. Aunque el modelo de Lange fue durante décadas la respuesta canónica al reto de Hayek y Mises, nunca llegó a implementarse con éxito en ninguna economía real, y el colapso de las economías de planificación centralizada del siglo XX se suele interpretar como una confirmación empírica del argumento hayekiano.

El desarrollo del *big data* y del aprendizaje automático ha revivido esta discusión con una pregunta nueva: si la limitación de Lange era la capacidad de cómputo y de recopilación de información, ¿no habría quedado superada por la inteligencia artificial contemporánea?

Fernández-Villaverde (2020) responde que no. Su argumento es que la ventaja de las reglas simples –la primera posesión, el intercambio voluntario en mercados libres– no reside únicamente en que ahorren información al planificador, sino en que generan los incentivos correctos de descubrimiento e innovación, algo que ningún algoritmo, por bien entrenado que esté sobre datos pasados, puede sustituir sin más. En sus propias palabras, el "socialismo digital" resultaría tan quimérico como el "socialismo analógico" de la planificación soviética: las reglas simples seguirán siendo el fundamento necesario para el diseño de instituciones y del entorno legal en el que operan los mercados, aunque la inteligencia artificial pueda ayudar en aspectos concretos de la política económica.

En la misma línea, pero con un análisis académico más extenso, Boettke y Candela (2023) examinan lo que denominan "tecnosocialismo" –la idea de que el *big data* y el aprendizaje automático resolverían el problema del cálculo económico– y concluyen que se trata, en sus palabras, de servir "vino viejo en una botella nueva irrelevante". Su argumento central es que el problema hayekiano no es de volumen de datos, sino de naturaleza dinámica: los precios de mercado no se limitan a agregar información ya existente, sino que son el resultado de un proceso continuo de descubrimiento empresarial que crea información nueva –sobre qué producir, con qué tecnología, a qué coste– que sencillamente no existe hasta que el propio proceso competitivo la genera. Un algoritmo, por muy sofis-

ticado que sea, solo puede procesar los datos que ya existen. No puede sustituir el proceso de mercado que los genera en primer lugar.

No todos los economistas comparten este escepticismo con la misma rotundidad. Acemoglu (2023) –galardonado con el Premio Nobel de Economía en 2024, precisamente por sus contribuciones al estudio de las instituciones– adopta una postura más matizada: reconoce que Hayek acertó en su crítica histórica a la planificación central, pero sugiere que el auge de la inteligencia artificial podría haber alterado la naturaleza del propio problema informativo que aquella crítica identificaba, sin llegar a concluir que ello haga viable un "comunismo asistido por IA". Su principal preocupación no es tanto la viabilidad técnica del cálculo, sino los problemas de incentivos y de elección pública –quién decide los objetivos del algoritmo y con qué controles– que cualquier forma de planificación, humana o algorítmica, sigue teniendo que resolver.

Este debate ilumina, en último término, una limitación importante del propio Segundo Teorema del Bienestar desarrollado en este tema: el teorema garantiza que cualquier asignación eficiente es alcanzable mediante una redistribución adecuada seguida de la actuación del mercado, pero da por supuesto que el planificador ya conoce qué redistribución es la correcta. La controversia sobre la inteligencia artificial y el cálculo económico no es sino la versión contemporánea de la pregunta que ese supuesto deja sin responder: de dónde sale, en la práctica, ese conocimiento.

 © 2026 Manuel Cabado García

Bibliografía

The Paretian System. https://www.hetwebsite.net/het/essays/paretian/paretocont.htm.

Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Princeton, NJ: Princeton University Press.

Acemoglu, D. (2023). Would AI-Enabled Communism Work? Project Syndicate, 28 de junio de 2023.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Arrow, K. J. (1964). The Role of Securities in the Optimal Allocation of Risk-bearing. *The Review of Economic Studies 31*(2), 91–96. Traducción al inglés de «Le rôle des valeurs boursières pour la répartition la meilleure des risques», Économétrie, 1953.

Baumol, W. J. (1952). *Welfare Economics and the Theory of the State*. Cambridge, MA: Harvard University Press.

Baumol, W. J., J. C. Panzar y R. D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne and Son.

Bergson, A. (1938). A Reformulation of Certain Aspects of Welfare Economics. *The Quarterly Journal of Economics 52*(2), 310–334.

Boadway, R. y N. Bruce (1984). *Welfare Economics*. Oxford: Basil Blackwell.

Boettke, P. J. y R. A. Candela (2023). On the Feasibility of Technosocialism. *Journal of Economic Behavior & Organization 205*, 44–54.

Bowley, A. L. (1924). *The Mathematical Groundwork of Economics: An Introductory Treatise*. Oxford: Oxford University Press.

Brecher, R. A. y C. F. Díaz-Alejandro (1977). Tariffs, Foreign Capital and Immiserizing Growth. *Journal of International Economics 7*(4), 317–322.

Corlett, W. J. y D. C. Hague (1953). Complementarity and the Excess Burden of Taxation. *The Review of Economic Studies 21*(1), 21–30.

Cramer, J. y A. B. Krueger (2016). Disruptive Change in the Taxi Business: The Case of Uber. *American Economic Review 106*(5), 177–182.

Davis, O. A. y A. B. Whinston (1965). Welfare Economics and the Theory of Second Best. *The Review of Economic Studies 32*(1), 1–14.

Debreu, G. (1959). *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*. núm. 17 de Cowles Foundation Monograph. New York: John Wiley and Sons.

Debreu, G. y H. Scarf (1963). A Limit Theorem on the Core of an Economy. *International Economic Review 4*(3), 235–246.

Edgeworth, F. Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London: C. Kegan Paul and Co.

Fernández-Villaverde, J. (2020). Simple Rules for a Complex World with Artificial Intelligence. PIER Working Paper 20-010, University of Pennsylvania, Penn Institute for Economic Research (PIER).

Foley, D. (1967). Resource Allocation and the Public Sector. *Yale Economic Essays 7*(1), 45–98.

Gillies, D. B. (1959). Solutions to general non-zero-sum games. En A. W. Tucker y R. D. Luce (eds.), *Contributions to the Theory of Games IV*, pp. 47–85. Princeton: Princeton University Press.

Glaeser, E. L. y J. Gyourko (2003). The Impact of Building Restrictions on Housing Affordability. *Economic Policy Review, Federal Reserve Bank of New York 9*(2), 21–39.

Hart, O. D. (1975). On the Optimality of Equilibrium when the Market Structure is Incomplete. *Journal of Economic Theory 11*(3), 418–443.

Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review 35*(4), 519–530.

Hicks, J. R. (1939). The Foundations of Welfare Economics. *The Economic Journal 49*(196), 696–712.

Johansson, P. O. (1991). *An introduction to modern welfare economics*. Cambridge: Cambridge University Press.

Kaldor, N. (1939). Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *The Economic Journal 49*(195), 549–552.

Lange, O. (1936). On the Economic Theory of Socialism, Part One. *The Review of Economic Studies 4*(1), 53–71.

 © 2026 Manuel Cabado García

Lipsey, R. G. y K. Lancaster (1956). The General Theory of Second Best. *The Review of Economic Studies 24*(1), 11–32.

Little, I. M. D. (1950). *A Critique of Welfare Economics*. Oxford University Press.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Pareto, V. (1906). *Manuale di Economia Politica, con una Introduzione alla Scienza Sociale*. núm. 13 de Piccola Biblioteca Scientifica. Milano: Società Editrice Libraria.

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

Prest, A. R. y R. Turvey (1965). Cost-Benefit Analysis: A Survey. *The Economic Journal 75*(300), 683–735.

Ramsey, F. P. (1927). A Contribution to the Theory of Taxation. *The Economic Journal 37*(145), 47–61.

Samuelson, P. A. (1950). Evaluation of Real National Income. *Oxford Economic Papers 2*(1), 1–29.

Scitovszky, T. (1941). A Note on Welfare Propositions in Economics. *The Review of Economic Studies 9*(1), 77–88.

Shubik, M. (1959). Edgeworth market games. En A. W. Tucker y R. D. Luce (eds.), *Contributions to the Theory of Games IV*, pp. 267–278. Princeton: Princeton University Press.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Strahan and Cadell.

Stigler, G. J. (1946). The Economics of Minimum Wage Legislation. *The American Economic Review 36*(3), 358–365.

Torres, J. L. (2010). Introducción al Equilibrio General Dinámico Macroeconómico. Universidad de Málaga, Departamento de Teoría e Historia Económica.

 © 2026 Manuel Cabado García

Tema 15

Economía del bienestar (II). Los fallos del mercado: externalidades y bienes públicos. Intervención y fallos del sector público.

15.1. Introducción ⏳ 1'

La Economía del bienestar es la rama de la teoría económica que intenta valorar, desde un punto de vista colectivo, las distintas situaciones en las que puede encontrarse un sistema económico. Estudia la deseabilidad social de los estados económicos alternativos a través de distintos criterios de valoración.

El criterio fundamental que ha surgido en el campo de la Economía del bienestar para valorar el bienestar social es el de eficiencia económica, entendida como mejor uso posible de los recursos limitados.

Pero, ¿qué sucede cuando el sistema de precios falla? ¿Cómo se determina el precio de algo que beneficia a toda la sociedad, como la defensa nacional, o de un servicio como la educación, que genera beneficios a largo plazo para todos, no solo para quienes estudian? Y, ¿qué pasa cuando una fábrica contamina un río, sin asumir el coste que genera para la comunidad?

Por consiguiente, el objetivo de esta exposición será estudiar las imperfecciones de mercado, con especial atención al caso de los bienes públicos y las externalidades y en sus posibles soluciones tanto públicas como privadas a través del siguiente esquema:

1. Concepto de eficiencia económica.
2. Optimalidad de la competencia perfecta.
3. Fallos de mercado: bienes públicos y externalidades.

4. Otros fallos de mercado.
5. Fallos del sector público.

15.2. Concepto de Eficiencia Económica ⏳ 3'

Comenzando con el concepto de eficiencia económica, la economía del bienestar considera que una asignación es eficiente en el sentido de Pareto cuando no es posible reorganizar los recursos de forma que algún individuo mejore su bienestar sin empeorar el bienestar de nadie. Además, para que una asignación sea eficiente en el sentido de Pareto tienen que cumplirse tres condiciones: óptimo de Pareto en producción, consumo y global.

Para su demostración, se van a utilizar los siguientes supuestos de partida:

1. Modelo 2x2x2: se van a utilizar dos factores productivos, capital, K, y trabajo, L, para producir dos bienes, X e Y, por medio de una función neoclásica de buen comportamiento, que van a consumir dos agentes, A y B.
2. Inexistencia de efectos externos y de bienes públicos.
3. Los consumidores y productores serán racionales.
4. Modelo cerrado, por lo que las rentas se agotan, es decir, se vacía el mercado.

Eficiencia en producción

Bajo estos supuestos, para conseguir la eficiencia en producción las condiciones de primer orden establecen que la relación marginal de sustitución técnica para la producción de X, $RMST^x_{K,L}$, debe coincidir con la relación marginal de sustitución técnica para la producción de Y, $RMST^y_{K,L}$, y a los precios relativos de los factores[1]:

$$RMST^x_{K,L} = RMST^y_{K,L} = \frac{w}{r} \tag{15.1}$$

es decir, deberá coincidir la tasa a la que se intercambian los factores productivos manteniendo constante el nivel de producción.

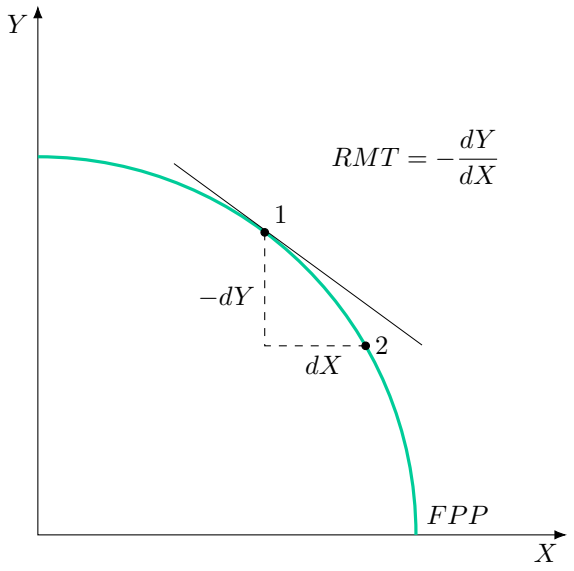
El resultado marcará las combinaciones eficientes en producción y que, gráficamente en el plano de dos bienes como el de la Figura 15.1, quedan recogidas en la FPP, cuya pendiente es la relación marginal de transformación, $RMT_{x,y}$ o el coste de oportunidad de sustituir un bien por otro en la producción y, a su vez, existen infinitos óptimos de Pareto en producción:

$$RMT = \frac{CMg_X}{CMg_Y} \tag{15.2}$$

[1] Como en competencia perfecta los precios de los factores son los mismos para todos los productores, la RMST será también la misma para todos los productores, que es la condición del óptimo en la producción (pues, dado que las valoraciones de los factores entre los productores son iguales, no será posible una reasignación de los factores que aumente la producción de un bien sin disminuir la del otro)

 © 2026 Manuel Cabado García

Figura 15.1: Frontera de posibilidades de producción



Eficiencia en consumo

Por otro lado, una asignación es eficiente en el consumo si la relación marginal de sustitución para el individuo A, $RMS^A_{x,y}$, coincide con la relación marginal de sustitución para el individuo B, $RMS^B_{x,y}$, y a su vez se igualan a los precios relativos de los bienes[2]:

$$RMS^A_{x,y} = RMS^B_{x,y} = \frac{P_X}{P_Y} \tag{15.3}$$

es decir, la valoración relativa que realizan los consumidores de los bienes X e Y en términos del otro bien debe ser la misma para ambos consumidores o, lo que es lo mismo, deben coincidir las tasas a las que los individuos pueden intercambiar un bien por otro manteniendo constante su utilidad.

Eficiencia global

Por último, para que una asignación sea un óptimo global la relación marginal de transformación debe coincidir con la relación marginal de sustitución de ambos individuos:

$$RMT_{x,y} = RMS^A_{x,y} = RMS^B_{x,y} \tag{15.4}$$

[2]Como en competencia perfecta los precios de los bienes son los mismos para todos los consumidores, la RMS será también la misma para todos los consumidores, que es precisamente la condición del óptimo en el consumo (pues, dado que las valoraciones de los bienes entre los consumidores son iguales, no será posible una redistribución de los bienes que aumente la utilidad de un consumidor sin disminuir la del otro)

es decir, debe existir eficiencia simultánea tanto en la producción como en el consumo, por lo que la valoración relativa de los bienes en el consumo coincidirá con la valoración relativa en la producción.

15.3. Optimalidad de la competencia perfecta ⏳ 1'

Una vez definido el criterio de eficiencia, la economía del bienestar analiza si existe alguna estructura del mercado capaz de generar eficiencia en los tres sentidos sin intervención pública. Bajo determinados supuestos estos resultados pueden obtenerse en un sistema competitivo de competencia perfecta, es decir, cuando:

1. Los mercados sean completos, es decir, todos los bienes se intercambian en el mercado a precios públicamente conocidos.
2. Información perfecta, no existan fallos de información.
3. Agentes precio-aceptantes.
4. Todas las acciones quedan determinadas por los precios.

Bajo estos supuestos, un equilibrio competitivo se define como un vector de precios y de asignaciones de bienes y factores tal que los agentes maximizan su utilidad –por el lado del consumo– y los beneficios –por el lado de la producción– y todos los mercados se vacían[3].

A partir de esta noción de equilibrio competitivo se formulan los dos teoremas fundamentales de la Economía del Bienestar:

- Primer Teorema de la Economía del Bienestar: en ausencia de fallos de mercado, todo equilibrio general competitivo es un óptimo de Pareto.
- Segundo Teorema de la Economía del Bienestar, cualquier asignación eficiente en sentido de Pareto puede alcanzarse como un equilibrio competitivo a través de la redistribución de las dotaciones iniciales.

Por lo tanto, el primer teorema sirve de defensa del libre mercado –limitando el papel del Estado a la corrección de imperfecciones–, mientras que el segundo teorema abre la puerta a la intervención pública ex ante para que, mediante la redistribución –y, posteriormente, la acción libre de los agentes en el mercado–, se alcance un determinado equilibrio deseado, separando así las dimensiones de eficiencia y equidad.

[3] Y es que, en un sistema perfectamente competitivo, cada individuo, al perseguir su propio interés personal, es guiado por una "mano invisible" hacia un curso de acción que garantiza una asignación eficiente global (esto es, que garantiza que el tamaño de la "tarta" sea máximo, pero no necesariamente que se maximice el bienestar social).

Sin embargo, la competencia perfecta es sólo una manera de alcanzar un óptimo de Pareto. Un sistema socialista de planificación central, por ejemplo, del tipo Lerner-Lange en el que el Gobierno ha estimado, de algún modo, los precios sombra que impone a los agentes, podría alcanzar, en teoría, los mismos resultados que un sistema perfectamente competitivo.

 © 2026 Manuel Cabado García

15.4. Fallos de mercado: bienes públicos y externalidades ⏳ 11'

No obstante, si alguno de esos supuestos no se cumple, aparece un fallo de mercado, es decir, una situación en la que el equilibrio competitivo deja de ser eficiente en el sentido de Pareto. Algunos de estos fallos son:

- Bienes públicos.
- Externalidades.
- Competencia imperfecta.
- Información imperfecta.

El análisis se centrará en los bienes públicos y en las externalidades[4].

Esto se debe a que en el caso de las externalidades y los bienes públicos, el objetivo de la intervención ya no es la solución competitiva cuando existen estos fallos de mercado, ya que no es la óptima, y por tanto el objetivo de la intervención deberá ser intentar establecer un equilibrio que sí lo sea.

Por ello, tanto en el caso de los bienes públicos como de las externalidades se estudian dos cuestiones básicas: i) Diferencia entre la asignación competitiva –es decir, con fallos de mercado– y la asignación óptima. ii) Mecanismos correctores.

Bienes públicos

Un bien público se diferencia de los bienes privados en que:

1. No existe rivalidad en su consumo, es decir, el hecho de que un agente aumente su consumo no implica una disminución de los beneficios que reporta el consumo del bien al resto de individuos, es decir, el coste marginal de un usuario adicional es nulo. A su vez, se puede distinguir entre

   a) No rivalidad estricta, propia de los bienes públicos puros.

   b) No rivalidad parcial, propia de los bienes públicos impuros, que empiezan a ser rivales cuando se congestionan.

2. No son excluibles, es decir, no es posible, o resulta demasiado costoso, impedir su consumo a quien no paga por ellos.

[4] A lo largo de la presente exposición el análisis se centra únicamente en los dos primeros fallos de mercado. ¿Por qué esta distinción? Porque existe una clara diferencia entre los dos primeros fallos de mercado y los dos últimos: y es que en el caso de la información imperfecta y la competencia imperfecta (rendimientos crecientes), el objetivo de la intervención es restablecer el equilibrio general competitivo, que sigue siendo un óptimo de Pareto: lo deseable es estar en una situación competitiva, y se intenta alcanzarla a través de la intervención.

Cuadro 15.1: Tipos de bienes

	Exclusión + ←	**Exclusión** − →
Rivalidad[5] + ↑	Bienes privados[6]: mercancías	Recursos comunes o comunales: pastos, caladeros
Rivalidad[5] − ↓	Bienes de club: TV de pago, wifi, autopistas	Bien público puro: defensa nacional, alumbrado público, aire

De esta forma, ante bienes públicos puros el sistema de precios dejará de ser válido porque solo reflejará la utilidad del individuo que más valora el bien, dado que el resto actuará como free-rider. Por lo tanto, se genera un equilibrio general competitivo que no es eficiente en el sentido de Pareto.

Así, desaparece la vinculación entre precio y utilidad que constituye el punto fundamental de la optimalidad de la competencia perfecta[7].

Diferencia entre la asignación competitiva y la asignación óptima

A continuación, se estudiará analíticamente el caso de los bienes públicos en el marco del modelo de Samuelson (1954), en el que existen dos individuos, A y B, que consumen dos bienes, un bien público, G, y un bien privado, X. La condición de optimalidad será la solución al siguiente problema de maximización de la utilidad de A, U_A, que consume bien privado y bien público:

$$\text{máx} \quad U_A\left(X_A, G\right) \tag{15.5}$$

sujeto a una utilidad dada del individuo B, una cantidad dada del bien X y una cantidad de bien público, G, que coincidirá en el consumo tanto para el individuo A como el individuo B, ya que se asume que el bien público es no rival en el consumo, y la última

[5]Nótese que la no rivalidad de las autopistas y del aire es relativa, ya que, en realidad, sí se trata de bienes rivales: el espacio de autopista ocupado por un coche no puede ser ocupado por otro, y las moléculas de oxígeno del aire respiradas por un individuo no pueden ser respiradas por nadie más. A pesar de ello, la razón por la que no se consideran rivales es por su abundancia relativa, pero empezarán a ser rivales cuando se congestionen.

[6]También existen bienes privados que el sector público provee o fomenta porque considera que, si cada individuo decidiera por sí solo, se consumirían en una cantidad inferior a la socialmente deseable, como la sanidad o la educación. Son los llamados bienes preferentes.

[7]La clave está en la agregación de demandas individuales: en el caso de los bienes privados, la demanda de mercado es la suma horizontal de las demandas individuales. En los bienes públicos, en cambio, la agregación es vertical como consecuencia de la no rivalidad: todos consumen la misma cantidad, y lo que se suma es la disposición a pagar de cada individuo por cada unidad.

 © 2026 Manuel Cabado García

restricción que impone que la solución esté sobre la FPP

$$\bar{U}_B = U_B\left(X_B, G\right) \tag{15.6}$$
$$X = X_A + X_B \tag{15.7}$$
$$G = G_A = G_B \tag{15.8}$$
$$T\left(X, G\right) = 0 \tag{15.9}$$

Por lo tanto:

- La condición de asignación eficiente del problema determinará lo que se denomina como condición de Samuelson, es decir, la relación marginal de transformación del bien X y el bien público G, será igual a la suma de las relaciones marginales de sustitución de todos los individuos:

$$RMT_{X,G} = \sum_{i=1}^{n} RMS^i_{X,G} \tag{15.10}$$

  es decir, el coste de oportunidad de producir una unidad más del bien público debe igualarse a la suma de lo que esa unidad vale para cada individuo, todo ello en términos del bien privado.
- Esto difiere de la solución competitiva, que hubiera conducido a una igualación de las relaciones marginales de sustitución, no alcanzando pues la solución óptima[8].
- Así, la provisión de un bien público debe hacerse hasta que la suma de las valoraciones marginales individuales del bien público –es decir, la valoración marginal social– se iguale al coste marginal de dicho bien, todo en términos del bien privado. O dicho de otra manera: la provisión óptima es aquella en la que el coste de producir la última unidad del bien público es igual a la utilidad que esta unidad reporta a la colectividad.

Gráficamente, en un plano de consumo del bien privado, X, y del bien público, G, como el de la Figura 15.2[9]:

- Dada la curva de FPP y la curva de indiferencia de B, U_B que corresponde a un nivel fijo de utilidad de dicho individuo, $\bar{U}_B$, dada la cantidad consumida total del bien X, X_T, y del bien G, el individuo B consumirá la cantidad X^*_B y G^*.

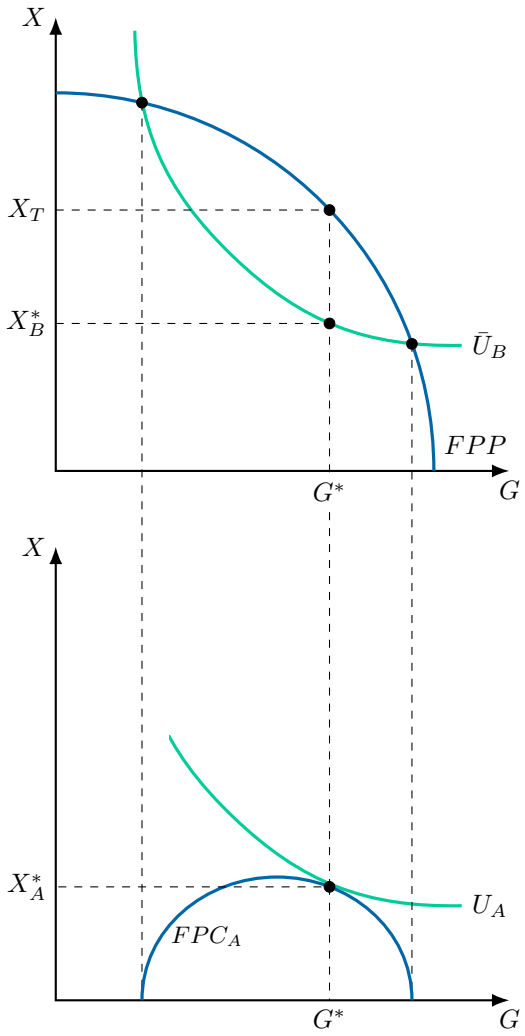
[8] En Competencia Perfecta $RMT = RMS_i$. La asignación competitiva no cumple la condición básica de optimalidad en presencia de bienes públicos. En principio, en un entorno de competencia perfecta, sólo demandaría el bien el consumidor que más lo valora, con lo que el bien público estaría siendo infraproducido, al estarse produciendo como un bien privado. Y es que por muy grande que sea la RMS del individuo que más valora el bien, siempre será inferior al sumatorio y siempre se estará infraproduciendo el bien público, como se verá.

[9] El gráfico reproduce el planteamiento de Gravelle y Rees (2004) (fig. 14.5) y de Stiglitz y Rosengard (2015) (fig. 5.10).

- A su vez, proyectando en un plano de consumo del bien X y del bien G, la diferencia entre la FPP y la utilidad del individuo B, se podrá determinar la frontera de posibilidades de consumo del individuo A, FPC_A. De esta forma, el individuo A consumirá la misma cantidad que el individuo B del bien público, G^*, ya que no existe rivalidad en el consumo y la cantidad de X que B no consuma, X_A^*.
  Nótese que la condición de eficiencia exige que A consuma allá donde su frontera de posibilidades de consumo sea tangente a la curva de indiferencia más alejada del origen, consumiendo X_A^* y G^*.

En el óptimo ambos consumidores consumen la misma cantidad del bien público, G^*, pero su valoración marginal, es decir, su RMS, es distinta. Con un bien privado ocurre lo contrario: cada individuo consume una cantidad diferente, pero, por la segunda ley de Gossen, iguala las utilidades marginales por euro de todos los bienes, es decir, su RMS al precio relativo, y como todos se enfrentan a los mismos precios, sus RMS coinciden.

Figura 15.2: Condición de Samuelson



 © 2026 Manuel Cabado García

Infraproducción Sin embargo, el mercado, por sí mismo, no proveerá una cantidad suficiente del bien público: la empresa no producirá hasta que el coste marginal iguale la utilidad marginal social, sino hasta que iguale las utilidades marginales reveladas, que serán menores que las reales.

Se da, pues, una ineficiencia que aleja al equilibrio del óptimo de Pareto, ya que existe una cantidad no producida para la cual la valoración de la sociedad es mayor que el coste de la producción.

Esta infraproducción del bien público se puede ver gráficamente a partir del análisis de Bowen (que no es más que el equivalente al modelo de Samuelson bajo un enfoque de equilibrio parcial):

- Gráficamente, en un plano de RMS y RMT, y G como el de la Figura 15.3[10] se representan
  - Las RMS de cada individuo que denotan la utilidad que se deriva de la producción del bien público, y que representan la disposición al pago para cada uno de los consumidores, es decir, la demanda (si el bien fuera privado), y
  - La oferta que vendrá dada por la RMT, que será el coste marginal de producir el bien público.
  - El mercado producirá allí donde la RMT coincida con la máxima RMS privada, es decir, la de A, produciendo G'.
- No obstante, la RMS colectiva, que se calcula a través de la suma vertical de las RMS, ya que el consumo de una unidad de G es disfrutada por todos[11], conduciría a un nivel superior de producción: $G^* > G'$.
- Así, el mercado, al no tener en cuenta el carácter no rival del bien público, infraproducirá G, dado que no es capaz de recoger el hecho de que el CMg de dar acceso al otro consumidor es 0. Además, si se incluye el carácter de no exclusión, ningún consumidor racional va a estar dispuesto a pagar por el bien al que puede tener acceso de forma gratuita, dando lugar al problema del free-rider.

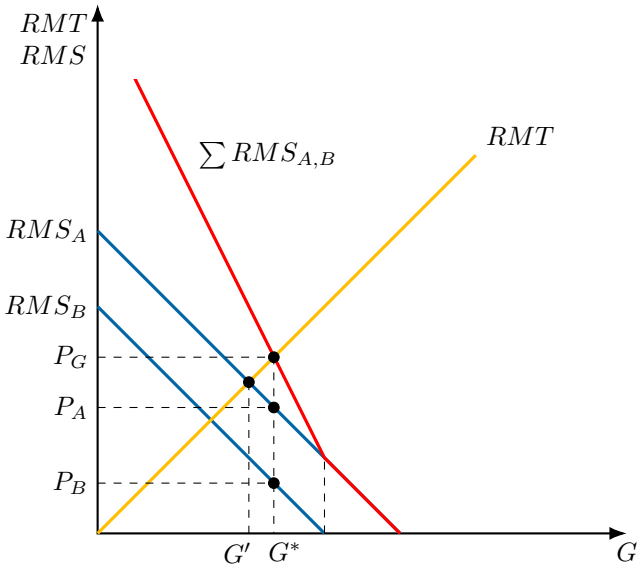
Mecanismos correctores

Por consiguiente, el Estado ocupa una posición idónea para resolver este problema mediante la provisión pública –que no quiere decir necesariamente producción pública–, ya que a priori, guía sus acciones en base al interés general y es el único agente capaz de recaudar fondos de manera coactiva. De esta forma, el sector público debe diseñar mecanismos correctores para corregir la infraprovisión.

[10] El gráfico sigue el planteamiento de Pindyck y Rubinfeld (2010) (cap. 18).

[11] Si fuesen bienes privados, la demanda total se obtendría por la suma horizontal de la disposición a pagar por cada unidad de bien privado.

Figura 15.3: Infraprovisión del bien público



Así, se enfrenta a dos cuestiones:

1. Cuánto bien público proveer, lo que requerirá recabar información entre los agentes económicos sobre sus preferencias por el bien público, para determinar el volumen de producción óptimo (aquel en el que el coste de producir una unidad adicional es igual a la utilidad que esta unidad reporta a la colectividad).
   El problema es que los agentes, como se ha dicho, pueden negarse a revelar sus verdaderas preferencias.
2. Cómo distribuir el coste de provisión entre los ciudadanos. En este sentido, el sector público podría:

   a) Atender a la capacidad de pago de cada individuo (vía, por ejemplo, impuestos progresivos). De esta manera, el sector público aprovecharía la necesidad de financiación del bien público para llevar a cabo, además, una redistribución.

   b) Atender al principio de beneficio (cada individuo paga en función de su valoración), esto es, detraer más renta a aquellos que valoren más el bien público. Esto recibe el nombre de "precios personalizados" o "precios de Lindahl", y consiste en sumar todas las utilidades marginales de todos los consumidores y producir hasta que dicha suma se iguale a los costes marginales. Después, cada individuo pagaría por unidad un precio igual a su valoración marginal del bien, de modo que todos consumen la misma cantidad, pero pagan precios distintos cuya suma cubre el coste marginal. Pero, de nuevo, el sector público dependería de la revelación de las preferencias.

 © 2026 Manuel Cabado García

Precios de Lindahl Así, uno de esos mecanismos es el método de los precios de Lindahl (1919)[12], según el cual se pagará el bien público en función de la valoración marginal por parte de los consumidores.

Siendo α la proporción del precio del bien público que paga A, el precio para A será:

$$P_A = \alpha \cdot P_G \tag{15.11}$$

por lo que A maximizará su utilidad cuando el ratio de sus precios se iguale con la relación marginal de sustitución para A:

$$\frac{\alpha \cdot P_G}{P_X} = RMS^A_{X,G} = \frac{UMg^A_G}{UMg^A_X} \tag{15.12}$$

Por su parte, el individuo B maximizará su utilidad igualando el ratio de precios a su relación marginal de sustitución:

$$\frac{(1-\alpha) \cdot P_G}{P_X} = RMS^B_{X,G} = \frac{UMg^B_G}{UMg^B_X} \tag{15.13}$$

A su vez, los productores maximizarán beneficios cuando el ratio de precios del mercado coincida con la relación marginal de transformación:

$$\frac{P_G}{P_X} = RMT_{X,G} \tag{15.14}$$

Dadas las condiciones de optimalidad, la condición de eficiencia afirma que la suma de las relaciones marginales de sustitución debe igualarse a la relación marginal de transformación, por lo que sumando los ratios de precios se cumpliría la relación de eficiencia de Samuelson:

$$RMS^A_{X,G} + RMS^B_{X,G} = \frac{P_G}{P_X} = RMT_{X,G} \tag{15.15}$$

es decir, la introducción de los precios de Lindahl sería una solución eficiente, tal y como puede observarse ahora en la Figura 15.3.

Sin embargo, para aplicar el mecanismo de precios de Lindahl el sector público dependería de la revelación de las preferencias de los individuos, los cuales no estarán dispuestos a revelarlas sin tener que pagar en función de las mismas, generando un comportamiento de free-rider. Por consiguiente, esta solución solo sería de carácter conceptual.

Mecanismo para la revelación de preferencias (Vickrey, Clarke y Groves) Una posible solución para conseguir la revelación de las preferencias verdaderas sería utilizar

[12]El método de precios de Lindahl atiende al principio de beneficio, esto es, detraer más renta a aquellos que valoren más el bien público. Por otro lado, se encuentra el principio de capacidad de pago, vía, por ejemplo, impuestos progresivos. De esta manera, el sector público aprovecharía la necesidad de financiación del bien público para llevar a cabo, además, una redistribución.

el mecanismo de Vickrey-Clarke-Groves (Vickrey, 1961; Clarke, 1971; Groves, 1973). Así, es eficiente suministrar un bien público si la suma de los valores es superior al coste o igual. Para solucionar el problema del free-rider se puede establecer que todos los agentes paguen lo mismo. Sin embargo dependiendo de si los agentes quieren que se provea el bien o no, pueden tender a exagerar o minimizar sus valoraciones. En todos los métodos vistos hasta ahora el problema es que no cuesta nada ocultar la verdad.

De esta forma:

- El mecanismo de VCG parte de que existen unos agentes bisagra que son aquellos que con su estimación hacen que se provea o no el bien público. El impuesto de Clarke se hace recaer sobre dichos agentes bisagra por el sumatorio de las valoraciones netas de los demás agentes (coste – valoración) (todas las valoraciones menos la suya). El problema de dicho impuesto es que incentiva a revelar las preferencias, pero la recaudación puede no ser suficiente para proveer el bien público. Parte además de que las decisiones de consumo no se modifican al introducir dicho impuesto.
- Problemas: El impuesto de Clarke, aunque tiene características positivas, plantea algunos problemas.
  - En primer lugar, solo funciona cuando las preferencias son cuasi-lineales, debido a que la cantidad que tiene que pagarse no debe influir en la cantidad demandada del bien[13]. Es importante que solo haya un nivel óptimo de bien público.
  - En segundo lugar, no genera un resultado eficiente de Pareto: el nivel del bien público es el óptimo, pero lo recaudado con el impuesto no puede devolverse a los participantes sin alterar sus incentivos, por lo que el consumo privado

[13]La votación por mayoría, incluso aunque dé lugar a un resultado bien definido, no incentiva necesariamente a la gente a que revele sinceramente sus verdaderas preferencias. Una manera de asegurarse de que la gente revela verdaderamente sus preferencias hacia un bien público es por medio de una especie de mercado o de "subasta". Desgraciadamente, este método requiere que las preferencias tengan una característica especial, a saber, que sean cuasilineales. El problema de estos métodos es que no cuesta nada ocultar la verdad. Sin un incentivo para declarar el verdadero valor del bien público de forma sincera, hay incentivos a subestimarlo o sobreestimarlo.

Una subasta Vickrey es un tipo de subasta de puja sellada, donde los oferentes presentan ofertas por escrito sin conocer la oferta de las otras personas en la subasta, y en la que gana el postor más alto, pero el precio que paga éste es la segunda oferta más alta. Este tipo de subasta fue descrito por primera vez por el profesor William Vickrey en la Universidad de Columbia en 1961, aunque había sido utilizada por coleccionistas de sellos desde 1893. Este tipo de subasta es estratégicamente similar a una subasta inglesa e incentiva a que los oferentes presenten ofertas iguales a su verdadera valoración del objeto subastado (ya que no pagan el precio que pujan, pero si infravaloran sus preferencias pierden oportunidades de ganar la subasta).

El mecanismo Vickrey-Clarke-Groves parte de que el bien pujado (es decir, provisión de un bien público) se asigna a la puja más alta, y el pago será el daño marginal causado al resto de individuos debido a su puja (a sobre cerrado determinada por las preferencias de cada agente).

 © 2026 Manuel Cabado García

queda por debajo del eficiente[14][15].

- Finalmente el último problema se refiere a la disyuntiva entre equidad y eficiencia inherente al impuesto de Clarke.

Nuevas líneas de investigación

- Aplicación de la teoría de juegos[16]. Se plantea un juego con dilema del prisionero en el que la estrategia de los individuos versa sobre la contribución a realizar para la provisión del bien público, y donde los resultados vienen dados por el volumen de bien público que se puede proveer a partir de las contribuciones ofrecidas por los agentes económicos. Si el juego se repite indefinidamente, la cooperación no llega a ser una estrategia dominante, porque si el otro jugador nunca coopera tampoco a uno le conviene hacerlo, pero la amenaza de castigo futuro permite sostenerla como equilibrio. Los trabajos de la politóloga Elinor Ostrom (Premio Nobel de Economía en 2009) documentaron que muchas comunidades logran en la práctica esa cooperación mediante reglas propias de vigilancia y sanción.
- Bienes públicos locales. Tiebout (1956) defendió que la solución a la infraprovisión

[14] Sin embargo, los impuestos solo tienen que pagarse si una persona actúa de bisagra. Si la decisión afecta a muchas personas, la probabilidad de que una de ellas haga de bisagra probablemente sea pequeña, por lo que cabe esperar que la recaudación también lo sea.

[15] Ejemplo práctico del mecanismo VCG: supóngase que tres vecinos deben decidir si instalar alumbrado público con un coste de 1.000€. Sus valoraciones son: Vecino A (600€), B (350€) y C (200€).

- La suma total de valoraciones (1.150€) supera el coste, por lo que el proyecto es viable.
- El mecanismo VCG identifica a los agentes bisagra, aquellos cuya participación es crucial para que el proyecto se realice. En este caso, todos son bisagra, ya que si alguno no participara, la suma de las valoraciones restantes no alcanzaría los 1.000€.
- A cada agente bisagra se le cobra un impuesto de Clarke, igual al beneficio que los demás pierden por su no participación. Por ejemplo, si el vecino A no estuviera, la suma de las valoraciones de los demás (550€) es inferior al coste. El "daño" a los demás es de 450€ (1.000€-550€), por lo que A pagaría 450€. El mecanismo incentiva la honestidad, pero no siempre es suficiente para cubrir los costes totales.

[16] Teoría del diseño de mecanismos: Este enfoque se debe a los autores Hurwicz, Maskin y Myerson, galardonados con el Premio Nobel de Economía en 2007. La teoría parte de que aunque todos los agentes se comporten siguiendo sus intereses personales y utilicen la información privada de forma estratégica, se consiga que las interacciones entre ellos conduzcan a un resultado adecuado para el conjunto de la sociedad. Es decir, se trata de diseñar un mecanismo que, cuando los individuos interactúan a través de él, tienen los incentivos para elegir las acciones en función de sus intereses propios que llevan a un resultado socialmente deseable. Es decir, se incorporan las restricciones de incentivos a las tradicionales restricciones de recursos.

Hurwicz introdujo un concepto fundamental, el de "compatibilidad de incentivos": un mecanismo es compatible con los incentivos cuando a cada individuo le resulta óptimo revelar honestamente su información privada, de modo que decir la verdad es un equilibrio. Esta propiedad no garantiza por sí sola la eficiencia en sentido de Pareto, como muestra el resultado de imposibilidad que se recoge a continuación.

Myerson (1979) formalizó el principio de revelación: cualquier resultado alcanzable con un mecanismo arbitrario puede replicarse con un mecanismo directo en el que a cada agente le resulte óptimo revelar honestamente su información privada. Por su parte, Hurwicz (1973) mostró un resultado de imposibilidad: no existe ningún mecanismo simultáneamente eficiente, compatible con los incentivos y descentralizado informacionalmente –lo que explica, por ejemplo, por qué el mecanismo de Clarke visto antes no logra el equilibrio presupuestario.

de bienes públicos pasaba por la creación de un cuasi-mercado de éstos, de forma que los individuos revelarían sus preferencias por los bienes públicos desplazándose a aquellas zonas donde las políticas públicas se aproximasen más a sus preferencias (el conocido como "voto con los pies")

- Mecanismos de decisión del sector público. Numerosos estudios demuestran que cuanto mayor es la participación del grupo objetivo en la elección del bien público a proveer, mayor predisposición tienen los individuos a contribuir. A modo de ejemplo, en 2010 economistas del MIT llevaron a cabo un experimento en 49 aldeas indonesias a las que se les daba a elegir un bien público (pozo, alcantarillado, alumbrado, etc.). Algunas aldeas tomaron la decisión a través de comités elegidos por sus habitantes, y otras, por referéndum directo. Los habitantes de las aldeas que decidieron vía plebiscito directo tenían mayor predisposición a contribuir a la financiación, construcción y protección del bien.
- Economía de la felicidad. La conocida como economía de la felicidad ha despertado un gran interés en los últimos años. La economía de la felicidad se focaliza en determinar los factores de los que depende la felicidad de los individuos. Recientemente, se ha utilizado para valorar bienes públicos. La idea es llevar a cabo encuestas en las que se pregunta a los individuos su nivel de felicidad (por ejemplo, del 1 al 10), su nivel de renta, y cualquier otra variable observable. En paralelo se obtienen datos del nivel de bienes públicos de los que disfruta cada agente (por ejemplo, gasto acumulado en infraestructuras en su barrio). Entonces se estima la felicidad de los individuos como una función de su renta, del nivel de bien público y del resto de variables. Manteniendo el nivel de felicidad constante, se puede determinar el nivel de renta que los individuos estarían dispuestos a renunciar (por ejemplo, vía impuestos) a cambio de una unidad adicional de bien público.

La literatura sobre los bienes públicos de las últimas décadas tiende a relativizar el concepto de bien público, considerando que el bien público puro hasta ahora analizado es una rareza, y que en realidad suelen presentarse bienes con una mayor o menor concurrencia de externalidades.

De esta forma, los problemas relacionados con su provisión óptima son los propios de los efectos externos, en los que el mercado puede tener un papel importante.

Externalidades

Las externalidades abarcan aquellas situaciones en las que la producción o el consumo de un bien incide (positiva o negativamente) en la utilidad o beneficio de otro agente económico, sin que esa incidencia tenga reflejo en los precios ni en los costes.

De esta forma, de nuevo los precios dejan de ser señales válidas, pues solo reflejarán la utilidad que el bien reporta a quien lo consume y los costes que genera a quien lo produce,

 © 2026 Manuel Cabado García

por lo que generan un equilibrio general competitivo que no es Pareto eficiente.

Hay por tanto dos elementos característicos de las externalidades:

1. No se reflejan directamente en el sistema de precios, lo cual las diferencia de las economías pecuniarias externas y
2. No son evitables por quien las sufre ni buscadas por quien se beneficia de ellas.

Al igual que para los bienes públicos, para las externalidades se estudian la diferencia entre la asignación competitiva y la asignación óptima, y los posibles mecanismos correctores.

Diferencia entre la asignación competitiva y la asignación óptima

Analíticamente, para el caso de una externalidad en el consumo de X que realiza A, la condición de optimalidad vendría de la resolución del siguiente problema de maximización de la utilidad de A, que deriva del consumo de X y de Y:

$$\text{máx} \quad U_A\left(X_A, Y_A\right) \tag{15.16}$$

sujeto a

$$\bar{U}_B = U_B\left(X_B, Y_B, X_A\right) \tag{15.17}$$

$$\bar{X} = X_A + X_B \tag{15.18}$$

$$\bar{Y} = Y_A + Y_B \tag{15.19}$$

Las condiciones de primer orden del problema establecen que

$$RMS_B = RMS_A + \frac{\partial U_B/\partial X_A}{\partial U_B/\partial Y_B} \tag{15.20}$$

Así:

- La solución competitiva/de libre mercado hubiera conducido a una igualación de las relaciones marginales de sustitución, no alcanzando, pues, la solución óptima.
- Mientras que las RMS denotarían la valoración relativa que los consumidores A y B realizan de su propio consumo de los bienes X e Y, el factor adicional estaría denotando la valoración, en términos del bien Y, que B realiza del consumo de X que A lleva a cabo.
- Este término podrá ser positivo (externalidad positiva) o negativo (externalidad negativa).

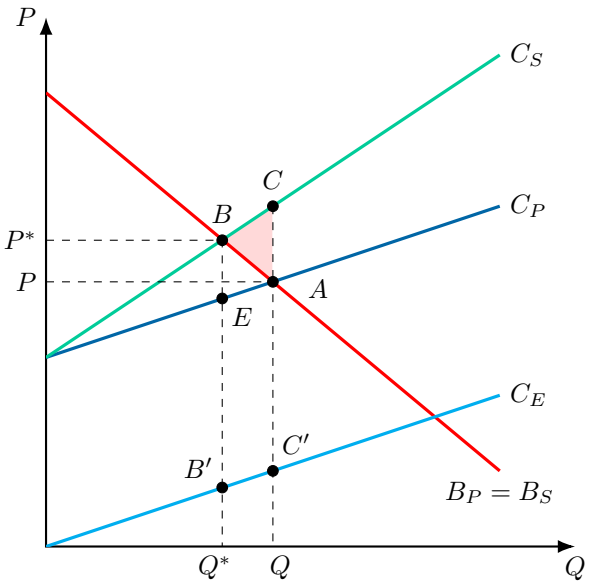
La causa radica en que se generaría una ineficiencia de la solución de mercado porque el precio no recogería la externalidad. Por consiguiente, los precios de mercado, al no incluir las valoraciones marginales de los perjudicados o beneficiados por un efecto externo,

dejarían de ser señales que orientarían eficientemente las decisiones de producción y consumo y los agentes emisores sobreproducirán las actividades que generan externalidades negativas e infraproducirán las que generan externalidades positivas.

Gráficamente, en un plano de precios y cantidades de un bien X como el de la Figura 15.4, se va a analizar una externalidad negativa en la producción:

- Un bien producido por una empresa en cuyo proceso se genera una externalidad negativa, por ejemplo, en el proceso productivo emite humos que contaminan el aire y perjudican a los residentes en una ciudad cercana a dicha empresa.
- Se representarán las curvas de oferta que reflejan el coste marginal privado, C_P, la curva de demanda, $B_P = B_S$ –ya que en este caso el beneficio privado y el social coinciden–, que representa el beneficio marginal privado que obtienen los consumidores de cada unidad del bien X.
- La curva C_E que representa el coste marginal externo o daño marginal que causa cada unidad producida del bien X. Su pendiente positiva refleja el supuesto de que una cantidad creciente de contaminación implica perjuicios crecientes.
- La curva C_S, que representará la suma de C_P más C_E y representa el coste marginal social de producir una unidad adicional de X.

Figura 15.4: Externalidad negativa en la producción



De acuerdo al gráfico:

- El equilibrio competitivo se alcanzaría en el punto A, donde el coste marginal privado de la producción del bien se iguala al beneficio marginal privado. De este modo, la producción de equilibrio sería Q y el precio P.

 © 2026 Manuel Cabado García

- Sin embargo, este nivel de producción no es eficiente, ya que no tiene en cuenta el efecto de la externalidad. El nivel eficiente corresponde al punto B, punto de intersección entre la curva C_S y la función de demanda, $B_P = B_S$.
- El punto E, situado sobre C_P en ese mismo nivel de producción Q^*, no es un equilibrio: es un punto auxiliar que se usará más adelante para calcular el segmento BE, el importe del impuesto corrector.
- Por lo tanto el nivel eficiente de producción sería Q^* a un precio P^* que incluye el coste marginal privado más la valoración del daño que sufren los perjudicados, evaluados ambos costes marginales en el nivel de producción Q^*.
- Esto lleva a una producción excesiva del bien, a la que se asocia un coste de eficiencia igual al área ABC[17].

Medidas correctoras

La intervención pública puede corregir este fallo de mercado, tomando como base los propios mecanismos de mercado, es decir, que en lugar de asumir la provisión del bien (única forma posible de intervención en el caso de bienes públicos), el sector público puede permitir que las fuerzas del mercado jueguen libremente, pero incidiendo previamente en el precio para que éstos internalicen los efectos externos, y que reflejen así todos los costes y beneficios asociados a la producción y al consumo.

Por lo tanto, existen tres tipos de medidas correctoras:

1. Legislación, asignar derechos de propiedad y permitir la aplicación del Teorema de Coase (Coase, 1960), que afirma que, si los derechos de propiedad están bien definidos, los costes de negociación son nulos y la información es perfecta, las partes afectadas negociarán hasta internalizar el efecto externo[18]. De esta manera se alcanzará un óptimo de Pareto.
   Además, por extraño que parezca, en ausencia de efecto renta, la asignación de recursos será siempre la misma con independencia de a quién se asignen los derechos de propiedad, pues esto afectará a la equidad, pero no a la eficiencia.

[17] De esta forma:

- Al reducirse la producción de Q a Q^* , los consumidores y los productores pierden conjuntamente un excedente igual a ABE.
- Por su parte, los residentes en la ciudad ganan $B'C'QQ^*$ – que es la suma de los daños marginales que causaban las unidades que se dejan de producir–.
- Como por construcción esta área es igual a $BCAE$, la ganancia neta para la economía de reducir el efecto externo a su nivel óptimo será el área ABC – que es igual al área BCAE menos el área ABE–.

Así, desaparecería la ineficiencia

[18] La intuición es la de dos vecinos que discuten por el ruido de una terraza: si pueden hablar sin coste, acabarán pactando un horario, y quién pague a quién dependerá solo de quién tenga el derecho al silencio o a la música.

2. Regulación, si el sector público conociese el efecto externo podría establecer una normativa que prohibiese la producción por encima de los niveles de la asignación eficiente, es decir, mediante una cuota.
   Sin embargo, esto conlleva una serie de inconvenientes ya que:

   a) Requiere conocer la curva C_E, que no es observable.

   b) Puede que existan muchas empresas generadoras de externalidad, por lo que se plantea el problema de cómo distribuir las cuotas.

   c) Cuando se trata de una externalidad positiva, el problema sería a la inversa, es decir, debería promoverse o incentivarse la producción de dicho bien.

3. Financiación, mediante instrumentos de tipo pigouviano –denominados así en honor a Arthur Pigou (Pigou, 1920)–, es decir, impuestos –para reducir la externalidad negativa– y subvenciones –para estimular las externalidades positivas, como en el caso de subvención de la educación o del uso de los transportes públicos[19–20].
   Gráficamente, como se observa en la Figura 15.5, en ambos casos se desplazaría la curva de oferta hasta que Q coincidiese con Q^*, que se convertiría en $C_P + \tau$, donde $\tau = BE$, es decir, el daño marginal correspondiente al nivel de producción socialmente eficiente. Así, en un contexto de información perfecta, la imposición y la regulación conducen al mismo nivel de producción. La diferencia entre ambas formas de intervención radica en la distribución de las ganancias[21].

[19] En realidad, los subsidios también pueden emplearse como medio para corregir efectos externos negativos en la producción o en el consumo. Bastaría con subsidiar las unidades que dejasen de producirse/consumirse, lo que incentivaría a la empresa/consumidor a reducir la cantidad intercambiada en el mercado.

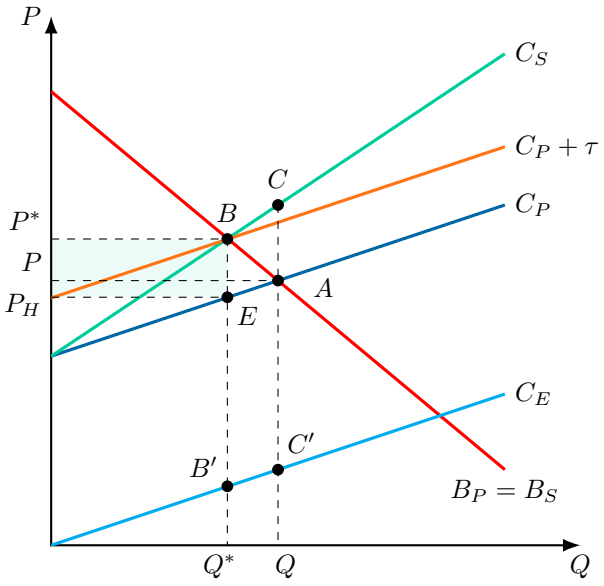
[20] Ejemplos de:

- Impuestos Pigouvianos Los impuestos pigouvianos son tributos que gravan una actividad que genera una externalidad negativa, buscando que el precio de mercado refleje el coste social total. En España y la UE, los más comunes son los impuestos ambientales.
  - En la UE: el Régimen de Comercio de Derechos de Emisión (RCDE UE) no es un impuesto, sino un mercado de derechos, pero también pone precio a las emisiones de carbono.
  - En España: El Impuesto sobre Gases Fluorados de Efecto Invernadero, la tasa por vertidos o el Impuesto sobre la Producción de Combustible Nuclear son ejemplos de impuestos pigouvianos.
- Subvenciones Pigouvianas Las subvenciones pigouvianas son ayudas económicas que el gobierno otorga para fomentar actividades que generan una externalidad positiva.
  - En la UE: Se destinan fondos a la investigación y desarrollo a través de programas como Horizonte Europa. Además, las subvenciones a las energías renovables, como las que recibe la energía solar o eólica, son un claro ejemplo.
  - En España: Las ayudas para la instalación de paneles solares, la financiación de la educación superior y las subvenciones a empresas para la contratación de trabajadores con discapacidad son ejemplos de subvenciones pigouvianas.

[21] Si, por ejemplo, en la Figura 15.4 el precio regulado se establece al nivel P^*, tanto los perjudicados como los consumidores de X estarán igual con la regulación y con los impuestos.
Los productores, sin embargo, preferirán la regulación porque con la fiscalidad tendrán que pagar impuestos por valor de τQ^* euros –área sombreada BEP_HP^*–, vendiendo a un precio P_H, normalmente inferior al que percibirían con la regulación.
La aceptación del principio de "quien contamina paga", combinada con el hecho de que con la re-

 © 2026 Manuel Cabado García

Figura 15.5: Corrección de la externalidad: Impuestos pigouvianos



Sin embargo, este mecanismo tiene inconvenientes similares a los de la regulación:

a) Establecer un impuesto óptimo requiere conocer la externalidad que normalmente no es observable y surgiendo así el problema de revelación de preferencias[22].

b) Los impuestos óptimos serían diferentes para cada empresa.

c) No se garantiza que el efecto externo se internalice si los objetivos de las empresas son otros que no sean la maximización de beneficios[23].

d) Las subvenciones correctoras deben financiarse con impuestos que sí son distorsionantes, y un impuesto que no se fije exactamente en el daño marginal introduce su propia distorsión, por lo que en la práctica ambos instrumentos tienden a resultados de second best[24].

caudación obtenida por el impuesto correctivo el Sector Público puede reducir otros impuestos que sí distorsionan, como los que gravan el trabajo –lo que Pearce (1991) denominó "doble dividendo": mejora ambiental y sistema fiscal más eficiente–, ha llevado a muchos economistas a considerar a la fiscalidad como un instrumento corrector de las externalidades negativas preferible a la regulación.

[22]Por ejemplo, a la hora de conocer el coste marginal que tiene la emisión de Gases de Efecto Invernadero (GEI), el informe Stern (2006) apuntaba a un coste de 300 dólares por tonelada de GEI mientras que el informe Nordhaus señalaba que dicho coste era de 30 dólares por tonelada de GEI. La discrepancia surge de la incertidumbre sobre la magnitud de los efectos, el uso de una diferente tasa de descuento a la hora de actualizar los flujos futuros, etc. En definitiva, si el sector público se equivoca en sus cálculos y fija un impuesto demasiado alto, el bienestar final podría incluso empeorar con respecto a la situación de partida.

[23]El empresario podría, por ejemplo, tratar de maximizar su participación en el mercado sujetándose a una rentabilidad mínima, o responder a objetivos políticos, sindicales o regionales, como con frecuencia ocurre en el caso de las empresas públicas.

[24]Un instrumento tipo Pigou no necesariamente tiene que elevar la carga impositiva de la sociedad en su conjunto, si su recaudación se emplea para reducir otros impuestos, o para financiar otras políticas o

4. Otras soluciones pasarían por crear unidades de decisión más grandes que englobasen a generadores y a receptores con el objetivo de que negocien una solución eficiente. Esto es debido a que los efectos externos dan lugar a asignaciones ineficientes porque sus generadores, al adoptar sus decisiones, no toman en cuenta el bienestar de los receptores [25]. Además, tratándose de efectos externos negativos, además de las soluciones analizadas, se podría recurrir a la subasta de los derechos de propiedad[26].

15.5. Otros fallos de mercado ⏳ 1'

A continuación se presentan otros posibles fallos de mercado.

Monopolio natural Un monopolio natural surge cuando una única empresa puede abastecer todo el mercado a un coste menor que si lo hicieran varias. Esto ocurre cuando los costes son subaditivos, es decir, cuando ningún reparto de la producción entre varias empresas resulta más barato que concentrarla en una. Las economías de escala, con un coste medio decreciente, son una condición suficiente, aunque no necesaria. Es típico en infraestructuras como redes eléctricas o ferroviarias, donde duplicar instalaciones sería ineficiente. Aunque esta empresa puede ofrecer precios más bajos inicialmente, su posición dominante puede llevarla a reducir la producción y aumentar los precios, generando ineficiencia. Para evitarlo, el Estado puede intervenir regulando tarifas, ofreciendo subsidios para cubrir pérdidas o incluso gestionando directamente el servicio, como ocurre con algunas empresas públicas de transporte.

Información asimétrica La información asimétrica aparece cuando una de las partes en una transacción conoce más que la otra, lo que distorsiona el funcionamiento del

gastos socialmente deseables.

[25] Supóngase, por ejemplo, que la construcción de una carretera en un municipio, al facilitar el tráfico, favorece a los residentes en otro municipio. Para adoptar una decisión óptima sería deseable crear un nivel de gobierno que tuviese jurisdicción en ambos municipios y al que se adjudicara la competencia decisoria en materia de carreteras.

[26] Por ejemplo, en el caso de la contaminación descrito en la Figura 15.4, se emitirían derechos de contaminación por una cantidad igual al nivel de producción eficiente Q^*, y se sacarían a la venta entre empresas competitivas. El precio de venta de estos derechos vendría determinado por el valor marginal de la última unidad de contaminación permitida, y éste equivale al daño marginal en el óptimo. No podrá ser inferior porque, de serlo, los perjudicados por la contaminación pujarían por el derecho para evitar que se ejercitara. Tampoco será superior, porque los productores no estarán dispuestos a pagar por ese derecho más que la diferencia entre el precio máximo que en el mercado puede alcanzar la última unidad producida y su coste de producción.

En el óptimo de contaminación, esta cantidad es precisamente igual a BE. El resultado de una subasta de derechos de contaminación es equivalente a un impuesto pigouviano, ya que el precio del derecho de contaminación será exactamente el mismo que el tipo impositivo del impuesto correctivo óptimo. Al igual que en el caso del impuesto, el Sector Público se queda con el importe de lo obtenido. No obstante, también se enfrenta a los mismos problemas que la regulación a la hora de determinar el nivel óptimo de contaminación.

 © 2026 Manuel Cabado García

mercado. Este problema se manifiesta de dos formas:

1. Selección adversa (ex ante): ocurre antes de cerrar el trato. Akerlof (1970) lo ilustró con el mercado de coches usados: el vendedor conoce defectos que el comprador no puede observar, lo que lleva a los compradores a desconfiar y reducir su disposición a pagar, afectando a todo el mercado.
2. Riesgo moral (ex post): aparece después de la transacción, cuando una de las partes, cuyas acciones la otra no puede observar, cambia su comportamiento porque ya no soporta todo el riesgo de sus decisiones —un concepto del mundo de los seguros que Arrow (1963) incorporó al análisis económico al estudiar la asistencia sanitaria—. Un ejemplo sería un empleado que, una vez contratado, reduce su esfuerzo al no estar supervisado, perjudicando a la empresa. Para mitigar estos efectos, se utilizan mecanismos como certificaciones, garantías, auditorías o incentivos que promuevan la revelación de información y alineen los intereses de ambas partes.

Información incompleta A diferencia de la asimétrica, la información incompleta implica que todos los agentes tienen el mismo nivel de conocimiento, pero este es insuficiente para tomar decisiones óptimas. Esto genera incertidumbre sobre escenarios futuros, como ocurre en mercados financieros ante eventos imprevisibles. Una solución teórica son los mercados contingentes, en los que hoy se compra un bien que solo se entrega si se produce un determinado estado de la naturaleza. Según el resultado de Arrow y Debreu, si existiera uno para cada estado posible, la asignación sería eficiente. Sin embargo, su implementación suele ser costosa y compleja, lo que limita su uso en la práctica. Por ejemplo, los seguros ante catástrofes naturales intentan cubrir este tipo de incertidumbre, pero requieren estimaciones difíciles y primas elevadas.

15.6. Fallos del sector público ⏳ 2'30"

Frente a estos fallos de mercado, cabría pensar que la intervención pública restaura siempre la eficiencia. Sin embargo, la propia intervención tiene sus propios costes: son los fallos del sector público, estudiados por la Escuela de Public Choice (Buchanan y Tullock, 1962), que aplica las herramientas del análisis económico a la toma de decisiones políticas partiendo de que también los votantes, los políticos y los burócratas actúan guiados por su interés propio.[27]

Los principales son:

[27] El Apéndice 15.A.1 desarrolla estos fallos con más detalle, junto con sus posibles soluciones –ligar la retribución a la productividad, mejorar la supervisión, o separar el diseño de la ejecución de las políticas–. En definitiva, los fallos de mercado justifican la intervención, pero los fallos del sector público obligan a hacerla con cautela.

1. Distorsiones en los precios: toda intervención pública altera el sistema de incentivos previamente existente, de manera que las ganancias de eficiencia obtenidas al corregir un fallo de mercado concreto pueden verse compensadas, total o parcialmente, por nuevas pérdidas de eficiencia generadas en otras partes de la economía.
2. Búsqueda de rentas –rent-seeking–, concepto de Tullock (1967) al que Krueger (1974) dio nombre: los grupos de presión dedican recursos a conseguir del sector público ventajas en su favor, como protecciones o exenciones, en lugar de dedicarlos a actividades productivas (Olson, 1965). Cuanto más pequeño y concentrado es el grupo, más fácil le resulta organizarse frente a colectivos grandes y difusos como los consumidores, que además tienen pocos incentivos para informarse dado el escaso peso de su voto individual –la llamada ignorancia racional del votante–. Ejemplos típicos son los sindicatos o las patronales, frente a colectivos como los consumidores en general, que raramente logran organizarse pese a ser, en conjunto, los más perjudicados.
3. Captura del regulador: según Stigler (1971), los grupos regulados tienen mucho más interés e información sobre la regulación que el público general, lo que les permite orientarla a su favor. La regulación pensada originalmente para proteger el interés general acaba, con el paso del tiempo, sirviendo sobre todo a los intereses del sector regulado. Cuando esto se combina con la búsqueda de rentas, da lugar a lo que Adams (1981) denominó el "triángulo de hierro" entre políticos, burócratas y grupos de interés.
4. Burocracia de Niskanen (1971): dado que su remuneración suele ser en gran parte fija, el burócrata maximiza su utilidad a través del tamaño del presupuesto que gestiona, ya que de ahí procede el poder y el prestigio que le otorga dirigir un departamento grande, compitiendo con otros departamentos por atraer fondos hacia sus programas de gasto.
5. Ciclo político de Nordhaus (1975): los gobiernos pueden aumentar el gasto público y reducir impuestos antes de unas elecciones, aunque ello no responda a criterios de eficiencia, aprovechando la miopía de unos votantes con información limitada. Esto se traduce en políticas expansivas en los meses previos a los comicios, seguidas de ajustes bruscos justo después de la cita electoral.
6. Inconsistencia temporal (Kydland y Prescott, 1977): un gobierno que anuncia una política óptima a largo plazo tiene luego incentivos para desviarse de ella una vez que los agentes privados ya han fijado sus expectativas conforme al anuncio, lo que explica por qué reglas fijas e instituciones independientes –como un banco central autónomo– mejoran el resultado frente a la discrecionalidad.
7. Ilusión fiscal, concepto de Puviani (1903) desarrollado por Buchanan y Wagner (1977): los votantes infravaloran la carga fiscal que soportan, lo que permite au-

 © 2026 Manuel Cabado García

mentar el tamaño del sector público sin que perciban un aumento equivalente de la presión fiscal –por ejemplo, mediante retenciones en origen o imposición indirecta–.

8. Sesgo deficitario de Alesina y Tabellini (1990): a diferencia de la ilusión fiscal, aquí el mecanismo no depende de que los votantes perciban mal la carga fiscal, sino de la alternancia en el poder. Un gobierno que prevé perder las elecciones frente a otro de distinta ideología tiene incentivos a endeudarse en exceso para condicionar el gasto de su sucesor, sesgo tanto mayor cuanto mayor es la polarización política.

15.7. Conclusiones ⏳30"

Como conclusión, la existencia de fallos de mercado, como la provisión insuficiente de bienes públicos o la presencia de externalidades, justifica la intervención estatal para mejorar la eficiencia. No obstante, esta intervención no está exenta de riesgos: la ineficiencia administrativa, la captura por grupos de interés o la falta de incentivos adecuados pueden comprometer sus resultados. Reconocer los límites del sector público es esencial para diseñar políticas más efectivas.

Además, los bienes públicos globales —como la estabilidad climática o la salud mundial— plantean desafíos que trascienden las capacidades de los gobiernos nacionales. Su provisión exige una gobernanza internacional sólida, basada en la cooperación y la coordinación entre países.

Por tanto, la intervención pública debe ser estratégica y prudente, evaluando tanto los fallos del mercado como los del propio Estado. Solo así se podrá avanzar hacia soluciones sostenibles que respondan a problemas complejos y transnacionales, sin caer en simplificaciones que ignoren la realidad institucional y política.

© 2026 Manuel Cabado García

Apéndice

15.A.1. Los fallos del sector público

Como se analizó anteriormente, ante fallos de mercado, la intervención pública en la economía podía mejorar la eficiencia. No obstante, las imperfecciones del mercado son una condición necesaria pero no suficiente para la intervención gubernamental en la economía. Al fin y al cabo, los costes de la intervención no son nulos, y a los fallos de mercado se contraponen los fallos del sector público.

Gran parte de este análisis procede de la Escuela de Public Choice, fundada por Buchanan y Tullock (1962), que aplica las herramientas del análisis económico a la toma de decisiones políticas, partiendo de que también los votantes, los políticos y los burócratas actúan guiados por su interés propio. Un ejemplo es la ignorancia racional del votante: dado que el coste de informarse adecuadamente sobre cada política es desproporcionado frente a la influencia real de un único voto en el resultado electoral, al votante le resulta racional permanecer poco informado.

A continuación, se presentan los principales fallos del sector público:

1. Distorsiones en los precios que pueden generar ineficiencias: toda intervención altera el sistema de incentivos previamente existente, por lo que las ganancias de eficiencia pueden verse disminuidas por las pérdidas de eficiencia generadas en otras partes de la economía.
2. Búsqueda de rentas –rent-seeking–: los grupos de presión se pueden definir como un conjunto de individuos con intereses comunes que se organizan para interferir en las decisiones de los políticos con el fin de que éstos favorezcan sus intereses privados[28][29]. Las actividades de búsqueda de rentas que llevan a cabo los grupos de presión generan ineficiencias por dos motivos:

    a) Se dedican recursos a obtener el favor de los poderes públicos, en lugar de dedicarlos a actividades productivas.

[28] La regulación en forma de política comercial proteccionista –por ejemplo, contingentes– o en forma de política fiscal –por ejemplo, exenciones–, posibilita la aparición de rentas de monopolio, incentivando a los agentes a dedicar recursos para captar esas rentas. De este modo, determinadas medidas regulatorias podrían estar camuflando intereses particulares, convirtiéndose la regulación en una injustificada barrera de entrada al mercado.

[29] Una teoría que estudia la dinámica de los grupos de presión es la Teoría de la Acción Colectiva de Mancur Olson (1965). Según esta teoría, como las actividades de los grupos de interés tienen características de bien público –pues si un grupo promueve dicho interés, se benefician todos los que tengan dicho interés, aunque no perteneciesen al grupo y no soportasen, por lo tanto, los costes en términos de tiempo, dinero y esfuerzo–, existen incentivos para actuar como free-rider. Esto explica por qué solo los grupos de interés pequeños suelen constituirse en lobbies, ya que los grupos de interés grandes –por ejemplo, consumidores– tienen más dificultades para organizarse. De ahí que, como señala Carlos Rodríguez Braun, la redistribución del sector público puede que no se realice de ricos a pobres, sino de grupos desorganizados a grupos organizados.

 © 2026 Manuel Cabado García

b) Se toman decisiones favorables a los intereses de los grupos de presión, que no suelen coincidir con los de la colectividad.
Ejemplos de grupos de presión: sindicatos, patronales, etc.

3. Captura del regulador: según Stigler (1971), se produce porque los grupos regulados tienen mucho más interés e información sobre la regulación que el público general, lo que les permite influir en el regulador hasta conseguir que la regulación acabe sirviendo a sus propios intereses en lugar de al interés general. A esto se suma que el simple contacto continuado entre representantes de una empresa o sector y agentes del sector público acaba generando inevitablemente relaciones personales que crean un sesgo inconsciente en la regulación a favor del colectivo.
Tanto la búsqueda de rentas como la captura del regulador pueden dar lugar a lo que Adams (1981) denominó "el triángulo de hierro", teoría según la cual las decisiones políticas están determinadas por las relaciones entre las tres partes del triángulo: políticos, burócratas, y grupos de interés.
4. Teoría de la burocracia de Niskanen (1971): para Niskanen, el burócrata es un individuo que, como todos los demás, trata de maximizar su utilidad. En el caso de los burócratas, dado que su remuneración suele ser en gran parte fija, su utilidad procederá principalmente del poder y del prestigio que tengan, lo que a su vez dependerá del tamaño del presupuesto que maneje su departamento. Así, los burócratas competirán con otros departamentos por atraer fondos hacia los programas de gasto que ellos gestionan.
5. Aversión al riesgo de Downs (1967). Este autor considera que los burócratas se caracterizan por su aversión al riesgo. La aversión al riesgo va a suponer un menor ritmo de innovación y, en definitiva, una mayor resistencia al cambio, lo que puede generar ineficiencias.
6. Teoría del ciclo político de Nordhaus (1975): puede ocurrir que las decisiones de gasto público no respondan a cuestiones de eficiencia, sino que se vean influidas por el ciclo electoral, de manera que la clase política aumente el gasto público y disminuya los impuestos en los períodos previos a unas elecciones[30].
7. Inconsistencia temporal de Kydland y Prescott (1977): un gobierno puede anunciar una política óptima a largo plazo –por ejemplo, comprometerse a mantener la inflación baja–, pero una vez que los agentes privados ya han fijado sus expectativas conforme a ese anuncio, al gobierno le interesa a posteriori desviarse de lo anunciado para obtener beneficios a corto plazo –por ejemplo, generar una sorpresa inflacionis-

[30]La falta de información hace que los votantes sean miopes, de manera que solo tengan en cuenta los resultados recientes y sus expectativas sean estáticas. En este contexto, el Gobierno tendría incentivos en llevar a cabo políticas económicas expansivas en período pre-electoral, y en ajustar bruscamente las variables justo después de las elecciones. Algunas soluciones serían alargar las legislaturas, limitación de mandatos, independencia del banco central, etc.

ta que reduzca el desempleo–. Como los agentes racionales anticipan este incentivo a incumplir, la política anunciada deja de ser creíble, lo que explica por qué reglas fijas y la independencia de instituciones como el banco central pueden mejorar el resultado frente a la discrecionalidad.

8. Teoría de la ilusión fiscal de Buchanan y Wagner (1977). Los votantes tendrán ilusión fiscal si infravaloran la carga fiscal que soportan. Esto permite aumentar el tamaño del sector público sin que los votantes perciban un aumento equivalente de la presión fiscal. Vías:

   a) Retenciones en origen.

   b) Imposición indirecta[31].

   c) Opacidad del sistema impositivo.

   d) Ilusión de deuda de Niskanen[32].

9. Sesgo deficitario de Alesina y Tabellini (1990): a diferencia de la ilusión fiscal, aquí el mecanismo no es que los votantes perciban mal la carga fiscal, sino que el gobierno que prevé ser sustituido por otro de distinta ideología usa la deuda para condicionar el gasto de su sucesor. El resultado es un sesgo estructural hacia el déficit, tanto mayor cuanto más polarizados estén los partidos y más probable sea la alternancia en el gobierno.

Posibles soluciones a los fallos del sector público:

1. Ligar la retribución a la productividad. Sin embargo, esto acarrea problemas: i) ¿cómo se mide la productividad?[33] Esto es particularmente difícil en el sector público, donde la producción suele valorarse al coste de los factores, y ii) ¿quién mide la productividad?
2. Mejorar los mecanismos de supervisión y control.
3. Individualizar responsabilidades para evitar que éstas se diluyan.
4. Separar el diseño de la ejecución, por ejemplo, encargar el análisis coste-beneficio a una agencia independiente, de manera que el sector público no pueda ya subestimar los costes y sobrestimar los beneficios de los proyectos públicos.
5. Presupuesto de base cero en lugar de presupuesto incremental.

[31] De hecho, los países escandinavos construyeron su estado del bienestar a partir de los años 60 aumentando la imposición indirecta, de manera que la carga fiscal percibida no se disparó.

[32] La teoría de la ilusión de la deuda de Niskanen argumenta que los votantes no internalizan igual el aumento de los impuestos que el aumento de la deuda, pues los agentes pueden percibir en cierta medida la deuda como riqueza neta. No obstante, esta teoría depende en gran medida del grado de racionalidad de los agentes, pues agentes perfectamente racionales aplicarán la equivalencia ricardiana.

[33] Por ejemplo, en el caso de un juez, ¿se valora su productividad por el número de sentencias o por la calidad de las mismas?

 © 2026 Manuel Cabado García

6. Sustituir la producción pública por privada en aquellos casos en los que sea posible, manteniendo la provisión pública, esto es, que siga siendo pagado con fondos públicos.
7. Limitar legalmente el tamaño del sector público, es decir, incorporar el concepto de "constitución fiscal".

15.A.2. Tragedia de los comunes y de los anticomunes

La tragedia de los comunes es la sobreexplotación de un recurso rival y de libre acceso, porque cada usuario no tiene en cuenta el coste que su uso impone a los demás. Ejemplos de ello son el pasto común que se arruina por el pastoreo excesivo o el lago común que se agota por la sobrepesca. Aunque la acción de cada individuo es un intento racional de promover su propio interés a corto plazo, las acciones colectivas del grupo no son beneficiosas para la comunidad a largo plazo.

Así, este dilema plantea una cuestión política desconcertante y muy debatida: ¿Cómo se gestiona un recurso que formalmente no pertenece a nadie? Durante mucho tiempo, la respuesta común fue la conversión de los recursos comunes en propiedad privada o la regulación externa de estos recursos comunes:

- Si un recurso se convierte en propiedad privada, el propietario debería tener tanto los incentivos de rentabilidad como la protección de los derechos de propiedad para gestionar todo el recurso de forma responsable.
- Si el gobierno regula el recurso, podría imponer normas a los usuarios individuales por el bien común.

Sin embargo, en 1999, la politóloga Elinor Ostrom, ganadora del Premio Nobel, revisó estas soluciones, ya que a veces dejaban los bienes comunes en peor situación (Ostrom et al., 1999). Por un lado, el control privado de recursos que antes eran públicos podía crear los problemas habituales derivados de las prácticas empresariales monopolísticas, mientras que por otro, la supervisión externa podía generar normativas inadecuadas y mal aplicadas.

Así, Ostrom señaló una tercera respuesta a la tragedia de los bienes comunes, que documentó empíricamente basándose en muchos casos de todo el mundo: la utilización del capital social de la comunidad para idear soluciones locales creativas y eficaces.

Anticomunes

Un nuevo enfoque de este debate es la tragedia de los anticomunes, término acuñado por Michael Heller (Heller, 1998), entonces en la Facultad de Derecho de Michigan[34]. En

[34]Heller se trasladaría más tarde a la Facultad de Derecho de Columbia, donde ejerce en la actualidad.

contraste con los problemas creados por la propiedad ambigua de un recurso comunitario, se refiere a lo contrario: la propiedad privada excesiva de un recurso comunitario que impide alcanzar un resultado deseable para la sociedad, llevando a una infrautilización del recurso, al contrario que en la tragedia de los comunes donde se daba un sobreuso.

La tragedia de los anticomunes se aplica ahora comúnmente a los derechos de propiedad excesivos en ámbitos como la investigación biomédica, por ejemplo, en la implementación de la vacunación de la Covid-19 –discusiones sobre la patente podrían haber reducido el proceso de producción, limitado el acceso a la tecnología y la competencia entre empresas, redundando en una infrautilización por exclusión–.

Otro ejemplo claro podría ser la crisis de la vivienda en muchas ciudades. Los procesos de urbanización y planificación otorgan a los grupos vecinales, ecologistas y otras partes interesadas un poder significativo a la hora de decidir si los nuevos proyectos de vivienda siguen adelante y en qué condiciones. Esto crea una situación en la que muchas partes interesadas tienen derecho a excluir a otras del uso de un recurso privado.

En muchos casos, los procesos de urbanización y planificación hacen imposible que los propietarios privados construyan nuevas viviendas, especialmente viviendas plurifamiliares. Esto se traduce en precios elevados, una oferta reducida de viviendas y un aumento del riesgo de exclusión, de pobreza y de personas sin hogar.

15.A.3. Teorema de Coase

Como se analizó anteriormente, el Teorema de Coase consiste en asignar derechos de propiedad sobre la externalidad y dejar que las partes afectadas negocien libremente su uso. De esta manera, se alcanzará un óptimo de Pareto. Además, por extraño que parezca, en ausencia de efecto renta, la asignación de recursos será siempre la misma con independencia de a quién se asignen los derechos de propiedad, pues esto afectará a la equidad, pero no a la eficiencia.

Para su análisis se utilizará la Figura 15.4, con el ejemplo ahora de una acería que, al producir acero aguas arriba, contamina las aguas del río, de forma que la piscifactoría que se encuentra aguas abajo se ve perjudicada –es decir, externalidad negativa en la producción–. A continuación se observará que a quién se asignen los derechos sobre el agua influirá en los esquemas de pagos –es decir, equidad–, pero no en la asignación óptima –que será siempre Q^*–.

Si los derechos de propiedad se conceden a la piscifactoría, la acería tendrá que compensar a la piscifactoría en la cuantía $B - E$, que es la externalidad asociada al nivel de producción Q^* . A la acería no le interesará producir más –ni menos– de la cantidad Q^*: a partir de Q^*, una unidad adicional de producción generará menos beneficio marginal $(B_P - C_P)$ que el pago a satisfacer por la externalidad que genera esa unidad adicional, C_E.

 © 2026 Manuel Cabado García

Si los derechos de propiedad se conceden a la acería, la piscifactoría tendrá que compensar a la acería en la cuantía del beneficio por unidad que ésta deja de producir ($B_P - C_P$). La piscifactoría buscará una forma de convencer a la fábrica de acero para que disminuya su volumen de producción desde Q hasta Q^* . Y es que a la piscifactoría no le interesará que la acería produzca menos –ni más– de la cantidad Q^*: para cantidades menores de Q^* la cuantía que la piscifactoría tiene que pagar a la acería por el beneficio marginal que ésta deja de percibir ($B_P - C_P$) es mayor que el daño que deja de recibir la piscifactoría, C_E. La disminución de la producción de acero desde Q hasta Q^* perjudica a la fábrica de acero, que experimentará una disminución de su beneficio representada por el área ABE. Por su parte, la citada disminución de la producción de acero supondrá una mejora para la piscifactoría porque la externalidad disminuirá en el área $BCAE$. Así pues, con la disminución de la producción que se acaba de considerar, la pérdida que sufriría la fábrica de acero sería menor que la ganancia que experimentaría la piscifactoría. Esto implica que es posible alcanzar algún acuerdo entre ambas en aras del beneficio mutuo. Ambas empresas mejorarían si suscribiesen un acuerdo en virtud del cual la piscifactoría compensase a la fábrica de acero con una cantidad mayor que ABE pero menor que $BCAE$.

Así, la ventaja del mecanismo de Coase es que el resultado sería similar al logrado con los instrumentos pigouvianos, con la ventaja de que el Gobierno no necesita conocer la valoración del efecto externo, ya que éste se manifestará a través de la oferta y la demanda.

En cuanto a las desventajas:

1. Supone cuasilinealidad en las preferencias, es decir, supone que no hay efecto renta, es decir, que la demanda del bien que genera la externalidad es independiente de la renta. Esto permite que la solución no dependa de a quién se asignan los derechos de propiedad. Gráficamente, en una caja de Edgeworth, las curvas de indiferencia son todas ellas traslaciones horizontales unas de otras, de forma que la tangencia se situará siempre en el mismo nivel, es decir, la curva de contrato será horizontal.
2. Supone ausencia de costes de transacción.
3. Dificultad de identificación de los agentes.
4. Altos costes asociados a la puesta en marcha de un mercado de este tipo.

Una aplicación de los instrumentos tipo Coase es el mercado Cap and Trade de la UE: European Union Emissions Trading System (EU ETS). El ETS es la "piedra angular de la política climática europea" y tiene como objetivo la internalización de los costes de contaminación a través de la limitación de la cantidad de emisiones producidas en la UE, siguiendo la lógica del Teorema de Coase. Cada año se asigna un número limitado de derechos equivalentes a emitir una tonelada de CO_2. A continuación, se permite que las empresas comercien con los derechos, permitiendo que aquellas más eficientes se beneficien de la venta de derechos y que las más contaminantes internalicen el coste de la contaminación que producen.

15.A.4. Bienes públicos globales

Los bienes públicos globales son bienes o servicios que benefician a todas las naciones, clases sociales y generaciones sin excepción, un concepto sistematizado por Kaul et al. (1999). Al igual que los bienes públicos nacionales, se caracterizan por la no exclusión y la no rivalidad, pero lo que los define como globales es la escala: sus beneficios son intrínsecos y se extienden a todo el planeta, sin importar las fronteras.

- No exclusión: una vez que se provee el bien, es imposible o extremadamente costoso impedir que cualquier persona, en cualquier país, lo consuma —el aire limpio en la atmósfera es un buen ejemplo—.
- No rivalidad: el consumo del bien por parte de un país o individuo no reduce la cantidad disponible para el consumo de otros, como ocurre con el conocimiento científico: que un país use la información sobre el genoma humano no limita el acceso de los demás.

Entre los ejemplos más citados en la literatura están la estabilidad climática —las acciones de un país para reducir sus emisiones de gases de efecto invernadero benefician a todos los demás, sin importar dónde se encuentren—, la erradicación de enfermedades contagiosas como la viruela mediante campañas de vacunación coordinadas internacionalmente, el conocimiento científico básico —que, una vez compartido, se convierte en un recurso accesible para toda la humanidad—, y el mantenimiento de la paz y la no proliferación de armas de destrucción masiva, que reduce el riesgo de conflictos a gran escala para el conjunto de países.

Estos bienes no se agregan todos de la misma manera. Siguiendo la clasificación de Hirshleifer (1983), extendida al ámbito global por Sandler (1998), cabe distinguir tres tecnologías de agregación. En la de suma, el nivel global del bien es la suma de las contribuciones de cada país, como ocurre con la reducción de emisiones. En la del eslabón más débil, el resultado depende del país que menos contribuye, como sucede en la erradicación de una enfermedad contagiosa, donde el riesgo de rebrote persiste para todos mientras quede un solo país sin vacunar. Y en la del mejor tiro, basta con que un único país lo consiga, como en un avance científico decisivo.

Esta distinción importa porque el problema del "free-rider", ya presente a nivel nacional, no es igual de severo en los tres casos. En los bienes de suma se agrava en el ámbito global: un país puede decidir no contribuir esperando beneficiarse de las contribuciones de otros —por ejemplo, evitando la inversión en tecnologías limpias mientras se beneficia de la reducción de la contaminación que sí implementan otros países—. En los bienes de eslabón más débil, sin embargo, puede incluso incentivar lo contrario: ayudar al país rezagado, ya que el resultado global depende precisamente de él.

Dado que ni los mercados ni los gobiernos nacionales pueden resolver por sí solos este problema, su provisión exige mecanismos de gobernanza global: acuerdos y trata-

 © 2026 Manuel Cabado García

dos internacionales —como el Protocolo de Kioto o el Acuerdo de París—, organismos multilaterales —la Organización Mundial de la Salud en la coordinación de la respuesta a pandemias, o el FMI y el Banco Mundial en la promoción de la estabilidad financiera global—, y mecanismos de financiación como los impuestos globales o las ayudas al desarrollo.

15.A.5. Externalidades gráficamente

Las externalidades en el consumo ocurren cuando el consumo de un bien por parte de un individuo afecta directamente el bienestar de terceros.

- Negativas: El consumo impone un coste a terceros. Por ejemplo, el humo de un cigarrillo que afecta a quienes están cerca o el ruido de un concierto en un apartamento que molesta a los vecinos.
- Positivas: El consumo beneficia a terceros. La vacunación es un ejemplo claro, ya que la inmunidad de una persona reduce el riesgo de contagio para toda la comunidad. Otro caso es la educación, que alarga la vida, hace a la gente más feliz y la convierte en un ciudadano más informado.

A continuación, se presenta su representación gráfica siendo las ineficiencias las áreas sombreadas y recordando que:

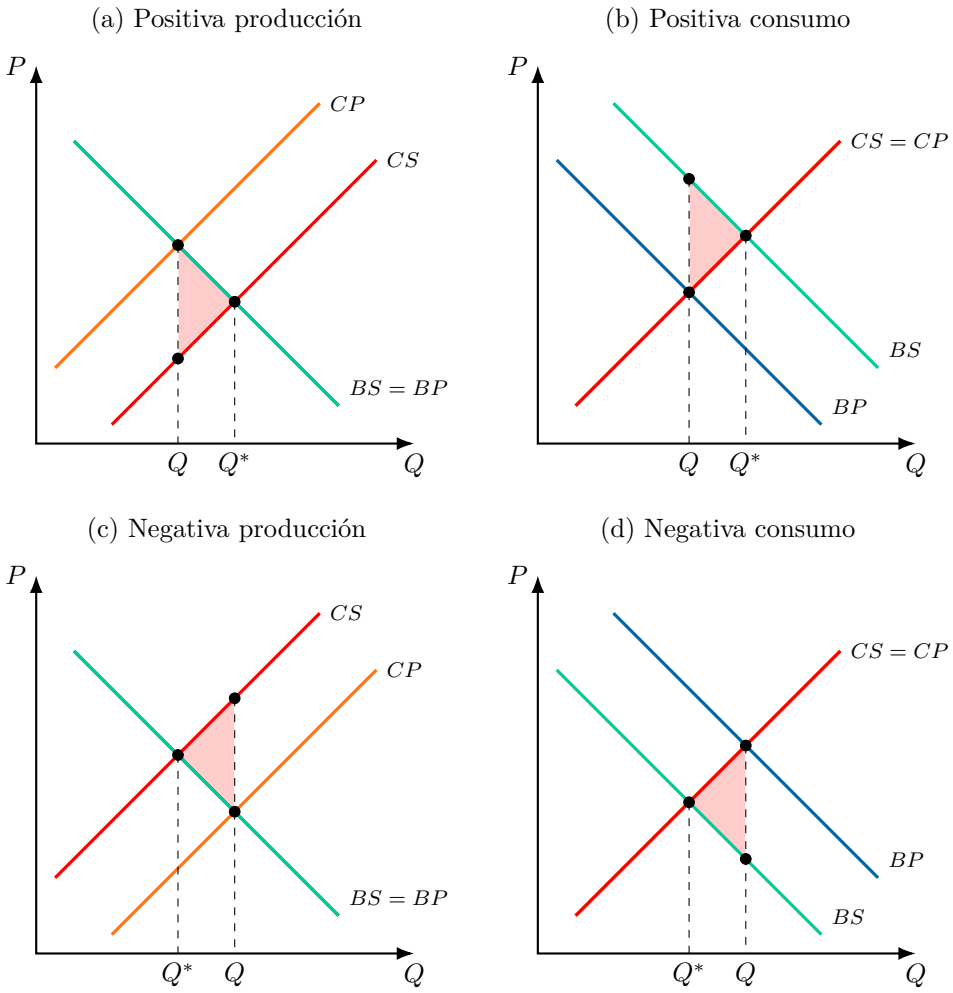
- Externalidad negativa en producción, $BPMg = BSMg$ pero $CPMg < CSMg$.
- Externalidad positiva en producción, $BPMg = BSMg$ pero $CPMg > CSMg$.
- Externalidad negativa en consumo, $CPMg = CSMg$ pero $BPMg > BSMg$.
- Externalidad positiva en consumo, $CPMg = CSMg$ pero $BPMg < BSMg$.

15.A.6. Provisión pública y externalidades

Si el sector público conoce las preferencias de los agentes con respecto al bien, así como los beneficios y/o costes marginales sociales del mismo, entonces podrá reservarse la producción del bien, y conocerá la cantidad que debe lanzar al mercado. Por ejemplo, si un bien genera una externalidad negativa, el sector público podría llegar a tomar la decisión de: prohibir su producción, si se valora que el coste marginal externo es lo suficientemente grande como para anular los beneficios de su producción (otra opción es controlar la cantidad producida, es decir, establecer una regulación de la cantidad producida).

Un ejemplo habitual es el del uso abusivo de los recursos pesqueros: cada pescador produce un efecto insignificante en las reservas marinas totales, pero su uso acumulado por todo el colectivo puede poner su supervivencia en peligro:

Figura 15.6: Tipos de externalidades



- En este caso, el sector público puede regular la cantidad de recursos que se pueden explotar: prohibiendo la entrada a nuevos pescadores en el sector, controlando el número de días que está permitido faenar, regulando el acceso de otras flotas a las aguas nacionales, etc.
- El valor actual de los beneficios del sector será superior bajo esta regulación, ya que se evita el uso abusivo al adoptar una visión intertemporal de la explotación óptima de los recursos.
- Sin embargo, esta regulación tendrá también efectos negativos, en términos de reducción del número de barcos existentes en el sector, descontento del colectivo, abandono de la actividad y necesidad de reestructuración, además de la ya mencionada necesidad de que el sector público disponga en realidad de la suficiente información como para saber cuál es el uso sostenible de los recursos pesqueros.

 © 2026 Manuel Cabado García

15.A.7. Elinor Ostrom y la Tragedia de los Comunes

Elinor Ostrom (politóloga estadounidense) obtuvo el Premio Nobel de 2009 investigando cómo las comunidades son o no capaces de gestionar recursos comunes (rivales, pero no excluibles) como tierras de cultivo o aguas de irrigación.

En una célebre publicación de Garrett Hardin en 1968, "The Tragedy of the Commons", este describía una situación en la cual unos campesinos compartían un terreno de pasto común, y decidían individualmente cuántas vacas poseer cada uno. El resultado era el característico del dilema del prisionero: aunque colectivamente lo que convenía era mantener un número reducido de vacas, individualmente cada uno tenía un incentivo a incrementar el suyo, puesto que esa era su estrategia dominante (la mejor respuesta tanto si el otro campesino respeta un acuerdo de pocas vacas como si no). Dicho de otro modo, el granjero que añadía una vaca adicional obtenía una ventaja sobre el resto de campesinos. Puesto que todos tenían incentivos a hacer lo mismo, como resultado se obtenía una sobreexplotación del terreno que perjudicaba a todos. Contrariamente a la lógica clásica, en esta circunstancia, cuando un agente actuaba guiado únicamente por su bienestar individual, ni se maximizaba el bienestar social, ni a largo plazo se maximizaba el individual.

Hasta los trabajos de Ostrom, las soluciones más populares establecían que la propiedad privada (es decir, hacer que el bien común sea excluible y por tanto se convierta en un bien privado como los demás) es necesaria para evitar el agotamiento de recursos limitados. Era por tanto una conclusión pesimista: no hay solución de mercado a los bienes comunes: o se privatizan, o se regula su uso por el Estado, con todas las ineficiencias que ello conlleva.

Por el contrario, Ostrom documentó numerosos casos alrededor del mundo en los que las comunidades diseñan mecanismos para "gobernar los comunes", asegurando su supervivencia durante generaciones, sin privatización y sin intervención directa del sector público. De hecho, sus trabajos constituyen una suerte de solución intermedia entre los defensores de la regulación pública y los del libre mercado: lo que realmente funciona son los acuerdos libres alcanzados por asociaciones voluntarias entre los usuarios directos de los recursos comunes, es decir, aquellos que se benefician directamente de su mantenimiento y saldrían perjudicados de su agotamiento.

Un ejemplo clásico se refiere a su investigación en un pueblo de Suiza donde los campesinos tienen tierras de cultivo privadas, pero comparten un prado común para alimentar a sus vacas. Aunque esto se parece enormemente al caso clásico de Tragedia de los comunes modelizado por Hardin, Ostrom descubrió que, en realidad, no se produce sobreconsumo. Esto se debe a que existe un acuerdo común entre los granjeros según el cual cada uno puede llevar al prado en verano tantas vacas como puedan cuidar durante el invierno,

una regla que lleva en pie desde 1517[35]. Ostrom documentó similares acuerdos en Kenia, Guatemala, Nepal, Turquía y Los Ángeles.

Basándose en sus investigaciones sobre el terreno, Ostrom ofreció ocho principios para que los comunes puedan ser gobernados de manera sostenible y justa en una comunidad.

1. Definir fronteras claras de quién pertenece a la comunidad
2. Adaptar las reglas a las necesidades y condiciones de la comunidad (no existen reglas aplicables a cada caso)
3. Asegurarse de que quienes fijan y modifican las reglas son los mismos afectados
4. Autonomía frente a autoridades externas
5. Monitorizar cumplimiento
6. Aplicación de sanciones graduales según gravedad del incumplimiento
7. Mecanismos rápidos y baratos de resolución de conflictos
8. Relaciones apropiadas con otros niveles de la administración

Estas normas y conclusiones fueron lo suficientemente poderosas para desafiar el conocimiento convencional y obtener el Premio Nobel de Economía, del que Ostrom fue la primera mujer galardonada. Una de sus contribuciones metodológicas fue la combinación de los últimos avances en teoría de juegos a modelos de ciencia política combinados con estudios empíricos.

Críticas a la solución de Ostrom: En comunidades muy grandes dichos acuerdos pueden no funcionar dados los costes de gobernanza.

15.A.8. Bienes públicos geográficos e intertemporales

Los bienes públicos geográficos e intertemporales son conceptos que extienden las características de los bienes públicos tradicionales —la no rivalidad y la no exclusión— a dimensiones espaciales y temporales. Estos bienes, a diferencia de los bienes públicos nacionales, son cruciales para entender los desafíos de la cooperación en un mundo globalizado.

Bienes Públicos Geográficos Los bienes públicos geográficos se refieren a servicios o recursos cuyos beneficios se limitan a un área geográfica específica. A menudo, se habla de ellos en términos de bienes públicos locales (BPL), que son provistos y consumidos por una comunidad delimitada por fronteras geográficas o administrativas. Es el mismo concepto que Tiebout (1956) formalizó como base de su modelo de "voto con los pies".

Un ejemplo claro es el alumbrado público en una calle de un barrio o un parque local.

[35] En la práctica, la necesidad de cuidar las vacas durante el invierno añade un coste a la decisión de aumentar el número de vacas, de manera que esta regla internaliza los costes que la vaca adicional genera a los demás en el pasto (Ostrom, 1990)

 © 2026 Manuel Cabado García

El beneficio de tener una calle bien iluminada es no rival y no excluyente, pero solo se aplica a quienes viven o transitan por esa zona. Si bien se parecen a los bienes públicos nacionales, los BPL resaltan cómo la provisión y el consumo están inherentemente ligados a la ubicación. La decisión de financiarlos suele recaer en los gobiernos locales, que recaudan impuestos a nivel de barrio o municipio para cubrir los costes, resolviendo así el problema del free-rider a una escala más pequeña.

Otro tipo de bien público geográfico, aunque más complejo, son los bienes públicos regionales, cuyos beneficios se extienden a múltiples países dentro de una región. Conviene no confundirlos con las externalidades transfronterizas: la gestión de una cuenca fluvial compartida, como la del río Danubio, es en realidad el segundo caso, porque el uso del agua es rival —lo que un país aguas arriba consume o contamina reduce la cantidad y calidad disponible aguas abajo— y en cierta medida excluible mediante presas o desvíos, de manera que no cumple la no rivalidad que exige un bien público en sentido estricto. Un ejemplo más ajustado de bien público regional es el control de enfermedades transmitidas por vectores que cruzan fronteras, como el Programa de Control de la Oncocercosis en África Occidental (1974-2002), que coordinó la fumigación en 11 países porque el mosquito vector migraba entre ellos: el resultado —eliminar el riesgo de ceguera del río para el conjunto de la región— es no rival y no excluible una vez logrado, y dependía, además, de que todos los países participasen —la misma tecnología de eslabón más débil vista en el apartado 15.A.4—.

Bienes Públicos Intertemporales Los bienes públicos intertemporales son aquellos cuyos beneficios se extienden a lo largo del tiempo, afectando no solo a la generación presente, sino también a las futuras. La característica clave aquí es la sostenibilidad y la preservación.

Un ejemplo fundamental es la protección de los recursos naturales o la estabilidad del clima. Las acciones de la generación actual para reducir las emisiones de carbono benefician a las generaciones futuras, que disfrutarán de un clima más estable y un medio ambiente más saludable. Sin embargo, esto plantea un desafío significativo: la generación actual puede tener un incentivo para sobreexplotar los recursos (como el petróleo) para obtener beneficios económicos a corto plazo, transfiriendo los costes medioambientales a las futuras generaciones, que no tienen voz ni voto en las decisiones actuales.

El problema del "free-rider" intertemporal ocurre cuando una generación se beneficia del consumo excesivo, mientras que las generaciones futuras soportan los costes. Esto justifica la intervención de los gobiernos para implementar políticas a largo plazo, como la inversión en energías renovables, la protección de la biodiversidad o la conservación de ecosistemas.

Conclusión La clasificación de los bienes públicos en categorías geográficas e intertemporales ayuda a comprender mejor la complejidad de su provisión. Mientras que los bienes públicos geográficos destacan la importancia de la cooperación a nivel local o regional para resolver problemas compartidos, los bienes públicos intertemporales obligan a considerar la equidad entre generaciones. Ambos conceptos demuestran que, sin una gobernanza adecuada y mecanismos de cooperación, tanto las comunidades locales como las generaciones futuras podrían sufrir las consecuencias del fallo del mercado.

15.A.9. Sistema de derechos de emisión

Los sistemas de derechos de emisión (SDE) son una herramienta económica diseñada para combatir la contaminación, principalmente las emisiones de gases de efecto invernadero (GEI). Funcionan como un mecanismo de mercado, siguiendo el principio de "limitar y comerciar" (cap and trade), que busca reducir las emisiones de la forma más eficiente posible.

Frente al impuesto pigouviano –que fija el precio de la externalidad y deja que la cantidad contaminada se ajuste–, los SDE fijan la cantidad y dejan que el precio se ajuste. Weitzman (1974) demostró que la elección entre ambos instrumentos no es indiferente bajo incertidumbre sobre los costes de reducir la contaminación: conviene fijar el precio –como el impuesto pigouviano– cuando la curva de coste marginal de reducir emisiones es más inclinada que la de beneficio marginal, mientras que conviene fijar la cantidad –como hacen los SDE– cuando es el beneficio marginal de reducir la contaminación el que cambia bruscamente, como ocurre cuando existen umbrales a partir de los cuales el daño se dispara[36].

Su funcionamiento se articula en tres fases:

1. Límite (Cap): En primer lugar, una autoridad gubernamental establece un límite total (o "tope") a la cantidad de emisiones que pueden liberar los sectores regulados (como la industria, la producción de energía y, en algunos casos, la aviación). Este límite se reduce progresivamente con el tiempo para asegurar que las emisiones disminuyan de forma constante.
2. Asignación de derechos: El total de derechos de emisión, donde cada derecho equivale a una tonelada de CO_2, se distribuye entre las empresas que participan en el sistema. Esta asignación puede hacerse de dos maneras:
   a) Subasta: Las empresas compran los derechos en un mercado primario. Esto garantiza que el precio del carbono refleje la oferta y la demanda, y los ingresos generados pueden reinvertirse en proyectos climáticos.

[36] La intuición es que, si rebasar cierto nivel de contaminación resulta muy dañino, lo prioritario es asegurar la cantidad aunque cumplirla salga más caro de lo previsto. Si el daño crece de forma suave, en cambio, es preferible acotar el coste con un precio.

 © 2026 Manuel Cabado García

b) Asignación gratuita: Los derechos se otorgan de forma gratuita a ciertas industrias, a menudo para prevenir la "fuga de carbono", es decir, que las empresas trasladen su producción a países con regulaciones ambientales menos estrictas.

3. Comercio (Trade): Una vez asignados, los derechos de emisión pueden ser comprados y vendidos en un mercado secundario. Esto crea un precio para el carbono, convirtiendo la contaminación en un coste para las empresas. Las empresas que logran reducir sus emisiones por debajo de su asignación inicial pueden vender sus derechos sobrantes, obteniendo un beneficio. Por el contrario, las empresas que superan su límite de emisiones deben comprar derechos adicionales en el mercado para cumplir con sus obligaciones.

Ventajas y desafíos

- La principal ventaja de este sistema es su eficiencia económica. Las empresas pueden elegir la opción más rentable: invertir en tecnología limpia para reducir sus emisiones o comprar derechos de emisión si les resulta más barato. Esto fomenta la innovación y asegura que las reducciones de emisiones se realicen donde el coste de hacerlo es menor. Además, el precio del carbono incentiva a los consumidores y a las empresas a tomar decisiones más sostenibles.
- Sin embargo, los sistemas de derechos de emisión también enfrentan desafíos. La volatilidad del precio del carbono puede dificultar la planificación a largo plazo para las empresas. Además, la asignación gratuita de derechos puede reducir la efectividad del sistema y crear un exceso de oferta, lo que hace caer los precios del carbono. Por último, la correcta aplicación del principio de "limitar y comerciar" depende de una gobernanza sólida para asegurar la transparencia y el cumplimiento de las normativas.

En la Unión Europea, este mecanismo se aplica a través del Régimen de Comercio de Derechos de Emisión (RCDE UE / EU ETS), ya descrito en el apartado 15.A.3 al explicar el Teorema de Coase: cubre aproximadamente el 40 % de las emisiones totales de gases de efecto invernadero de la región y es el mercado de carbono de mayor valor del mundo.

15.A.10. Ejemplos prácticos de externalidades en producción y en consumo

A continuación se recogen ejemplos prácticos de las cuatro combinaciones de externalidad ya analizadas en 15.A.5 –negativa o positiva, en producción o en consumo–, útiles para ilustrar con casos concretos el análisis formal desarrollado en el cuerpo del tema.

En la producción, las externalidades negativas incluyen:

- Sobreexplotación pesquera: cuando el recurso es de acceso libre, cada pescador impone a los demás el coste marginal externo de reducir las reservas –el mismo problema de bien común analizado en 15.A.2–, aunque aquí el énfasis está en la decisión individual de capturar una unidad más, no en las soluciones institucionales ya vistas en 15.A.6.
- Ruido de la construcción: una obra en un barrio genera ruido y molestias a los vecinos, afectando su bienestar sin que el promotor internalice ese coste.
- Generación de residuos tóxicos: una industria química que vierte residuos a un vertedero o al subsuelo puede contaminar las aguas subterráneas, dañando la salud pública y el medio ambiente sin coste para la empresa.

Y las positivas:

- Investigación y desarrollo: una empresa que invierte en I+D genera conocimiento que otras empresas pueden aprovechar, beneficiando a la industria en general sin que el innovador capture toda la rentabilidad social de su esfuerzo.
- Restauración de un edificio histórico: un promotor que restaura un edificio antiguo mejora la estética del barrio y aumenta el valor de las propiedades circundantes.
- Instalación de paneles solares: una empresa que instala paneles solares para su propio consumo contribuye a la red eléctrica local cuando produce excedentes, beneficiando a la comunidad al reducir la carga sobre la red.

En el consumo, las externalidades negativas incluyen:

- Fumar en público: quien fuma en un espacio público expone a los demás al humo de segunda mano, con un impacto negativo en su salud que el fumador no soporta.
- Conducción en la ciudad: el uso del coche en una zona urbana contribuye a la congestión del tráfico y a la contaminación del aire, afectando a todos los ciudadanos, no solo al conductor.
- Concierto ruidoso en un apartamento: un vecino que escucha música a un volumen muy alto molesta a los demás vecinos del edificio, alterando su tranquilidad sin compensarles por ello.

Y las positivas:

- Vacunación: quien se vacuna no solo se protege a sí mismo, sino que reduce el riesgo de contagio para toda la comunidad, generando inmunidad de rebaño.
- Educación: quien adquiere educación formal contribuye al desarrollo de la sociedad, favoreciendo una ciudadanía más informada y una menor delincuencia.
- Jardinería y embellecimiento de un jardín: quien cuida y embellece su jardín mejora el aspecto de la calle y aumenta el atractivo visual del vecindario para los demás.

 © 2026 Manuel Cabado García

Bibliografía

Adams, G. (1981). *The Iron Triangle: The Politics of Defense Contracting*. New York: Council on Economic Priorities.

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Albi Ibáñez, E., J. M. Gónzalez-Páramo, R. M. Urbanos y I. Zubiri (2017). *Economía Pública I: Fundamentos. Presupuestos y Gastos*. Barcelona: Ariel.

Alesina, A. y G. Tabellini (1990). A Positive Theory of Fiscal Deficits and Government Debt. *The Review of Economic Studies 57*(3), 403–414.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review 53*(5), 941–973.

Buchanan, J. M. y G. Tullock (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Buchanan, J. M. y R. E. Wagner (1977). *Democracy in Deficit: The Political Legacy of Lord Keynes*. New York: Academic Press.

Clarke, E. H. (1971). Multipart Pricing of Public Goods. *Public Choice 11*, 17–33.

Coase, R. H. (1960). The Problem of Social Cost. *The Journal of Law & Economics 3*, 1–44.

Downs, A. (1967). *Inside Bureaucracy*. Boston: Little, Brown.

Gravelle, H. y R. Rees (2004). *Microeconomics* (Third ed.). Addison Wesley.

Groves, T. (1973). Incentives in Teams. *Econometrica 41*(4), 617–631.

Heller, M. A. (1998). The Tragedy of the Anticommons: Property in the Transition from Marx to Markets. *Harvard Law Review 111*(3), 621–688.

Hirshleifer, J. (1983). From Weakest-Link to Best-Shot: The Voluntary Provision of Public Goods. *Public Choice 41*, 371–386.

Hurwicz, L. (1973). The Design of Mechanisms for Resource Allocation. *The American Economic Review 63*(2), 1–30.

Kaul, I., I. Grunberg y M. A. Stern (eds.) (1999). *Global Public Goods: International Cooperation in the 21st Century*. New York: Oxford University Press.

Krueger, A. O. (1974). The political economy of the rent-seeking society. *American Economic Review 64*(3), 291–303.

Kydland, F. E. y E. C. Prescott (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy 85*(3), 473–491.

Lindahl, E. (1919). Just Taxation–A Positive Solution. En R. A. Musgrave y A. T. Peacock (eds.), *Classics in the Theory of Public Finance (1958)*, pp. 170–176. London: Palgrave Macmillan UK.

Myerson, R. B. (1979). Incentive Compatibility and the Bargaining Problem. *Econometrica 47*(1), 61–73.

Niskanen, W. A. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.

Nordhaus, W. D. (1975). The Political Business Cycle. *The Review of Economic Studies 42*(2), 169–190.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ostrom, E., J. Burger, C. B. Field, R. B. Norgaard y D. Policansky (1999). Revisiting the Commons: Local Lessons, Global Challenges. *Science 284*(5412), 278–282.

Pearce, D. (1991). The Role of Carbon Taxes in Adjusting to Global Warming. *The Economic Journal 101*(407), 938–948.

Pigou, A. C. (1920). *The Economics of Welfare*. London: Macmillan.

Pindyck, R. S. y D. L. Rubinfeld (2010). *Microeconomía* (Séptima ed.). Madrid: Pearson Educación.

Puviani, A. (1903). *Teoria della illusione finanziaria*. Palermo: Remo Sandron.

Samuelson, P. A. (1954). The Pure Theory of Public Expenditure. *The Review of Economics and Statistics 36*(4), 387–389.

Sandler, T. (1998). Global and Regional Public Goods: A Prognosis for Collective Action. *Fiscal Studies 19*(3), 221–247.

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell Journal of Economics and Management Science 2*(1), 3–21.

 © 2026 Manuel Cabado García

Stiglitz, J. E. y J. K. Rosengard (2015). *Economics of the Public Sector* (Fourth ed.). WW Norton & Company.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy 64*(5), 416–424.

Tullock, G. (1967). The welfare costs of tariffs, monopolies, and theft. *Western Economic Journal 5*(3), 224–232.

Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance 16*(1), 8–37.

Weitzman, M. L. (1974). Prices vs. Quantities. *The Review of Economic Studies 41*(4), 477–491.

 © 2026 Manuel Cabado García

Tema 16

Economía del bienestar (III): teoría de la elección colectiva. El teorema de imposibilidad de Arrow. Funciones de bienestar social, reglas de votación y otros desarrollos posteriores.

16.1. Introducción ⏳ 1'

En el blog *Nada es Gratis*, el profesor José Luis Ferreira relató cómo los participantes del 15-M, al organizar sus asambleas, descubrieron por sí mismos los desafíos de la toma de decisiones colectiva. Sin saberlo, se enfrentaban a la misma complejidad que Kenneth Arrow abordó en su trabajo: la aparente sencillez del voto esconde una profunda complejidad.

Así, el Sector Público se enfrenta a un problema central: cómo agregar las diferentes preferencias individuales en una decisión colectiva coherente. El análisis de cómo se realiza esta agregación y de las propiedades de los diferentes mecanismos que la hacen posible se denomina Teoría de la Elección Colectiva. Dentro de este marco, se exploran los mecanismos de decisión del Sector Público, profundizando en las reglas de votación que traducen los votos en resultados, en la elección directa como un reflejo de la voluntad popular, y en la democracia representativa, que delega el poder de decisión en los representantes.

Por consiguiente, el objetivo de esta exposición será estudiar la Teoría de la Elección Colectiva, es decir, si se puede encontrar una buena regla de Elección Colectiva y cuáles son sus efectos a través del siguiente esquema:

1. El problema de la Elección Colectiva.

2. Criterios que debe cumplir la Regla de la Elección Colectiva.
3. Teorema de Imposibilidad de Arrow.
4. Elección mediante votación, donde se analizará la democracia directa y la democracia representativa.

16.2. El problema de la Elección Colectiva ⏳ 2'

Comenzando con el problema de la elección colectiva, este estudia la cuestión de la toma de decisiones del sector público analizando si es posible formular una regla de elección colectiva que permita agregar las preferencias individuales, respetándolas y cumpliendo unos requisitos de racionalidad y eficiencia.

Así, una regla de elección colectiva es un mecanismo que permite agregar preferencias individuales para conseguir una preferencia social, siendo los más utilizados:

- Las reglas de votación, que se basan en la ordenación de preferencias individuales.
- Funciones de Bienestar Social, que ponderan la intensidad de preferencias individuales en una función de utilidad.

La función de bienestar social fue planteada inicialmente por Bergson (1938) y Samuelson (1947) y se define como una función W que asigna un índice de utilidad a un vector de variables, $(Z_1, Z_2, \ldots, Z_n)$, que pueden afectar al bienestar social:

$$W = W(Z_1, Z_2, \ldots, Z_n) \tag{16.1}$$

donde las variables son índices de utilidad, de modo que la función de bienestar social permite elegir entre óptimos de Pareto. De esta manera, la función de bienestar social asigna un índice de utilidad social a cada vector de índices de utilidad.

La forma de la función de bienestar social –concavidad– reflejará las distintas preferencias sobre la distribución de la renta:

- La función de bienestar utilitarista, desarrollada por Bentham y Harsanyi (1955; 1789), queda definida como la suma lineal de utilidades donde se valora socialmente de forma idéntica la valoración de utilidad de cualquier individuo. No se preocupa por la igualdad –aversión a la desigualdad igual a 0 o agentes neutrales ante el riesgo–, siendo las curvas de indiferencia social rectas con pendiente -1.
- La función de bienestar social maxi-min, desarrollada por Rawls[1] (1971), donde el criterio maximin implica que el bienestar social es igual a la utilidad del individuo con menor utilidad. Según Rawls, cuando los agentes actúan bajo el "velo de la ignorancia", es decir, desconocen ex ante su lugar en la sociedad, se protegerán

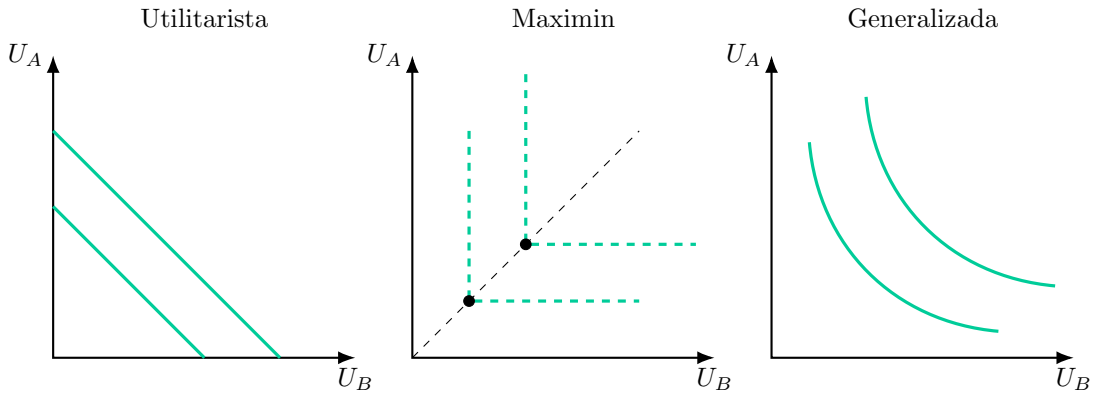
[1] Rawls heredó las ideas de Rousseau, Hobbes y Locke sobre el contrato social

 © 2026 Manuel Cabado García

frente al peor resultado posible, como si tuvieran una aversión infinita al riesgo, y por eso seguirán esta regla[2]. Así, las curvas de indiferencia social tendrán forma de L –tipo Leontief– y sus vértices se sitúan en la bisectriz, ya que el bienestar social solo aumenta si mejora la utilidad del individuo peor situado: aumentar la del mejor situado, sin mejorar la del otro, no añade bienestar.

- La función de bienestar social de Bernoulli y Nash (1738; 1950), que toma como referencia una especificación tipo Cobb-Douglas, el producto de las utilidades individuales, con curvas de indiferencia social convexas. De esta forma, se pasa de la aversión infinita a la desigualdad de Rawls a una aversión finita. Ahora se está dispuesto a intercambiar pérdidas de utilidad de los más favorecidos a cambio de ganancias de los más desfavorecidos hasta un límite.

Figura 16.1: Tipos de Función de Bienestar Social



Sin embargo, la Función de Bienestar Social plantea limitaciones al realizar comparaciones interpersonales de utilidad y tratarse de un enfoque cardinal[3].

Por ello, buena parte de la Teoría de la Elección Colectiva se ha centrado en el análisis de las Reglas de Votación, tanto porque evitan el debate sobre la posibilidad o no de com-

[2]Concretamente, Rawls indica que los individuos de la sociedad llegarían a un acuerdo social unánime basado en tres principios:

1. Principio de libertad, cada persona tiene derecho a la máxima libertad compatible con la libertad de los demás.
2. Principio de diferencia, las desigualdades económicas y sociales solo son legítimas si van ligadas a cargos y posiciones abiertos a todos en condiciones de igualdad equitativa de oportunidades y, además, benefician a los más desfavorecidos.
3. Prioridad entre principios, el principio de libertad prima sobre el de diferencia y, dentro de este, la igualdad de oportunidades prima sobre el beneficio de los más desfavorecidos.

[3]La Función de Bienestar Social exige poder realizar comparaciones interpersonales de utilidad de carácter cardinal ya que, por definición, comparar entre dos óptimos globales implicará el incremento de la utilidad de un agente y la reducción de la del otro. Para decidir la superioridad hace falta conocer la intensidad de estos incrementos en $u(x)$ y compararlos entre los agentes. Esto choca con la teoría de la utilidad, que ha rechazado casi unánimemente las funciones de utilidad cardinal y el empleo de comparaciones interpersonales de utilidad. A pesar de esto, si se quieren tener en cuenta los trade-offs entre agentes, es necesario hacer comparaciones cardinales.

parar utilidades, como por el hecho de que en las sociedades democráticas las decisiones se toman a través de votación.

16.3. Criterios que debe cumplir la Regla de la Elección Colectiva ⏳ 1'

En sociedades democráticas y racionales, se espera que las reglas de elección colectiva cumplan unos determinados criterios de coherencia respecto a las preferencias de los individuos. El Premio Nobel Kenneth Arrow (Arrow, 1951) estableció cinco propiedades básicas deseables que debería cumplir cualquier Regla de Elección Colectiva:

1. Dominio no restringido o universalidad (U), es decir, la regla debe estar definida para cualquier combinación de preferencias individuales.
2. Racionalidad (R), es decir, la preferencia social que genera la regla es transitiva: dadas tres alternativas, A, B, y C, si $A \succ B$ y $B \succ C$, entonces $A \succ C$. Implica que el resultado de la regla de elección social tiene coherencia lógica, evitando votos cíclicos.
3. Criterio de Pareto (optimalidad débil, P), es decir, si todos prefieren individualmente una alternativa A a otra B entonces socialmente también debe ser así, asegurando la eficiencia[4].
4. Ausencia de Dictador (AD), es decir, no debe existir ningún individuo cuya preferencia estricta entre dos alternativas se convierta siempre en la preferencia social, sean cuales sean las preferencias de los demás. Esto implica que nadie controla la regla de elección colectiva.
5. Independencia de las Alternativas Irrelevantes (IAI), es decir, la elección entre dos alternativas debe depender solo de las preferencias individuales entre ambas alternativas y no de la posición relativa frente a terceras[5]. Este axioma impide utilizar información sobre la intensidad de las preferencias, de forma que la regla solo puede basarse en cómo ordena cada individuo las dos alternativas en juego.

16.4. El Teorema de Imposibilidad de Arrow ⏳ 1'

En base a las anteriores propiedades, el propio Arrow (1950; 1951) planteó el teorema de imposibilidad, afirmando que, si existen al menos tres alternativas y el número de indivi-

[4]Si se exigiera Pareto optimalidad fuerte, entonces para que la alternativa A fuese preferida a la B no haría falta que todos los individuos la prefiriesen, sino que bastaría con que un individuo la prefiriese y otros estuviesen indiferentes.

[5]Por ejemplo, si la elección es entre gasto en sanidad y en educación, el resultado debe ser independiente de cómo valoren los individuos la sanidad y la educación en relación a un tercer bien (como, por ejemplo, la justicia).

 © 2026 Manuel Cabado García

duos es finito, no existe ninguna regla de elección colectiva que cumpla simultáneamente los cinco axiomas[67].

Por consiguiente, será imposible encontrar una Regla de Elección Colectiva que utilizando información exclusivamente ordinal acerca de las preferencias individuales se ajuste a unos principios mínimos de coherencia y respeto a preferencias individuales[8]. Por lo tanto, si se utiliza una regla de elección colectiva, se ha de renunciar a alguna de estas condiciones[9][10].

Sin embargo, la posibilidad de renunciar a R, P y AD parece poco aceptable en sociedades democráticas, por lo que la Economía Pública se ha centrado en los sistemas de elección colectiva que renuncian a la U, como la votación mayoritaria, o la IAI, como la función de bienestar social.

16.5. Elección mediante votación ⏳ 14'30"

Como se ha mencionado, buena parte de las teorías de Elección Social se basan en las reglas de votación para evitar comparar utilidades y porque, normalmente, se utilizan en sociedades democráticas. A continuación, se analizarán las reglas de democracia directa y democracia representativa.

Democracia Directa

Respecto a la democracia directa –donde las decisiones públicas se adoptan por votación entre los individuos[11]–, aunque casi no se aplica su estudio es importante ya que:

1. Revela los problemas inherentes a cualquier mecanismo de votación.

[6] Como no existe ninguna regla que cumpla todos los axiomas, hay que conformarse con reglas que renuncien a alguno de ellos.

[7] Lo que el Teorema de la imposibilidad de Arrow indica es que las reglas de votación no son un buen mecanismo de decisión, ya que todas ellas deben renunciar a alguna de las cinco propiedades mencionadas, y todas ellas se consideran propiedades básicas y deseables. Así, según Arrow, ningún sistema de votación es justo.

[8] Sin embargo, hubo autores de la Economía del Bienestar que criticaron la teoría de Arrow:

- Little criticó que la regla de elección colectiva de Arrow simplemente actúe como una máquina que agrega valores y permite alcanzar una solución, pero que esta solución no representaría necesariamente los valores de la sociedad.
- Bergson, por su parte, introdujo una distinción entre aconsejar a un individuo y aconsejar a un funcionario. El objetivo real de la Economía del bienestar, dirá Bergson, es el primero, mientras que la regla de elección colectiva de Arrow indicaría al funcionario lo que debe de hacer.

[9] Por su parte, una Función de Bienestar Social, como mecanismo de elección colectiva, si es individualista y anónima, cumple las condiciones U, AD, P y R del Teorema de Arrow. Sin embargo, al depender de los niveles de utilidad de los individuos –de la intensidad de sus preferencias– incumple la condición IAI.

[10] Por tanto, existe un "precio" por emplear una regla de elección colectiva, que implicará renunciar a alguna de las condiciones (ningún sistema de votos es justo). Como ha señalado Winston Churchill: "la democracia es el menos malo de todos los sistemas"

[11] Este sería el caso, por ejemplo, si se realizara un referéndum en un país sobre alguna cuestión pública.

2. Es aplicable en circunstancias económicas en las que el número de votantes es reducido, por ejemplo, votaciones en un Consejo de Ministros, en un Parlamento o en el seno del consejo directivo de una empresa.
3. Bajo ciertas condiciones, lleva a las mismas decisiones de provisión pública que la Democracia Representativa[12].

En Democracia Directa se parte de suponer una serie de preferencias individuales que tratan de traducirse en decisiones presupuestarias a través de un proceso regido por unas reglas de votación previamente establecidas. Se supone un individuo, un voto, y se analizan los aspectos de las reglas de votación: la distribución de votos y los criterios que determinan la decisión del voto ganador.

A continuación, se analizarán los principales criterios para determinar el voto ganador.

Regla de la Unanimidad

En primer lugar, el economista sueco Knut Wicksell (Wicksell, 1896) propuso la regla de la unanimidad que exige el acuerdo total de todos los individuos para aprobar una decisión. Wicksell la diseñó para resolver el problema de la financiación de los bienes públicos de forma eficiente.

Ventajas de la Regla

- Mejora de Pareto: solo se aprueban las propuestas que benefician al menos a una persona sin perjudicar a nadie, porque cualquier perjudicado podría vetarlas. Si se siguen aprobando propuestas mientras quede alguna mejora posible, se alcanza un óptimo de Pareto.
- Financiación consensuada: el coste del bien público se financia con impuestos que todos los individuos aceptan voluntariamente. Por ejemplo, todos los vecinos tendrían que acordar no solo construir una nueva acera, sino también la fórmula exacta para repartir el coste entre ellos.

Desventajas de la Regla Lentitud del Proceso:

- La negociación es extremadamente larga y compleja, haciendo que sea muy difícil, e incluso imposible, llegar a un acuerdo.
- Comportamientos Estratégicos: La regla puede generar tácticas manipuladoras, especialmente si el poder de negociación no es igual entre las partes:
    - Tiranía de la minoría (Baumol): Una minoría con poder de veto puede bloquear cualquier propuesta de la mayoría hasta conseguir concesiones que les

[12] La relación entre las elecciones en una democracia directa y en una representativa depende del comportamiento de los votantes y de los políticos, e incluso de la estructura del proceso de toma de decisiones dentro del Sector Público.

 © 2026 Manuel Cabado García

beneficien de forma desproporcionada. Por ejemplo, una pequeña comunidad que no quiere pagar impuestos para un hospital podría vetar la construcción a menos que el gobierno les ofrezca algo a cambio, como la exención del pago.

- Tiranía de la mayoría (Wicksell): La mayoría podría presionar a la minoría para que acepte una propuesta que no les beneficia, amenazando con negarle otros bienes o servicios. Por ejemplo, una mayoría que quiere construir una autopista podría amenazar a una minoría que se opone a ella con dejarla fuera del acceso a otros servicios públicos.

Regla de la mayoría óptima en cada elección

Debido a las limitaciones del criterio de unanimidad, Buchanan y Tullock (1962) plantean una mayoría óptima para cada decisión. Según estos autores, existiría una mayoría óptima en la elección que pondera dos tipos de costes:

- Costes Externos, CE, que se producen cuando la alternativa elegida va en contra de los intereses de los individuos. Por lo tanto, cuanto mayor es el porcentaje de votos requeridos para que una alternativa sea vencedora, menores son los costes externos.
- Costes de Decisión, CD, surgen por la paralización de las decisiones hasta que se alcanza una mayoría y, por tanto, son más elevados cuanto mayor es el porcentaje de votos requeridos.

Gráficamente, en un plano de costes y porcentajes de mayoría requerida como el de la Figura 16.2, los CE serán decrecientes y los CD serán crecientes. Por lo tanto, los Costes Totales , CT, que serán la suma de ambos, tendrán forma de U. En el mínimo de los CT, M^*, se encuentra la mayoría óptima, que no tiene por qué coincidir con el 50 % de la población.

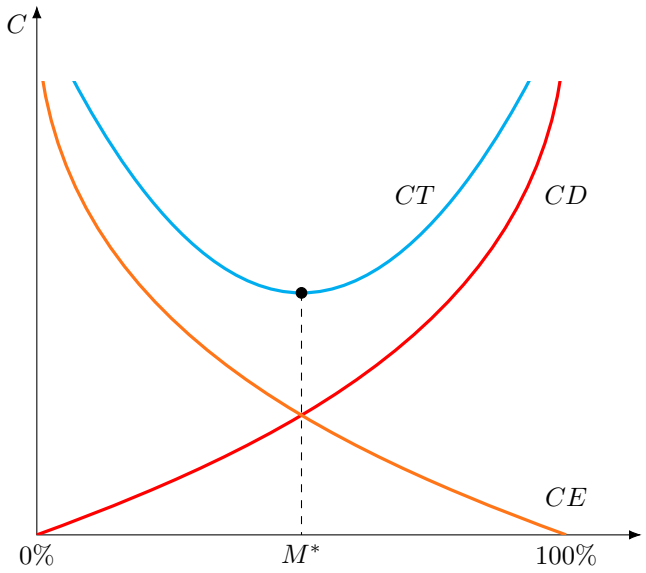
Sin embargo, para este criterio también se plantean una serie de problemas prácticos a la hora de fijar la mayoría óptima, ya que se deberán valorar los costes de transacción y, además, la estructura de costes será diferente para cada decisión. Por ello, la regla predominante es la Regla de la Mayoría Simple.

Regla de la Mayoría Simple

La Regla de la Mayoría Simple establece que una alternativa es elegida cuando recibe la mitad más uno de los votos cuando existan dos alternativas. Cuenta con grandes ventajas ya que:

- Evita largas deliberaciones, por lo que los costes de decisión son muy bajos.
- Es de fácil comprensión y goza de gran legitimidad entre las sociedades democráticas.

Figura 16.2: Regla de la mayoría óptima en cada elección



Por ello, la mayoría simple constituye el mecanismo de referencia para la elección directa. Sin embargo, no está exenta de inconvenientes relacionados con los criterios de Arrow:

- No garantiza una provisión eficiente, ya que no revela la intensidad de las preferencias. Esto no significa que incumpla el criterio de Pareto de Arrow, que sí cumple, porque si todos prefieren una alternativa esta gana la votación, sino que el nivel elegido puede no ser un óptimo de Pareto.
- Incentiva escasamente a la participación, dado que la posibilidad de que un voto sea decisivo es muy reducida.
- Tiene propiedades redistributivas que pueden convertirlo en un instrumento de explotación de la minoría por la mayoría.
- Posible incoherencia de resultados, que se explica a través de la paradoja de Condorcet (1785), es decir, ninguna alternativa resulta ganadora a causa de que cada alternativa puede ser batida por otra. Entonces se habla de mayorías cíclicas.

Incoherencia de resultados

Para explicar este último punto, si se plantea una situación en la que existen tres votantes, V_1, V_2 y V_3, y tres alternativas sobre niveles de gasto público, A, B, y C, y se establecen las siguientes preferencias:

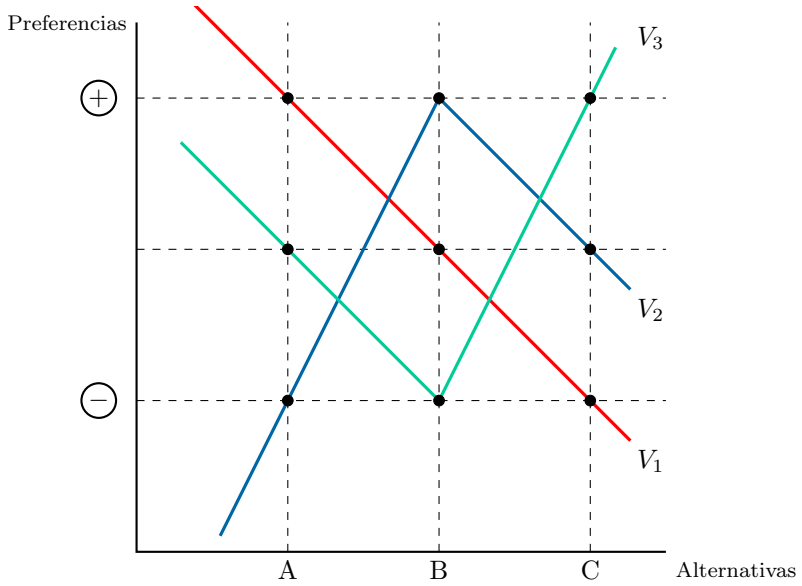
$$V_1 : A \succ B \succ C$$
$$V_2 : B \succ C \succ A$$
$$V_3 : C \succ A \succ B$$

 © 2026 Manuel Cabado García

votando por pares de alternativas.

Esto se puede observar en un plano de intensidad de preferencias y niveles de gasto público como el de la Figura 16.3.

Figura 16.3: Paradoja de Condorcet



Para adoptar una decisión social, se compararán las preferencias por pares:

- Entre A y B, se elige la alternativa A, preferida por V_1 y V_3.
- Entre B y C, se elige la alternativa B, preferida por V_1 y V_2.
- Entre A y C, se elige la alternativa C, preferida por V_2 y V_3.

Así, aunque cada votante tiene una ordenación transitiva, la preferencia social resultante es intransitiva ($A \succ B$, $B \succ C$ y $C \succ A$), por lo que ninguna sale ganadora, violando así el supuesto de racionalidad de Arrow[13].

Esto ocurre porque el votante 3, V_3, tiene preferencias multimodales (es decir, tiene dos máximos relativos, A y C, cuando las alternativas se ordenan según el nivel de gasto). La ciclicidad se evita, por ejemplo, cuando todos los individuos tienen preferencias idénticas o cuando todas las preferencias son unimodales. En este caso, en cambio, existe ciclicidad y no se verifica la transitividad social, dando lugar a la paradoja de Condorcet.

Aun así, existen vías para evitar la paradoja como:

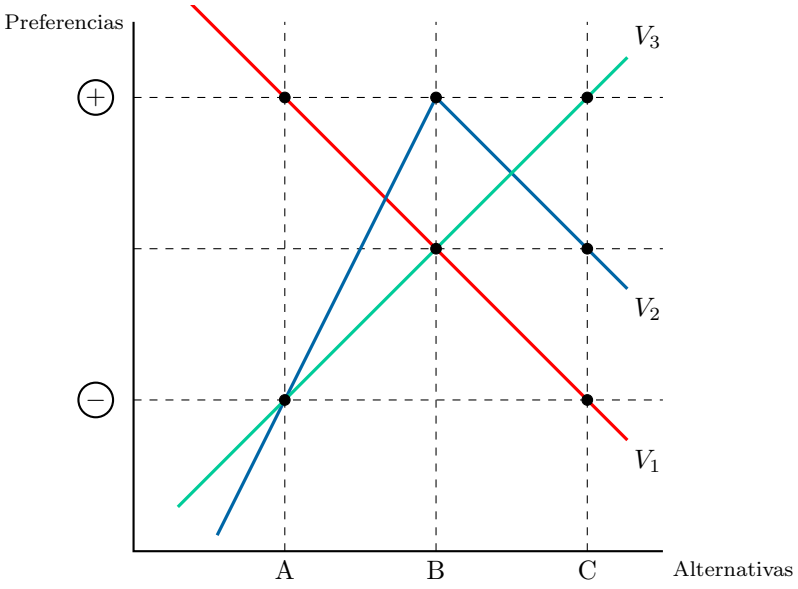
1. Todas las preferencias son idénticas. Sin embargo, en ese caso no sería necesaria una regla de elección colectiva.

[13]La intransitividad implica que no existe ningún equilibrio en votación mayoritaria ya que el proceso de votación se extendería indefinidamente sin que se alcanzara una decisión. Así, se vota entre A y B, se elegiría A y el individuo 3 pediría que se votase entre C y A. Entre estas dos alternativas saldría elegida la alternativa C y el votante 2 pediría que se votase entre C y B, siendo elegida la alternativa B, a continuación el votante 1 demandaría una votación entre A y B. De esta forma se volvería al punto de partida y la votación se repetiría de forma indefinida.

2. Eliminar las alternativas que han sido rechazadas alguna vez. Sin embargo, este método puede resultar en problemas de manipulación de la agenda por parte del seleccionador del orden de votación: el que diseña el orden de votación de las alternativas puede lograr el resultado final que desea, de modo que, en la práctica, actúa como un dictador aunque la regla no sea formalmente dictatorial.
3. Las preferencias son unimodales (Black), lo que implica renunciar al axioma de universalidad.
   El análisis se centra en esta opción. Además, deben considerarse preferencias unidimensionales (es decir, se decide solo sobre una dimensión, por ejemplo, el nivel de gasto). Si son multidimensionales también pueden llevar a ciclos, comparación de preferencias, etc.

En el ejemplo anterior, para hacer las preferencias unimodales, habría que modificar las preferencias del individuo 3: $C \succ B \succ A$, de manera que la alternativa ganadora sería la B, y en el gráfico existiría un solo pico –máximo– para cada individuo tal y como se observa en la Figura 16.4.

Figura 16.4: Preferencias unimodales



Pero, ¿hasta qué punto es restrictivo renunciar a la universalidad exigiendo preferencias unimodales? No es demasiado restrictivo exigir la unimodalidad, pues ésta implica que, a medida que un individuo se aleja de su opción preferida, debería estar siempre peor, de manera que las preferencias se modulen y no salten de extremo a extremo.

Así, por ejemplo, si se describen las preferencias de un individuo sobre sanidad pública, parece lógico que, si su opción preferida es mucho gasto público, su siguiente opción preferida sea gasto público medio, y su última opción sea gasto público nulo.

 © 2026 Manuel Cabado García

En definitiva, abandonar la condición de universalidad implica un cierto sacrificio (pues la regla deja de funcionar para algunas combinaciones de preferencias individuales), pero este sacrificio se juzga menos importante que el que resulta del incumplimiento de la condición de transitividad (es decir, la no existencia de una solución única).

Teorema del votante mediano de Black (1948) La principal implicación de imponer unimodalidad es que la elección vendrá determinada por las preferencias del votante mediano, es decir, aquel cuya alternativa preferida se encuentra en la mediana de las alternativas más preferidas[14], que en el caso anterior se daría sobre la opción B.

Este hecho se aprecia en una votación sobre niveles de gasto público, como se verá a continuación.

Provisión no eficiente

Se supone que los n individuos con preferencias unimodales se preocupan únicamente por el nivel de gasto y que tiene las características de bien público, el cual es financiado por impuestos de cuantía alícuota. El nivel de gasto se determina por la regla de votación mayoritaria.

Así, cada consumidor determinará su nivel de gasto público preferido maximizando su utilidad

$$\text{máx} \quad U\left(Y_d, G\right) \tag{16.2}$$

sujeto a su restricción presupuestaria

$$Y_d = Y - T\left(Y_i, G\right) \tag{16.3}$$

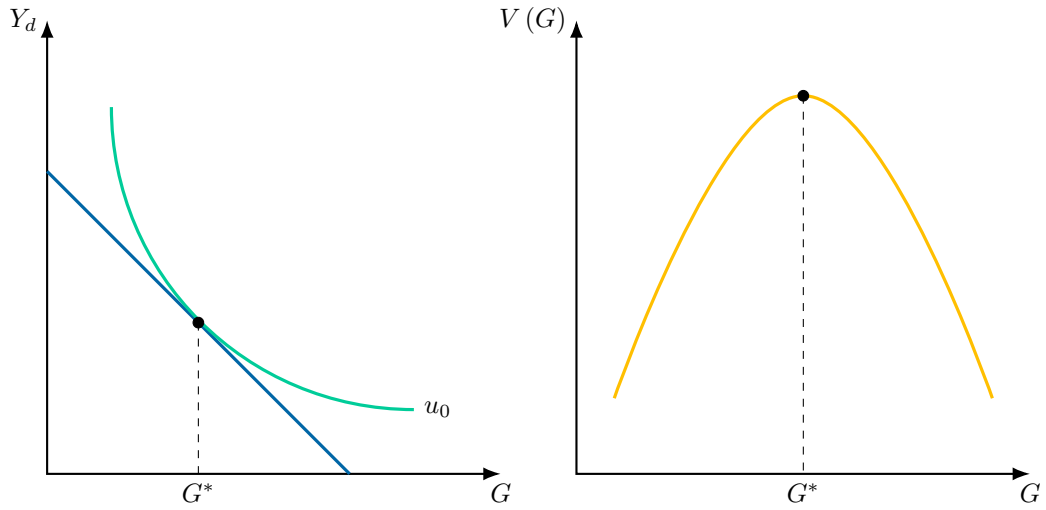
Por tanto, tal y como se observa en la Figura 16.5:

- Cada consumidor determinará su nivel de gasto público preferido maximizando su utilidad sujeto a su restricción presupuestaria, gráficamente en el punto de tangencia de la curva de indiferencia con la restricción presupuestaria.
- El nivel óptimo de gasto vendrá dado por G^*.
- Por consiguiente, la elección del bien público G por parte de los agentes, $V\left(G\right)$, depende de sus gustos, el nivel de renta y los impuestos (precio del bien público G).

Llevando este resultado a una votación, gráficamente en un plano de preferencias y niveles de gasto como el de la Figura 16.6, se suponen cinco individuos y se ordenan los niveles de gastos preferidos por los votantes de mayor a menor, donde el votante mediano será aquel cuyo nivel de gasto preferido, G_M, se encuentra en la mediana de la distribución de gastos óptimos.

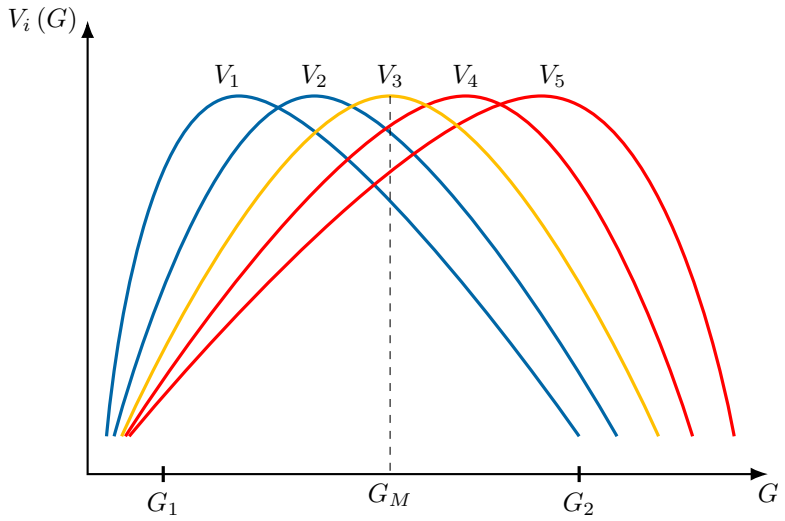
[14] El votante mediano y el medio coinciden si la distribución de las preferencias de los votantes es simétrica. En este tipo de distribución, la media (votante medio) y la mediana (votante mediano) se encuentran en el mismo punto, en el centro de la curva de preferencias.

Figura 16.5: Elección del bien público por parte de los agentes



- Si hay dos alternativas de gasto, G_1 y G_M, votarán por la alternativa G_M, tanto el votante mediano como los que están a su derecha, V_4 y V_5, y algunos de los que están a su izquierda, V_2.
- Ocurrirá lo mismo para la alternativa de gasto G_2.
- Dicho de otra manera, el gasto preferido por el votante mediano es el preferido en una votación.

Figura 16.6: Votación sobre el nivel de gasto



Este es un resultado general de la votación mayoritaria: cuando los individuos tienen preferencias unimodales y se votan alternativas unidimensionales, el resultado de la votación coincidirá con la alternativa preferida por el votante mediano.

 © 2026 Manuel Cabado García

Sin embargo, este resultado es válido si las alternativas son unidimensionales, es decir, solo se elige sobre el nivel de gasto, ya que como afirma Kramer (1973) con alternativas multidimensionales volvería a aparecer el problema de la ciclicidad[15].

Posible ineficiencia del resultado No obstante, aun habiendo superado el problema de la ciclicidad, la regla de la mayoría simple no revela la intensidad de las preferencias por lo que no es eficiente en el sentido de Pareto. Esta premisa se va a demostrar con el análisis de la eficiencia, donde el nivel de gasto preferido por el votante mediano no tiene por qué coincidir con el nivel eficiente de gasto.

Gráficamente, en un plano de relaciones marginales de sustitución y transformación, RMS y RMT, y niveles de provisión de un bien público, G, como el de la Figura 16.7, se van a representar

- Las diferentes relaciones marginales de sustitución, que representan las preferencias marginales de los individuos y
- La relación marginal de transformación, RMT, que representa el coste marginal de producción, que se va a asumir constante por simplicidad.

La regla de eficiencia de Samuelson (1954) para la provisión de bienes públicos asegura que la suma de relaciones marginales de sustitución de los individuos, i, debe coincidir con la relación marginal de transformación[16]:

$$\sum RMS_i = RMT \quad i = 1, 2, 3 \tag{16.4}$$

Sin embargo, el resultado de la votación no es independiente de la financiación del gasto público. Por lo tanto, asumiendo que la financiación del gasto público es a través de un impuesto de capitación, el nivel de impuesto que paga cada individuo, τ_i, será:

$$\tau_i = \frac{G}{3} \cdot RMT \tag{16.5}$$

por lo tanto, para un individuo i el nivel de gasto público G preferido será aquel que iguale su relación marginal de sustitución al coste marginal que le supone una unidad más de gasto público[17], es decir:

$$RMS_i = \frac{\partial \tau_i}{\partial G} = \frac{RMT}{3} \tag{16.6}$$

Gráficamente, en Figura 16.7, si se representa la suma de las relaciones marginales de sustitución, $\sum RMS_i$, el nivel de gasto óptimo se determinaría en el punto de intersección

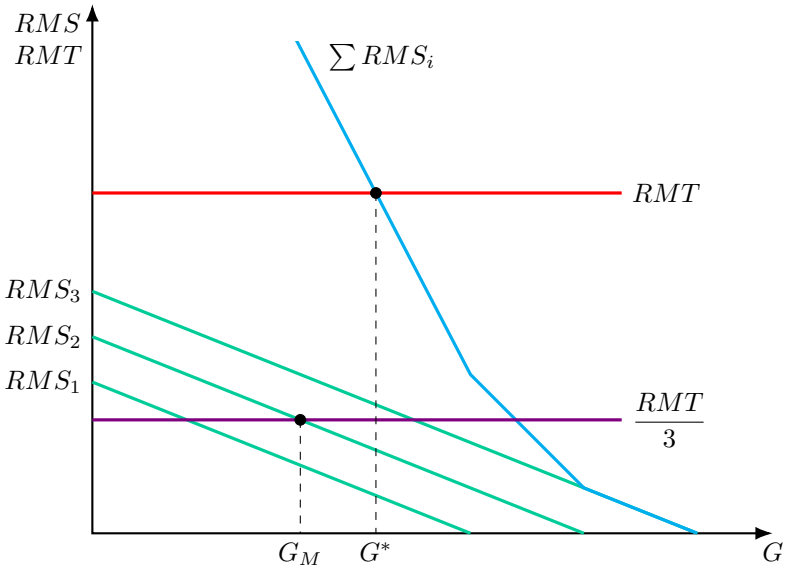
[15] Esto es, exceptuando situaciones triviales en las que existe unanimidad, los ciclos son una realidad inherente a cualquier proceso de votación multidimensional. El modelo de votación, entonces, solo producirá resultados definidos y predecibles con alternativas unidimensionales.

[16] Por definición, la RMT entre un bien y la renta es igual al coste marginal de producción de este bien.

[17] El nivel de gasto más preferido por un individuo será aquel para el que se iguale el beneficio marginal de la última unidad provista al coste (fiscal) marginal que tiene esa unidad para él.

entre la RMT y la suma de las RMS, cumpliéndose así la condición de eficiencia de Samuelson. Por lo tanto, el nivel óptimo de gasto público sería G^*.

Figura 16.7: Nivel de provisión ineficiente de gasto público



Sin embargo, como se ha visto, a través de la regla de votación cada votante igualaría su RMS a su parte del coste marginal, es decir, a $\frac{RMT}{3}$. Como hay tres votantes y el votante 2 es el votante mediano, el nivel de gasto público preferido sería igual al nivel de gasto público preferido por el votante mediano, G_B, que sería menor que el gasto eficiente, G^*, que era uno de los inconvenientes relacionados con los criterios de Arrow.

La intensidad de preferencias Esto se debe a que no se considera la intensidad de preferencias, que es la que determina el nivel óptimo de provisión de bien público:

- Si los individuos pudieran negociar antes de la votación, la minoría, que tiene preferencias más intensas –y que por lo tanto obtiene mayores ganancias de la provisión del bien público–, podría compensar monetariamente a la mayoría para que votase a favor del nivel eficiente de gasto.
- Esto implica relajar IAI, para lo que se plantean dos posibilidades: el intercambio de votos y la regla de Borda.

Intercambio de votos o logrolling Por un lado, para analizar el método de intercambio de votos se supondrá que existen dos proyectos, A y B, y tres votantes, V_1, V_2 y V_3, suponiendo que el proyecto A solo beneficia a V_1 y el proyecto B solo beneficia a V_2, si los tres individuos votan para ver si salen o no los proyectos, V_2 y V_3 votarán en contra del proyecto A y V_1 y V_3 votarán en contra del proyecto B. Por lo tanto, no se aceptarían los proyectos.

 © 2026 Manuel Cabado García

Sin embargo, si V_1 y V_2 negocian y se comprometen a intercambiar votos ambos proyectos saldrían adelante, ya que solo V_3 votaría en contra de ambos proyectos. No obstante, este método tiene una serie de inconvenientes ya que:

1. Nada garantiza que se cumpla el acuerdo, a no ser que V_1 y V_2 presenten ambos proyectos en una única propuesta conjunta, de modo que ninguno pueda votar a favor del suyo y en contra del otro.
2. Existen elevados costes de negociación.
3. Este método puede aumentar el gasto público por encima del nivel eficiente.
4. Si V_3 no se ve representado en ningún proyecto, se produce una explotación de la minoría.

Regla de Borda Por otro lado, la regla de Borda (1781) supone que si existen cuatro alternativas, A, B, C y D, y tres votantes, V_1, V_2 y V_3, cada votante ordena las alternativas y asigna 4 puntos a la que más prefiere, 3 a la segunda, 2 a la tercera y 1 a la última. Se elige la alternativa que obtenga más puntos en total.

Los inconvenientes de este mecanismo son:

1. La posibilidad de que se dé un empate.
2. El resultado es manipulable: un votante puede falsear el orden de las alternativas que no le importan, por ejemplo relegando al último lugar a la principal rival de su opción preferida, para inclinar el resultado a su favor.

Teorema de Gibbard-Satterthwaite Respecto al último punto, el teorema de Gibbard-Satterthwaite (Gibbard, 1973; Satterthwaite, 1975) establece que, si existen al menos tres alternativas, no existe ninguna regla de elección colectiva no manipulable o a prueba de estrategia, es decir, en la que declarar las verdaderas preferencias sea una estrategia dominante para cada individuo, que cumpla simultáneamente los criterios de U, P, AD y R. Dicho de otra manera, con tres o más alternativas, un mecanismo de elección que cumpla R, P y U o es dictatorial o es manipulable[18].

En suma, respecto a las reglas de elección colectiva analizadas: la votación mayoritaria es manipulable cuando cumple U, porque puede incumplir R al existir ciclicidad. Si se restringe U, considerando preferencias unimodales sobre una sola dimensión, deja de ser

[18]Una regla de elección colectiva es no manipulable –o a prueba de estrategias– si ningún individuo puede conseguir un resultado mejor para él declarando unas preferencias falsas. Toda regla no manipulable cumple la propiedad de Asociación Positiva (AP), que establece que si la regla de elección colectiva elige determinada alternativa x_i y las preferencias de un individuo cambian de tal forma que x_i mejora en su escala de valoración, si la ordenación relativa del resto de alternativas no varía, entonces la regla de elección debe seguir siendo x_i. En otras palabras, si una persona pasa a valorar más una alternativa previamente escogida no debe cambiar la elección colectiva. Formalmente, si hay tres alternativas y para el individuo j $x_1 \succ x_2 \succ x_3$, y la elección colectiva es x_2, si las preferencias de j pasan a ser $x_2 \succ x_1 \succ x_3$, la elección colectiva debe seguir siendo x_2.

manipulable. En general, todos los mecanismos que se basan en la intensidad de las preferencias incumplen IAI y son manipulables.

Democracia Representativa

Por último, otro criterio de representación social puede ser el de Democracia Representativa, que supone votar sobre decisiones presupuestarias indirectamente a través de los parlamentarios o representantes, reduciendo los costes de negociación y simplificando el proceso.

$$\underset{\text{Eligen a los políticos}}{\text{Ciudadanos}} \longrightarrow \underset{\text{Eligen los programas}}{\text{Políticos}} \longrightarrow \underset{\text{Aplican los programas}}{\text{Burócratas o funcionarios}}$$

Para analizar la democracia representativa hay que estudiar el comportamiento de tres tipos de agentes: los políticos, los votantes y los grupos de interés.

Así, la teoría de la "Public Choice" resalta que el proceso de adopción de las decisiones públicas es más complejo de lo planteado por la teoría de la elección colectiva. En él intervienen diferentes tipos de agentes económicos, no solo votantes, sino también políticos, funcionarios, grupos de presión, con objetivos particulares, y debido a la presencia de información asimétrica son capaces de sesgar el proceso de elección colectiva hacia sus intereses.

Políticos

El comportamiento de los políticos se puede analizar mediante el modelo de Downs-Hotelling[19]. El modelo prevé que los gobiernos llevarán a cabo políticas con rentabilidad electoral y no social, de la misma forma que los partidos políticos en la oposición.

Los supuestos en los que se basa el modelo son los siguientes:

1. Sistema bipartidista maximizador de votos.
2. Votantes con preferencias unimodales.
3. Se vota sobre una variable unidimensional, que es la ideología.

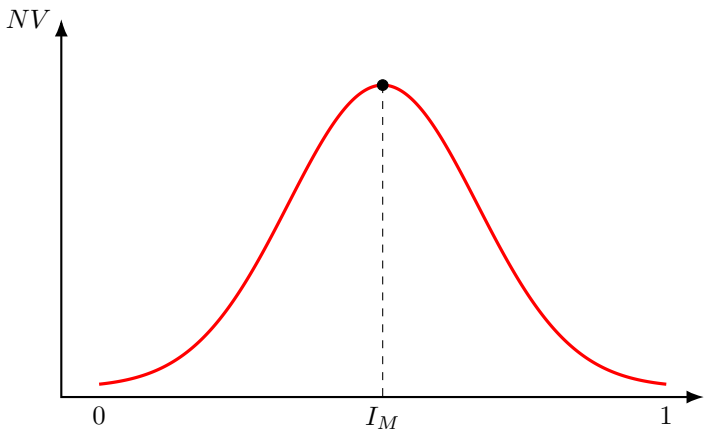
Estos supuestos coinciden con la existencia de un votante mediano, por lo que ganará las elecciones aquel partido cuya ideología coincida con la del votante mediano.

Gráficamente, en un plano de número de votos, NV, y posiciones ideológicas, I, donde 0 es la posición de extrema izquierda y 1 la de extrema derecha como el de la Figura 16.8, se representa cómo se distribuyen los votantes según su posición ideológica (función de densidad), observándose que los partidos tenderán a aproximar su ideología a la del votante mediano, I_M, que será la que dé la victoria en las elecciones.

[19] Fue elaborado por Downs (1957), pero se le añade el apellido Hotelling porque su funcionamiento y resultados son muy similares a los del modelo de localización espacial de Hotelling (1929).

 © 2026 Manuel Cabado García

Figura 16.8: Resultado del modelo Downs-Hotelling



No obstante, el modelo Downs-Hotelling suscita una serie de limitaciones[20]:

- La votación se realiza sobre una variable unidimensional, la ideología, cuando en la realidad esto no ocurre ya que se suele votar sobre variables multidimensionales[21].
- Pueden existir partidos cuyo objeto no sea la maximización de votos.
- No todos los posibles votantes se comportan de la misma manera (por ejemplo, por la abstención), de modo que el votante mediano efectivo puede diferir del teórico.
- Las preferencias pueden no ser unimodales.

Con respecto al último punto, en el caso de que las preferencias fueran bimodales se daría que:

- Las preferencias de los electores se polarizarían en dos o más campos.
- Los partidos tendrían posiciones muy diferenciadas.
- La democracia sería mucho más inestable y un cambio de Gobierno podría desembocar en una crisis social.

[20] El modelo de Downs-Hotelling muestra cómo, en un sistema bipartidista con votación unidimensional en el espectro ideológico, los dos partidos tenderán a adoptar una ideología de centro, proponiendo un programa lo más similar posible a la ideología del votante mediano (principio de diferenciación mínima).
Limitaciones del modelo:

- Cuando existen más de dos partidos "voto útil o estratégico".
- No solo cuenta ideología del partido (variables multidimensionales y problema de preferencias cíclicas).
- Abstención y el votante mediano no da victoria (votante mediano "efectivo" (los que votan) difiere del "real" (si vota toda la población)). Ejemplo: burócratas y funcionarios. Conclusión: Con más de dos partidos, es difícil determinar si gana o no votante mediano. Probablemente, los partidos se sitúen en torno a él (ideología de "centro").

[21] Por ejemplo, si los individuos a la hora de votar tomaran en cuenta no solo el nivel del presupuesto sino también la estructura del mismo, es decir, no solo cuánto se gasta sino también en qué se gasta. Cada alternativa sería entonces un vector en el que se especificaría cuánto se gasta en cada uno de los programas de gasto y cómo se financia el gasto.

- Se podría desembocar en un sistema multipartidos cuando las preferencias sean multimodales.
- Entonces, la estabilidad de los gobiernos no depende, por tanto, del número de partidos, sino de las preferencias de los votantes.

Votantes

Respecto a la actuación de los votantes, este modelo estudia también el fenómeno de la decisión de voto, acentuando la influencia provocada por el conocimiento imperfecto y la información costosa. Dado que los agentes son racionales, decidirán si votan, y cuánta información recaban, comparando el beneficio esperado de su voto con su coste.

Ese beneficio esperado depende de lo que gana el votante si vence su opción y de la probabilidad de que su voto sea decisivo. La escasa probabilidad de que un voto sea decisivo y la indivisibilidad de los beneficios hacen que no exista estímulo para lograr una buena información. Este escaso estímulo para informarse es la llamada ignorancia racional del votante.

Por consiguiente, la participación electoral aumenta si:

- La probabilidad de vencer tiende a igualarse.
- Los programas divergen más, ya que aumenta el ratio de beneficios/costes.
- Los individuos con un mayor nivel de renta y educación tendrán más incentivos a votar, dado que disminuirán los costes y aumentarán los beneficios derivados del voto.

Por otro lado, la Teoría del Ciclo Político considera que los votantes valoran las actuaciones recientes del Gobierno –como aumentos del gasto público o bajadas de impuestos– a la hora de decidir el sentido de su voto. Esto supone que existiría cierta miopía o comportamiento no plenamente racional, lo que, además contribuiría a explicar los ciclos económicos (por las expansiones del gasto asociadas a la proximidad de las elecciones).

Grupos de interés

Los grupos de interés son asociaciones de individuos que se unen con un objetivo común: influir en las decisiones públicas para que les sean favorables. Un ejemplo clásico son los sindicatos, que buscan mejorar las condiciones laborales para sus afiliados y no afiliados.

Para lograr sus metas, estos grupos utilizan diversos instrumentos de manera simultánea:

- Intercambio de favores: Pueden intercambiar votos por programas políticos o financiar campañas a cambio de favores.

 © 2026 Manuel Cabado García

- Influencia directa: Buscan mejorar la información que llega a los legisladores, burócratas y público para posicionar su causa.

Los efectos económicos de la actividad de estos grupos dependen de sus objetivos. Generalmente, los más activos y poderosos no son altruistas, sino aquellos con intereses económicos privados. Sus acciones, a menudo, perjudican el bienestar general de la sociedad.

La teoría de los grupos de interés de Mancur Olson El economista Mancur Olson (1965) analizó los grupos de interés desde la perspectiva de los bienes públicos. Para él, la actividad de un grupo de interés tiene carácter de bien público para todos los que comparten ese interés, ya que si el grupo logra promover su causa, todos se benefician, independientemente de si contribuyeron o no:

- El problema del free-rider: La afiliación a un grupo de interés tiene un coste (tiempo, dinero, etc.), pero el beneficio es gratuito para quienes no se afilian. Esto crea un incentivo para que los individuos se conviertan en "usuarios gratuitos" (free-riders).
- Grupos grandes vs. grupos pequeños: la tendencia a ser free-rider es mayor en los grupos grandes, porque la aportación de cada miembro apenas cambia el resultado y los costes de organización son altos, y también en los que defienden a las rentas bajas, que soportan peor el coste de participar. En cambio, los grupos pequeños y de rentas elevadas están mejor organizados y son más efectivos. A estos se les conoce como Grupos de Interés Directamente Influyentes, y Olson denomina “privilegiados” a los grupos pequeños y “latentes” a los grandes.

El triángulo de hierro y los buscadores de rentas Un concepto relacionado es el "triángulo de hierro", una alianza entre políticos, burócratas y grupos de presión que actúan como "buscadores de rentas". Su objetivo es obtener beneficios económicos no a través de la producción, sino a través de la manipulación del entorno político y regulatorio. Esta dinámica perjudica a los votantes (que representan el interés general) porque:

- Son menos numerosos que los votantes y, por lo tanto, tienen una mayor posibilidad de organizarse y ejercer presión para la asignación de Gasto Público en su beneficio.
- Los votantes, al ser un grupo tan numeroso, enfrentan el problema de la acción colectiva, haciendo que su organización para contrarrestar la influencia de los buscadores de rentas sea extremadamente difícil.

Ampliaciones y otras teorías

A su vez, existen ampliaciones que estudian fenómenos como:

- Voto útil y comportamiento estratégico: la teoría original supone que los votantes eligen a su candidato preferido. Sin embargo, en la realidad, el voto útil ocurre cuando un votante prefiere a un candidato pero vota por otro que tiene más probabilidades de ganar. Esto puede hacer que el resultado final se desvíe del punto del votante mediano.
- La abstención, el modelo del votante mediano supone que todos los individuos participan, pero la abstención rompe esa premisa. Al estudiar la abstención, se reconoce que el votante mediano del total de la población puede ser diferente al votante mediano de quienes realmente votan. Por ejemplo, si los votantes más extremos (los de los polos ideológicos) están más motivados para participar, el resultado final de la elección se alejará del centro político, contradiciendo la predicción del teorema.
- Los objetivos privados de los políticos, los políticos no siempre se mueven por la voluntad popular. Muchos modelos incluyen los objetivos personales de los políticos (como la reelección, el poder o la búsqueda de rentas) como un factor clave que moldea las políticas públicas.
- La votación probabilística, donde la existencia de incertidumbre dentro del proceso de decisión electoral –donde los votantes no podrán estimar con exactitud la situación de los candidatos en el espacio político– hace necesaria la incorporación explícita de este factor en el modelo del votante. Por un lado, los individuos en el momento de votar, no conocen el grado de realización de las promesas electorales del candidato en el caso de que llegue al poder. Por lo tanto, existe un grado de aleatoriedad derivado del desfase temporal entre la celebración de las elecciones y la aplicación de políticas por el partido ganador.
- Voto significativo, se refiere a la capacidad de un votante para influir en el resultado final. En un sistema de votación mayoritaria, un voto tiene un peso significativo si puede romper un empate o cambiar la victoria de un candidato a otro. Esta es una variable que los votantes racionales consideran.
- La relación entre el nivel de renta y las preferencias políticas. Por ejemplo, los votantes de bajos ingresos suelen favorecer políticas de redistribución de la riqueza o mayor gasto social, mientras que los votantes de altos ingresos tienden a preferir recortes de impuestos y menor intervención del Estado.
- Ventana de Overton. La ventana de Overton encaja en estas teorías al explicar cómo las reglas de votación y los mecanismos de elección se ven afectados por la percepción pública. Esta ventana representa el rango de ideas políticas aceptables para el público en un momento dado, desde lo "impensable" hasta la "política vigente". En las democracias, las reglas de votación no operan en un vacío. Los candidatos y grupos de interés manipulan la ventana de Overton para que sus propuestas se consideren aceptables por la mayoría. Así, por muy bien diseñado que esté un mecanismo de de-

 © 2026 Manuel Cabado García

cisión, el resultado final siempre estará influenciado por lo que la sociedad considera posible, desplazando el debate y el voto hacia un conjunto limitado de opciones.

16.6. Conclusiones ⏳ 30"

Como conclusión, el Sector Público se enfrenta a un desafío inherente: nunca dispone de toda la información necesaria para agregar perfectamente las preferencias de la ciudadanía. Esta limitación fundamental puede llevar a fallos del sector público, resultando en un aumento del gasto o del tamaño del Estado por encima de un nivel eficiente.

Por ello, las sociedades no buscan un mecanismo de decisión perfecto, sino que recurren a Reglas de Elección Colectiva sencillas. Estas reglas, si bien son imperfectas, se complementan con límites constitucionales que actúan como contrapesos para proteger a la sociedad de los riesgos de la ineficiencia y de las tiranías de la mayoría o la minoría.

© 2026 Manuel Cabado García

Apéndice

16.A.1. El modelo de Salop como extensión de Downs-Hotelling: más de dos partidos

El modelo de Downs-Hotelling desarrollado en el cuerpo del tema explica la convergencia de los partidos hacia el votante mediano bajo un supuesto muy concreto: un sistema bipartidista. Esta restricción no es solo una simplificación cómoda, sino una condición necesaria para que el resultado se sostenga. Osborne (1995) recoge la prueba formal de que, cuando hay tres o más partidos compitiendo sobre la misma línea ideológica, el modelo generalmente no tiene equilibrio de Nash en estrategias puras: para casi cualquier distribución de las preferencias de los votantes, no existe ninguna configuración de posiciones en la que ningún partido tenga incentivo a desplazarse. La intuición es que, con dos partidos, cada uno se enfrenta al otro desde ambos lados de la línea, lo que empuja a los dos hacia el centro. Con tres o más partidos, en cambio, siempre queda al menos uno situado entre otros dos que puede mejorar su resultado desplazándose ligeramente, lo que rompe cualquier posición de equilibrio antes de que llegue a estabilizarse. El propio Osborne concluye que extender directamente el modelo lineal a más de dos partidos aporta poca capacidad explicativa sobre lo que ocurre realmente en sistemas multipartidistas.

El problema no es exclusivo de la política. Salop (1979) se había encontrado con una dificultad estructuralmente idéntica en la competencia monopolística: en el modelo de Hotelling, las empresas situadas en los extremos de la línea disfrutan de una ventaja artificial —solo compiten por un lado—, lo que distorsiona el análisis de la entrada de nuevas empresas. Su solución fue sustituir la línea por una circunferencia: al no haber extremos, todas las posiciones del espacio son simétricas entre sí, y el número de empresas que pueden coexistir en equilibrio depende únicamente de la relación entre el coste fijo de entrada y el coste de "desplazamiento" que asumen los consumidores para llegar hasta la empresa más cercana en el círculo.

Peeters, Saran y Yüksel (2016) trasladan esta misma solución geométrica a la competencia electoral. En su modelo, el espacio ideológico ya no es un segmento con dos extremos, sino una circunferencia unitaria sobre la que se distribuyen tanto los votantes como las posibles posiciones de los partidos. Sobre este espacio, los autores caracterizan los equilibrios de Nash en estrategia pura bajo dos supuestos de comportamiento del votante: votación sincera (cada votante apoya siempre al partido ideológicamente más cercano) y votación estratégica (el votante puede apoyar a un partido distinto de su preferido si así evita que gane el que menos le convence). En ambos casos exigen además que el equilibrio sea "a prueba de deserción": ningún subgrupo de partidos debe tener incentivo a reorganizarse conjuntamente abandonando sus posiciones actuales.

Sus resultados muestran que, bajo votación sincera, estos equilibrios sostienen confi-

 © 2026 Manuel Cabado García

guraciones de dos o tres partidos efectivos, mientras que bajo votación estratégica solo se sostienen una o dos. Este resultado es parcialmente consistente con la Ley de Duverger (1951), que predice una tendencia al bipartidismo bajo la regla de pluralidad, pero va más allá: identifica con precisión en qué condiciones puede sobrevivir en equilibrio un tercer partido, algo que la representación lineal de Downs-Hotelling no permite explicar por la propia inexistencia de equilibrio que demuestra Osborne.

La circunferencia no es, por tanto, un simple giro geométrico sobre el modelo original, sino la vía formal que permite responder a una pregunta que el modelo lineal deja sin resolver: cómo pueden coexistir de forma estable más de dos partidos cuando los votantes mantienen preferencias unimodales sobre el espacio ideológico.

16.A.2. La paradoja liberal de Amartya Sen

Es imposible ser liberal y paretiano a la vez: no es posible respetar mínimamente la libertad de los individuos y al mismo tiempo alcanzar la optimalidad paretiana.

La paradoja del liberal paretiano, paradoja liberal o paradoja de imposibilidad de Sen es un resultado de imposibilidad propuesto por Amartya Sen (1970), en la línea del teorema de Arrow. Afirma que, si la regla de elección social admite cualquier combinación de preferencias individuales, es imposible respetar a la vez el criterio del óptimo de Pareto y un principio de "libertad mínima", según el cual al menos dos individuos deben ser decisivos sobre algún asunto de su esfera personal, de modo que su preferencia entre esas dos alternativas se convierta en la preferencia social[22].

El aspecto más polémico es, por un lado, la contradicción con la noción del liberalismo económico, de que el mecanismo del mercado es suficiente para producir un óptimo de Pareto y por otro, la sugerencia de que la capacidad de elección y libertad deben ser el rasgo definitorio de ese mecanismo de mercado, en lugar del criterio de bienestar económico de la economía del bienestar.

La paradoja liberal cuestiona un axioma concreto de Arrow —la optimalidad paretiana débil— frente a la libertad individual, pero mantiene intacto el marco informativo del welfarismo: sigue evaluando el bienestar social únicamente a partir de niveles de utilidad. Sen (1979) fue más allá y cuestionó ese propio marco: ¿es la utilidad, entendida como satisfacción subjetiva de preferencias, el espacio informativo adecuado para juzgar el bienestar de una persona? Su respuesta fue sustituir la utilidad por los funcionamientos, los distintos "seres y haceres" que una persona logra efectivamente alcanzar, desde funcio-

[22] El ejemplo canónico de esta paradoja, presente ya en el propio artículo original de Sen, es el de dos individuos, "Lascivo" y "Mojigato", que deben decidir quién lee un único ejemplar de *El amante de Lady Chatterley*. Mojigato preferiría que el libro no lo leyera nadie, pero, si tiene que leerlo alguien, prefiere leerlo él mismo antes que verlo en manos de Lascivo. Lascivo disfrutaría leyéndolo, pero disfrutaría aún más viendo a Mojigato obligado a leerlo. Respetar la libertad de cada uno para decidir sobre su propia lectura conduce a un resultado que ambos consideran peor que la alternativa que señalaría el criterio de Pareto, lo que ilustra la incompatibilidad entre libertad individual y optimalidad paretiana.

namientos elementales como estar bien nutrido hasta otros más complejos como participar en la vida de la comunidad. Dos individuos con la misma cesta de bienes pueden lograr funcionamientos muy distintos si uno de ellos padece, por ejemplo, una enfermedad que le impide aprovecharlos con normalidad.

El conjunto de capacidad de una persona es el conjunto de todas las combinaciones de funcionamientos que le resulta factible alcanzar dada su dotación de bienes y sus circunstancias. Mide, por tanto, no lo que la persona efectivamente hace o es, sino la libertad real de la que dispone para elegir entre distintas formas de vida valiosas —una distinción que separa el enfoque de Sen tanto del welfarismo, que solo mira el resultado final, como de la disponibilidad de bienes primarios de Rawls (1971), que no tiene en cuenta que distintos individuos convierten los mismos recursos en funcionamientos con distinta eficacia—. Esta idea tuvo una consecuencia práctica de primer orden: si lo relevante no es la renta ni la utilidad, sino la capacidad real de alcanzar funcionamientos valiosos, los índices de bienestar deben construirse sobre indicadores multidimensionales —salud, educación, participación política— y no sobre una única variable monetaria agregada, como hace el Índice de Desarrollo Humano del Programa de las Naciones Unidas para el Desarrollo, elaborado desde 1990 con la colaboración directa de Sen.

16.A.3. Análisis coste-beneficio

El análisis coste-beneficio (ACB) es una técnica de asignación de recursos con un recorrido mucho más largo de lo que su uso extendido en la segunda mitad del siglo XX podría sugerir. Su origen teórico se remonta al ingeniero francés Dupuit (1844), que ya proponía valorar una obra pública sumando la disposición a pagar de sus usuarios. Esta idea se trasladó pronto a la práctica administrativa: en Estados Unidos, la Rivers and Harbors Act de 1902 exigió justificar los proyectos del Cuerpo de Ingenieros del Ejército por sus beneficios, y la Flood Control Act de 1936 formalizó el criterio de aprobar un proyecto siempre que sus beneficios, recayeran sobre quien recayeran, superasen los costes estimados. El aparato teórico que sostiene esta regla se consolidó poco después con el criterio de compensación de Kaldor (1939) y Hicks (1939), que es precisamente el que fundamenta el ACB tal y como se describe más abajo. Toda esta tradición, con más de 120 años de recorrido a esas alturas, quedó recogida en la gran revisión clásica del campo, Prest y Turvey (1965), que sistematizó sus fundamentos y sus principales problemas prácticos justo antes de que el ACB se generalizara como herramienta habitual de evaluación de políticas públicas.

El ACB parte de la existencia de diferencias entre los costes/beneficios privados y los sociales. Esta discrepancia lleva a que la asignación privada de los recursos, basada en la maximización de beneficios particulares, no sea óptima.

Así, la función del ACB es la de considerar todos los costes y beneficios, privados y

 © 2026 Manuel Cabado García

públicos, derivados de la provisión de un bien o servicio.

El ACB, tal y como está planteado, tiene su base en el criterio de compensación de Kaldor-Hicks: al decidir sobre la conveniencia de un proyecto, se analiza el valor presente de beneficios y costes sociales, de manera que se aceptará el proyecto si los beneficios son mayores que los costes, aunque no recaigan sobre las mismas personas.

El ACB es especialmente útil para la asignación de recursos públicos, dado que a menudo los bienes y servicios ofrecidos por este sector tienen externalidades importantes.

El problema más grave del ACB proviene de la Teoría del second best. Sea una economía de mercado, con unos precios que no reflejan todos los costes y beneficios, pero que son los que emplean los agentes. Si el sector público utiliza precios sombra representativos de dichos costes y beneficios, la economía se habrá acercado en un sector (el público) al óptimo paretiano, pero, según Lipsey y Lancaster (1956), un movimiento en dirección al óptimo dejando aún imperfecciones en otros sectores, no implica necesariamente una mejor situación que la de partida, e incluso puede empeorarla.

16.A.4. La tiranía de la mayoría y la tiranía de la minoría: el diseño de las reglas de votación

Como se ha visto al analizar la regla de la unanimidad, tanto la tiranía de la mayoría —que ya preocupaba a Wicksell, aunque la expresión procede de Tocqueville y Mill— como la tiranía de la minoría —atribuida a Baumol— son riesgos inherentes a cualquier regla de decisión colectiva. Buchanan y Tullock (1962) formalizan este dilema en su cálculo del consentimiento: a medida que aumenta la mayoría exigida para aprobar una decisión, disminuyen los costes externos, el riesgo de que una mayoría imponga su voluntad a una minoría, pero aumentan los costes de decisión, el tiempo y los recursos necesarios para alcanzar el acuerdo, y el riesgo de que una minoría bloquee la decisión. La regla óptima, ya vista anteriormente, es la que minimiza la suma de ambos costes, y no tiene por qué ser la misma para todas las decisiones.

Las democracias constitucionales aplican precisamente este cálculo, calibrando el umbral de mayoría según la gravedad de la decisión que ha de aprobarse. En España, el artículo 167 de la Constitución exige una mayoría de tres quintos en cada Cámara para una reforma constitucional ordinaria, ampliable a dos tercios del Congreso si no hay acuerdo inicial. El artículo 168, en cambio, exige dos tercios en ambas Cámaras, disolución de las Cortes, nuevas elecciones y ratificación por dos tercios de las Cámaras resultantes, además de referéndum obligatorio, para reformar el Título Preliminar o los derechos fundamentales. El umbral crece con la irreversibilidad de la decisión y con el riesgo de tiranía de la mayoría que supondría adoptarla con una mayoría simple.

El riesgo simétrico, la tiranía de la minoría, aparece cuando ese mismo umbral elevado se convierte en poder de veto. En el Consejo de la Unión Europea, una minoría de

bloqueo formada por al menos cuatro Estados miembros que representen más del 35 % de la población de la Unión puede impedir la aprobación de una decisión por mayoría cualificada, aunque el resto de Estados esté de acuerdo. La misma regla que protege a las minorías de una imposición mayoritaria es, al mismo tiempo, la que les permite paralizar la voluntad de la mayoría.

16.A.5. Funciones de elección social no manipulables en dominios restringidos

El teorema de Gibbard-Satterthwaite muestra que, en un dominio universal de preferencias y con tres o más alternativas, toda regla de elección que no sea dictatorial es manipulable. El resultado deja abierta, sin embargo, una pregunta constructiva: si se restringe el dominio de preferencias admisibles, ¿es posible recuperar mecanismos de elección social eficientes y no manipulables? Cuando los individuos tienen preferencias unimodales sobre una única dimensión, el mismo supuesto que sostiene el teorema del votante mediano de Black (1948) ya visto en el cuerpo del tema, la respuesta es afirmativa.

Una función de elección social que selecciona siempre la alternativa preferida por el votante mediano resulta no manipulable: si un individuo declara unas preferencias distintas de las suyas, en el mejor de los casos consigue desplazar la mediana hacia una alternativa que le gusta menos que la mediana original, nunca hacia una que le guste más. Moulin (1980) generaliza este resultado: toda función de elección social que satisface anonimato, unanimidad y no manipulabilidad en un dominio de preferencias unimodales es una mediana generalizada. Si n es el número de individuos, el mecanismo fija de antemano $n-1$ "votos fantasma" —posiciones ficticias, independientes de los votos reales— y elige la mediana del conjunto formado por los n picos declarados más esos $n-1$ votos fantasma.

La posición de los votos fantasma determina el mecanismo resultante. Situándolos todos en el extremo inferior del dominio se elige el menor de los picos declarados, y situándolos todos en el superior, el mayor. Repartiéndolos entre ambos extremos se obtiene cualquier pico intermedio: con $n-k$ votos fantasma abajo y $k-1$ arriba se elige el k-ésimo pico ordenado de menor a mayor. En particular, con un número impar de individuos, situar la mitad de los votos fantasma en cada extremo recupera la mediana simple de los votos reales, la regla más habitual en la práctica. El resultado muestra, en definitiva, que restringir el dominio de preferencias no es solo una simplificación técnica: es la vía que permite escapar de la imposibilidad de Gibbard-Satterthwaite sin renunciar ni a la eficiencia ni a la ausencia de manipulación.

 © 2026 Manuel Cabado García

16.A.6. El mecanismo de Vickrey-Clarke-Groves y la revelación de preferencias

Wicksell, como se ha visto en el cuerpo del tema, propuso la regla de la unanimidad para que nadie tuviera que financiar un bien público que no valora, pero esa misma regla invita a ocultar la propia valoración para pagar menos o arrancar concesiones. La teoría del diseño de mecanismos ofrece una solución distinta al mismo problema: en vez de exigir el acuerdo de todos, diseña las reglas del juego de forma que decir la verdad sea la mejor estrategia de cada individuo, sin necesidad de que este conozca ni confíe en el comportamiento de los demás.

El ejemplo más sencillo es la subasta de Vickrey (1961): una subasta a sobre cerrado en la que cada postor presenta su oferta sin conocer las de los demás, gana quien puja más alto, pero paga el importe de la segunda oferta más alta. Si v_1 es la verdadera valoración del postor 1 y r_2 la oferta declarada por su rival, la ganancia del postor 1 si gana es $v_1 - r_2$, una cantidad que no depende de lo que él mismo declare mientras siga ganando: declarar su valoración real es, por tanto, una estrategia dominante, sin ningún incentivo a mentir.

El mecanismo de Vickrey-Clarke-Groves (VCG) traslada esta misma idea a la provisión de un bien público. Cada individuo i declara su valoración $r_i(x)$ de cada nivel x del bien público, y se elige el nivel que maximiza la suma de las valoraciones declaradas. Para que declarar la verdad siga siendo la estrategia dominante de cada individuo, el mecanismo le cobra un impuesto —el llamado impuesto de Clarke— igual a la diferencia entre el bienestar máximo que alcanzarían los demás si ese individuo no participase y el bienestar que efectivamente alcanzan con su participación: solo pagan, por tanto, los individuos cuya presencia cambia realmente el resultado. El mecanismo VCG selecciona siempre el nivel eficiente del bien público y hace que decir la verdad sea una estrategia dominante para cada individuo, pero no está exento de limitaciones: en general genera un superávit presupuestario que no puede devolverse a los individuos sin alterar sus incentivos, por lo que la asignación final no es eficiente en el sentido de Pareto, no es inmune a la colusión entre varios de ellos, y exige revelar la valoración completa de cada alternativa posible, algo que puede resultar inabordable en problemas de gran escala.

16.A.7. Funciones de bienestar social tipo CES e índices de desigualdad

La función de bienestar utilitarista y la maxi-min de Rawls, ya vistas en el cuerpo del tema, no son alternativas aisladas: son los dos casos extremos de una familia paramétrica más amplia, las funciones de bienestar colectivo de elasticidad constante (CES),

$$W^{\rho}(u) = \left(\sum_{i \in N} u_i^{1-\rho} \right)^{\frac{1}{1-\rho}} \quad \text{si } \rho \neq 1, \qquad W^{\rho}(u) = \sum_{i \in N} \ln(u_i) \quad \text{si } \rho = 1$$

donde $\rho \geq 0$ mide el grado de aversión social a la desigualdad: $\rho = 0$ devuelve la función utilitarista, $\rho \to \infty$ devuelve la maxi-min de Rawls (1971), y el caso intermedio $\rho = 1$ se corresponde con la función de Nash, $W^N(u) = \prod_{i \in N} u_i$, la misma especificación tipo Cobb-Douglas ya presentada en el cuerpo como función de Bernoulli-Nash.

A cada función de esta familia que respeta el principio de Pigou-Dalton —una transferencia de utilidad de un individuo más rico a otro más pobre que no invierte su posición relativa nunca puede reducir el bienestar social— puede asociársele un índice de desigualdad. Las funciones CES generan, en concreto, los índices de Atkinson (1970), con el mismo parámetro ρ interpretado ahora como el grado de aversión a la desigualdad del índice. Junto a ellos, el índice de Gini —definido a partir del área entre la curva de Lorenz y la línea de equidistribución perfecta— se corresponde con una función de bienestar que pondera más la utilidad de los individuos peor situados, aunque no pertenece a la familia CES.

16.A.8. La ventana de Overton: cómo cambia lo políticamente posible

Los modelos de elección colectiva vistos en este Apéndice y en el cuerpo del tema —votante mediano, funciones de bienestar social, mecanismos de revelación de preferencias— comparten un supuesto que rara vez se cuestiona: el espacio de alternativas sobre el que se vota está dado de antemano. Sin embargo, ese espacio no es fijo, sino que se desplaza con el tiempo. El marco de la ventana de Overton, formulado por Joseph Overton y desarrollado tras su muerte por Russell (2006), describe precisamente ese fenómeno: el rango de políticas que una sociedad considera aceptables en un momento dado, desde lo impensable hasta la política vigente, y cómo ese rango se desplaza a lo largo del tiempo hasta que una opción antes marginal termina aprobada por mayoría. Cinco episodios históricos ilustran cómo ha operado este desplazamiento.

La abolición de la esclavitud A comienzos del siglo XIX, abolir la esclavitud era una posición radical incluso en los círculos más progresistas. William Lloyd Garrison fundó en 1833 la American Anti-Slavery Society, y en las décadas siguientes campañas como la suya fueron desplazando la opinión pública del norte de Estados Unidos, de la tolerancia hacia el rechazo moral de la esclavitud, hasta que la abolición pasó de ser una posición marginal a convertirse en política aprobada en un país tras otro (Drescher, 2009).

El sufragio femenino A mediados del siglo XIX, conceder el voto a las mujeres era una propuesta marginal incluso entre los círculos reformistas más avanzados de su tiempo. El desplazamiento hacia su aceptación no siguió un único camino: Nueva Zelanda se convirtió en 1893 en el primer país autogobernado del mundo en reconocer el voto femenino,

 © 2026 Manuel Cabado García

tres décadas antes que la mayoría de las democracias occidentales. En el Reino Unido, la Women's Social and Political Union de Emmeline Pankhurst, fundada en 1903, optó por la confrontación directa —manifestaciones, rotura de escaparates, huelgas de hambre en prisión—, hasta el punto de que Emily Davison murió en 1913 al lanzarse al paso del caballo del rey durante el Derby de Epsom (Purvis, 2002). En Estados Unidos, en cambio, Carrie Chapman Catt condujo desde 1916 el llamado *Winning Plan*, una estrategia institucional de coordinación estado por estado y presión legislativa sostenida sobre el Congreso (Baker, 2002). Confrontación y negociación institucional operaron así como dos estrategias distintas hacia el mismo desplazamiento, que se generalizó en la mayoría de las democracias occidentales en las primeras décadas del siglo XX.

Los derechos de los trabajadores Durante buena parte del siglo XIX, la protección laboral —jornadas máximas, edad mínima de acceso al empleo, condiciones de seguridad— no formaba parte del debate político: el trabajo infantil en fábricas y minas se aceptaba como un hecho normal de la vida económica. La Revolución Industrial expuso estas condiciones y dio inicio a una campaña de reforma que tardó décadas en traducirse en ley: en Estados Unidos, dos intentos de prohibición federal del trabajo infantil (1916 y 1919) fueron anulados por el Tribunal Supremo, y no fue hasta la Fair Labor Standards Act de 1938 —ya con el país sumido en la Gran Depresión— cuando la protección laboral moderna, incluida esa prohibición, se consolidó como política vigente (Schmidt, 2010).

El matrimonio igualitario A finales del siglo XX, el reconocimiento legal de las parejas del mismo sexo era una posición marginal incluso en las democracias más consolidadas. España es un ejemplo especialmente claro de lo rápido que puede desplazarse la ventana: una encuesta de junio de 2004 ya mostraba un 66 % de apoyo social al matrimonio igualitario, un año antes de que la Ley 13/2005 convirtiera al país en el tercero del mundo en reconocerlo, con una votación parlamentaria de 187 votos a favor. El proceso se repitió, con ritmos distintos, en buena parte de Europa durante la misma década (Calvo, 2010).

La revolución keynesiana El desplazamiento no afecta solo a los derechos civiles, también a las propias ideas económicas que sustentan la política pública. Antes de la Gran Depresión, la intervención estatal en la economía se consideraba peligrosa bajo el consenso de laissez-faire dominante. Tras 1929, la obra de Keynes desplazó ese consenso: la intervención estatal pasó de ser vista como una amenaza a convertirse en la herramienta estabilizadora estándar de la política económica occidental hasta bien entrados los años setenta (Hall, 1989).

En los cinco casos, el mecanismo de elección colectiva en sí —votación por mayoría, representación parlamentaria— no cambió: lo que cambió fue el conjunto de alternativas que ese mecanismo consideraba elegibles. Esto recuerda que, por muy bien diseñada

que esté una regla de decisión colectiva, su resultado depende siempre de un espacio de alternativas que la propia sociedad redefine con el tiempo.

 © 2026 Manuel Cabado García

Bibliografía

Albi Ibáñez, E., J. M. Gónzalez-Páramo, R. M. Urbanos y I. Zubiri (2017). *Economía Pública I: Fundamentos. Presupuestos y Gastos*. Barcelona: Ariel.

Arrow, K. J. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economy 58*(4), 328–346.

Arrow, K. J. (1951). *Social Choice and Individual Values*. Yale University Press.

Atkinson, A. B. (1970). On the Measurement of Inequality. *Journal of Economic Theory 2*(3), 244–263.

Baker, J. H. (ed.) (2002). *Votes for Women: The Struggle for Suffrage Revisited*. Oxford University Press.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne and Son.

Bergson, A. (1938). A Reformulation of Certain Aspects of Welfare Economics. *The Quarterly Journal of Economics 52*(2), 310–334.

Bernoulli, D. (1738). Specimen Theoriae Novae de Mensura Sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae 5*, 175–192.

Black, D. (1948). On the Rationale of Group Decision-making. *Journal of Political Economy 56*(1), 23–34.

Borda, J.-C. d. (1781). Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 657–665.

Buchanan, J. M. y G. Tullock (1962). *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor, MI: University of Michigan Press.

Calvo, K. (2010). Reconocimiento, ciudadanía y políticas públicas hacia las uniones homosexuales en Europa. *Revista Española de Investigaciones Sociológicas 129*, 37–59.

Condorcet, M. d. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*.

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy 65*(2), 135–150.

Drescher, S. (2009). *Abolition: A History of Slavery and Antislavery*. New York: Cambridge University Press.

Dupuit, J. (1844). De la mesure de l'utilité des travaux publics. *Annales des Ponts et Chaussées 8*, 332–375.

Duverger, M. (1951). *Les Partis Politiques*. Paris: Armand Colin.

Gibbard, A. (1973). Manipulation of Voting Schemes: A General Result. *Econometrica 41*(4), 587–601.

Hall, P. A. (ed.) (1989). *The Political Power of Economic Ideas: Keynesianism across Nations*. Princeton, NJ: Princeton University Press.

Harsanyi, J. C. (1955). Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy 63*(4), 309–321.

Hicks, J. R. (1939). The Foundations of Welfare Economics. *The Economic Journal 49*(196), 696–712.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal 39*(153), 41–57.

Kaldor, N. (1939). Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *The Economic Journal 49*(195), 549–552.

Kramer, G. H. (1973). On a Class of Equilibrium Conditions for Majority Rule. *Econometrica 41*(2), 285–297.

Lipsey, R. G. y K. Lancaster (1956). The General Theory of Second Best. *The Review of Economic Studies 24*(1), 11–32.

Moulin, H. (1980). On Strategy-Proofness and Single Peakedness. *Public Choice 35*(4), 437–455.

Nash, J. F. (1950). The Bargaining Problem. *Econometrica 18*(2), 155–162.

Olson, M. (1965). *The Logic of Collective Action*. Harvard, Massachusetts: Harvard University Press.

Osborne, M. J. (1995). Spatial Models of Political Competition under Plurality Rule: A Survey of Some Explanations of the Number of Candidates and the Positions They Take. *Canadian Journal of Economics 28*(2), 261–301.

Peeters, R., R. Saran y A. M. Yüksel (2016). Strategic Party Formation on a Circle and Duverger's Law. *Social Choice and Welfare 47*(3), 729–759.

Prest, A. R. y R. Turvey (1965). Cost-Benefit Analysis: A Survey. *The Economic Journal 75*(300), 683–735.

Purvis, J. (2002). *Emmeline Pankhurst: A Biography*. New York: Routledge.

 © 2026 Manuel Cabado García

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: The Belknap Press of Harvard University Press.

Russell, N. J. (2006). An Introduction to the Overton Window of Political Possibilities. Mackinac Center for Public Policy.

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *The Bell Journal of Economics 10*(1), 141–156.

Samuelson, P. A. (1947). *Foundations of Economic Analysis*. Cambridge, MA: Harvard University Press.

Samuelson, P. A. (1954). The Pure Theory of Public Expenditure. *The Review of Economics and Statistics 36*(4), 387–389.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory 10*(2), 187–217.

Schmidt, J. D. (2010). *Industrial Violence and the Legal Origins of Child Labor*. New York: Cambridge University Press.

Sen, A. K. (1970). The Impossibility of a Paretian Liberal. *Journal of Political Economy 78*(1), 152–157.

Sen, A. K. (1979). Personal Utilities and Public Judgements: Or What's Wrong with Welfare Economics? *The Economic Journal 89*(355), 537–558.

Vickrey, W. (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance 16*(1), 8–37.

Wicksell, K. (1896). A New Principle of Just Taxation. En R. A. Musgrave y A. T. Peacock (eds.), *Classics in the Theory of Public Finance (1958)*, pp. 72–118. London: Palgrave Macmillan UK.

 © 2026 Manuel Cabado García

Tema 17

El gasto público. Razones de su crecimiento. Teoría de la burocracia y constitución fiscal. El debate sobre el tamaño del Sector Público, en lo referente al gasto.

17.1. Introducción ⏳ 1'

La Economía del Sector Público estudia la intervención del Sector Público en una economía de mercado fundamentada en programas de ingresos y gastos con el objetivo de cumplir con sus funciones

1. Asignativa, dirigida a corregir fallos de mercado.
2. Distributiva, para conseguir equidad.
3. Crecimiento y estabilización, para suavizar las fluctuaciones cíclicas.

Para llevar a cabo estas funciones, el Sector Público tiene que decidir los niveles de gasto e impuestos que deben resultar óptimos desde el punto de vista colectivo.

Concretamente, durante gran parte del siglo XX, la expansión del Estado del Bienestar fue la norma, impulsada por la creencia de que un mayor gasto público en servicios sociales era esencial para la equidad y la provisión de bienes preferentes.

Sin embargo, a partir de la crisis de los años 70 y 80, esta expansión comenzó a ser cuestionada. El estancamiento económico y el aumento de la deuda desataron una crítica a la ineficiencia del Estado y a la sostenibilidad de sus finanzas, lo que llevó a la necesidad de controles fiscales más estrictos.

No obstante, la reciente crisis de la COVID-19 ha reavivado este debate de manera drástica:

- Por un lado, la intervención sin precedentes del sector público, a través de medidas como los ERTEs, los fondos Next Generation y el programa REACT-EU, demostró la capacidad del Estado para responder a emergencias.
- Por otro lado, esta masiva movilización de recursos ha disparado la deuda pública de muchos países, como España, por encima del 100 % del PIB.

Por tanto, el objetivo de la presente exposición será el estudio del crecimiento del gasto público, con especial referencia a la teoría de la burocracia y al mecanismo corrector de la Constitución Fiscal a través del siguiente esquema:

1. Teorías explicativas del aumento del gasto público, desde una óptica de demanda y oferta.
2. Debate sobre el tamaño del Sector Público, donde se realizará una revisión histórica sobre el tamaño óptimo.
3. La Constitución fiscal, como límite al crecimiento desmedido del Sector público.

17.2. Teorías que explican las causas del aumento del Sector Público ⏳ 10'

A lo largo del siglo XX se ha observado que la mayoría de los países, a medida que han ido progresando en su desarrollo económico, han aumentado el porcentaje de la renta que dedican a usos públicos. En la literatura hacendística se han sugerido diversas teorías acerca de por qué crece el gasto público. A grandes rasgos, estas teorías se pueden dividir en dos grandes grupos: teorías de demanda y teorías de oferta.

Teorías de la Demanda

Se comenzará el análisis con las teorías de la demanda.

Leyes de Wagner

Las leyes de Wagner (Wagner, 1890)[1], revisadas por Bird (1971), explican el crecimiento del Gasto Público como consecuencia de dos leyes:

1. Reestructuración de la Sociedad: la transformación de una sociedad tradicional en una sociedad industrializada moderna conlleva la sustitución de actuaciones privadas

[1] Para más información sobre Wagner y su obra, consultar Jaén García (2004)

 © 2026 Manuel Cabado García

–institución de familias o grupos sociales– por actuaciones públicas. En sociedades avanzadas aparecen nuevas necesidades (demanda de bienes públicos y el mantenimiento de la ley y el orden) y la necesidad de corregir externalidades y monopolios naturales.

2. Aparición de nuevas necesidades (públicas): el aumento de la renta per cápita estimula el consumo de ciertos bienes y servicios provistos por el Sector Público –educación, sanidad, carreteras...– cuya demanda se incrementa más que proporcionalmente, es decir, con una elasticidad-renta superior a la unidad, por lo que aumenta la proporción del PIB que se dedica a ellos. Por consiguiente, estos bienes tendrían el carácter de bienes de lujo —el reverso, en términos de reparto del presupuesto, de la Ley de Engel (Engel, 1857), que muestra cómo la proporción de la renta dedicada a bienes de primera necesidad, como la alimentación, desciende a medida que la renta aumenta.

Sin embargo, la teoría de Wagner suscitó críticas como la de Timm (1961), que señala su carácter valorativo, ya que Wagner parte de que la actuación pública es preferible a la privada y la necesidad de intervención es discutible, por lo que la teoría descansa en juicios de valor. Por otro lado, no se explica a partir de qué momento el gasto público tiende a crecer ni las variables que han intervenido en el proceso.

Transformaciones de Musgrave

Al igual que Wagner, Musgrave (1959) señala que el desarrollo económico de una sociedad genera una serie de transformaciones estructurales que tienden a elevar el gasto público:

- Transformaciones demográficas: la reducción de la natalidad y el aumento de la esperanza de vida modifican la pirámide poblacional, presionando al alza partidas como las pensiones o la sanidad.
- Transformaciones tecnológicas: los avances técnicos pueden requerir la presencia del sector público en ámbitos antes inexistentes —el desarrollo del automóvil exigió, por ejemplo, una red pública de carreteras, del mismo modo que el despliegue de la banda ancha ha exigido la extensión pública de la fibra óptica a las zonas rurales para evitar la brecha digital.
- Transformaciones urbanísticas: la mayor urbanización exige equipamientos —zonas verdes, saneamiento, redes de recarga para vehículos eléctricos— innecesarios en un entorno rural.

Efecto desplazamiento

Por su parte, en un estudio para Reino Unido Peacock y Wiseman (1961) señalan que el gasto público aumenta de forma discontinua como consecuencia de conmociones o per-

turbaciones sociales –guerras o crisis– o naturales –catástrofes– tras las cuales el gasto no retorna a su magnitud inicial. Un ejemplo reciente a escala global fue la pandemia de la Covid-19.

En este proceso identifican tres efectos:

1. Desplazamiento: ante una crisis o "conmoción social", los contribuyentes están dispuestos a aceptar niveles impositivos que no habrían aceptado en condiciones normales, lo que permite financiar un mayor gasto público, produciéndose un efecto desplazamiento. Con la vuelta a la normalidad, la tolerancia a impuestos más elevados no regresa a su nivel previo y el gasto público continúa en un nivel superior al inicial, ya que ha posibilitado el acceso a niveles de gasto público deseados pero anteriormente reprimidos por la resistencia a un aumento de la presión fiscal.
2. Inspección: las perturbaciones sociales hacen que los ciudadanos descubran necesidades sociales y ámbitos en los que la actuación del Sector Público era insuficiente, por lo que están dispuestos a soportar una mayor carga fiscal para atenderlos.
3. Concentración: ante dichas perturbaciones se produce un transvase de funciones de las administraciones locales a la Central por considerar que esta actúa de manera más eficaz.

Redistribución de la renta

Por otro lado, autores como Downs (1957) o Romer y Rosenthal (1979) plantean que el gasto público en una sociedad democrática tiende a crecer por la competencia entre los partidos políticos por los votos. En un contexto de distribución desigual de la renta, el votante mediano, con una renta inferior a la renta media, demandará políticas de redistribución, por lo que los partidos políticos tenderán a aproximar sus programas electorales a las preferencias del votante mediano.

Grupos de interés

Adicionalmente, Buchanan y Tullock (1962) plantean que los grupos de interés –caracterizados por ser pequeños y organizados, con intereses comunes y fuerte preferencia y beneficios de la consecución de sus intereses y objetivos– tratarán de modificar la actuación del Sector Público hacia sus intereses particulares, aumentando el gasto público en programas de naturaleza cuasiprivada, pero que son financiados por toda la población.

Ilusión fiscal

Finalmente, la teoría de la ilusión fiscal de Buchanan y Wagner (1977) explica que el aumento del gasto público se debe a que los ciudadanos están bajo ilusión fiscal, es decir, consideran que el coste del gasto público en términos de impuestos es inferior al coste

 © 2026 Manuel Cabado García

efectivo, debido a los costes de información y la complejidad del sistema impositivo y fiscal.

Teorías de Oferta y Demanda

Por otro lado, antes de pasar a analizar las teorías de oferta, se analizará como caso híbrido particular la teoría de la enfermedad de Baumol (1967) que explicaría el crecimiento del gasto público debido al efecto de los precios relativos crecientes y la existencia de una demanda con una reducida elasticidad-precio.

Así, parte de considerar que la mayor parte de los servicios prestados por el sector público son intensivos en trabajo, como podría ser la educación o la justicia:

1. Como consecuencia, los aumentos de productividad en el sector público son menores que en el sector privado, en el que es más fácil aumentar la productividad sustituyendo trabajo por capital.
2. Así, la productividad crece más despacio en el sector público. Sin embargo, los salarios del sector público crecen a un ritmo similar a los del sector privado.
3. Derivado de los dos puntos anteriores, el precio relativo de los bienes y servicios públicos aumenta progresivamente respecto a los privados. Esto es lo que se conoce como "enfermedad de Baumol".[2]

No obstante, el aumento del precio relativo no es suficiente, en principio, para explicar el aumento del gasto público, pues podría ocurrir que, al aumentar el precio, la demanda disminuyese y el gasto total incluso se redujese. Sin embargo, Baumol señala que la demanda de los bienes y servicios públicos tiene una reducida elasticidad-precio. Por tanto, al aumentar el precio, la cantidad demandada apenas disminuye, lo que provoca un aumento del gasto.

Como crítica, esta teoría explica el aumento del gasto público nominal y, si la demanda es suficientemente inelástica, también el de su peso en el PIB nominal, pero no el aumento del gasto público en términos reales, es decir, de la cantidad de servicios públicos efectivamente provista.

Teorías de Oferta

A continuación se analizarán las teorías del crecimiento del gasto público por el lado de la oferta.

[2]La intuición es que un cuarteto de cuerda necesita hoy los mismos músicos y el mismo tiempo que hace dos siglos para interpretar una obra, pero sus salarios han subido al ritmo de los del resto de la economía, por lo que su coste relativo no deja de crecer. Es el ejemplo de Baumol y Bowen (1966).

Presupuesto incrementalista

Comenzando por la teoría del presupuesto incrementalista de Wildavsky (1964), esta se basa en la idea de que la asignación de fondos para los programas de gasto público se realiza tomando como referencia el presupuesto del año anterior para decidir si se incrementa o se reduce el importe del gasto. Por consiguiente, se trata de una regla de decisión incrementalista que explica el aumento del gasto público en función de las decisiones pasadas.

Analíticamente, la unidad administrativa al realizar su petición de fondos para el año t, x_t, toma como referencia los fondos que recibió en el ejercicio anterior, y_{t-1}:

$$x_t = \beta \cdot y_{t-1} + u_t, \quad \beta > 1 \tag{17.1}$$

donde β es superior a 1 y representa el parámetro con el que la unidad administrativa ajusta su petición de fondos a la dotación del año anterior, reflejando el interés en incrementar el presupuesto. Además, el término u_t es una perturbación aleatoria que refleja factores no previstos que hacen que se incremente el presupuesto.

Por su parte, el Ministerio de Hacienda , cuando recibe los anteproyectos parciales de cada departamento para elaborar el presupuesto agregado, y_t, ajusta a la baja dichas solicitudes, concediendo una fracción α, que es inferior a 1, de las mismas:

$$y_t = \alpha \cdot x_t + v_t, \quad \alpha < 1 \tag{17.2}$$

donde v_t representa otra perturbación aleatoria.

El resultado, sustituyendo la regla de petición de fondos en la de dotación de fondos, es que el presupuesto dependerá del valor de $\gamma = \alpha \cdot \beta$, que será más probable que sea superior a la unidad y por lo tanto incrementalista cuanto mayores sean las solicitudes de fondos respecto al año anterior, β, y cuanto mayor sea la permisividad del Ministerio de Hacienda, α:

$$y_t = \gamma \cdot y_{t-1} + z_t \tag{17.3}$$

donde $z_t = \alpha \cdot u_t + v_t$ combina las dos perturbaciones anteriores, la primera ponderada por la fracción concedida por Hacienda.

No obstante, el enfoque de Wildavsky plantea limitaciones. Concretamente, Williamson (1967) critica que las reglas de fijación de gasto solo dependen de las decisiones pasadas y no tienen en cuenta la existencia de procesos económicos y políticos que afectarán a los valores de los parámetros α y β y que no están contemplados en la teoría.

 © 2026 Manuel Cabado García

Teorías de la burocracia

Por otro lado, las teorías de la burocracia se centran en el papel de los burócratas y funcionarios en generar incrementos de gasto público.

Comenzando con la teoría de la burocracia de Niskanen (1971), esta parte de considerar que, al igual que en el sector privado la empresa trata de maximizar beneficios, en el sector público, los burócratas tendrán como objetivo maximizar su utilidad privada.

Sin embargo, a diferencia del caso de una empresa, dada la naturaleza de su trabajo y las restricciones a las que se enfrenta el burócrata no podrá beneficiarse del mercado para elevar su renta. Por el contrario, obtendrá utilidad de otros factores que se pueden sintetizar suponiendo que la utilidad del burócrata dependerá del tamaño del presupuesto: a mayor presupuesto mayor utilidad, pues un mayor presupuesto le reportará un mayor prestigio, reputación e influencia. Así, los burócratas utilizarán cierto poder de monopolio para imponer sus preferencias sobre gasto a un legislador poco informado, al existir información asimétrica.

Gráficamente, la teoría de la burocracia puede analizarse en un plano como el de la Figura 17.1[3], que representa en el eje vertical el beneficio para el político de la provisión de un bien público, B, y su coste de provisión, C, y en el eje horizontal el nivel de producción del bien público, G, que suministra un departamento. La función de costes del bien público, $C(G)$, es creciente con el nivel de producción, con un coste marginal positivo y creciente. Dicha función no es conocida por el político y solo el burócrata conoce el proceso de provisión del bien público y, por tanto, su coste.

Por otro lado, la función de beneficios para el político de la provisión del bien público, $B(G)$, es creciente pero con un beneficio marginal positivo pero decreciente. Dicha función de beneficios representaría la utilidad del consumo del bien público que obtienen los votantes y que redundaría en mayores posibilidades de reelección del político.

Con información perfecta, es decir, si el político conociese los costes de provisión, el equilibrio se alcanzaría para un nivel del bien público G^* en el que se iguale el coste marginal de provisión, dado por la pendiente de la tangente a la curva de costes para dicho nivel de gasto público, con el beneficio marginal, representado, a su vez, por la pendiente de la tangente a la curva de beneficios.

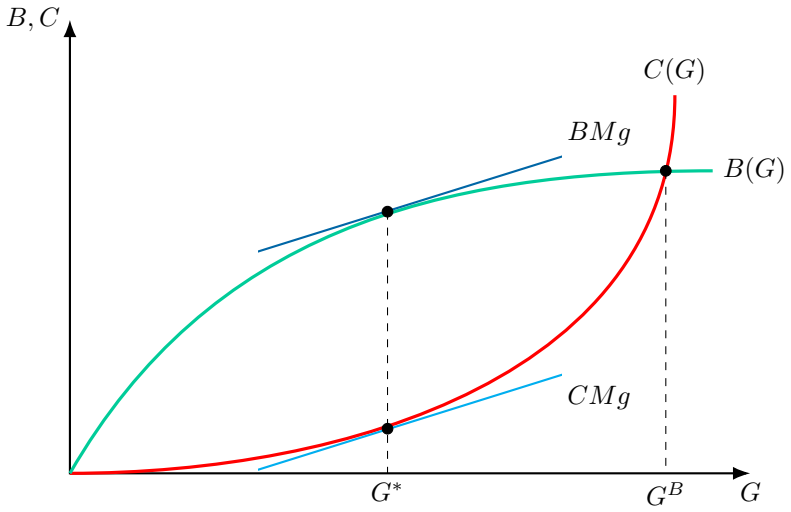
Sin embargo, el burócrata, cuyo objetivo es maximizar el presupuesto, no tiene incentivos para revelar la información sobre el coste de provisión, lo que impide que el nivel de provisión sea G^*. Ahora bien, se enfrenta a la restricción de que el coste de provisión no supere al beneficio que se obtiene del bien público, ya que el político no quiere un proyecto con beneficio neto negativo, $B(G) \geq C(G)$.

Por otro lado, si el burócrata elige un nivel de provisión en el que el coste es inferior al beneficio, $B(G) > C(G)$, estaría renunciando a un aumento del presupuesto. Por ello,

[3]Elaboración propia a partir del modelo de Niskanen (1971). Para una representación equivalente, véase Rosen y Gayer (2010), Figura 6.4, p. 122.

dadas estas restricciones, el burócrata elige el nivel de provisión del bien público G^B, en el que el beneficio total iguala al coste total, $B\left(G^B\right) = C\left(G^B\right)$, y que es superior al que se suministraría si no existiera información asimétrica.

Figura 17.1: Teoría de la burocracia de Niskanen



Por lo tanto, el comportamiento de la burocracia llevaría a un aumento del gasto público y del presupuesto, que se vería incrementado aún más por la sobrestimación de los beneficios y de las estimaciones de costes.

Este crecimiento excesivo del gasto público generaría una ineficiencia en la asignación de recursos, ya que se destinan recursos, como consecuencia del comportamiento de la burocracia, que no producen resultados lo suficientemente valiosos. Además, se produce la llamada ineficiencia X, debido a la falta de presión competitiva que hace que no se utilicen las técnicas más eficientes para producir, generándose así una ineficiencia productiva.

Las soluciones a estos problemas de ineficiencia pasarían por:

1. Reforzar las mayorías necesarias para aprobar los programas de gasto público.
2. Introducir competencia entre el sector público y el sector privado.
3. Remunerar al burócrata y funcionarios de acuerdo a su productividad marginal. Sin embargo, esto resulta difícil ya que la producción del sector público no se vende en el mercado, sino que suele valorarse al coste de los factores.
4. Separar el diseño de la ejecución (por ejemplo, encargar el análisis coste-beneficio a una agencia independiente, de manera que el sector público no pueda ya subestimar los costes y sobrestimar los beneficios de los proyectos públicos).
5. Individualizar las responsabilidades de los burócratas para evitar que se diluyan.
6. Presupuestos de base cero, en lugar de presupuesto incremental.

 © 2026 Manuel Cabado García

7. Limitar el tamaño del sector público, mediante la implementación de la constitución fiscal que se verá más adelante.

También cabe señalar que la teoría de la burocracia asume que no existen restricciones al comportamiento de los burócratas. Sin embargo, estas pueden venir dadas por presiones políticas o demandas de los ciudadanos.

Además, hay que destacar otras dos teorías de la burocracia que analizan a los empleados públicos como votantes:

- Bush y Denzau (1977) plantean la importancia del comportamiento como votantes de los empleados públicos, los cuales tenderán a favorecer a los partidos cuyos programas propongan un aumento del gasto público, de los salarios públicos o una mayor estabilidad en el empleo público. La importancia de este efecto para explicar el incremento del gasto público dependerá del porcentaje que representen los empleados públicos sobre el total de votantes. Un ejemplo de esto se encuentra en España, con las crecientes Ofertas de Empleo Público[4].
- Frey y Pommerehne (1982) destacan que este efecto se puede ver amplificado ya que los empleados públicos tenderían a tener unas mayores tasas de participación electoral, de forma que tienden a tener un mayor peso entre los votantes que entre la población general.

Teorías del ciclo político

Las teorías del ciclo político destacan la influencia de los procesos electorales y el objetivo de conseguir la reelección sobre la evolución del gasto público.

El planteamiento inicial de Nordhaus (1975) considera que si los votantes son miopes, es decir, deciden su voto en función de los resultados económicos obtenidos en el pasado reciente y sus expectativas son estáticas, el gobierno en el poder podría tener un comportamiento oportunista para aplicar políticas expansionistas. Dichas políticas darían lugar a incrementos del gasto público para generar una fase expansiva que corresponda con la duración de la legislatura, con una fase máxima de expansión justo antes de las elecciones.

No obstante, la capacidad del gobierno de aplicar políticas de demanda expansivas con objetivos electorales dependerá de la frecuencia electoral y de la existencia de poderes compensadores fuertes como un banco central independiente, limitación de mandatos, mayorías cualificadas o parlamentos fragmentados. Por otro lado, esta teoría se critica ya que implica que el gobierno tiene un gran control sobre la amplitud y la sincronización de los ciclos económicos.

[4]En el contexto español, esto se manifiesta en la justificación de nuevas contrataciones para "mejorar la calidad de los servicios", "aumentar la eficiencia", "cubrir necesidades crecientes", o "el relevo generacional" aunque el número de empleados pueda exceder el nivel óptimo.

Empíricamente, Frey y Schneider (1981) encuentran que existe evidencia según la cual unas condiciones económicas positivas mejoran la popularidad del gobierno en el poder, pero Alesina et al. (1997) concluyen que no hay evidencia de que el output y el desempleo sigan un ciclo político, pero la inflación sí tiende a incrementarse antes de las elecciones, lo que se puede explicar por las políticas expansivas para mejorar las perspectivas electorales.

Ideología del partido gobernante

Por otro lado, Cameron (1978) y Castles (1982) plantean que las diferencias ideológicas entre los partidos progresistas y conservadores reflejan una diferencia en las preferencias colectivas por la provisión pública.

Concretamente, los modelos de ideología del partido gobernante consideran que los partidos tienen diferentes funciones de pérdida dependiendo de la ideología[5]:

- Progresistas: tienden a favorecer un mayor gasto público, especialmente en áreas como el bienestar social, la sanidad y la educación, para lograr una mayor redistribución de la riqueza y reducir la desigualdad.
- Conservadores: tienden a abogar por un menor gasto público, una reducción de impuestos y la liberalización económica, priorizando la eficiencia del mercado.

Sin embargo, esta teoría no está exenta de críticas. Así, en los países desarrollados existen restricciones institucionales, las preferencias del votante mediano[6] o las coaliciones, donde la ideología del partido mayoritario puede verse diluida por las exigencias de sus socios.

Descentralización del poder político

Por su parte, Tarschys (1975) propone que la descentralización puede dar lugar a un crecimiento del gasto público, ya que si no existe una división clara entre las competencias de los distintos niveles de gobierno se pueden producir solapamientos y duplicidades.

Además, la descentralización facilitaría el incremento del gasto público por la dependencia de los gobiernos subcentrales de las transferencias a través del llamado efecto papel matamoscas, que implica que una transferencia incondicional de una unidad monetaria a una jurisdicción da lugar a un aumento del gasto público mayor que el que generaría un aumento de una unidad monetaria en la renta privada de sus residentes.

Adicionalmente, si existe ilusión fiscal, de modo que el aumento de gasto público del gobierno subcentral no incrementa el precio de su provisión y la población no conozca

[5] Esta teoría está estrechamente relacionada con la idea de que los partidos modifican sus políticas económicas no solo por sus principios ideológicos, sino también en función del ciclo electoral. Por ejemplo, un partido puede aumentar el gasto público antes de unas elecciones para ganar votantes, independientemente de sus promesas de campaña a largo plazo.

[6] Esto puede llevar a que un gobierno de derechas implemente políticas de gasto social que, en teoría, serían más propias de la izquierda, y viceversa.

 © 2026 Manuel Cabado García

el coste real del gasto al financiarse mediante transferencias, se reforzará la tendencia al crecimiento del gasto público.

No obstante, cabe señalar como crítica que la descentralización del poder político podría ayudar a reducir el gasto público al permitir que se alcance una asignación de recursos más eficiente, pues los niveles inferiores de Gobierno conocen mejor las preferencias de los individuos.

17.3. Debate sobre el tamaño del Sector Público⏳4'30"

El análisis de todas las teorías anteriores no constituye un todo integrado, pero ofrece distintas explicaciones a la tendencia creciente del gasto. Esto hace que surja un debate sobre si el crecimiento del gasto público es excesivo y sobre cuál es el tamaño óptimo del sector público.

Dicho debate está presente en toda la historia del pensamiento sin que exista una conclusión única, sino que las distintas escuelas de pensamiento con distinta relevancia histórica plantean diferentes enfoques.

Fisiócratas

El objetivo de los fisiócratas del siglo XVIII era el desarrollo económico y centraban su atención en la agricultura, actividad que consideraban como generadora de crecimiento y riqueza. Por consiguiente, el papel del sector público sería:

- La organización de cultivos mediante la sustitución de terratenientes por burguesía para capitalizar las explotaciones
- La defensa de las libertades económicas.
- Los gastos deberían ser solamente los indispensables para fomentar la producción agraria, como la seguridad o las obras públicas, y dicho gasto se financiaría mediante impuestos sobre la agricultura, se trataría de un único impuesto sobre el sector que se considera productivo.

La Teoría de la Hacienda Pública Clásica

La escuela clásica considera que el crecimiento económico a largo plazo solo es posible debido a la acción del libre mercado. Por ello, el Estado solo debería intervenir para permitir y fomentar la actividad del mercado es decir, la competencia y la acumulación de capital que actúan como fuerzas impulsoras del crecimiento.

Adam Smith en el Libro V de su obra *La riqueza de las naciones* (Smith, 1776) establece los principios que considera deberían guiar la intervención del sector público:

- Respecto al gasto público, Smith plantea que el sector público debería realizar gastos de defensa y diplomacia así como los necesarios para mantener la seguridad y el orden público y la financiación de la justicia. Además, también debería financiar las obras públicas no rentables para la iniciativa privada.
- Por lo que se refiere a los impuestos deben ser neutrales, para no afectar al crecimiento económico, y exigirse con certeza, economicidad y comodidad. Smith se muestra contrario a la imposición sobre el ahorro, que afectaría a la financiación de la acumulación de capital, y define la justicia impositiva de acuerdo a:
  - El principio de beneficio: según el cual el pago de impuestos tiene que estar en función del beneficio disfrutado bajo la protección del Estado.
  - El principio de capacidad de pago: según el cual la contribución debe realizarse en función de la renta, independientemente de los beneficios que se obtengan de la intervención del Estado.

Finalmente, Smith propugna el equilibrio presupuestario entre gastos e ingresos públicos lo que limitará el tamaño del Estado. La necesidad del equilibrio presupuestario se justifica por dos tipos de razones:

- Razones económicas: evitar que la existencia de déficit pueda condicionar las decisiones de acumulación de capital y de trabajo.
- Razones políticas ya que el déficit sería un elemento distorsionador en las decisiones políticas.

No obstante, plantea como excepción la regla de oro de la Hacienda Pública clásica que admite la existencia de déficit para financiar las inversiones que sean autoliquidables, es decir, aquellas cuya rentabilidad permite pagar intereses y reembolsar los préstamos.

Cabe destacar que dentro de la escuela clásica otros autores plantean ciertos matices sobre los postulados de Smith:

- Malthus(1820) plantea la posibilidad de que existan insuficiencias de la demanda efectiva, que podría corregirse con el incremento del gasto público.
- John Stuart Mill(1848) y Bentham(1789) destacan la importancia de asegurar la igualdad de oportunidades, para lo que sería necesario introducir figuras impositivas como la imposición sobre la herencia.

La Hacienda Pública Alemana

Del mismo modo, la Hacienda Pública Alemana acepta los principios presupuestarios clásicos pero plantea una distribución del gasto diferente, ya que se opone al individualismo

 © 2026 Manuel Cabado García

liberal y propugna un programa de reformas sociales y políticas fiscales distributivas como medio para incrementar el bienestar social. Así, el gasto se debería dirigir también al impulso al desarrollo y la corrección de las desigualdades sociales.

La Escuela Marginalista y la Neoclásica

Por otro lado, la teoría de la economía del bienestar desarrollada por Marshall, Pigou y Edgeworth define la necesidad de la intervención del Estado en aquellas situaciones en las que la competencia no consiga maximizar el bienestar.

Así, el tamaño del sector público estaría determinado por el hecho de que los ingresos y gastos públicos deberían ser suficientes para contribuir a corregir los fallos de la asignación de recursos y a mejorar la distribución de la renta. Esto exige conocer los efectos de impuestos y gastos públicos para evitar los efectos desincentivos y aplicar un criterio de justicia fiscal basada en la maximización del bienestar social y en el principio de igualdad de sacrificio.

La Hacienda Pública Keynesiana

Keynes en su obra *Teoría General de la Ocupación, el Interés y el Dinero* (Keynes, 1936) plantea el rechazo de la Ley de Say, aceptada por los economistas clásicos, así como de otros postulados clásicos y defiende la necesidad de la intervención del sector público para estabilizar la economía mediante el incremento de la demanda agregada, a través del aumento del gasto público que actuaría como un mecanismo para incrementar la renta y reducir el desempleo.

Por consiguiente, la estabilización macroeconómica justificaría el aumento del gasto público y del tamaño del sector público. Keynes considera que la acción única de las fuerzas de mercado no conduce por sí sola a un equilibrio de pleno empleo y que la política fiscal sería más efectiva que otras políticas económicas para estabilizar la economía y generar expansiones de la renta, vía efecto multiplicador.

No obstante, la teoría keynesiana se enfrenta a una serie de críticas como la existencia de efectos crowding-out, por los que la actividad del sector público expulsa la actividad privada, las limitaciones del sector público por el funcionamiento de la burocracia, los problemas de información y de revelación de preferencias y las dudas de la efectividad de la política fiscal en un contexto de expectativas racionales.

A pesar de las críticas, la teoría keynesiana constituyó la corriente dominante de pensamiento económico desde la Segunda Guerra Mundial hasta los años 70, por lo que la política fiscal se convierte en elemento de estabilización de la economía y, en conjunción con la política monetaria, se utiliza para evitar la inestabilidad del mercado que genera inflación y desempleo y estimular la demanda efectiva de modo que aumente la renta y el empleo.

La síntesis de Musgrave

Por otro lado, Musgrave (1959) incorpora el pensamiento económico keynesiano y la economía del bienestar para justificar que el sector público debe intervenir en la economía por tres razones:

1. Función de asignación, es decir, alcanzar la eficiencia corrigiendo fallos de mercado, de acuerdo con la visión neoclásica.
2. Función de distribución, para conseguir la equidad, según la visión de autores como Mill, Wagner o Wicksell.
3. Función de estabilización, para suavizar las fluctuaciones cíclicas, de acuerdo con la visión keynesiana.

Este enfoque de Musgrave es la base del desarrollo del Estado del Bienestar y del incremento del gasto público en las economías occidentales en los años 60 y 70.

El enfoque del crecimiento económico

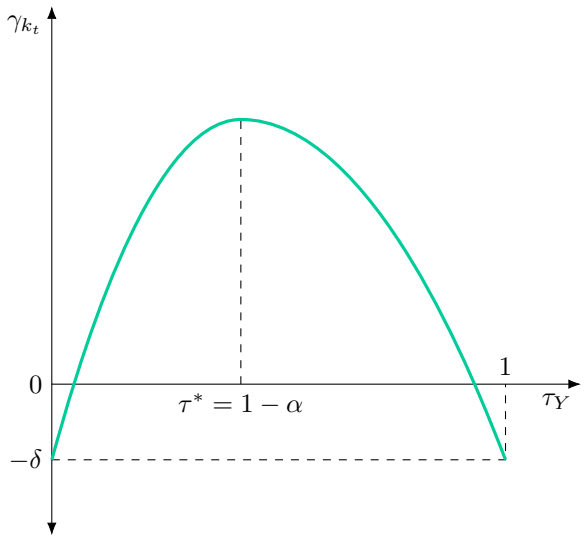
Frente al enfoque normativo de Musgrave sobre las funciones del Sector Público, la literatura del crecimiento económico ofrecería, ya en los años 90, una respuesta positiva y formal a la pregunta del tamaño óptimo. Barro (1990) amplía un modelo de crecimiento endógeno incorporando los servicios públicos, g, como un input productivo de la función de producción privada, de forma que la tasa de crecimiento del capital, γ_{k_t}, depende de la ratio de gasto público sobre el PIB, $\tau_Y = g/Y$.

Un aumento de τ_Y tiene dos efectos de signo contrario sobre el crecimiento: por un lado, eleva la productividad marginal del capital privado al complementarlo con más servicios públicos, y, por otro, exige una mayor imposición distorsionadora para financiarlo. El primer efecto domina cuando el Sector Público es pequeño y el segundo cuando es grande, por lo que la relación entre τ_Y y γ_{k_t} tiene forma de U invertida (Figura 17.2). Bajo una tecnología Cobb-Douglas, $y = A \cdot k^{\alpha} \cdot g^{1-\alpha}$, con α la elasticidad-producto del capital privado, el tamaño del Sector Público que maximiza el crecimiento coincide con la elasticidad-producto de los servicios públicos, $\tau_Y^* = 1 - \alpha$: el Estado maximiza el crecimiento fijando su participación en el PIB igual a la que tendrían los servicios públicos si se contratasen de forma competitiva.

Esta relación en forma de U invertida entre tamaño del Sector Público y crecimiento se popularizaría después de forma más empírica como curva de Armey, o curva BARS —por Barro, Armey, Rahn y Scully—, que estima directamente el punto de máximo crecimiento a partir de datos reales. En esta línea, Vedder y Gallaway (1998) estiman para Estados Unidos un tamaño óptimo del Sector Público en torno al 17,5 % del PIB, sensiblemente inferior al gasto público efectivo del país.

 © 2026 Manuel Cabado García

Figura 17.2: Tamaño del Sector Público y crecimiento económico: el modelo de Barro (1990)



17.4. La Constitución Fiscal ⏳ 4'

Precisamente, la crisis económica de los años 70 hace que se cuestione el tamaño del sector público. Es la denominada por Tarschys (1983) como "crisis de las tijeras"[7]:

- Por el lado de los impuestos, se reducen los ingresos debido a la evasión, a los costes de control y los efectos desincentivos.
- Por el lado del gasto, aumenta el gasto público debido a la crisis económica que incrementa los gastos del Estado del Bienestar, lo que junto al reducido crecimiento de la productividad marginal del sector público lleva a problemas de sostenibilidad y a la llamada crisis del Estado de Bienestar.

Por tanto, los problemas del Estado de Bienestar plantean el debate sobre la expansión excesiva del gasto público y del tamaño del sector público.

Así, Brennan y Buchanan (1980) plantean que el sector público se convertirá en un moderno Leviatán[8] lo que dará lugar a la sobre-expansión de su actividad financiera y fiscal, generando la expulsión de la actividad privada ante la excesiva carga impositiva.

[7]No confundir con la crisis de las tijeras de la Unión Soviética (1922-1923), que fue un fenómeno económico caracterizado por un gran diferencial entre los precios de los productos industriales y agrícolas. Este desequilibrio provocó dificultades económicas, llevando al gobierno a tomar medidas como la reducción de plantillas y la racionalización de la producción para superar la crisis. La situación recibió su nombre debido a la forma gráfica que representaba la divergencia de precios, similar a las hojas de una tijera abriéndose.

[8]Las imperfecciones vistas anteriormente pueden dar lugar a lo que Brennan y Buchanan, tomando la imagen de Thomas Hobbes, llaman "Estado leviatán": Hobbes la empleó para justificar un Estado fuerte capaz de evitar la anarquía, pero aquí designa, metafóricamente, una gran criatura que se comería al sector privado si no se le imponen límites. Es, pues, una idea de sobredimensionamiento del sector público.

Como se ha analizado anteriormente, la sobre-expansión del sector público se derivaría del proceso de elección social, por la toma de decisiones de acuerdo a la regla de mayoría simple, la existencia de ilusión financiera y fiscal, el poder de monopolio de burócratas y la redistribución de la renta a través del presupuesto.

Brennan y Buchanan (1980) abordan la cuestión de la sobre-expansión de la actividad financiera y fiscal del sector público y, ante la incapacidad del votante para tener un control directo y regular sobre la actividad de los burócratas, plantean la necesidad de establecer límites constitucionales al "poder fiscal" del Estado.

De esta forma, la constitución fiscal es un mandato legal (a veces, constitucional) que establece límites a la discrecionalidad de la política fiscal (Brennan y Eusepi, 2005) y protege, además, la equidad intergeneracional, al impedir que las generaciones presentes trasladen cargas excesivas a las futuras. Principios que deben guiar la constitución fiscal:

1. Realismo.
2. Consenso.
3. Flexibilidad.
4. Claridad y sencillez.

Modalidades de Constitución Fiscal

Existen diferentes modalidades:

1. Techo de déficit público: los presupuestos deben contemplar, en su elaboración y ejecución, un nivel de gastos que no sea superior al de ingresos (el artículo 135 de la Constitución Española recoge una versión más flexible, referida al déficit estructural). Equilibrio presupuestario defendido por los clásicos y por Buchanan, siendo esta la regla y teoría más antigua[9].
2. Techo de gasto público. Se puede definir un límite en términos absolutos o al crecimiento relacionado con algún índice (por ejemplo, la directriz agrícola del Paquete

[9] Críticas:

- Excesiva rigidez.
- Teoría de la política fiscal anticíclica. Los defensores de la política fiscal anticíclica argumentan que el equilibrio presupuestario no debería garantizarse cada año, sino cada ciclo económico (de modo que permita registrar déficits en períodos de recesión a cambio de lograr superávits en períodos de expansión, logrando un efecto estabilizador de la actividad económica). No obstante, esta teoría es compatible con una constitución fiscal si ésta se formula en términos de déficit estructural.
- Teoría de la suavización fiscal (Barro, 1979). La ineficiencia generada por los impuestos se minimiza con un tipo impositivo uniforme, por lo que es preferible mantener el tipo constante, dejando actuar a los estabilizadores automáticos, por lo que se incurrirá en déficit en recesión y en superávit en expansión. O lo que es lo mismo: el déficit público será contracíclico (es decir, ahorro público, procíclico). En este contexto, el papel de la deuda será, pues, el de absorber los shocks para permitir la deseada suavización impositiva que minimice las pérdidas de eficiencia.

 © 2026 Manuel Cabado García

Delors de 1988, la regla de gasto de la LOEPSF de 2012 o la regla de gasto del PEC).

3. Techo a la capacidad de gravamen Brennan y Buchanan (1980). Plantean que el control sobre los burócratas por el lado del gasto es muy difícil, por lo que el control deberá ejercerse mediante límites al diseño de la estructura impositiva, para lo que proponen[10]:
   - Bases imponibles reducidas y elásticas de impuestos, que deben preferirse sobre bases amplias y menos sensibles.
   - Uniformidad impositiva, mediante una estructura proporcional para limitar los poderes del burócrata que puede alterar la progresividad para extraer excedente a los votantes.
   - Para asegurar la provisión eficiente de bienes públicos las bases impositivas deben ligarse a un servicio público complementario. Por ejemplo, el gasto en carreteras a la imposición sobre los carburantes.[11]
   - Principio de presupuesto equilibrado y límites a la imposición sobre el capital, para limitar la capacidad de los gobiernos actuales de extraer recursos de generaciones futuras.
4. Veto del Gobierno a partidas de aumento de gasto. Requisito de mayorías cualificadas.
5. Federalismo fiscal: correspondencia fiscal en los distintos niveles administrativos, de forma que tengan el mismo grado de autonomía por el lado del gasto que por el lado de los ingresos.
6. Competencia fiscal entre las administraciones del mismo nivel administrativo (por ejemplo, entre Comunidades autónomas, entre ayuntamientos, etc.), de manera que el "voto con los pies" de los ciudadanos les incite a limitar el crecimiento del gasto público para poder ofrecer una fiscalidad ventajosa.
7. Autoridades fiscales independientes. Al igual que han surgido autoridades monetarias independientes como los bancos centrales, ajenas al ciclo político, no sometidas a los grupos de presión y con un objetivo determinado, Alan Blinder (Blinder, 2004) propone crear autoridades fiscales independientes con perfil técnico (por ejemplo, la AIReF o el Two-Pack).

Formalmente, y por analogía directa con el "banquero central conservador" de Rogoff (1985), cabe modelizar a esta autoridad fiscal conservadora como un agente que minimiza

[10] No obstante, cabe destacar que estas condiciones contrastan con los principios de imposición óptima: eficiencia económica, equidad, sencillez administrativa, flexibilidad y suficiencia.

[11] La intuición es que, si lo recaudado con los carburantes solo puede gastarse en carreteras, al Gobierno le conviene construir las carreteras que la gente vaya a usar de verdad, porque solo así aumenta el consumo de carburante y, con él, su recaudación.

una función de pérdida cuadrática,

$$\min_{d_t} L = (Y_t - Y^*)^2 + \delta \cdot (d_t - d^*)^2 ,$$

donde d_t es el déficit público del período, d^* su objetivo de largo plazo, e Y_t el nivel de actividad, con su objetivo Y^*. El parámetro δ recoge la aversión de la autoridad al desvío respecto de su objetivo de déficit, igual que el parámetro análogo en el modelo de Rogoff recoge la aversión del banquero central al desvío de su objetivo de inflación: cuanto mayor sea δ, más se aproximará el déficit de equilibrio a su objetivo d^*, corrigiendo el sesgo deficitario de la política discrecional, a costa de sacrificar parte de la capacidad estabilizadora automática del presupuesto ante perturbaciones coyunturales de la actividad.

Ejemplos de Constitución Fiscal

Albi Ibáñez et al. (2017) señalan los cambios legislativos efectuados en España, que incluyen:

- La reforma del artículo 135 de la Constitución Española en 2011, que otorga rango constitucional al principio de estabilidad presupuestaria, que se vincula al cumplimiento de los límites máximos establecidos por ley para el déficit estructural del Estado y de las Comunidades Autónomas en relación con el PIB, y a la garantía de equilibrio presupuestario de las Entidades Locales. Asimismo, el artículo 135 establece que el volumen de deuda pública del conjunto de las Administraciones Públicas en relación con el PIB no podrá superar los límites establecidos por el Tratado de Funcionamiento de la Unión Europea, y concede prioridad absoluta al pago de los intereses y el capital de la deuda pública. Por último, la reforma constitucional introduce una cláusula de salvaguarda, por la cual los límites de déficit estructural y deuda pública pueden superarse con el apoyo de la mayoría absoluta de los miembros del Congreso en caso de circunstancias excepcionales (catástrofes naturales, recesión económica o situaciones de emergencia extraordinaria), que pongan en peligro la sostenibilidad económica o social del Estado.
- La aprobación de la Ley Orgánica de Estabilidad Presupuestaria y Sostenibilidad Financiera (LOEPSF) en 2012, consagra los principios de estabilidad presupuestaria y sostenibilidad financiera y establece como principio rector el de plurianualidad, que exige que los presupuestos públicos se encuadren en un marco presupuestario a medio plazo (y del que se deriva la ya citada regla de gasto)[12].

[12] Esta Ley establece limitaciones en tres ámbitos: déficit público, gasto público, y deuda pública.

- Déficit público (art. 11): ninguna Administración pública podrá incurrir en déficit estructural. Excepciones al mandato de déficit estructural nulo:
  - Reformas estructurales. Si se llevan a cabo reformas estructurales con efectos presupuestarios a largo plazo, podrá alcanzarse en el conjunto de las Administraciones Públicas un déficit

 © 2026 Manuel Cabado García

- La creación de la Autoridad Independiente de Responsabilidad Fiscal (AIReF) en 2013, creada como instrumento dentro del marco de gobernanza reforzada de la UE , una de sus tareas consiste en la evaluación continua del ciclo presupuestario, como forma de velar por el estricto cumplimiento de los principios de estabilidad presupuestaria y sostenibilidad financiera, y somete a escrutinio la actividad fiscal y presupuestaria de las AA.PP.

En el ámbito europeo junto con la reciente reforma de las reglas fiscales en 2024[13]:

- Pacto de Estabilidad y Crecimiento (PEC reformado 2024) El nuevo marco del Pacto de Estabilidad y Crecimiento busca un equilibrio entre la disciplina fiscal y la capacidad de los Estados miembros para invertir y realizar reformas. Su funcionamiento se basa en dos brazos, pero con un enfoque renovado.
- Brazo Preventivo: El objetivo sigue siendo prevenir políticas fiscales insostenibles, pero ahora se centra en planes fiscales individualizados. Los países deben presentar Planes Fiscales y Estructurales de Medio Plazo que detallen una senda de ajuste fiscal.
    - Senda de gasto neto: La variable principal de seguimiento ya no es el MTO, sino el gasto público neto. Este se establece para cada país en función de su nivel de deuda y se acuerda con la Comisión Europea. El plan tiene una duración de cuatro años, ampliable a siete si se incluyen inversiones y reformas estratégicas.
    - Regla de gasto: La nueva regla de gasto es el pilar del brazo preventivo, ya que el control del gasto neto es la forma en que los países deben garantizar la sostenibilidad de sus finanzas.
- Brazo Correctivo: Este brazo se activa cuando un país ya ha superado los umbrales de déficit o deuda.

estructural del 0,4 % del PIB nominal.

- Catástrofes naturales, recesión económica grave o situaciones de emergencia. En estos casos, se podrá incurrir en déficit estructural si se aprueba por mayoría absoluta del Congreso de los Diputados.

- Regla de gasto (art. 12): la tasa de crecimiento del gasto público computable del conjunto de las Administraciones Públicas no podrá superar la tasa de crecimiento del PIB de medio plazo. Corresponde al Ministerio de Asuntos Económicos calcular la tasa de referencia de crecimiento del Producto Interior Bruto de medio plazo de la economía española.
- Deuda pública (art. 13): la deuda pública del conjunto de las Administraciones Públicas no podrá superar el 60 % del PIB nominal. Este límite se distribuye de acuerdo a los siguientes porcentajes:
    - Administración central: 44 puntos porcentuales.
    - Comunidades Autónomas: 13 puntos porcentuales .
    - Corporaciones Locales: 3 puntos porcentuales.

[13] Recordar que durante la Pandemia se suspendieron las reglas fiscales para poder hacer frente a la crisis.

  - Déficit y Deuda: Se mantienen los límites de referencia del 3 % del PIB para el déficit y del 60 % del PIB para la deuda. Si un país incumple estos límites, el Consejo puede abrirle un Procedimiento por Déficit Excesivo (PDE).
  - Ajuste estructural: El PDE exige un ajuste estructural mínimo del 0,5 % del PIB anual hasta que el déficit se corrija. Sin embargo, la activación del PDE ahora también puede ocurrir si un país se desvía de su senda de gasto neto acordada.

- Fiscal Compact (Tratado de Estabilidad, Coordinación y Gobernanza) El Fiscal Compact, que nació en 2012 para reforzar la disciplina fiscal, también se ha visto afectado por la reforma de 2024.
  - Objetivos del Fiscal Compact: Las nuevas reglas fiscales integran gran parte de los objetivos del Fiscal Compact en el marco general de la UE, incluyendo la regla de equilibrio presupuestario estructural. Esto significa que las reglas ahora tienen un estatus legal más sólido en el derecho de la UE.
  - Autoridades Fiscales Independientes: El rol de las autoridades fiscales independientes, como la AIReF en España, se ha reforzado. Estas instituciones son cruciales para supervisar el cumplimiento de las sendas de gasto y dar credibilidad a los planes fiscales de los gobiernos.
  - Mecanismo corrector automático: El principio del mecanismo corrector se mantiene, pero ahora se materializa a través de las sendas de gasto netas. Las desviaciones de esta senda activan mecanismos de corrección dentro del brazo preventivo o correctivo del nuevo marco fiscal.

17.5. Conclusiones ⏳30"

Como conclusión, en el debate sobre el tamaño del sector público existen dos posturas:

1. Optimista: sostiene que el Estado es un garante esencial de servicios (sanidad, educación), corrige fallos de mercado y fomenta el progreso técnico y el capital humano. Por tanto, es clave para el desarrollo económico y la igualdad.
2. Pesimista: señala el problema del crowding-out (desplazamiento de la inversión privada), ineficiencias burocráticas y el riesgo moral derivado de los diferentes grupos de interés, por lo que el gasto público no siempre responde al interés general.

No obstante, la clave está en la evaluación del gasto y en la constante rendición de cuentas, por lo que el debate debe centrarse en la eficacia de las políticas públicas.

 © 2026 Manuel Cabado García

Apéndice

17.A.1. Reducción de gastos partida a partida

La reducción de gastos partida a partida consiste en el recorte selectivo de líneas concretas del presupuesto. Es el método más utilizado en la práctica —resulta política y administrativamente más sencillo que rediseñar el presupuesto en su conjunto—, pero también uno de los menos eficaces: la mayor parte del gasto público está comprometido de antemano por obligaciones legales o contractuales —pensiones, salarios públicos, servicio de la deuda—, por lo que los recortes recaen sobre una fracción reducida y discrecional del presupuesto. Las reducciones que resultan practicables son, en consecuencia, de escasa cuantía, y no logran por sí solas constituir un freno real al crecimiento global del gasto.

17.A.2. Presupuesto plurianual

La técnica del presupuesto plurianual sustituye la confección anual del presupuesto —uno de los principios de la Hacienda Pública clásica— por una programación a varios años vista. Frente al presupuesto anual, presenta la ventaja de visibilizar los costes futuros de los programas de gasto que se ponen en marcha, corrigiendo el sesgo de visión a corto plazo propio de la presupuestación año a año.

No obstante, la experiencia británica del Public Expenditure Survey Committee (PESC) (Clarke, 1978) muestra que la presupuestación plurianual tampoco está exenta de inconvenientes:

- Los compromisos de reducción del gasto adquiridos para los ejercicios futuros del ciclo plurianual rara vez se cumplen, por lo que raramente se traducen en una reducción efectiva del gasto del año en curso.
- El papel de la inflación cobra especial importancia, sobre todo cuando la presupuestación se realiza en términos de volumen.
- Resulta más difícil prescindir de un gasto determinado una vez que ya ha sido incluido en el presupuesto.
- Cuanto más prolongado es el periodo de programación, menor es el control efectivo sobre el gasto.
- La estabilidad de los organismos públicos exige, en ocasiones, renunciar a imponer límites estrictos en un año dado.

17.A.3. Presupuesto Base Cero

El Presupuesto Base Cero (PBC) es una técnica presupuestaria propuesta originalmente por Pyhrr (1970) a partir de su experiencia en Texas Instruments y adoptada después

por distintas administraciones públicas, entre ellas la del estado de Georgia durante el mandato de Jimmy Carter como gobernador. Frente al presupuesto incrementalista de Wildavsky, ya analizado, en el que cada partida se dota tomando como referencia la del ejercicio anterior, el PBC exige justificar cada partida desde cero en cada ejercicio presupuestario, ordenando los programas de gasto por prioridad. Esto permite recortar con seguridad los programas de menor prioridad, sin depender de la inercia de decisiones pasadas que caracteriza al presupuesto incrementalista.

17.A.4. Determinación del volumen del presupuesto

Determinar el presupuesto en términos de volumen de bienes y servicios prestados, en lugar de en términos nominales, evita que la inflación por sí sola incremente el gasto público sin que aumente realmente la provisión de bienes y servicios: si el presupuesto se fija en términos nominales, cualquier repunte de precios eleva automáticamente el gasto contable aunque la cantidad real provista permanezca constante, generando una suerte de ilusión fiscal por el lado del gasto —el gasto parece crecer sin que la sociedad reciba más a cambio—. Fijarlo en términos de volumen corrige este sesgo: solo un incremento efectivo de la cantidad o calidad de los servicios prestados justificaría un mayor presupuesto, con independencia de la evolución de los precios.

17.A.5. Clasificación del gasto público

El gasto público puede definirse como el conjunto de flujos de recursos que las Administraciones Públicas destinan a la producción de bienes y servicios y a la redistribución de la renta, financiados fundamentalmente mediante ingresos coactivos —impuestos y cotizaciones sociales. Sin embargo, antes de clasificar el gasto conviene delimitar con precisión su ámbito, distinguiendo entre las Administraciones Públicas propiamente dichas y el Sector Público en sentido amplio.

El Sistema Europeo de Cuentas (SEC-2010), aprobado por el Reglamento (UE) n.º 549/2013, define las Administraciones Públicas (AAPP, sector S.13) como el conjunto de unidades institucionales cuya función principal es producir bienes y servicios no destinados a la venta y efectuar operaciones de redistribución de la renta, financiadas fundamentalmente mediante pagos obligatorios (impuestos y cotizaciones sociales). El Sector Público en sentido amplio añade además las empresas públicas cuya producción sí es de mercado —aquellas cuyas ventas cubren más del 50 % de sus costes de producción, el llamado criterio del 50 %—, que quedan fuera del perímetro de las AAPP. Esta distinción no es meramente formal: cuando se habla de déficit o deuda pública, la Contabilidad Nacional se refiere siempre al sector AAPP, nunca al Sector Público en sentido amplio.

Delimitado así el gasto de las AAPP, puede clasificarse según un triple criterio, cada uno respondiendo a una pregunta distinta:

 © 2026 Manuel Cabado García

- Clasificación orgánica — responde a "¿quién gasta?". Identifica la unidad responsable de la gestión y ejecución de cada crédito, agrupando el gasto por secciones y servicios entre los distintos centros gestores: la Administración Central (Estado y organismos autónomos), las Administraciones Territoriales (Comunidades Autónomas y Corporaciones Locales) y los Fondos de la Seguridad Social.
- Clasificación económica — responde a "¿cómo se gasta?". Es la adoptada por los Presupuestos Generales del Estado (regulada por la Ley General Presupuestaria, Ley 47/2003) mediante capítulos, con un segundo nivel de artículo y un tercero de concepto: operaciones corrientes no financieras (personal, bienes y servicios, gastos financieros, transferencias corrientes, fondo de contingencia) y operaciones de capital (inversiones reales, transferencias de capital), a las que se añaden las operaciones financieras (activos y pasivos financieros).
- Clasificación funcional — responde a "¿para qué se gasta?". Agrupa el gasto en políticas y áreas de gasto (servicios públicos básicos, actuaciones de protección y promoción social, producción de bienes públicos de carácter preferente, actuaciones de carácter económico y actuaciones de carácter general). Su equivalente armonizado a nivel internacional es la clasificación COFOG (*Classification of the Functions of Government*, Naciones Unidas, 1999), que distingue diez funciones —servicios públicos generales, defensa, orden público y seguridad, asuntos económicos, protección del medio ambiente, vivienda, sanidad, ocio y cultura, educación y protección social— y resulta especialmente útil para las comparaciones internacionales, al no depender de la clasificación presupuestaria propia de cada país. Conviene señalar que la correspondencia entre la clasificación funcional española y la COFOG no es exacta —por ejemplo, la COFOG separa "vivienda y servicios comunitarios" como función propia, mientras que la clasificación española integra el acceso a la vivienda dentro de las "actuaciones de protección y promoción social"—, razón por la que las comparaciones internacionales de composición del gasto se realizan siempre sobre datos COFOG.

17.A.6. Los *spending reviews* como instrumento de evaluación del gasto

Junto a los mecanismos de limitación del gasto ya vistos, la disciplina presupuestaria moderna ha incorporado los *spending reviews* —exámenes sistemáticos de la eficacia y eficiencia de las políticas de gasto, con el objetivo de identificar partidas prescindibles y mejorar la asignación de recursos— como complemento a los controles puramente cuantitativos sobre el déficit y el gasto agregado. La OCDE, a través de su Grupo de Trabajo de Altos Funcionarios Presupuestarios (desde 1980), incluye la realización de *spending reviews* regulares como uno de los diez principios de su marco de buena institucionali-

dad presupuestaria, y promueve el análisis de *value for money*: si las políticas cumplen los objetivos para los que fueron diseñadas y al menor coste posible. Según su informe *Government at a Glance* (2023), 34 de los 35 países miembros encuestados —el 97 %— ya habían implantado esta práctica, lo que confirma su carácter prácticamente universal entre las economías avanzadas.

El origen del término se sitúa en el Reino Unido, con el *Comprehensive Spending Review* introducido por el entonces canciller Gordon Brown en 1998. Dinamarca y Países Bajos son pioneros, con prácticas sistemáticas desde los años 80 —si bien en el caso danés los informes se publican solo raramente, a diferencia de la transparencia pública que caracteriza al modelo español—. Irlanda los introdujo en 2009, tras la crisis financiera, a través de una red propia de especialistas (el IGEES), y Francia estableció sus *revues de dépenses* en 2008, hoy obligatorias con periodicidad anual. En Estados Unidos coexisten dos instituciones con papeles distintos que conviene no confundir: la *Congressional Budget Office* (CBO), del poder legislativo y no partidista, se limita a estimar el coste presupuestario de las leyes en tramitación y a elaborar proyecciones macro-fiscales —mira hacia adelante, no evalúa el gasto ya ejecutado—, por lo que su función se asemeja más a la de vigilancia presupuestaria de una autoridad fiscal independiente que a la de un *spending review* propiamente dicho. La evaluación de políticas de gasto ya en marcha recae, en cambio, en la *Government Accountability Office* (GAO), brazo investigador del Congreso que examina el uso real de los fondos públicos, y en la ya extinta *Program Assessment Rating Tool* (PART) de la Office of Management and Budget —instaurada en 2002, llegó a evaluar más de mil programas federales, el 98 % del presupuesto, pero fue discontinuada por la siguiente Administración—, lo que muestra que este tipo de instrumentos no está exento de fragilidad política.

A nivel de la Unión Europea, el Semestre Europeo incorpora recomendaciones específicas por país para que los Estados miembro implanten o refuercen sus *spending reviews*. En España, este impulso se concretó en el Componente 29 del Plan de Recuperación, Transformación y Resiliencia —financiado con fondos Next Generation—, que creó dentro de la AIReF una División de Evaluación del Gasto Público de carácter permanente y ligó el desembolso de fondos europeos al cumplimiento de dos reformas: el propio proceso de revisión del gasto y el alineamiento de los Presupuestos Generales del Estado con sus resultados.

La AIReF ha desarrollado, con esta nueva capacidad, dos ciclos de *spending review*: el primero (2018-2021, en tres fases) evaluó, entre otras políticas, las políticas activas de empleo y el gasto farmacéutico y de alta tecnología del Sistema Nacional de Salud. El segundo (2022-2026) ha analizado la incapacidad temporal, la ayuda oficial al desarrollo y el ingreso mínimo vital, y tiene ya en preparación una tercera fase, con evaluaciones previstas sobre la regla de gasto en pensiones y la eficacia de los incentivos a la contratación. El seguimiento de estas evaluaciones revela también los límites del instrumento: en

 © 2026 Manuel Cabado García

torno al 16 % de las recomendaciones formuladas en el primer ciclo seguía "en proceso" de implementación años después de su publicación, lo que confirma que la evaluación por sí sola no garantiza el cambio efectivo de las políticas —de ahí que, como se apunta en la Conclusión, la clave última esté en la voluntad política de traducir esa evaluación en rendición de cuentas real.

Bibliografía

Albi Ibáñez, E., J. M. Gónzalez-Páramo, R. M. Urbanos y I. Zubiri (2017). *Economía Pública I: Fundamentos. Presupuestos y Gastos*. Barcelona: Ariel.

Alesina, A., N. Roubini y G. D. Cohen (1997). *Political Cycles and the Macroeconomy*. The MIT Press.

Barro, R. J. (1979). On the Determination of the Public Debt. *Journal of Political Economy 87*(5), 940–971.

Barro, R. J. (1990). Government Spending in a Simple Model of Endogenous Growth. *Journal of Political Economy 98*(5), 103–126.

Baumol, W. J. (1967). Macroeconomics of Unbalanced Growth: the Anatomy of Urban Crisis. *The American Economic Review*, 415–426.

Baumol, W. J. y W. G. Bowen (1966). *Performing Arts, The Economic Dilemma: A Study of Problems Common to Theater, Opera, Music and Dance*. New York: Twentieth Century Fund.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. London: T. Payne and Son.

Bird, R. (1971). Wagner's Law of Expanding State Activity. *Public Finance 26*(1), 1–26.

Blinder, A. S. (2004). The Case Against the Case Against Discretionary Fiscal Policy. CEPS Working Paper 100, Princeton University, Center for Economic Policy Studies. Publicado en Kopcke, R. W., Tootell, G. M. B. y Triest, R. K. (eds.) (2006), *The Macroeconomics of Fiscal Policy*, MIT Press.

Brennan, G. y J. M. Buchanan (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge University Press.

Brennan, G. y G. Eusepi (2005). Fiscal Constitutionalism. En J. G. Backhaus y R. E. Wagner (eds.), *Handbook of Public Finance*, pp. 53–76. Springer.

Buchanan, J. M. y G. Tullock (1962). *The Calculus of Consent: Logical Foundations for Constitutional Democracy*. Michigan: The University of Michigan Press.

Buchanan, J. M. y R. E. Wagner (1977). *Democracy in deficit: The political legacy of Lord Keynes*, vol. 4. Academic Press New York.

Bush, W. C. y A. T. Denzau (1977). The Voting Behavior of Bureaucrats and Public Sector Growth. En T. E. Borcherding (ed.), *Budgets and Bureaucrats: The Sources of Government Growth*, pp. 90–99. Durham, NC: Duke University Press.

 © 2026 Manuel Cabado García

Cameron, D. R. (1978). The Expansion of the Public Economy: A Comparative Analysis. *American Political Science Review 72*(4), 1243–1261.

Castles, F. G. (1982). *The Impact of Parties: Politics and Policies in Democratic Capitalist States*.

Clarke, R. (1978). *Public Expenditure, Management and Control: The Development of the Public Expenditure Survey Committee (PESC)*. London: Palgrave Macmillan.

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy 65*(2), 135–150.

Engel, E. (1857). Die Produktions—und Consumtionsverhältnisse des Königreichs Sachsen. *Zeitschriftdes statistischen Büreaus des Königlich Sächsischen Ministeriums des Innern*.

Frey, B. S. y W. W. Pommerehne (1982). How powerful are public bureaucrats as voters? *Public Choice 38*(3), 253–262.

Frey, B. S. y F. Schneider (1981). A Politico-Economic Model of the U.K.: New Estimates and Predictions. *The Economic Journal 91*(363), 737.

Jaén García, M. (2004). *La ley de Wagner: un análisis sintético*. núm. 6/04 de IEF papeles de trabajo. Serie Economía. Madrid: Instituto de Estudios Fiscales.

Keynes, J. M. (1936). *The General Theory of Employment Interest and Money*. London: Macmillan.

Malthus, T. R. (1820). *Principles of Political Economy Considered with a View to Their Practical Application*. London.

Mill, J. S. (1848). *Principles of political economy: with some of their applications to social philosophy*, vol. 1. Longmans, Green, Reader, and Dyer.

Musgrave, R. A. (1959). *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw Hill.

Niskanen, W. A. (1971). *Bureaucracy & representative government*. Routledge.

Nordhaus, W. D. (1975). The Political Business Cycle. *The Review of Economic Studies 42*(2), 169.

Peacock, A. T. y J. Wiseman (1961). *The Growth of Public Expenditure in the United Kingdom*. National Bureau of Economic Research, Inc.

Pyhrr, P. A. (1970). Zero-Base Budgeting. *Harvard Business Review 48*(6), 111–121.

Rogoff, K. (1985). The Optimal Degree of Commitment to an Intermediate Monetary Target. *The Quarterly Journal of Economics 100*(4), 1169–1189.

Romer, T. y H. Rosenthal (1979). The Elusive Median Voter. *Journal of Public Economics 12*(2), 143–170.

Rosen, H. S. y T. Gayer (2010). *Public Finance* (9 ed.). New York: McGraw-Hill/Irwin.

Say, J.-B. (1803). *Traité d'économie politique, ou simple exposition de la manière dont se forment, se distribuent, et se composent les richesses*. Paris: Crapelet.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*.

Tarschys, D. (1975, enero). The Growth of Public Expenditures: Nine Modes of Explanation. *Scandinavian Political Studies 10*(A10), 9–31.

Tarschys, D. (1983). The Scissors Crisis in Public Finance. *Policy Sciences 15*(3), 205–224.

Timm, H. (1961). Das Gesetz der wachsenden Staatsausgaben. *FinanzArchiv / Public Finance Analysis 21*(2), 201–247.

Vedder, R. K. y L. E. Gallaway (1998). Government Size and Economic Growth. Informe técnico, Joint Economic Committee, United States Congress, Washington, D.C.

Wagner, A. (1883,1890). Finanzwissenchaft. En R. A. Musgrave y A. T. Peacock (eds.), *Classics in the Theory of Public Finance (1958)*, pp. 1–15. London: Palgrave Macmillan UK.

Wildavsky, A. B. (1964). Politics of the budgetary process.

Williamson, O. E. (1967). A Rational Theory of the Federal Budgeting Process. *Papers on Non-Market Decision Making 2*, 71–89.

 © 2026 Manuel Cabado García

Tema 18

Poder de mercado y regulación óptima: instrumentos, efectos y limitaciones. Desarrollos en presencia de información asimétrica. Aplicaciones prácticas.

18.1. Introducción ⏳ 1'

En octubre de 2014 la Real Academia Sueca de Ciencias concedió el Premio de Ciencias Económicas en memoria de Alfred Nobel a Jean Tirole "por su análisis del poder de mercado y su regulación" (The Royal Swedish Academy of Sciences, 2014). El premio reconocía una línea de trabajo que responde a una pregunta clásica: qué debe hacer el sector público cuando en un mercado hay una sola empresa, o muy pocas, capaces de fijar precios por encima de su coste.

El primer teorema fundamental de la economía del bienestar garantiza que un mercado competitivo asigna los recursos de forma eficiente, pero solo si se cumplen sus supuestos. Cuando una empresa tiene poder de mercado, la asignación deja de ser eficiente y la intervención pública puede estar justificada. El caso extremo es el monopolio natural, propio de actividades como el transporte y la distribución de electricidad, gas o agua, en las que resulta más barato que una sola empresa abastezca todo el mercado. En ellas la competencia no solo no es posible, sino que tampoco sería deseable, y la defensa de la competencia deja paso a la regulación. Surgen así varias cuestiones:

- ¿Qué instrumentos tiene el regulador para acercar el resultado al óptimo social y qué efectos y limitaciones tiene cada uno?
- ¿Cómo cambia el problema cuando la empresa regulada sabe más que el regulador

sobre sus costes y su esfuerzo?

- ¿Qué puede hacer el regulador cuando hay varias empresas, ya sea para compararlas entre sí, para hacerlas competir por el mercado o para ordenar su acceso a la red?
- ¿Cómo se aplican estos principios en la regulación real de las redes energéticas, las telecomunicaciones o las plataformas digitales?

Por consiguiente, el objetivo del presente tema es responder a las cuestiones anteriores, para lo que se estudiarán:

1. El poder de mercado y los instrumentos de la regulación con información completa, con sus efectos y limitaciones.
2. Los desarrollos de la nueva economía de la regulación en presencia de información asimétrica, incluida la regulación por incentivos.
3. La regulación óptima cuando hay varias empresas: competencia por comparación, competencia por el mercado, acceso a las redes y plataformas.
4. Las aplicaciones prácticas en España, la Unión Europea y el Reino Unido.

18.2. Poder de mercado y regulación con información completa: instrumentos, efectos y limitaciones ⏳ 6'

Comenzando por los fundamentos, antes de comparar instrumentos conviene precisar en qué mercado se mide el poder de mercado, qué es, por qué preocupa y en qué casos la respuesta adecuada es regular y no solo defender la competencia.

Poder de mercado y justificación de la regulación

Mercado relevante El mercado relevante es el conjunto de productos y de zonas geográficas que compiten entre sí lo bastante como para disciplinar los precios de una empresa.[1] Para delimitarlo se usa el test del monopolista hipotético, o SSNIP (por las siglas en inglés de una subida pequeña pero significativa y no transitoria del precio), que introdujeron las directrices de concentraciones de Estados Unidos (U.S. Department of Justice, 1982). Se pregunta si a un monopolista hipotético de un conjunto candidato de productos le resultaría rentable una subida del orden del 5 al 10 %. Si no le resulta rentable, porque

[1] La intuición es que solo tiene sentido hablar de poder de mercado frente a lo que los clientes ven como alternativas reales: si pueden cambiar fácilmente a otro producto o comprar en otra zona, la empresa no puede subir el precio aunque sea la única que vende el suyo.

 © 2026 Manuel Cabado García

los clientes se pasarían a otros productos, estos se añaden al conjunto y se repite el test.[2] El mercado relevante es el marco en el que se miden las cuotas y el poder de mercado, y el primer paso de la regulación ex ante de sectores como las telecomunicaciones.

Poder de mercado El poder de mercado es la capacidad de una empresa para fijar de forma rentable y duradera un precio superior a su coste marginal.[3] Su medida habitual es el índice que propuso Lerner (1934), IL, que relaciona el margen entre el precio, P, y el coste marginal, CMg, con el propio precio. En el óptimo de un monopolista ese margen es igual a la inversa del valor absoluto de la elasticidad-precio de la demanda, $|\varepsilon|$:

$$IL = \frac{P - CMg}{P} = \frac{1}{|\varepsilon|}$$

El índice vale cero en competencia perfecta y es mayor cuanto más rígida es la demanda que afronta la empresa.

El poder de mercado genera tres tipos de ineficiencia, que sirven después para valorar cada instrumento de regulación:

1. Ineficiencia asignativa: al fijar un precio superior al coste marginal se produce menos de lo socialmente óptimo y aparece una pérdida irrecuperable de bienestar, el triángulo de Harberger (1954).
2. Ineficiencia productiva: sin la presión de los competidores, la empresa no tiene incentivos para producir al mínimo coste, lo que Leibenstein (1966) denominó ineficiencia X.
3. Ineficiencia dinámica: el poder de mercado persistente puede frenar la innovación y la inversión, aunque la evidencia empírica sobre este efecto no es concluyente.

La respuesta pública depende del origen del poder de mercado (Braeutigam, 1989). Si procede de conductas estratégicas de las empresas o de barreras legales, la competencia sigue siendo posible y el instrumento adecuado es la defensa de la competencia, que actúa ex post sancionando conductas y controlando concentraciones. Si procede de la propia estructura de costes de la actividad, la competencia deja de ser deseable y la respuesta es la regulación económica, que actúa ex ante sobre los precios, la entrada y la calidad.

Monopolio natural Un monopolio natural es un mercado en el que la función de costes es subaditiva, es decir, en el que producir toda la cantidad demandada con una

[2] El test falla si se parte de un precio que ya es de monopolio, porque entonces una subida adicional puede ahuyentar a los clientes hacia sustitutos lejanos y el mercado parece más amplio de lo que es: es la falacia del celofán. Para servicios de precio nulo, como muchas plataformas digitales, la Comunicación de la Comisión de 2024 sobre la definición del mercado de referencia admite sustituir la subida de precio por una disminución pequeña pero significativa y no transitoria de la calidad.

[3] La intuición es que una empresa con poder de mercado puede subir su precio sin perder a todos sus clientes, porque estos no encuentran sustitutos cercanos. En competencia perfecta, en cambio, cualquier subida de precio desplaza toda la demanda hacia los rivales.

sola empresa cuesta menos que repartirla entre varias.[4] Formalmente, si $C(Q)$ es el coste de producir la cantidad total, Q, y q_i la producción de cada una de las n empresas entre las que podría repartirse:

$$C(Q) < \sum_{i=1}^{n} C(q_i), \qquad Q = \sum_{i=1}^{n} q_i$$

Con un solo producto, las economías de escala son condición suficiente, pero no necesaria, para la subaditividad (Baumol, 1977; Baumol et al., 1982).[5]

No obstante, el monopolio natural plantea un dilema: mantener una sola empresa preserva la eficiencia productiva, pero dejarla sin regular sacrifica la eficiencia asignativa. Además, puede no ser sostenible, porque a un entrante puede resultarle rentable servir solo una parte de la demanda aunque con ello se pierda la ventaja de costes. Por eso la regulación debe ocuparse también de las condiciones de entrada, ya sea prohibiéndola o subastando el derecho de exclusiva. Resolver ese dilema de la mejor manera posible es, precisamente, el objeto de la regulación óptima.

Regulación óptima La regulación óptima es la elección de las reglas de precios, entrada y calidad que maximizan el bienestar social, medido como la suma del excedente de los consumidores y el beneficio de la empresa, sujeta a que la empresa regulada sea viable y a lo que el regulador puede observar.[6] Para ello dispone de varios instrumentos: la fijación de precios, el control de la entrada, la exigencia de calidad y, como alternativa a regular una empresa privada, la propiedad pública de la empresa. Cada instrumento se valora por sus efectos sobre las tres eficiencias y por sus limitaciones, que proceden de la información, del compromiso del regulador y del riesgo de captura.

A continuación, el análisis comenzará por el caso más favorable al regulador, aquel en que conoce todo sobre la empresa, e introducirá después las limitaciones de información y de compromiso que aparecen en la práctica.

Tarificación con información completa

La teoría tradicional supone que el regulador conoce los costes y la demanda de la empresa, y que esta produce al mínimo coste. La pregunta es entonces qué precio fijar.

[4] La intuición es que la mayor parte del coste es una infraestructura que sirve igual a pocos que a muchos usuarios, como una red de tuberías o de cables, y duplicarla solo añade coste.

[5] En el caso multiproducto ni siquiera son suficientes: la subaditividad depende también de las complementariedades de costes entre los productos y de las economías de alcance (Braeutigam, 1989, pp. 1295-1296).

[6] La intuición es que el regulador busca el mejor resultado posible dentro de dos límites: la empresa no puede tener pérdidas permanentes, porque dejaría de prestar el servicio, y el regulador solo puede condicionar sus reglas a lo que es capaz de comprobar.

 © 2026 Manuel Cabado García

Precio igual al coste marginal

La primera respuesta, que se remonta a Dupuit (1844) y que formalizó Hotelling (1938), es la regla de eficiencia de la competencia perfecta, $P = CMg$. Con ella se alcanza el óptimo de primer orden, porque cada unidad se produce mientras el valor que le dan los consumidores supera su coste. El problema aparece en el monopolio natural típico, en el que la producción eficiente se sitúa en el tramo decreciente del coste medio, CMe. Allí el coste marginal es menor que el coste medio y la empresa incurre en pérdidas, como muestra la Figura 18.1. Si, en cambio, la producción eficiente se situara en el tramo creciente del coste medio, el precio igual al coste marginal no causaría pérdidas, pero el monopolio podría no ser sostenible y la regulación tendría que limitar la entrada.

Hotelling proponía cubrir ese déficit con impuestos de suma fija, pero la propuesta tiene tres limitaciones.

1. Los impuestos realmente disponibles distorsionan otros mercados, así que cada euro de subvención cuesta a la sociedad más de un euro.
2. Hacer pagar a los contribuyentes un servicio que no usan es discutible en términos de equidad.
3. Por último, una empresa que sabe que sus pérdidas serán cubiertas pierde el incentivo a controlar sus costes.

Además, Coase (1946) añadió una crítica de fondo: si el precio no cubre el coste total, el regulador nunca sabe si los usuarios valoran el servicio lo suficiente como para justificar su existencia.

Precio igual al coste medio

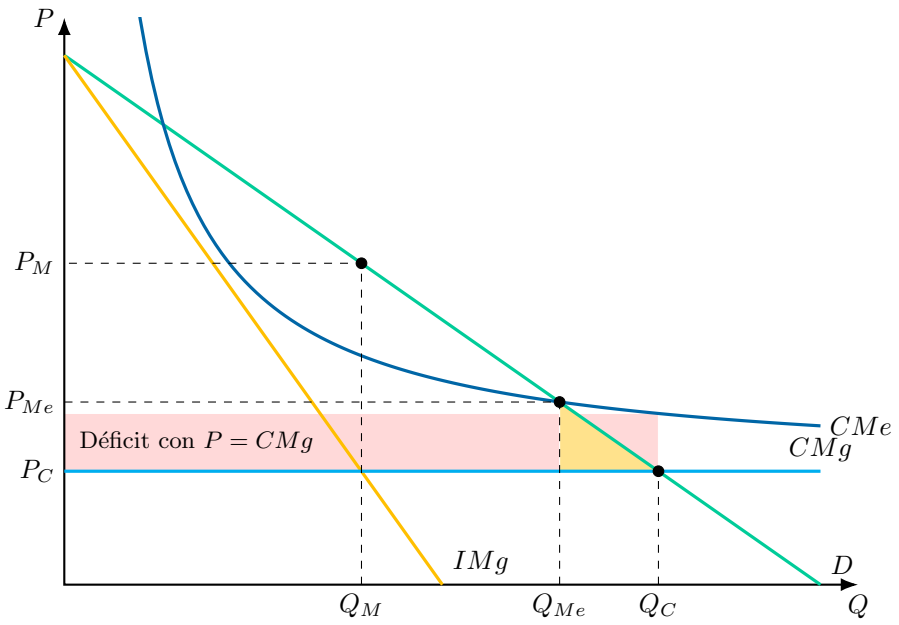
La alternativa más simple es exigir que el precio cubra exactamente el coste medio, $P = CMe$, de modo que la empresa no tenga pérdidas ni beneficios extraordinarios. Es un óptimo de segundo orden: la producción queda por debajo de la eficiente y reaparece una pérdida de eficiencia asignativa, aunque menor que la del monopolio sin regular. Su limitación principal es de incentivos, porque si el precio se revisa para cubrir cualquier coste, la empresa no gana nada reduciéndolo.

Gráficamente, en un plano de precio, P, y cantidad, Q, como el de la Figura 18.1, se representan la demanda, D, el ingreso marginal, IMg, el coste medio decreciente, CMe, y el coste marginal constante, CMg. Así:

- Sin regular, el monopolio produce la cantidad Q_M, en la que el ingreso marginal iguala al coste marginal, y cobra el precio P_M.
- Con la regla del coste medio, produce Q_{Me} al precio P_{Me}, con una pérdida de eficiencia pequeña (el triángulo sombreado).

- Con la regla del coste marginal, produce la cantidad eficiente, Q_C, al precio P_C, con un déficit igual al rectángulo sombreado, que coincide con el coste fijo porque el coste marginal es constante.

Figura 18.1: Reglas de precio en el monopolio natural



Tarifas no lineales

Una tercera vía renuncia a cobrar un precio único por unidad. En la tarifa en dos partes (Coase, 1946; Oi, 1971), cada usuario paga una cuota fija, A, que reparte el coste fijo de la empresa, CF, entre sus n usuarios, y un precio por unidad, p, igual al coste marginal. La factura, T, de un usuario que consume q unidades es entonces:

$$T = A + p \cdot q, \qquad A = \frac{CF}{n}, \qquad p = CMg$$

En el margen el precio es eficiente y la cuota cubre el déficit, como lo haría un impuesto de suma fija pagado solo por los usuarios.

Su limitación es la exclusión, porque los usuarios cuyo excedente es menor que la cuota abandonan el mercado. Por eso suele ofrecerse un menú de tarifas, en el que cada usuario elige entre pagar solo un precio por unidad superior al coste marginal o pagar una cuota fija más alta a cambio de un precio por unidad más bajo, que puede llegar al coste marginal. Si la cuota es lo bastante alta como para que la empresa ingrese con cada usuario que cambia de opción al menos lo que ingresaba con el precio único, nadie empeora y los grandes consumidores mejoran, por lo que el menú supone una mejora

 © 2026 Manuel Cabado García

paretiana (Braeutigam, 1989).[7] La factura eléctrica española, con un término de potencia y un término de energía, responde a esta lógica.

Por otro lado, cuando una misma capacidad atiende demandas que varían a lo largo del día o del año, como ocurre con la electricidad, la tarificación según la carga formalizada por Steiner (1957) carga el coste de la capacidad sobre los usuarios de las horas punta, que son quienes obligan a instalarla, y cobra en las horas valle solo el coste de operación.[8]

Precios de Ramsey-Boiteux

Cuando la empresa produce varios bienes y no puede discriminar entre usuarios, la cuestión es cómo repartir el coste fijo entre los precios. La respuesta de Boiteux (1956), análoga a la que Ramsey (1927) había dado para la imposición óptima, consiste en maximizar el bienestar sujeto a que la empresa no tenga pérdidas.

Baumol y Bradford (1970) difundieron después el resultado y mostraron la equivalencia entre ambos problemas: si las demandas de los bienes son independientes, el margen del precio de cada bien, P_i, sobre su coste marginal, CMg_i, debe ser inversamente proporcional al valor absoluto de la elasticidad-precio de su demanda, $|\varepsilon_i|$, en una proporción que depende del multiplicador de la restricción de beneficio, λ, que mide cuánto bienestar se pierde por exigir que la empresa cubra sus costes:

$$\frac{P_i - CMg_i}{P_i} = \frac{\lambda}{1+\lambda} \cdot \frac{1}{|\varepsilon_i|} \tag{18.1}$$

El cociente $\lambda/(1+\lambda)$ es el número de Ramsey, comprendido entre cero y uno.[9] Que la restricción obligue significa que, sin ella, el regulador fijaría precios distintos. Cuanto mayor es el coste fijo que hay que cubrir, más tiene que alejarse de los precios eficientes y más bienestar cuesta cada euro de beneficio que exige a la empresa, que es lo que mide λ:

- Si la restricción apenas obliga, λ tiende a cero y el precio tiende al coste marginal.
- Si obliga mucho, el número de Ramsey tiende a uno y los precios se acercan a los de monopolio.[10]

[7] La intuición es que el gran consumidor solo elige la cuota si le compensa por las unidades adicionales que compra a un precio más bajo, mientras que la empresa sigue cobrándole, a través de la cuota, el margen que antes obtenía de él. Willig (1978) demostró que cualquier precio uniforme distinto del coste marginal puede mejorarse en sentido de Pareto con una tarifa no lineal ofrecida como opción adicional (Braeutigam, 1989, p. 1329).

[8] El desarrollo formal, incluido el caso en que la punta se desplaza al otro periodo, figura en el Apéndice 18.A.2. El principio de cargar toda la capacidad sobre la punta depende de la tecnología: con tecnologías de proporciones variables o periodos de distinta duración, los usuarios del valle también contribuyen a pagar la capacidad (Braeutigam, 1989, pp. 1316-1320).

[9] La intuición es que cubrir el coste fijo exige alejar los precios del coste marginal, y la pérdida de eficiencia es menor si el margen se carga donde los consumidores apenas reducen su consumo cuando sube el precio. Braeutigam (1989, p. 1323) reserva el nombre de número de Ramsey para el producto del margen relativo por la elasticidad-precio con su signo, cuyo valor es $-\lambda/(1+\lambda)$, y el cociente del texto es su valor absoluto.

[10] La intuición es que λ mide cuánto bienestar cuesta exigir a la empresa un euro más de beneficio. Si

Así, esta es la regla de la elasticidad inversa aplicada a los precios públicos, que carga más margen sobre los bienes de demanda más rígida.[11]

Gráficamente, en dos planos de precio, P_i, y cantidad, Q_i, como los de la Figura 18.2, se representan un mercado de demanda rígida, D_1, y otro de demanda elástica, D_2, con el mismo coste marginal constante, CMg. Así:

- En el mercado 1, el precio de Ramsey-Boiteux, P_1^R, queda muy por encima del coste marginal.
- En el mercado 2, el margen es menor.
- En ambos, las cantidades de Ramsey-Boiteux, Q_1^R y Q_2^R, se reducen en la misma proporción respecto a las eficientes, Q_1^C y Q_2^C, pero en el mercado de demanda rígida eso exige un margen mucho mayor.
- Los rectángulos son los ingresos que cubren el coste fijo y los triángulos, las pérdidas de eficiencia.

Sin embargo, los precios de Ramsey-Boiteux tienen tres limitaciones.

1. Son regresivos, porque los bienes de demanda rígida suelen ser de primera necesidad y pesan más en el gasto de los hogares de menor renta.
2. Exigen conocer elasticidades y costes que, en la práctica, solo conoce bien la propia empresa.
3. Por último, solo son óptimos si no hay otras distorsiones en la economía, pues en caso contrario la teoría del segundo óptimo (Lipsey y Lancaster, 1956) obliga a tenerlas en cuenta.

Regulación de la tasa de rendimiento y teoría positiva

En la práctica, ningún regulador conoce costes y elasticidades con la precisión que exigen las reglas anteriores. Durante buena parte del siglo XX, sobre todo en Estados Unidos, se optó por un mecanismo mucho más sencillo: limitar la rentabilidad de la empresa.

Regulación de la tasa de rendimiento

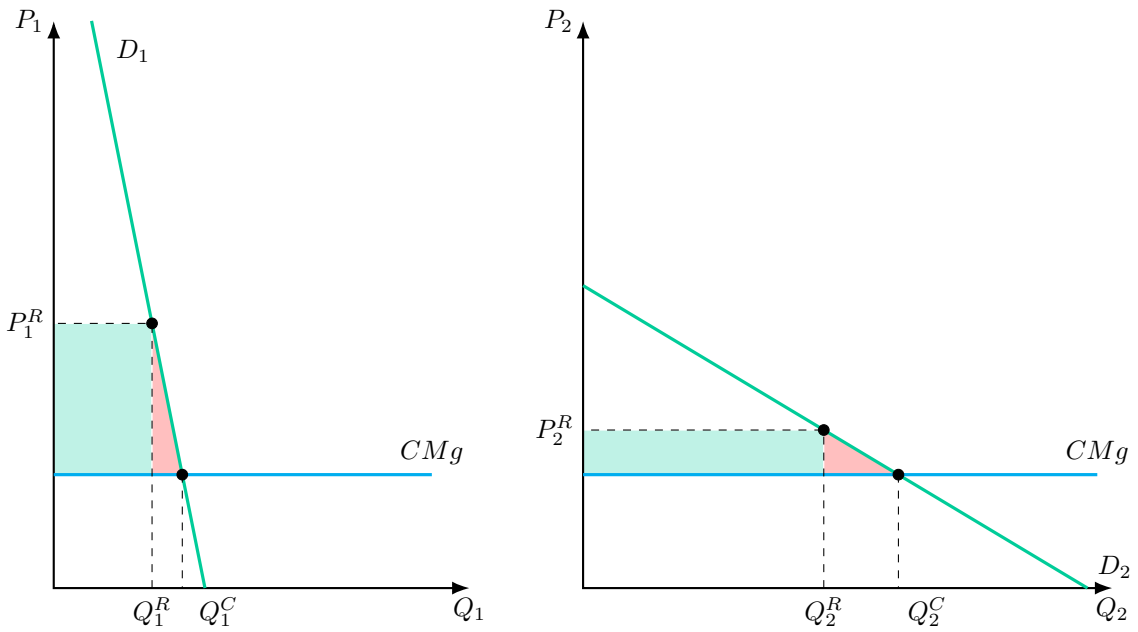
La regulación de la tasa de rendimiento, o por coste del servicio, es la que fija precios que permiten a la empresa recuperar sus costes de operación y obtener sobre el capital invertido, K, una tasa de rendimiento máxima, s, superior al coste de oportunidad del

el coste fijo que hay que cubrir es pequeño, basta con subir un poco los precios sobre el coste marginal, y ese euro apenas cuesta bienestar, por lo que λ es casi nulo. Si es grande, hay que subir mucho los precios, incluso en los mercados de demanda elástica. En el límite, solo los precios de monopolio, que maximizan el beneficio, consiguen cubrir los costes, y el número de Ramsey se acerca a uno.

[11] La derivación, también con demandas relacionadas, figura en el Apéndice 18.A.1.

 © 2026 Manuel Cabado García

Figura 18.2: Precios de Ramsey-Boiteux con dos mercados



Demanda rígida

Demanda elástica

capital, r, pero inferior a la que obtendría sin regular. Si los ingresos totales, $IT(K, L)$, dependen del capital y del trabajo, L, cuyo precio es el salario, w, la restricción puede escribirse también en términos del beneficio económico, $\pi = IT(K, L) - w \cdot L - r \cdot K$:

$$\frac{IT(K, L) - w \cdot L}{K} \leq s \quad \Longleftrightarrow \quad \pi \leq (s - r) \cdot K$$

Averch y Johnson (1962) demostraron que, como cada unidad adicional de capital eleva en $s-r$ el beneficio máximo permitido, la empresa elige una combinación de factores con más capital del que minimiza los costes. Es el efecto Averch-Johnson o sobrecapitalización.[12]

Gráficamente, en un plano de capital, K, y trabajo, L, como el de la Figura 18.3, se representan la isocuanta q_0 y tres isocostes. Así:

- La combinación que minimiza los costes de producir q_0 es E, donde la isocuanta es tangente al isocoste CT_E, cuya pendiente es el cociente entre el salario y el coste del capital.
- La empresa regulada se comporta como si el capital fuera más barato y elige la combinación AJ, tangente a un isocoste más inclinado, $\widetilde{CT}$, con más capital y menos trabajo.
- A los precios reales de los factores, esa combinación cuesta CT_{AJ}, más que CT_E.[13]

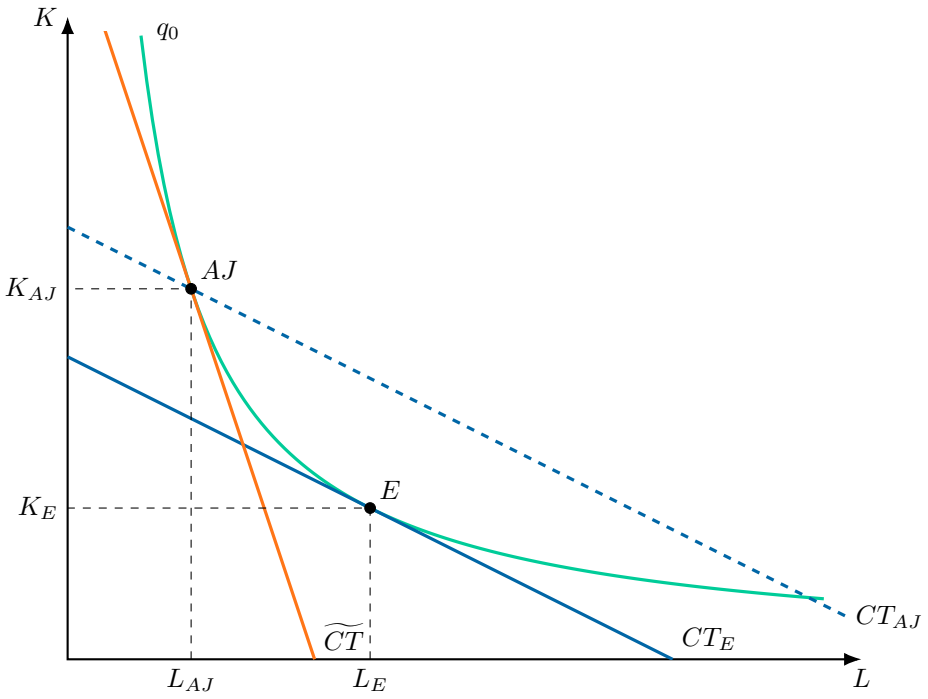
[12] Laffont y Martimort (2002, p. 18) consideran que la importancia de este resultado se sobrevaloró.

[13] La intuición es que, si el beneficio permitido es un porcentaje del capital, invertir de más no es un despilfarro para la empresa, sino una forma de ampliar la base sobre la que puede ganar dinero. La

Su ventaja es informativa, porque solo exige observar beneficios contables y capital, comprobables mediante auditoría, y da a la empresa seguridad para recuperar inversiones hundidas.[14]

Sus limitaciones son la ineficiencia productiva, la ausencia de incentivos a reducir costes, porque cualquier coste se traslada al precio, y una estructura de precios que no sigue la lógica de Ramsey-Boiteux. Tampoco incentiva la innovación, porque cualquier ahorro de costes se traslada al precio en la siguiente revisión. Además, el regulador debe fijar s sin conocer la tecnología de la empresa.

Figura 18.3: Efecto Averch-Johnson en el plano capital-trabajo



Teoría positiva de la regulación

Los instrumentos anteriores suponen un regulador benevolente, y la Escuela de Chicago puso en duda ese supuesto. Stigler (1971) sostuvo que la regulación es un bien que se intercambia en un mercado político y que, por regla general, la adquiere la propia industria y se diseña en su beneficio, porque los productores son pocos, están bien organizados y tienen mucho en juego, mientras que los consumidores están dispersos. Es la teoría de la captura del regulador.[15]

derivación con el lagrangiano figura en el Apéndice 18.A.3.

[14] Baron (1989, p. 1352) reinterpreta el modelo de Averch y Johnson como un mecanismo de regulación basado en los dos únicos datos que observa el regulador, el beneficio contable y el capital.

[15] La intuición es que un pequeño grupo que gana mucho con una norma tiene más motivos para presionar que millones de consumidores que pierden poco cada uno, como ya había señalado Olson (1965) con la lógica de la acción colectiva.

 © 2026 Manuel Cabado García

Peltzman (1976) generalizó el argumento: el regulador maximiza su apoyo político y reparte las rentas entre productores y consumidores, con lo que el precio regulado queda entre el competitivo y el de monopolio. Becker (1983) lo modelizó como una competencia entre grupos de presión en la que los costes de eficiencia de la regulación limitan su propio alcance, porque cuanto mayor es la pérdida irrecuperable, más se resiste el grupo perjudicado. La implicación de política es que la limitación de la regulación no es solo informativa sino también institucional, y justifica reguladores independientes, reglas transparentes y la evaluación de resultados.

18.3. Desarrollos en presencia de información asimétrica ⏳ 5'30"

Los instrumentos anteriores tropiezan con una misma dificultad: la empresa regulada conoce mejor que el regulador sus costes y el esfuerzo que dedica a reducirlos. Desde finales de los años setenta, la nueva economía de la regulación trató este problema como una relación entre un principal, el regulador, y un agente, la empresa, y se preguntó cuál es la mejor regulación posible dada esa asimetría de información (Laffont y Tirole, 1993; Armstrong y Sappington, 2007).

Información asimétrica en la regulación Hay información asimétrica en la regulación cuando la empresa regulada conoce variables relevantes que el regulador no puede observar. El problema adopta dos formas:

1. Selección adversa cuando la empresa conoce su tecnología, por ejemplo si es de coste alto o bajo, y puede declarar costes superiores a los reales.
2. Riesgo moral cuando el regulador no observa el esfuerzo de los gestores por reducir costes, y entonces un coste alto puede deberse a una mala tecnología o a la falta de esfuerzo.

En los dos casos, el regulador solo consigue que la empresa revele lo que sabe o se esfuerce si le deja una parte del excedente, que es la renta informativa.

Renta informativa La renta informativa es el beneficio extraordinario que el regulador debe dejar a la empresa eficiente para que no le compense hacerse pasar por ineficiente.[16] Toda la teoría gira en torno a un mismo conflicto: reducir esa renta exige distorsionar la producción o el esfuerzo de la empresa, y mejorar la eficiencia exige dejarle más renta.

[16] La intuición es que una empresa de costes bajos siempre puede declarar costes altos y quedarse con la diferencia, de modo que el regulador solo consigue que diga la verdad si le paga, al menos, lo que ganaría mintiendo.

Selección adversa: de Loeb y Magat a Baron y Myerson

Loeb y Magat (1979) fueron los primeros en plantear la regulación como un problema de información:

- Proponían dejar que la empresa fijara libremente su precio y pagarle una subvención igual al excedente de los consumidores.[17]
- Al apropiarse de todo el excedente, la empresa elige por sí misma un precio igual al coste marginal, cualquiera que sea su coste y sin necesidad de revelarlo.
- El mecanismo es eficiente, pero transfiere a la empresa todo el excedente y exige conocer la demanda, por lo que solo es aceptable si los fondos públicos no tienen coste social y no importa redistribuir renta hacia la empresa (Laffont y Martimort, 2002, p. 18).

Por su parte, Baron y Myerson (1982) convirtieron el problema en uno de segundo óptimo al suponer que el regulador valora menos el beneficio de la empresa que el excedente de los consumidores, de modo que dejarle rentas tiene un coste.

En la versión más sencilla, la empresa produce una cantidad, q, que los consumidores valoran en $S(q)$, con un coste marginal que solo ella conoce y que puede ser bajo, $\underline{\theta}$, con probabilidad ν, o alto, $\overline{\theta}$, con probabilidad $1-\nu$.

Llamando $\Delta\theta = \overline{\theta} - \underline{\theta}$ a la diferencia de costes, el regulador ofrece un menú de contratos, uno para cada tipo, con una cantidad y una transferencia, $(\underline{q}, \underline{t})$ y $(\overline{q}, \overline{t})$. La renta de cada tipo es lo que cobra menos lo que le cuesta producir, $\underline{U} = \underline{t} - \underline{\theta} \cdot \underline{q}$ y $\overline{U} = \overline{t} - \overline{\theta} \cdot \overline{q}$. El regulador elige el menú que maximiza el bienestar esperado sujeto a dos restricciones:

- Compatibilidad de incentivos: cada tipo prefiere el contrato pensado para él. Para la empresa de coste bajo, $\underline{U} \geq \overline{U} + \Delta\theta \cdot \overline{q}$, porque si eligiera el contrato de la de coste alto cobraría su transferencia y produciría $\overline{q}$ con un coste menor.
- Participación: cada tipo acepta su contrato, $\underline{U} \geq 0$ y $\overline{U} \geq 0$.

Si, para simplificar, el regulador no da ningún peso al beneficio de la empresa, en el óptimo solo son activas la compatibilidad de la empresa de coste bajo y la participación de la de coste alto, de modo que la de coste alto no obtiene renta y la de coste bajo obtiene una renta informativa igual a lo que ganaría produciendo con su coste más bajo la cantidad pensada para la de coste alto (Laffont y Martimort, 2002, pp. 41-43):

$$\overline{U} = 0, \qquad \underline{U} = \Delta\theta \cdot \overline{q}$$

Sustituyendo estas rentas, el regulador elige las dos cantidades para maximizar el exce-

[17] Es una aplicación a la regulación de los mecanismos de Groves, en los que decir la verdad es una estrategia dominante (Laffont y Martimort, 2002, p. 18).

 © 2026 Manuel Cabado García

dente esperado menos la renta que deja a la empresa de coste bajo:

$$\max_{\underline{q},\overline{q}} \ \nu \cdot \left[S(\underline{q}) - \underline{\theta} \cdot \underline{q}\right] + (1-\nu) \cdot \left[S\left(\overline{q}\right) - \overline{\theta} \cdot \overline{q}\right] - \nu \cdot \Delta\theta \cdot \overline{q}$$

Como esa renta crece con la cantidad que se pide a la empresa de coste alto, el regulador la reduce por debajo de la eficiente. La empresa de coste bajo produce la cantidad de información completa, $S'(\underline{q}) = \underline{\theta}$, lo que se conoce como ausencia de distorsión en el extremo superior, y la de coste alto produce hasta que el valor marginal para los consumidores, $S'\left(\overline{q}\right)$, iguala su coste marginal más un recargo informativo:

$$S'\left(\overline{q}\right) = \overline{\theta} + \frac{\nu}{1-\nu} \cdot \Delta\theta \tag{18.2}$$

El recargo es mayor cuanto más probable es que la empresa sea eficiente y cuanto mayor es la diferencia de costes. De ahí que, con información asimétrica, la regulación óptima sea un menú de contratos que acepta cierta ineficiencia asignativa para extraer rentas, con un precio de la empresa menos eficiente superior a su coste marginal.[18]

Selección adversa y riesgo moral: Laffont y Tirole

Laffont y Tirole (1986) añadieron el riesgo moral y una justificación distinta del coste de las rentas: cada euro transferido a la empresa se financia con impuestos distorsionantes y cuesta a la sociedad $1+\lambda$ euros, donde λ es el coste sombra de los fondos públicos. El coste observado de prestar el servicio, C, depende de un parámetro de eficiencia que solo conoce la empresa, β, y del esfuerzo de sus gestores, e, que el regulador no observa y que les supone una desutilidad creciente y convexa, $\psi\left(e\right)$:

$$C = \beta - e$$

El regulador observa el coste, pero no puede separar qué parte se debe a la tecnología y cuál al esfuerzo: un coste bajo puede proceder de una empresa eficiente o de una ineficiente que se esfuerza mucho.

La solución es un menú de contratos lineales. La empresa anuncia un coste previsto, C^a, y recibe un pago total, TR, formado por un pago fijo, a, y una fracción, b, de la desviación entre el coste real y el anunciado, ambos dependientes del anuncio:

$$TR = a\left(C^a\right) + b\left(C^a\right) \cdot \left(C - C^a\right)$$

De esta forma:

[18] La derivación completa, incluida la comprobación de que las restricciones omitidas se cumplen, figura en el Apéndice 18.A.4.

- Con $b = 0$, contrato de precio fijo, la empresa asume todo el sobrecoste y se queda con todo el ahorro, lo que da el máximo incentivo al esfuerzo.
- Con $b = 1$, contrato de reembolso de costes o cost plus, el regulador cubre cualquier desviación y la empresa no tiene incentivo alguno a reducir costes.

En el óptimo, la empresa eficiente elige un contrato próximo al de precio fijo, ejerce el esfuerzo eficiente y obtiene una renta informativa. La ineficiente elige un contrato con mayor reembolso de costes, no obtiene renta y se le induce un esfuerzo inferior al eficiente, porque exigirle más esfuerzo aumentaría la renta que hay que dejar a la eficiente (Laffont y Tirole, 1993).

Laffont y Tirole mostraron además que, bajo ciertas condiciones de separabilidad de los costes, la fijación de precios y los incentivos pueden tratarse por separado: los precios siguen una regla de tipo Ramsey-Boiteux, en la que el coste de los fondos públicos hace de multiplicador, y la extracción de rentas se resuelve con el menú de transferencias. Los esquemas reales se interpretan con este menú: el price cap es un contrato de alto poder, próximo al de precio fijo, y la regulación por coste del servicio es uno de bajo poder, próximo al cost plus.[19]

Implicaciones: regulación por incentivos

Estos resultados dan fundamento teórico al paso, desde los años ochenta, de la regulación por coste del servicio a mecanismos que desvinculan el precio de los costes declarados por la empresa, un paso impulsado también por la sobrecapitalización y por las privatizaciones británicas.

Precio máximo o price cap

El price cap es una regulación que fija una senda de precios máximos para una cesta de servicios que crece con la inflación menos un factor de eficiencia, X, durante un periodo regulatorio de varios años. Lo propuso Littlechild (1983) para British Telecom. Si P_t es el precio máximo en el año t e IPC la tasa de inflación:

$$P_t = P_{t-1} \cdot (1 + IPC - X)$$

Como el precio no depende de los costes que la empresa logre, cada euro que ahorra durante el periodo es un euro de beneficio. El price cap incentiva así la eficiencia productiva y traslada las ganancias al consumidor a través de X y de las revisiones periódicas, en las que el precio vuelve a alinearse con los costes.

Cuanto más largo es el periodo, mayor es el incentivo, pero más tarda el consumidor en beneficiarse. Cuando el tope se aplica a la media ponderada de una cesta, la empresa

[19] El desarrollo figura en el Apéndice 18.A.5.

 © 2026 Manuel Cabado García

elige los precios relativos y, si las ponderaciones son las cantidades vendidas en el periodo anterior, tiende a una estructura próxima a la de Ramsey-Boiteux que el regulador no necesita calcular (Armstrong y Sappington, 2007, pp. 1617-1618).[20]

Sus limitaciones son la dificultad de fijar el precio inicial y X sin información sobre los costes, la tentación de empeorar la calidad para ahorrar costes, que obliga a regularla por separado, y el efecto trinquete: si la empresa anticipa que sus ahorros se trasladarán a un X más exigente en la siguiente revisión, reduce su esfuerzo al acercarse el final del periodo. Además, la empresa puede ahorrar aplazando inversiones de larga vida y, si obtiene beneficios elevados, el regulador puede verse presionado a revisar la senda antes de tiempo, lo que resta credibilidad al esquema.[21]

Esquemas intermedios

Entre el coste del servicio, que no da incentivos pero tampoco deja rentas, y el price cap, que da incentivos máximos pero puede dejar rentas elevadas, existen esquemas intermedios. En el reparto de beneficios o escala móvil, la empresa retiene solo una parte de los beneficios que superan una banda de rentabilidad y comparte el resto con los consumidores.

Compromiso y captura

El modelo básico supone que el regulador se compromete a respetar el contrato y que es benevolente. Relajar estos supuestos explica buena parte de los problemas que se observan en la práctica.

Compromiso y efecto trinquete El efecto trinquete es la tendencia de una empresa regulada a ocultar su eficiencia o reducir su esfuerzo cuando anticipa que sus buenos resultados endurecerán el contrato siguiente. Si el contrato se revisa periódicamente, la empresa que revela costes bajos hoy recibirá mañana un contrato más exigente.
Anticipándolo, la empresa eficiente oculta su eficiencia o modera su esfuerzo, y el regulador debe dejarle rentas mayores para que revele su tipo (Weitzman, 1980; Freixas et al., 1985).
Es un problema de inconsistencia temporal como el que Kydland y Prescott (1977) analizaron en la política económica, y su remedio es el compromiso: periodos regulatorios largos, reglas fijadas de antemano y reguladores con reputación que proteger. El intervalo

[20] La intuición es que, con las cantidades ya vendidas como ponderaciones, la empresa solo gana si sube los precios donde la demanda reacciona poco, que es lo que pide la regla de la elasticidad inversa. Si se ponderase con las cantidades del propio periodo, la empresa podría relajar el tope vendiendo más de los productos baratos, y los consumidores saldrían perdiendo (Armstrong y Sappington, 2007, pp. 1612-1614).

[21] Un esquema de alto poder traslada además a la empresa más riesgo que la regulación de la tasa de rendimiento, lo que puede elevar su coste de capital (Cabral, 2000, pp. 77-78).

entre revisiones, o retraso regulatorio, actúa en el mismo sentido, porque mientras el precio no se revisa la empresa se queda con los ahorros. Por eso un price cap que se reajusta a los costes observados funciona como una regulación por coste del servicio con un retraso fijado de antemano, y la frecuencia óptima de las revisiones equilibra el incentivo a reducir costes con la eficiencia asignativa (Armstrong y Sappington, 2007, pp. 1624-1625).

Captura del regulador Laffont y Tirole (1991) formalizaron la captura como una colusión entre la empresa y el regulador, que dispone de información que el legislador no tiene. Cuanto más poderosos son los incentivos, mayores son las rentas en juego y mayor el riesgo de que la empresa pague al regulador para que oculte lo que sabe. Por eso, cuando el riesgo de captura es alto, conviene usar contratos de menor poder, más próximos al reembolso de costes, aunque se pierda eficiencia productiva.

18.4. Regulación óptima con varias empresas ⏳ 4'

Hasta aquí se ha supuesto una única empresa regulada. Cuando hay varias, el regulador dispone de una fuente adicional de información: al igual que una auditoría de costes, la comparación entre empresas similares o la competencia por obtener la concesión reducen las rentas que hay que dejar a cada una (Armstrong y Sappington, 2007). La presencia de varias empresas plantea además un problema nuevo, el del acceso de los rivales a la red del antiguo monopolista.

Competencia por comparación

La competencia por comparación (Shleifer, 1985) aprovecha la existencia de varias empresas reguladas en mercados similares, como las distribuidoras eléctricas de distintas zonas, y fija el precio de cada una según los costes de las demás, no según los suyos. Ninguna puede mejorar su retribución inflando sus costes y solo gana si es más eficiente que la media. Así se induce la reducción de costes sin que el regulador necesite conocer la tecnología. Sus limitaciones son el riesgo de colusión entre las empresas comparadas y la heterogeneidad de sus entornos, que obliga a corregir las comparaciones. Shleifer mostró que el mecanismo se puede extender a empresas heterogéneas siempre que sus diferencias sean observables, ajustando la referencia por esas características. Él mismo citaba como ejemplo el pago de Medicare a los hospitales estadounidenses por grupos de diagnóstico, que fija un importe por caso a partir de los costes de hospitales comparables. La comparación funciona porque las empresas similares comparten perturbaciones de coste comunes, así que permite separar la parte del coste que no depende del esfuerzo de cada una (Armstrong y Sappington, 2007).

 © 2026 Manuel Cabado García

Competencia por el mercado

Demsetz (1968) cuestionó la necesidad misma de regular: si no puede haber competencia en el mercado, puede haberla por el mercado. Basta con subastar el derecho a prestar el servicio en exclusiva durante un periodo y adjudicarlo a quien ofrezca el precio más bajo para unas condiciones de calidad dadas.[22] Si se puja por un precio único, la competencia lleva el precio al coste medio, un óptimo de segundo orden, sin que el regulador necesite conocer los costes. La subasta selecciona además a la empresa más eficiente, que es la que puede ofrecer el precio más bajo. Produce competencia ex ante y su repetición periódica, competencia ex post, y la renta del ganador es menor cuanto más licitadores compiten (Armstrong y Sappington, 2007, p. 1648).

El diseño de la subasta también importa. Laffont y Tirole (1987) unieron la subasta y la regulación por incentivos: si lo que se subasta es un contrato de incentivos, el ganador recibe un contrato lineal cuyo pago fijo depende de la segunda mejor oferta, de modo que la competencia entre licitadores reduce su renta informativa, y la parte de los sobrecostes que soporta es mayor cuanto más eficiente se declara. En las subastas de valor común, además, el ganador tiende a ser quien más infravaloró sus costes, la llamada maldición del ganador, lo que lleva a los licitadores a pujar con prudencia. El diseño de formatos para bienes complejos, como las frecuencias del espectro radioeléctrico, valió el Premio Nobel de Economía a Milgrom y Wilson (The Royal Swedish Academy of Sciences, 2020), este último autor de la teoría de las subastas de valor común.

Sus limitaciones son conocidas:

- Puede haber colusión entre los licitadores.
- Además, como señaló Williamson (1976), cuando el servicio exige inversiones específicas de larga duración es muy difícil especificar un contrato completo y valorar y transferir esos activos al renovar la concesión, y el adjudicatario acumula una ventaja informativa que reduce la competencia en la renovación.
- La duración de la concesión plantea una disyuntiva. Si es corta, el concesionario no tiene incentivos a invertir en activos de larga vida ni a mantenerlos. Si es larga, desaparece la competencia periódica y es más difícil prever en el contrato todas las contingencias.

Relaciones verticales y precio de acceso

Cuando la red es un monopolio natural pero los servicios que se prestan sobre ella pueden ser competitivos, como en la electricidad, el gas, el ferrocarril o las telecomunicaciones, el regulador debe decidir cómo acceden los rivales a la red del antiguo monopolista. La regla

[22] La intuición es que la competencia entre los aspirantes antes de la adjudicación sustituye a la competencia que no puede existir después, porque cada uno ofrece el precio más bajo con el que todavía cubre sus costes.

del precio eficiente de los componentes fija el peaje en el coste de dar acceso más el margen que el incumbente pierde en el mercado final, de modo que solo entran rivales al menos tan eficientes como él, aunque perpetúa ese margen si el precio final es de monopolio (Willig, 1979; Baumol y Sidak, 1994; Armstrong et al., 1996).[23] Armstrong, Doyle y Vickers (1996) mostraron que la regla solo es óptima en condiciones restrictivas y que, en general, el precio de acceso óptimo depende de las elasticidades y tiene una estructura de tipo Ramsey-Boiteux. Frente al acceso regulado sin separación, la alternativa estructural es separar verticalmente la red de los servicios, desde la separación operativa hasta la separación de la propiedad, con una disyuntiva entre las economías de integración y el incentivo del propietario de la red a discriminar a sus rivales.[24]

Mercados de dos caras

Un mercado de dos caras es aquel en el que una plataforma reúne a dos grupos de usuarios, como compradores y vendedores, y el valor que obtiene cada grupo depende del número de usuarios del otro. Las plataformas digitales son el caso típico (Rochet y Tirole, 2003): por esos efectos de red cruzados, el precio de cada lado no refleja solo su coste y puede ser óptimo cobrar un precio nulo a uno de ellos.[25] Por eso el margen de un lado, medido con el índice de Lerner, no basta para diagnosticar el poder de mercado. Además, los efectos de red, los costes hundidos y el control de los datos hacen que estos mercados no sean disputables en el sentido de Baumol, Panzar y Willig (1982), y la regulación tiende a centrarse en las conductas de acceso y no en los precios.

Balance En definitiva, la regulación óptima no elimina el poder de mercado sin coste, sino que equilibra eficiencia, rentas y credibilidad del regulador. La tabla 18.1 resume los efectos y limitaciones de los instrumentos estudiados.

18.5. Aplicaciones prácticas ⏳ 3’

Los instrumentos estudiados tienen aplicaciones de política económica. Las redes energéticas, las telecomunicaciones y las plataformas digitales se regulan hoy con combinaciones de ellos, y su evolución reciente reproduce la secuencia del tema: de la tasa de rendimiento a los incentivos, y de la regulación de precios a la regulación del acceso.[26]

[23] Por ejemplo, si el incumbente vende el servicio a 10 euros, dar acceso le cuesta 5 y el resto del servicio 2, la regla fija un peaje de $5 + (10 - 5 - 2) = 8$ euros, el coste del acceso más el margen que pierde el incumbente, y un entrante solo cubre costes si su coste propio no supera 2 euros, es decir, si es al menos tan eficiente como el incumbente. El desarrollo de la regla y de su alternativa, los precios de acceso óptimos de tipo Ramsey-Boiteux, figura en el Apéndice 18.A.6.

[24] Las distintas formas de separación vertical se describen en el Apéndice 18.A.7.

[25] La intuición es que un buscador gratuito atrae a muchos usuarios, y son los anunciantes, que quieren llegar a ellos, quienes pagan el servicio.

[26] El marco normativo de referencia de estas aplicaciones figura en el Apéndice 18.A.8.

 © 2026 Manuel Cabado García

Cuadro 18.1: Efectos y limitaciones de los instrumentos de regulación

Instrumento	**Eficiencia asignativa**	**Eficiencia productiva**	**Eficiencia dinámica**	**Limitación principal**
$P = CMg$ con subvención	Óptima	Débil: las pérdidas se cubren	Débil	Coste de los fondos públicos
$P = CMe$	Pérdida pequeña	Débil	Débil	Sin incentivo a reducir costes
Tarifa en dos partes	Óptima en el margen	Débil	Débil	Exclusión de usuarios
Ramsey-Boiteux	La mejor sin pérdidas	Débil	Débil	Regresividad e información
Tasa de rendimiento	Precios sin lógica de Ramsey	Sobrecapitalización	Escasa	Efecto Averch-Johnson
Subasta de Demsetz	$P = CMe$	Alta al pujar	Depende de la duración	Activos específicos al renovar
Price cap	Precios relativos próximos a Ramsey	Alta	Alta si el periodo es largo	Calidad, trinquete e infrainversión
Competencia por comparación	Óptima si las empresas son comparables	Alta	Media	Colusión y heterogeneidad
Menú de Laffont y Tirole	Distorsión en el tipo ineficiente	Alta en el tipo eficiente	Depende del compromiso	Rentas informativas

Redes eléctricas y gasistas en España

En España, el transporte y la distribución de electricidad y de gas son actividades reguladas cuya retribución fija la Comisión Nacional de los Mercados y la Competencia (CNMC), creada por la Ley 3/2013 al integrar la autoridad de competencia y los reguladores sectoriales:

- Tasa de rendimiento: cada empresa recibe la amortización de sus activos más una tasa de retribución financiera (TRF) aplicada a su valor neto.
- Cálculo de la TRF: es el coste medio ponderado del capital expresado antes de impuestos, a partir de la tasa libre de riesgo, la prima de riesgo del mercado, la beta del sector, el coste de la deuda y el apalancamiento (Circular 2/2019). Fue del 5,58 % para las redes eléctricas en 2020-2025, y la Circular 9/2025 la elevó al 6,58 % para 2026-2031, tras una propuesta inicial del 6,46 %.[27]
- Incentivos: la retribución de la distribución eléctrica premia o penaliza la calidad del suministro, medida por el tiempo y el número de interrupciones (TIEPI y NIEPI),

[27] Para el transporte y la regasificación de gas, la Circular 2/2019 fijó una TRF del 5,44 % para 2021-2026, y la Circular 5/2026 la elevó al 6,46 % para el periodo regulatorio 2027-2032, cuyo primer año de gas comienza el 1 de octubre de 2026. La Circular 9/2025 ha sido recurrida ante la Audiencia Nacional.

y la reducción de las pérdidas en la red (Circular 6/2019).

- Enfoque TOTEX: la Circular 8/2025 retribuye la distribución en 2026-2031 comparando la inversión y los gastos de operación de forma conjunta con unos costes de referencia. El periodo se divide en dos semiperiodos, 2026-2028 y 2029-2031, y el incentivo que reparte con el sector los márgenes obtenidos frente a esa referencia solo se aplica en el segundo.
- Lectura teórica: es un sistema híbrido. La tasa de rendimiento da seguridad a la inversión en redes, mientras que los incentivos de calidad y el TOTEX dan más poder al contrato y reducen el sesgo hacia el capital que predice el modelo de Averch y Johnson. El nivel de la TRF sigue siendo la decisión más disputada, porque fija el reparto entre rentabilidad e inversión.

El price cap en la práctica: RPI–X, RIIO y Aena

- Origen: tras la privatización de British Telecom en 1984 se aplicó el tope que había propuesto Littlechild, RPI–3, es decir, el índice de precios minoristas (RPI) menos tres puntos, y la fórmula se extendió al gas, a los aeropuertos, al agua y a la electricidad.
- Del RPI–X al RIIO: a partir de 2013, el regulador energético, Ofgem, sustituyó el RPI–X de las redes por el modelo RIIO, cuyo nombre resume su fórmula (ingresos iguales a incentivos más innovación más resultados). RIIO fija los ingresos permitidos para periodos de varios años, trata conjuntamente la inversión y los gastos de operación y reparte con los consumidores las desviaciones de gasto mediante una tasa de incentivo. La decisión final de RIIO-3, de diciembre de 2025, cubre de abril de 2026 a marzo de 2031, con una rentabilidad de los fondos propios del 5,70 % para el transporte eléctrico y del 6,12 % para el gas.
- Tope minorista: desde enero de 2019 Ofgem aplica también un tope a las tarifas por defecto de electricidad y gas de los hogares, en virtud de la Domestic Gas and Electricity (Tariff Cap) Act 2018, un caso de regulación de precios en un mercado minorista formalmente liberalizado.
- Aena: la Ley 18/2014 fija para cada quinquenio, en el Documento de Regulación Aeroportuaria (DORA), un ingreso máximo anual por pasajero, $IMAP_t$, que se actualiza con un índice de los precios de los insumos que la empresa no controla, P_{t-1}, y con un componente, X:

$$IMAP_t = IMAP_{t-1} \cdot \left(1 + \frac{P_{t-1} + X}{100}\right)$$

El DORA 2027-2031, aprobado el 15 de septiembre de 2026, parte de 10,52 euros

 © 2026 Manuel Cabado García

por pasajero en 2026, fija X en el 0,33 % anual y prevé incentivos y penalizaciones por calidad de hasta un 2 % del ingreso máximo.

- Lectura teórica: en los dos países el periodo regulatorio de cinco años equilibra el incentivo a reducir costes con el traslado de las ganancias al consumidor, y los incentivos de calidad corrigen la tentación de empeorarla que la teoría atribuye al price cap. En el caso de Aena, el componente X se calcula para que los ingresos del quinquenio cubran sus costes, de modo que actúa más como un ajuste que cuadra ingresos y costes que como un descuento por productividad.

Regulación del acceso en las telecomunicaciones

- Marco: el Código Europeo de las Comunicaciones Electrónicas, aprobado por la Directiva (UE) 2018/1972 y traspuesto en España por la Ley 11/2022, General de Telecomunicaciones, ordena una regulación ex ante en tres pasos: definir los mercados relevantes, comprobar que cumplen tres criterios (barreras de entrada importantes y no transitorias, una estructura que no tiende hacia la competencia efectiva e insuficiencia del derecho de la competencia por sí solo) e imponer obligaciones a las empresas con peso significativo en el mercado, es decir, con una posición de fuerza que les permite actuar con independencia apreciable de sus competidores y clientes, como el acceso, la no discriminación o el control de precios orientados a costes.
- Aplicación en España: la CNMC obligaba a Telefónica a ofrecer acceso mayorista a su red de fibra (los servicios NEBA) con precios sometidos a un test de replicabilidad económica, que comprueba que un operador alternativo eficiente puede igualar sus ofertas minoristas con los precios mayoristas regulados.
- Retirada gradual: a medida que avanzó la competencia en infraestructuras, en 2021 la CNMC declaró competitiva una zona de 696 municipios, con cerca del 70 % de la población, en la que levantó la mayoría de las obligaciones mayoristas, y en agosto de 2025 eliminó la obligación de ofrecer NEBA en condiciones reguladas, manteniendo solo el acceso a la obra civil (canalizaciones, conductos y postes).
- Lectura teórica: es el recorrido que predice la teoría del acceso. Se regula mientras existe un cuello de botella, con precios que permitan entrar solo a rivales eficientes, y la regulación se retira cuando la competencia en infraestructuras se hace sostenible.

Topes de emergencia y plataformas digitales

Excepción ibérica Fue el mecanismo temporal con el que España y Portugal limitaron el precio del gas usado para producir electricidad durante la crisis energética de 2022, una intervención atípica que se resume así:

- Problema: el mercado eléctrico mayorista aplica la regla de precio igual al coste

marginal, ya que la central más cara que resulta necesaria, a menudo una de gas, fija el precio que cobran todas. Esta regla es eficiente a corto plazo, pero con el gas muy caro deja grandes rentas a las tecnologías de menor coste.

- Mecanismo: con la autorización de la Comisión Europea como ayuda de Estado, España y Portugal limitaron el precio del gas utilizado para producir electricidad (Real Decreto-ley 10/2022). El precio de referencia del gas partía de 40 €/MWh y subía 5 €/MWh al mes a partir del séptimo mes, y las centrales recibían un ajuste por la diferencia, financiado con las rentas de congestión y con un cargo a los compradores del mercado.
- Vigencia: se aplicó desde el 14 de junio de 2022 y el Real Decreto-ley 3/2023 lo prorrogó hasta el 31 de diciembre de 2023.[28]
- Lectura teórica: no corrige el poder de mercado de una empresa, sino la traslación al conjunto del mercado del coste de la tecnología marginal. Es sobre todo una medida redistributiva, que reduce esas rentas sin corregir una ineficiencia, y al alejarse de la señal del coste marginal tuvo costes de segundo óptimo: al abaratar la electricidad sin abaratar el gas, aumentó la generación con ciclos combinados de gas y las exportaciones de electricidad a Francia crecieron más del 80 % (Hidalgo-Pérez et al., 2024). Su carácter temporal y tasado limita el riesgo de que los inversores lo interpreten como una expropiación de rentas futuras, el mismo problema de compromiso que plantea el efecto trinquete.

Digital Markets Act La Digital Markets Act (Reglamento (UE) 2022/1925, aplicable desde mayo de 2023) traslada la lógica de la regulación ex ante a las grandes plataformas digitales, que son mercados de dos caras no disputables:

- Por qué ex ante: la defensa de la competencia actúa ex post y caso a caso, y en mercados que cambian deprisa llega tarde. La DMA traslada a las plataformas la lógica de la regulación del acceso de las industrias de red, pero sobre conductas y no sobre precios.
- Designación con información limitada: en lugar de definir mercados caso a caso, el Reglamento designa como guardianes de acceso a las plataformas que superan unos umbrales de tamaño y de usuarios, criterios que el regulador puede comprobar sin conocer los costes ni la demanda de cada plataforma.[29] En septiembre de 2023 la

[28] Según el preámbulo del Real Decreto-ley 3/2023, el precio medio del mercado diario español desde la entrada en vigor del mecanismo hasta febrero de 2023 fue de 127,89 €/MWh, frente a 273,23 €/MWh en Francia, 240,48 €/MWh en Alemania y 306,18 €/MWh en Italia.

[29] Se presume que una empresa es guardián de acceso si su volumen de negocios en la Unión es de al menos 7.500 millones de euros en cada uno de los tres últimos ejercicios, o su capitalización media de al menos 75.000 millones, opera en al menos tres Estados miembros y tiene al menos 45 millones de usuarios finales activos al mes y 10.000 usuarios profesionales activos al año (artículo 3.2 del Reglamento).

 © 2026 Manuel Cabado García

Comisión designó a Alphabet, Amazon, Apple, ByteDance, Meta y Microsoft, y en 2024 a Booking.

- Obligaciones: les impone obligaciones y prohibiciones directas, como no favorecer sus propios servicios, que es la misma no discriminación que se exige al incumbente integrado de una red, o permitir que los usuarios profesionales dirijan a sus clientes hacia ofertas fuera de la plataforma.
- Sanciones: las multas pueden alcanzar el 10 % del volumen de negocios mundial, y el 20 % en caso de reincidencia, y las primeras decisiones de incumplimiento llegaron en abril de 2025, con 500 millones de euros para Apple y 200 millones para Meta.
- Limitaciones: persisten las de toda regulación, porque las plataformas saben más que el regulador sobre sus algoritmos y sus costes, y tienen un fuerte incentivo a influir en cómo se aplican las obligaciones, el riesgo de captura que describe la teoría.

18.6. Conclusiones ⏳ 30"

Como conclusión, este tema responde a una pregunta que no envejece: cómo lograr que una empresa sin competidores sirva al interés general. Su lección de fondo es que la regulación óptima no es una fórmula, sino un equilibrio entre eficiencia, incentivos y credibilidad, que depende tanto de la teoría como de las instituciones que la aplican.

Esta lección está hoy más viva que nunca:

- En la energía, la transición exige grandes inversiones en redes que solo llegarán si la regulación es estable y creíble, y la crisis de 2022 mostró que incluso los mercados liberalizados pueden necesitar intervenciones de emergencia, con sus costes.
- En la economía digital, los nuevos cuellos de botella ya no son cables ni tuberías, sino datos, plataformas y capacidad de cálculo. Si el acceso a los modelos de inteligencia artificial se convierte en una instalación esencial, ¿bastará con vigilar conductas, como hace la Digital Markets Act, o habrá que regular el acceso como se hizo con las redes?
- En las instituciones, ¿tendrán los reguladores la independencia y la capacidad técnica que, como defendieron Laffont y Tirole, exigen estos retos?

Apéndice

18.A.1. Derivación de los precios de Ramsey-Boiteux

Sea una empresa que produce n bienes con precios P_i, demandas $q_i(P_1, \ldots, P_n)$ y costes marginales CMg_i. El bienestar es la suma del excedente de los consumidores, $EC(P)$, cuya derivada respecto a cada precio es $-q_i$, y del beneficio de la empresa, $\pi(P)$. El regulador maximiza el bienestar sujeto a que el beneficio no sea negativo, con un multiplicador $\lambda \geq 0$:

$$\mathcal{L} = EC(P) + \pi(P) + \lambda \cdot \pi(P) = EC(P) + (1+\lambda) \cdot \pi(P)$$

Como $\partial\pi/\partial P_i = q_i + \sum_j (P_j - CMg_j) \cdot \partial q_j/\partial P_i$, la condición de primer orden respecto a P_i es:

$$-q_i + (1+\lambda) \cdot \left[q_i + \sum_j (P_j - CMg_j) \cdot \frac{\partial q_j}{\partial P_i} \right] = 0 \quad \Longrightarrow \quad \sum_j (P_j - CMg_j) \cdot \frac{\partial q_j}{\partial P_i} = -\frac{\lambda}{1+\lambda} \cdot q_i$$

Si las demandas son independientes, solo queda el término $j = i$. Dividiendo por q_i y usando el valor absoluto de la elasticidad-precio, $|\varepsilon_i| = -(\partial q_i/\partial P_i) \cdot P_i/q_i$, se obtiene la regla de la elasticidad inversa (18.1). Ambos lados son adimensionales: un margen relativo frente al cociente entre el número de Ramsey y una elasticidad. Si $\lambda = 0$, el precio es igual al coste marginal, y si $\lambda \to \infty$, $\lambda/(1+\lambda) \to 1$ y se obtiene la regla del monopolista, $IL = 1/|\varepsilon|$.

Con dos bienes relacionados, llamando $\varepsilon_{ji} = (\partial q_j/\partial P_i) \cdot P_i/q_j$ a la elasticidad cruzada de la demanda del bien j respecto al precio del bien i, la condición anterior se reescribe:

$$\frac{P_i - CMg_i}{P_i} = \frac{1}{|\varepsilon_i|} \cdot \left[\frac{\lambda}{1+\lambda} + \frac{P_j - CMg_j}{P_j} \cdot \varepsilon_{ji} \cdot \frac{P_j \cdot q_j}{P_i \cdot q_i} \right]$$

Si los bienes son sustitutivos ($\varepsilon_{ji} > 0$), subir el precio del bien i desvía demanda hacia el bien j, que también tiene margen, y el precio del bien i es mayor que con demandas independientes. Si son complementarios ($\varepsilon_{ji} < 0$), ocurre lo contrario. Para desviaciones pequeñas respecto a los precios iguales al coste marginal y sin efectos renta, la condición general equivale a que todas las cantidades se reduzcan en la misma proporción respecto a las de tarificación a coste marginal. Es la regla de la reducción equiproporcional que Ramsey (1927) obtuvo para la imposición óptima y que Baumol y Bradford (1970) trasladaron a los precios.

18.A.2. Precios pico-valle

Siguiendo a Steiner (1957), se divide el ciclo de demanda en un periodo de punta, con demanda q_D, y un periodo valle, con demanda q_N, y para cualquier precio común la

 © 2026 Manuel Cabado García

demanda de punta es mayor. La producción no es almacenable y ambas se atienden con una misma capacidad, K, cuyo coste por unidad y ciclo es C_K, mientras que producir cada unidad cuesta C_E. Si $S_D(q_D)$ y $S_N(q_N)$ son los valores brutos para los consumidores en cada periodo, con $S'_D = P_D$ y $S'_N = P_N$, el regulador resuelve:

$$\mathcal{L} = S_D(q_D) + S_N(q_N) - C_E \cdot (q_D + q_N) - C_K \cdot K + \lambda_D \cdot (K - q_D) + \lambda_N \cdot (K - q_N)$$

donde λ_D y λ_N son los precios sombra de la capacidad en cada periodo. Las condiciones de primer orden son:

$$P_D = C_E + \lambda_D, \qquad P_N = C_E + \lambda_N, \qquad \lambda_D + \lambda_N = C_K$$

junto con las de holgura complementaria. En el caso de punta firme, la capacidad solo se agota en la punta, de modo que $\lambda_N = 0$ y $\lambda_D = C_K$:

$$P_D = C_E + C_K, \qquad P_N = C_E$$

Los usuarios de la punta pagan toda la capacidad y los del valle solo el coste de producir. Si con esos precios la demanda del valle superase a la de la punta, la punta se desplazaría. En ese caso, la capacidad se agota en ambos periodos y su coste se reparte según las disposiciones a pagar, con $P_D + P_N = 2 \cdot C_E + C_K$ y $P_D > P_N$. Ambas reglas cubren exactamente el coste total con precios eficientes en cada periodo.

18.A.3. Modelo de Averch-Johnson

La empresa regulada elige el capital, K, y el trabajo, L, para maximizar su beneficio económico, sujeta a que la rentabilidad de su capital no supere la tasa permitida, s, con $s > r$:

$$\max_{K,L} \ \pi = IT(K,L) - w \cdot L - r \cdot K \qquad \text{s.a.} \qquad IT(K,L) - w \cdot L - s \cdot K \leq 0$$

Con la restricción activa y su multiplicador, μ, el lagrangiano es:

$$\mathcal{L} = IT(K,L) - w \cdot L - r \cdot K - \mu \cdot [IT(K,L) - w \cdot L - s \cdot K]$$

Llamando IT_K e IT_L al ingreso del producto marginal de cada factor, las condiciones de primer orden son:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial L} &= (1-\mu) \cdot (IT_L - w) = 0 \quad &\Longrightarrow \quad IT_L = w \\ \frac{\partial \mathcal{L}}{\partial K} &= (1-\mu) \cdot IT_K - r + \mu \cdot s = 0 \quad &\Longrightarrow \quad IT_K = r - \frac{\mu}{1-\mu} \cdot (s - r) \end{aligned}$$

Averch y Johnson (1962) demostraron que, cuando la restricción es activa, $0 < \mu < 1$. Como el cociente entre los ingresos del producto marginal es igual a la relación marginal de sustitución técnica, $RMST = F_K/F_L$:

$$RMST = \frac{IT_K}{IT_L} = \frac{r}{w} - \frac{\mu}{1-\mu} \cdot \frac{s-r}{w} < \frac{r}{w}$$

La minimización de costes exigiría $RMST = r/w$. La empresa regulada se comporta como si el capital costara $\tilde{r} = r - \mu \cdot (s-r) \,/\, (1-\mu) < r$, y por eso emplea más capital por unidad de trabajo del que minimiza los costes, que es la sobrecapitalización de la Figura 18.3. Todos los términos son coherentes en dimensiones: r, s e IT_K son rendimientos por unidad de capital y periodo, w e IT_L por unidad de trabajo y periodo. Si $s = r$, la restricción solo permite un beneficio económico nulo y la combinación de factores queda indeterminada.

18.A.4. Selección adversa con dos tipos: Baron y Myerson

Se sigue la versión de Laffont y Martimort (2002, pp. 32-43). La empresa tiene un coste marginal bajo, $\underline{\theta}$, con probabilidad ν, o alto, $\overline{\theta}$, con probabilidad $1-\nu$, y $\Delta\theta = \overline{\theta} - \underline{\theta}$. El regulador ofrece un contrato para cada tipo, con una cantidad y una transferencia, $\left(\underline{q}, \underline{t}\right)$ y $(\overline{q}, \overline{t})$, y llama renta a la diferencia entre la transferencia y el coste, $\underline{U} = \underline{t} - \underline{\theta} \cdot \underline{q}$ y $\overline{U} = \overline{t} - \overline{\theta} \cdot \overline{q}$. Las restricciones de compatibilidad de incentivos y de participación son:

$$\begin{aligned}
&\underline{U} \geq \overline{U} + \Delta\theta \cdot \overline{q} && \text{(compatibilidad del tipo eficiente)} \\
&\overline{U} \geq \underline{U} - \Delta\theta \cdot \underline{q} && \text{(compatibilidad del tipo ineficiente)} \\
&\underline{U} \geq 0, \quad \overline{U} \geq 0 && \text{(participación)}
\end{aligned}$$

Sumando las dos primeras se obtiene $\underline{q} \geq \overline{q}$: la empresa eficiente debe producir al menos tanto como la ineficiente. En el óptimo solo son activas la compatibilidad del tipo eficiente y la participación del ineficiente, así que $\overline{U} = 0$ y $\underline{U} = \Delta\theta \cdot \overline{q}$. Si el regulador maximiza el valor esperado de lo que recibe menos lo que paga, el problema queda:

$$\max_{\underline{q}, \overline{q}} \; \nu \cdot \left[S\left(\underline{q}\right) - \underline{\theta} \cdot \underline{q}\right] + (1-\nu) \cdot \left[S\left(\overline{q}\right) - \overline{\theta} \cdot \overline{q}\right] - \nu \cdot \Delta\theta \cdot \overline{q}$$

Las condiciones de primer orden son:

$$S'\left(\underline{q}\right) = \underline{\theta}, \qquad (1-\nu) \cdot \left[S'\left(\overline{q}\right) - \overline{\theta}\right] = \nu \cdot \Delta\theta$$

La segunda es la ecuación (18.2): aumentar la cantidad del tipo ineficiente gana eficiencia en ese estado, con probabilidad $1-\nu$, pero aumenta la renta del eficiente en $\Delta\theta$ por unidad, con probabilidad ν. Las restricciones omitidas se cumplen porque $\underline{q} > \overline{q}$. Si ν es muy alto, puede ser óptimo cerrar la empresa de coste alto, $\overline{q} = 0$, para no dejar ninguna

 © 2026 Manuel Cabado García

renta a la eficiente.

18.A.5. Laffont y Tirole: esfuerzo, renta y menú de contratos

En el modelo de Laffont y Tirole (1986), con un proyecto de tamaño dado, el coste observado es $C = \beta - e$, el gestor obtiene $U = TR - C - \psi(e)$, donde TR es el pago total que recibe la empresa, y cada euro pagado cuesta $1 + \lambda$ a la sociedad. El bienestar, para un valor del proyecto S, es:

$$W = S - (1 + \lambda) \cdot [C + \psi(e)] - \lambda \cdot U$$

Con información completa, $U = 0$ y el esfuerzo iguala su coste marginal a su beneficio marginal, $\psi'(e^*) = 1$. Con dos tipos, $\underline{\beta}$ con probabilidad ν y $\overline{\beta}$ con probabilidad $1 - \nu$, y $\Delta\beta = \overline{\beta} - \underline{\beta}$, la empresa eficiente puede alcanzar el coste de la ineficiente con un esfuerzo menor en $\Delta\beta$, de modo que su renta es:

$$\underline{U} = \Phi(\overline{e}) = \psi(\overline{e}) - \psi(\overline{e} - \Delta\beta), \qquad \Phi' > 0$$

Maximizando el bienestar esperado se obtiene:

$$\psi'(\underline{e}) = 1, \qquad \psi'(\overline{e}) = 1 - \frac{\lambda}{1 + \lambda} \cdot \frac{\nu}{1 - \nu} \cdot \Phi'(\overline{e}) < 1$$

La empresa eficiente ejerce el esfuerzo de primer óptimo y la ineficiente un esfuerzo menor, porque cada unidad de esfuerzo exigida a esta eleva la renta de aquella. Si $\lambda = 0$, las rentas no tienen coste social y desaparece la distorsión. La asignación se implementa con el menú de contratos lineales del texto, $TR = a(C^a) + b(C^a) \cdot (C - C^a)$: la empresa eficiente elige el contrato de precio fijo, $b = 0$, con el que se queda con todo el ahorro que logra con su esfuerzo, y la ineficiente uno con reembolso parcial de costes (Laffont y Tirole, 1993).

En un contexto dinámico, si el regulador no puede comprometerse a no revisar el contrato, la empresa eficiente anticipa que revelar su tipo empeorará el contrato siguiente, y puede ser óptimo un contrato que agrupe a ambos tipos en el primer periodo. Es el efecto trinquete (Freixas et al., 1985), que hace que la falta de compromiso encarezca la regulación.

18.A.6. Precio de acceso a la red

Sea un incumbente que vende un servicio final al precio P_I y que necesita su propia red para prestarlo, con un coste del acceso por unidad, c_a, y un coste propio del resto del servicio, c_I. Un entrante con coste propio c_E necesita comprar acceso al precio A. La regla del precio eficiente de los componentes (Willig, 1979; Baumol y Sidak, 1994) fija el precio de acceso en el coste del acceso más el coste de oportunidad del incumbente, es decir, el

margen que deja de ganar por cada unidad que le quita el entrante:

$$A = c_a + [P_I - (c_a + c_I)] = P_I - c_I$$

En general, el coste de oportunidad se multiplica por el ratio de desplazamiento, σ, que mide cuántas ventas pierde el incumbente por cada unidad de acceso que vende, $A = c_a + \sigma \cdot (P_I - c_I - c_a)$, y la fórmula anterior es el caso $\sigma = 1$ (Armstrong y Sappington, 2007, p. 1674). El entrante solo es rentable si $A + c_E \leq P_I$, es decir, si $c_E \leq c_I$. Así solo entran rivales al menos tan eficientes como el incumbente y se evita el descremado ineficiente del mercado. Su limitación es que toma el precio final como dado, por lo que, si este es de monopolio, el margen se traslada al precio de acceso. Armstrong, Doyle y Vickers (1996) mostraron que la regla solo es óptima en condiciones restrictivas, como productos homogéneos, imposibilidad de sortear la red y un precio final regulado, y que en general el precio de acceso óptimo depende de las elasticidades y, si la empresa debe cubrir costes, tiene una estructura de tipo Ramsey-Boiteux.

18.A.7. Formas de separación vertical en las industrias de red

Una industria de red combina un componente con rasgos de monopolio natural, la red, con actividades potencialmente competitivas. Para que la competencia en estas últimas no se vea frenada por el dueño de la red, existen varias soluciones:

- Acceso regulado sin separación: el incumbente conserva la red y debe dar acceso a sus rivales en condiciones reguladas, como en la red fija de telecomunicaciones. Mantiene las economías de integración, pero conserva el incentivo a discriminar a los rivales.
- Separación de la propiedad: la red pertenece a una empresa que no compite en los servicios, como la gestión de la infraestructura ferroviaria por Adif, separada de la operación de los trenes, que ha permitido la entrada de operadores distintos de Renfe en la alta velocidad. Elimina el incentivo a discriminar, pero pierde las economías de integración.
- Propiedad compartida: la red pertenece conjuntamente a los operadores que la usan. Reduce la discriminación entre ellos, pero facilita la colusión frente a los nuevos entrantes.
- Separación operativa: la propiedad se mantiene, pero la gestión de la red se encomienda a un operador independiente. Evita la discriminación, aunque debilita los incentivos a invertir.
- Redes recíprocas: cuando el poder procede de los efectos de red, como en la telefonía móvil, cada operador necesita acceder a las redes de los demás, y la regulación se limita a los precios de interconexión.

 © 2026 Manuel Cabado García

18.A.8. Marco normativo de referencia

- Ley 3/2013, de 4 de junio, de creación de la Comisión Nacional de los Mercados y la Competencia.
- Circular 2/2019 de la CNMC, de 12 de noviembre, metodología de la tasa de retribución financiera, modificada por la Circular 9/2025, de 22 de diciembre (6,58 % para 2026-2031).
- Circulares 5/2019 y 6/2019 de la CNMC, de 5 de diciembre, metodologías de retribución del transporte y de la distribución de electricidad, la segunda sustituida desde 2026 por la Circular 8/2025, de 22 de diciembre (enfoque TOTEX), y la primera modificada por la Circular 7/2025, de 16 de diciembre.
- Circular 3/2020 de la CNMC, de 15 de enero, metodología de los peajes de transporte y distribución de electricidad.
- Circulares 9/2019 y 4/2020 de la CNMC, retribución del transporte y la regasificación y de la distribución de gas, sustituidas por las Circulares 4/2026 y 6/2026, de 8 de septiembre.
- Circular 5/2026 de la CNMC, de 8 de septiembre, tasa de retribución financiera del gas.
- Circular 6/2020 de la CNMC, de 22 de julio, metodología de los peajes gasistas.
- Directiva (UE) 2018/1972, Código Europeo de las Comunicaciones Electrónicas, y Ley 11/2022, de 28 de junio, General de Telecomunicaciones.
- Real Decreto-ley 10/2022, de 13 de mayo, mecanismo temporal de ajuste de costes de producción (excepción ibérica), prorrogado por el Real Decreto-ley 3/2023, de 28 de marzo.
- Reglamento (UE) 2022/1925, de 14 de septiembre de 2022, sobre mercados disputables y equitativos en el sector digital.
- Ley 18/2014, de 15 de octubre, y Documento de Regulación Aeroportuaria 2027-2031, aprobado por Acuerdo del Consejo de Ministros de 15 de septiembre de 2026.
- Domestic Gas and Electricity (Tariff Cap) Act 2018, del Reino Unido.

Bibliografía

Armstrong, M., C. Doyle y J. Vickers (1996). The Access Pricing Problem: A Synthesis. *The Journal of Industrial Economics 44*(2), 131–150.

Armstrong, M. y D. E. M. Sappington (2007). Recent Developments in the Theory of Regulation. En M. Armstrong y R. H. Porter (eds.), *Handbook of Industrial Organization*, vol. 3, cap. 27, pp. 1557–1700. Amsterdam: North-Holland.

Averch, H. y L. L. Johnson (1962). Behavior of the Firm Under Regulatory Constraint. *The American Economic Review 52*(5), 1052–1069.

Baron, D. P. (1989). Design of Regulatory Mechanisms and Institutions. En R. Schmalensee y R. D. Willig (eds.), *Handbook of Industrial Organization*, vol. 2, cap. 24, pp. 1347–1447. Amsterdam: North-Holland.

Baron, D. P. y R. B. Myerson (1982). Regulating a Monopolist with Unknown Costs. *Econometrica 50*(4), 911–930.

Baumol, W. J. (1977). On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry. *The American Economic Review 67*(5), 809–822.

Baumol, W. J. y D. F. Bradford (1970). Optimal Departures from Marginal Cost Pricing. *The American Economic Review 60*(3), 265–283.

Baumol, W. J., J. C. Panzar y R. D. Willig (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich.

Baumol, W. J. y J. G. Sidak (1994). The Pricing of Inputs Sold to Competitors. *Yale Journal on Regulation 11*(1), 171–202.

Becker, G. S. (1983). A Theory of Competition Among Pressure Groups for Political Influence. *The Quarterly Journal of Economics 98*(3), 371–400.

Boiteux, M. (1956). Sur la gestion des monopoles publics astreints à l'équilibre budgétaire. *Econometrica 24*(1), 22–40.

Braeutigam, R. R. (1989). Optimal Policies for Natural Monopolies. En R. Schmalensee y R. D. Willig (eds.), *Handbook of Industrial Organization*, vol. 2, cap. 23, pp. 1289–1346. Amsterdam: North-Holland.

Cabral, L. M. B. (2000). *Introduction to Industrial Organization*. Cambridge, MA: MIT Press.

Coase, R. H. (1946). The Marginal Cost Controversy. *Economica 13*(51), 169–182. New Series.

 © 2026 Manuel Cabado García

Demsetz, H. (1968). Why Regulate Utilities? *The Journal of Law and Economics 11*(1), 55–65.

Dupuit, J. (1844). De la mesure de l'utilité des travaux publics. *Annales des Ponts et Chaussées 8*, 332–375. 2e série.

Freixas, X., R. Guesnerie y J. Tirole (1985). Planning under Incomplete Information and the Ratchet Effect. *The Review of Economic Studies 52*(2), 173–191.

Harberger, A. C. (1954). Monopoly and Resource Allocation. *The American Economic Review 44*(2), 77–87. Papers and Proceedings.

Hidalgo-Pérez, M., N. Collado, J. Galindo y R. Mateo (2024). The Iberian Exception: Estimating the Impact of a Cap on Gas Prices for Electricity Generation on Consumer Prices and Market Dynamics. *Energy Policy 188*, 114092.

Hotelling, H. (1938). The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates. *Econometrica 6*(3), 242–269.

Kydland, F. E. y E. C. Prescott (1977). Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy 85*(3), 473–491.

Laffont, J.-J. y D. Martimort (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton, NJ: Princeton University Press.

Laffont, J.-J. y J. Tirole (1986). Using Cost Observation to Regulate Firms. *Journal of Political Economy 94*(3), 614–641.

Laffont, J.-J. y J. Tirole (1987). Auctioning Incentive Contracts. *Journal of Political Economy 95*(5), 921–937.

Laffont, J.-J. y J. Tirole (1991). The Politics of Government Decision-Making: A Theory of Regulatory Capture. *The Quarterly Journal of Economics 106*(4), 1089–1127.

Laffont, J.-J. y J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

Leibenstein, H. (1966). Allocative Efficiency vs. "X-Efficiency". *The American Economic Review 56*(3), 392–415.

Lerner, A. P. (1934). The Concept of Monopoly and the Measurement of Monopoly Power. *The Review of Economic Studies 1*(3), 157–175.

Lipsey, R. G. y K. Lancaster (1956). The General Theory of Second Best. *The Review of Economic Studies 24*(1), 11–32.

Littlechild, S. C. (1983, febrero). Regulation of British Telecommunications' Profitability: Report to the Secretary of State. Informe técnico, Department of Industry, London. Publicado por HMSO.

Loeb, M. y W. A. Magat (1979). A Decentralized Method for Utility Regulation. *The Journal of Law and Economics 22*(2), 399–404.

Oi, W. Y. (1971). A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly. *The Quarterly Journal of Economics 85*(1), 77–96.

Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press.

Peltzman, S. (1976). Toward a More General Theory of Regulation. *The Journal of Law and Economics 19*(2), 211–240.

Ramsey, F. P. (1927). A Contribution to the Theory of Taxation. *The Economic Journal 37*(145), 47–61.

Rochet, J.-C. y J. Tirole (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association 1*(4), 990–1029.

Shleifer, A. (1985). A Theory of Yardstick Competition. *The RAND Journal of Economics 16*(3), 319–327.

Steiner, P. O. (1957). Peak Loads and Efficient Pricing. *The Quarterly Journal of Economics 71*(4), 585–610.

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell Journal of Economics and Management Science 2*(1), 3–21.

The Royal Swedish Academy of Sciences (2014, octubre). The Prize in Economic Sciences 2014: Press Release. Nota de prensa, 13 de octubre de 2014. Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2014 a Jean Tirole "for his analysis of market power and regulation".

The Royal Swedish Academy of Sciences (2020, octubre). The Prize in Economic Sciences 2020: Press Release. Nota de prensa, 12 de octubre de 2020. Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2020 a Paul R. Milgrom y Robert B. Wilson "for improvements to auction theory and inventions of new auction formats".

U.S. Department of Justice (1982). Merger Guidelines.

Weitzman, M. L. (1980). The "Ratchet Principle" and Performance Incentives. *The Bell Journal of Economics 11*(1), 302–308.

 © 2026 Manuel Cabado García

Williamson, O. E. (1976). Franchise Bidding for Natural Monopolies—in General and with Respect to CATV. *The Bell Journal of Economics 7*(1), 73–104.

Willig, R. D. (1978). Pareto-Superior Nonlinear Outlay Schedules. *The Bell Journal of Economics 9*(1), 56–69.

Willig, R. D. (1979). The Theory of Network Access Pricing. En H. M. Trebing (ed.), *Issues in Public Utility Regulation*, pp. 109–152. East Lansing, MI: Michigan State University Press.

 © 2026 Manuel Cabado García

Tema 19

El Estado del Bienestar y los instrumentos de redistribución y protección social. Medición e indicadores del impacto redistributivo.

19.1. Introducción ⏳ 1'

Según el último informe de Fedea (López Laborda et al., 2026) sobre distribución de impuestos para el año 2023, el presupuesto público redujo la desigualdad de la renta de mercado en España un 31,9 %. Además, más de la mitad de los hogares españoles fueron perceptores netos del Estado, al recibir en prestaciones públicas más de lo que aportaron vía impuestos.

Así, una de las mayores transformaciones que ha experimentado el sector público en los últimos 60 años en la mayoría de países desarrollados ha sido el cambio en las funciones que este ejerce, marcado por la mayor importancia concedida a la función distributiva:

- El medio principal utilizado por el sector público para cumplir con esta función es el presupuesto, entendido como un programa de ingresos y de gastos públicos.
- El resultado principal de dicha función es la construcción del Estado del Bienestar.

Por consiguiente, el objetivo del tema es analizar los instrumentos redistributivos del Estado de Bienestar, sus efectos y sostenibilidad a través del siguiente esquema:

1. Justificación.
2. Prestaciones económicas.

3. Provisión de bienes preferentes.
4. Impacto redistributivo.

19.2. Justificación 1'30"

Comenzando con la justificación, la Economía del Bienestar muestra que las asignaciones Pareto-eficientes resultantes de un mercado perfectamente competitivo no tienen por qué ser las que maximicen el bienestar social. Es decir, la optimalidad de Pareto es una condición necesaria pero no suficiente para la maximización del bienestar.

El Segundo Teorema del Bienestar establece que, con preferencias y tecnologías convexas, cualquier asignación eficiente puede alcanzarse como equilibrio competitivo mediante una redistribución previa vía transferencias de suma fija, dejando después actuar a los mecanismos del mercado. Por tanto, puede alcanzarse otra solución eficiente que, además, maximice el bienestar social[1].

Estado del Bienestar De este modo, para garantizar su función distributiva, el Sector Público ha construido el Estado del Bienestar, que puede definirse como un conjunto de instrumentos redistributivos que el sector público utiliza para garantizar un cierto nivel de equidad y cobertura mínima de necesidades en la sociedad. No obstante, el sector público debe tratar de garantizar un equilibrio que no siempre es fácil de lograr: tiene que ser eficaz, eficiente y sostenible.

Pilares de la redistribución Así, en los países desarrollados, la función redistributiva del sector público se suele diseñar desde una perspectiva socio-liberal basada en tres pilares[2]:

1. Equidad horizontal: no discriminación entre iguales.
2. Equidad vertical: discriminación positiva entre diferentes (a favor del menos dotado).
3. Igualdad de oportunidades (que es diferente de igualdad de resultados).

Instrumentos Respecto a los instrumentos, aunque el sector público tiene tanto instrumentos presupuestarios como no presupuestarios[3], se opta por los primeros debido al alto coste de eficiencia de los segundos.

[1] Trade-off eficiencia-equidad: no obstante, es prácticamente imposible introducir un sistema de impuestos y transferencias que no distorsione las decisiones y el comportamiento de los agentes económicos, por lo que pueden surgir pérdidas de eficiencia y aparecer con fuerza el trade-off equidad vs. eficiencia.

[2] Diferencia entre igualdad y equidad: la igualdad consiste en dar a todos lo mismo, mientras que la equidad reparte bienes o servicios en función de los méritos, necesidades o características de cada individuo.

[3] Se trata de la regulación. Su mayor presencia se encuentra en el mercado de trabajo: salario mínimo, regulación de entrada (edad mínima) y salida (jubilación), etc. No obstante, también suele regular en busca de la equidad en otros ámbitos: alimentos, vivienda, energía, etc.

 © 2026 Manuel Cabado García

Distorsiones Además, la mayor parte de la función redistribuidora del presupuesto se hace a través del gasto y no de los impuestos, pues la concepción moderna de la imposición (aportada fundamentalmente por la Nueva Macroeconomía Clásica) recomienda que esta sea neutral para evitar distorsiones en la toma de decisiones.

A continuación, se analizarán los dos principales instrumentos de gasto del Estado de Bienestar: las prestaciones económicas y la provisión de bienes preferentes.

19.3. Prestaciones económicas ⏳ 6'

Se trata de transferencias monetarias de libre disposición (pensiones, prestaciones por desempleo, prestaciones asistenciales, etc.). El receptor hace el uso que quiera de las mismas.

Rasgos generales

Comenzando con los rasgos generales, las prestaciones se clasifican:

- Según el tipo de cobertura:
  - Profesional (o bismarckiano): la cobertura se limita fundamentalmente a los asalariados. En puridad, el sistema propone una estricta proporcionalidad entre las cotizaciones y las prestaciones.
  - Universal, heredero del Informe Beveridge (1942), la cobertura se ofrece a toda la población.
- Según el modo de financiación:
  - No contributivo (o asistencial): se financia con cargo a los Presupuestos Generales del Estado, sin exigir cotizaciones previas. Su principal ventaja es la contribución a la equidad y su principal inconveniente es la presión que ejerce sobre el déficit.
  - Contributivo: se financia a través de las cotizaciones de los afiliados.

Categorías de prestaciones Asimismo, las prestaciones económicas pueden agruparse en dos categorías:

1. Programas de sustitución de rentas: ofrecen prestaciones que compensan la pérdida de rentas como consecuencia de determinados acontecimientos (pensiones contributivas, prestaciones por desempleo, etc.). Suelen ser profesionales y contributivas.
2. Programas de reducción de la pobreza: ofrecen prestaciones que tratan de paliar situaciones de pobreza extrema, proporcionando unos ingresos mínimos a individuos cuyos ingresos son muy bajos o nulos (pensiones no contributivas, subsidios por

desempleo, etc.). No exigen cotizaciones previas, se dirigen a quien acredita falta de recursos y se financian con impuestos generales.

Obligatoriedad de participación Por otro lado, la obligatoriedad de participación en los programas contributivos se debe a cuatro motivos:

1. Paternalismo: se basa en la creencia de que el Estado conoce mejor los beneficios de tener un seguro, pues los individuos suelen subestimar los riesgos de un acontecimiento y podrían no asegurarse ante éste.[4]
2. Información asimétrica: selección adversa. Las aseguradoras no conocen la probabilidad de que una persona pierda su puesto de trabajo, por lo que fijarán unas primas en función del riesgo medio, con lo que las personas de bajo riesgo no contratarán el seguro. Las aseguradoras recalcularán entonces al alza la prima, repitiéndose el proceso hasta que los individuos con menor riesgo son expulsados completamente del mercado de seguros.
3. Eficiencia: por la ley de los grandes números (risk pooling), cuanto mayor sea el número de asegurados con riesgos independientes, menor será el riesgo por asegurado, porque la pérdida media del grupo se vuelve casi predecible aunque no se sepa a quién afectará el siniestro.[5]
4. Externalidad negativa: una pobreza generalizada puede generar tensión social, inestabilidad, delincuencia, inseguridad, etc.

Justificación del carácter público Además, existe justificación del carácter público de los programas contributivos. Así, la provisión pública de los programas contributivos se debe a dos motivos:

1. Riesgos sociales. Se trata de riesgos que, de producirse, afectan a gran parte de la población, por lo que una aseguradora privada podría verse abocada a la quiebra (por ejemplo, desempleo). El sector público sí puede asegurar porque dispone de ingresos adicionales a las primas.
2. Equidad. La provisión pública asegura que los individuos pagarán en función de su renta, y no en función de su riesgo (que sería la variable clave para las aseguradoras privadas).

A continuación se analizan las principales prestaciones económicas.

[4] La economía del comportamiento ha dado una versión moderna de este argumento, el paternalismo libertario de Thaler y Sunstein. Por ejemplo, afiliar automáticamente a los trabajadores a un plan de pensiones, con libertad para darse de baja, elevó la afiliación del 49 % al 86 % en una empresa estadounidense (Thaler y Sunstein, 2003, p. 177). El apartado 4.A.8 del Apéndice del Tema 4 lo desarrolla.

[5] La intuición es que una aseguradora no sabe quién enfermará este año, pero sí, con mucha precisión, cuántos de cada mil asegurados lo harán. Formalmente, si cada uno de los N asegurados afronta una pérdida independiente de varianza σ^2, la varianza de la pérdida media es σ^2/N, que tiende a cero al crecer N.

 © 2026 Manuel Cabado García

Pensiones de jubilación

Definición: prestación económica periódica que compensa la pérdida de la capacidad del trabajador para generar rentas laborales por vejez. Las pensiones constituyen un pilar fundamental dentro del Estado de Bienestar[6].

Objetivos Los sistemas de pensiones, entendidos como el conjunto de programas de transferencias de renta y de los correspondientes mecanismos financieros, persiguen fundamentalmente tres objetivos:

1. Prevenir el riesgo de pobreza entre adultos de mayor edad.
2. Permitir el mantenimiento del estándar de vida después de la jubilación.
3. Promover la cohesión social.

Sistemas de pensiones

Las pensiones de jubilación pueden seguir dos sistemas:

- Capitalización: los trabajadores realizan contribuciones forzosas a una cuenta de ahorro de la cual el trabajador obtendrá su pensión cuando se jubile (ahorro forzoso, pero no redistribución). Sistema de Capitalización, actúa como un fondo privado de pensiones, donde el sector público invierte las cotizaciones, C, recaudadas en un activo sin riesgo y, llegado el momento de la jubilación, se pagan las cotizaciones más los rendimientos que se hayan generado. El importe total disponible, A, vendrá dado por
$$A = C \cdot \sum_{t=0}^{J} (1+r)^{J-t} \tag{19.1}$$
y dependerá fundamentalmente del rendimiento obtenido que viene dado por el tipo de interés.
- De reparto (pay as you go): basado en un contrato implícito intergeneracional según el cual las prestaciones del período se cubren con las cotizaciones de ese mismo período (ahorro forzoso y redistribución intergeneracional). El sector público se compromete a pagar las pensiones futuras pero no el nivel de las mismas. La cuantía de la pensión media, P_t, en cada momento será igual a los fondos disponibles, que serán el tipo de cotización, τ, por el salario medio, w_t, y por el número de ocupados, L_t, todo ello dividido entre el número de jubilados, J_t:
$$P_t = \frac{\tau \cdot w_t \cdot L_t}{J_t} \tag{19.2}$$

[6] En España, constituyen la prestación económica más importante, situándose en torno al 12 % del PIB y 4 de cada 10 euros de los PGE

¿Cuál elegir? La elección entre un sistema de capitalización y uno de reparto depende, en gran medida, de factores demográficos y económicos:

- En un sistema de reparto, el rendimiento de las pensiones está vinculado al crecimiento de la masa salarial y del empleo, descontando el crecimiento del número de jubilados. Es decir, la pensión media crece más cuanto más rápido aumenten los ingresos por cotizaciones respecto al número de beneficiarios.
- En un sistema de capitalización, el rendimiento depende exclusivamente del tipo de interés que generan las inversiones realizadas con las cotizaciones acumuladas.

Por tanto, el sistema de reparto será más ventajoso que el de capitalización siempre que la siguiente expresión se cumpla:

$$\frac{\dot{w}}{w} + \frac{\dot{L}}{L} - \frac{\dot{J}}{J} > r \tag{19.3}$$

donde $\frac{\dot{w}}{w}$ es la tasa de crecimiento de los salarios, $\frac{\dot{L}}{L}$ la del empleo, $\frac{\dot{J}}{J}$ la del número de jubilados, y r el tipo de interés. Esta desigualdad refleja que, si el crecimiento económico y demográfico supera el rendimiento financiero, el sistema de reparto ofrece una rentabilidad mayor que la capitalización y puede mejorar el bienestar de todas las generaciones.[7]

Efectos sobre crecimiento y ahorro

Bajo el sistema de reparto, los agentes pueden modificar sus decisiones de ahorro por distintos motivos:

- Efecto sustitución de la riqueza: los individuos perciben que las cotizaciones a la Seguridad Social se transformarán en renta futura, lo que les lleva a reducir su ahorro privado. Sin embargo, dado que en los sistemas de reparto las cotizaciones no se acumulan, sino que se destinan directamente al pago de pensiones actuales, este comportamiento puede provocar una disminución del ahorro nacional.
- Efecto jubilación: las pensiones públicas pueden inducir a jubilarse antes, y un periodo de jubilación más largo, financiado con menos años de trabajo, obliga a ahorrar más durante la vida activa.[8] Además, la incertidumbre sobre la cuantía de la pensión futura puede elevar el ahorro por precaución, especialmente si se perciben riesgos en la sostenibilidad del sistema.
- Efecto herencia: como el sistema de reparto transfiere renta de los hijos, que cotizan, a los padres, que cobran la pensión, los individuos que desean dejar un legado a sus descendientes tienden a incrementar su ahorro para compensarles.

[7] La inmigración eleva temporalmente el crecimiento del empleo, $\frac{\dot{L}}{L}$, y con él la rentabilidad implícita del reparto, pero los inmigrantes también acaban jubilándose. El apartado 19.A.3 del Apéndice desarrolla este efecto y la evidencia sobre el saldo fiscal de los inmigrantes.

[8] La intuición es que, si la pensión se pierde al seguir trabajando, prolongar la vida laboral apenas compensa, de modo que muchos se retiran antes y necesitan haber ahorrado para más años sin salario.

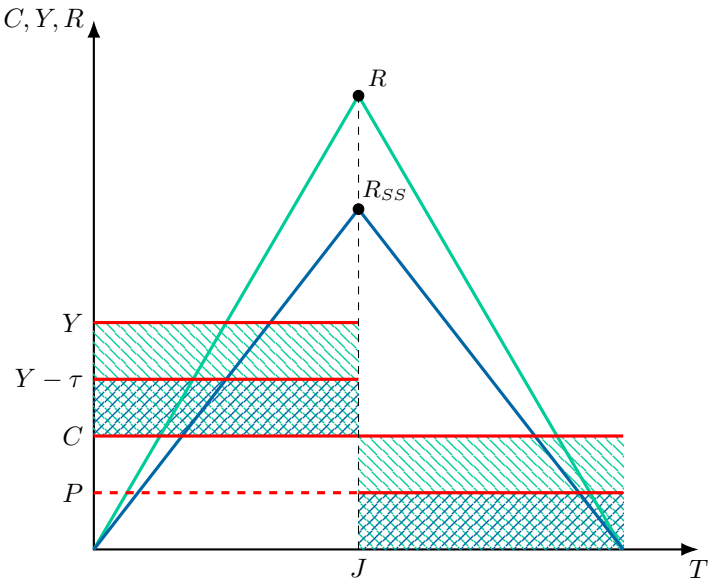
 © 2026 Manuel Cabado García

Análisis gráfico Gráficamente, en un plano de consumo, renta y riqueza a lo largo del tiempo como el de la Figura 19.1, se analizará el efecto sobre el ahorro de los sistemas de reparto y capitalización:

- Se parte de un modelo de ciclo vital, donde los agentes desean suavizar el consumo, C, a lo largo de su vida, por lo que distribuirán el ahorro para conseguirlo.
- Recibirán una renta exógena, Y, que consumirán y ahorrarán, por lo que acumularán riqueza, R, hasta que se jubilen en el momento J.
- Cuando se jubilan, al no recibir la renta laboral, comenzarán a desahorrar, por lo que irán disminuyendo paulatinamente la riqueza hasta el final de sus días.
- En un sistema de reparto, el proceso es el mismo pero mientras trabaja cotizará, así que la renta será menor, $Y - \tau$, ahorrando menos hasta el momento J.
- En el momento de jubilarse no solo consumirá lo ahorrado, sino que recibirá una pensión, P, gracias a sus cotizaciones mientras trabajaba.

Así, bajo un sistema de reparto el ahorro nacional disminuye respecto al sistema de capitalización.

Figura 19.1: Evolución de la riqueza a lo largo del ciclo vital



Concepto de eficiencia dinámica Esta menor acumulación de ahorro y capital bajo el sistema de reparto no es necesariamente negativa: su efecto sobre el bienestar depende de si la economía se encontraba o no en una situación de eficiencia dinámica.

La eficiencia dinámica es la situación en la que no es posible aumentar el consumo de una generación sin reducir el de otra, lo que exige que el capital por trabajador no

supere al de la Regla de Oro. Según la Regla de Oro del modelo de Solow, existe un nivel de capital por trabajador que maximiza el consumo por trabajador sostenible en el largo plazo.

Así, la clave está en si la economía se encuentra o no en un régimen de sobreacumulación, es decir, si el nivel de capital por trabajador supera al de la Regla de Oro. Solo en ese caso, una reducción del ahorro agregado como la que provoca el sistema de reparto frente al de capitalización mejora el bienestar:

- En un sistema de capitalización, el ahorro agregado puede situarse por encima del nivel eficiente si la economía entra en sobreacumulación, en cuyo caso deja de ser dinámicamente eficiente.
- En cambio, en una economía en sobreacumulación, el sistema de reparto, al reducir el ahorro privado, acerca la economía al nivel de capital de la Regla de Oro y mejora el bienestar.

La pregunta relevante, por tanto, es si las economías desarrolladas se encuentran realmente en régimen de sobreacumulación. Abel, Mankiw, Summers y Zeckhauser (1989) contrastaron la condición de eficiencia dinámica para Estados Unidos y el resto de países del G7 y concluyeron que estas economías son, en general, dinámicamente eficientes, ya que la inversión bruta no supera de forma sistemática a los beneficios del capital. Sin embargo, este resultado no está exento de controversia: Blanchard (2019), en su discurso presidencial ante la American Economic Association, sostuvo que la eficiencia dinámica debe evaluarse con el rendimiento del capital productivo y no con el tipo de interés libre de riesgo, lo que matiza el criterio original. En la misma línea, Geerolf (2018) recalculó el criterio de AMSZ incorporando las rentas de la tierra –excluidas del capital productivo en el cálculo original– y no encontró evidencia de eficiencia dinámica en ninguna de las economías avanzadas analizadas, señalando además una sobreacumulación de capital inequívoca en Japón y Corea del Sur.

Por tanto, el argumento de la sobreacumulación no puede descartarse como una mera posibilidad teórica: sigue siendo objeto de debate empírico abierto entre la literatura clásica (Abel et al., 1989) y revisiones más recientes (Blanchard, 2019; Geerolf, 2018) que cuestionan si las economías desarrolladas son realmente dinámicamente eficientes.

Problemas y reformas

El principal desafío de los sistemas de pensiones en países desarrollados sigue siendo su sostenibilidad, especialmente en regímenes de reparto o mixtos. Las causas son múltiples:

- El desempleo reduce las cotizaciones.
- La devaluación salarial limita la base contributiva.

 © 2026 Manuel Cabado García

- El envejecimiento poblacional incrementa el número de pensionistas (se prevén 15 millones en 2050)[9].
- El efecto sustitución de las pensiones, es decir, la sustitución de las pensiones que se extinguen por altas de mayor cuantía, eleva el gasto, ya que las nuevas pensiones son un 30 % superiores a las que se extinguen.
- La robotización transforma la base productiva, reduciendo el empleo asalariado.

España ha abordado el problema desde el lado del gasto con reformas en 2011 y 2013: aumento de la edad de jubilación, ampliación del periodo de cálculo, y un factor de sostenibilidad y un índice de revalorización que la Ley 21/2021 derogó. En 2025, se ha revalorizado un 2,8 % la pensión contributiva y se han endurecido los requisitos para la jubilación anticipada. Por el lado de los ingresos, desde 2023 se aplica el Mecanismo de Equidad Intergeneracional (MEI), una cotización adicional que sube del 0,6 % al 1,2 % del salario hasta 2029, y desde 2025 la cuota de solidaridad sobre los salarios que superan la base máxima de cotización, que además sube de forma gradual. Otras propuestas son revisar bonificaciones y fomentar planes privados complementarios (modelo británico o mochila austriaca).

También se propone permitir la compatibilidad plena entre pensión y empleo, incentivar la prolongación de la vida laboral y endurecer el acceso a la jubilación anticipada.

Cuentas nocionales Además, se plantea implantar un sistema de cuentas nocionales, como en Suecia o Italia, que simula cuentas individuales vinculadas a las cotizaciones y a la esperanza de vida. Este modelo combina reparto y capitalización, mejora la equidad intergeneracional y refuerza la sostenibilidad del sistema. En concreto, simulan cuentas individuales como en la capitalización, pero sin acumular fondos reales, ya que se financian colectivamente como en el sistema de reparto.

Prestaciones por desempleo

Las prestaciones por desempleo son uno de los pilares del Estado del Bienestar en los países desarrollados[10]. Se definen como ayudas económicas periódicas destinadas a compensar la pérdida de ingresos por desempleo. Su existencia se justifica por razones de equidad, paternalismo y para corregir fallos de mercado como la selección adversa.

Modelos de búsqueda de empleo Desde la perspectiva de los modelos de búsqueda de empleo, estas prestaciones favorecen un mejor emparejamiento entre vacantes y

[9] La AIReF proyecta que el gasto en pensiones pasará del 12,9 % del PIB en 2023 al 16,1 % en 2050, con cerca de 6 millones de pensionistas adicionales. Fuente: AIReF, *Informe de evaluación de la regla de gasto de pensiones*, 2025.

[10] En 2024, España destinó cerca de 22.000 millones de euros a prestaciones por desempleo, impulsado por el aumento del SMI, mayor cuantía media por beneficiario y persistencia del desempleo estructural.

trabajadores, lo que contribuye a mejorar la productividad agregada.

Problemas No obstante, presentan varios problemas.

- Por un lado, pueden generar una "trampa de pobreza": como la prestación se pierde al encontrar empleo, aceptar un trabajo apenas aumenta la renta disponible, lo que equivale a un tipo marginal implícito cercano al 100 % y eleva el salario de reserva. Además, su financiación reduce el salario neto de los ocupados, dado que las cotizaciones sociales suelen recaer sobre los trabajadores, y el desempleo prolongado deteriora el capital humano y la empleabilidad.
- Por otro lado, existe el riesgo de fomentar la economía sumergida, ya que algunos desempleados optan por trabajos no declarados para no perder la ayuda.

Trade-off entre eficiencia y equidad El diseño de los programas de desempleo enfrenta un equilibrio complejo entre eficiencia y equidad. La tasa de reposición, definida como $\frac{w^*}{w} \cdot 100$, donde w^* es la prestación por desempleo y w el salario anterior, representa el porcentaje del salario anterior cubierto por la prestación. Si esta tasa es elevada, el salario de reserva también lo será, lo que puede desincentivar la búsqueda de empleo y prolongar el tiempo en paro. Para evitar este efecto, es esencial que el salario ofrecido en el mercado supere significativamente al ingreso por desempleo. El objetivo no debe ser solo proteger al desempleado, sino fomentar su reincorporación laboral.

Soluciones Entre las soluciones propuestas destacan:

- El rediseño de las prestaciones (ajustando duración y cuantía, o adaptándolas al ciclo económico) y el refuerzo de políticas activas de empleo, como formación obligatoria o incentivos fiscales a empresas que contraten desempleados.
- Según de Kerchove, Gabszewicz y Gérard-Varet (1995), los subsidios al empleo generan mayor bienestar que las prestaciones pasivas.
- También se plantea recurrir a empresas privadas de colocación para mejorar la eficiencia del sistema.
- Como innovación, la mochila austriaca, introducida en Austria en 2003, es un sistema que sustituye la indemnización por despido por una cuenta individual a la que el empleador aporta cada mes un porcentaje del salario, en torno al 1,5 % (Brogueira de Sousa et al., 2019), y que el trabajador conserva al cambiar de empleo. Al no perder lo acumulado al cambiar de empresa, fomenta la movilidad laboral, y al igualar el coste del despido entre contratos reduce la dualidad del mercado laboral. Complementa, sin sustituirla, la prestación por desempleo, y lo no utilizado puede destinarse a la jubilación.

 © 2026 Manuel Cabado García

En conjunto, convertir el seguro de desempleo en un verdadero seguro de empleo es clave para lograr un sistema más justo, dinámico y sostenible.

Programas de reducción de la pobreza

Además de las prestaciones contributivas que sustituyen rentas laborales, los sistemas de Seguridad Social incluyen prestaciones no contributivas destinadas a garantizar una renta mínima a personas con ingresos nulos o muy bajos. Estas ayudas responden a objetivos de equidad, redistribución y lucha contra la pobreza.

Tipos Las prestaciones no contributivas suelen replicar las contributivas (jubilación, desempleo, viudedad, invalidez, dependencia), pero se dirigen a quienes no cumplen los requisitos de cotización.

Problemas Sin embargo, presentan problemas similares a los subsidios por desempleo: generan la llamada "trampa de pobreza" (cada euro adicional ganado reduce la ayuda en casi un euro, lo que desincentiva trabajar) y fomentan la economía sumergida. Además, no abordan la raíz del problema, ya que no mejoran la empleabilidad ni la capacidad de generar ingresos.

Solución Una solución es complementar estas prestaciones con programas de apoyo: incentivos fiscales a la contratación de colectivos vulnerables, formación profesional o estímulos a la educación.

Renta básica universal Una propuesta más ambiciosa es la renta básica universal (RBU), que consiste en otorgar a cada ciudadano una cantidad periódica de dinero sin condiciones. Sus ventajas incluyen equidad, la garantía de un mínimo vital incluso para quien no fue previsor y reducción de las externalidades negativas de la pobreza. Además, responde a tres desafíos clave:

1. Automatización: La Cuarta Revolución Industrial amenaza millones de empleos poco cualificados. El reciclaje laboral es lento y, en muchos casos, inviable. La RBU evitaría la exclusión social de amplios sectores. Finlandia probó este modelo entre 2017 y 2018, pagando 560 €/mes a 2.000 personas. El experimento mejoró el bienestar, aunque no la empleabilidad[11].
2. Simplificación administrativa: Al ser universal, la RBU podría sustituir el complejo entramado de ayudas actuales, facilitando su gestión.

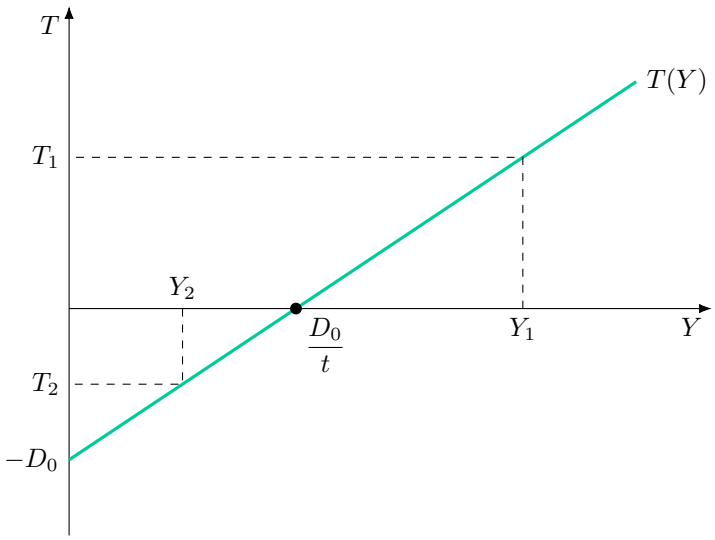
[11] El grupo tratamiento trabajó de media 78 días frente a unos 72 del grupo de control (efecto empleo casi nulo, confirma el texto), pero declaró mayor satisfacción vital (7,3 frente a 6,8 sobre 10) y menos síntomas de malestar psicológico que el grupo de control. Fuente: Kela, resultados finales del experimento, 2020.

3. Eliminación de la trampa de pobreza: La RBU no se retira al encontrar empleo, lo que incentiva la actividad laboral legal.

Problemas renta básica universal No obstante, plantea desafíos importantes:

- Efecto llamada: Podría atraer migración económica, comprometiendo la sostenibilidad del sistema.[12]
- Financiación: Su coste sería elevado, salvo que sustituya al sistema actual mediante un "impuesto negativo sobre la renta" (Friedman, 1962; Tobin et al., 1967), combinando la RBU con impuestos progresivos.
  Gráficamente, en un plano de impuestos y renta como el de la Figura 19.2, se representa la función impositiva $T(Y) = t \cdot Y - D_0$, donde D_0 es el dividendo social que recibe todo ciudadano y t el tipo impositivo. En este contexto, si la renta supera el umbral D_0/t, como en Y_1, pagará un impuesto neto como T_1, y si es menor, recibirá una transferencia neta como T_2.
  Alternativamente, podría financiarse con una subida de impuestos, lo que implicaría distorsiones, pero también beneficios como mayor transparencia y afloramiento de empleo sumergido.
- En definitiva, la RBU representa una transformación profunda del Estado del Bienestar, con potencial para reducir la pobreza estructural, aunque requiere un diseño fiscal riguroso y consenso político amplio.

Figura 19.2: Impuesto negativo sobre la renta



[12] La evidencia sobre el saldo fiscal de los inmigrantes y su efecto sobre la sostenibilidad del Estado del Bienestar se recoge en el apartado 19.A.3 del Apéndice.

 © 2026 Manuel Cabado García

19.4. Provisión de bienes preferentes ⏳5'

Hasta ahora se han analizado las prestaciones económicas del Estado del Bienestar, junto con sus desafíos y posibles soluciones. A continuación se aborda otra vía clave para garantizar equidad y redistribución: la provisión de bienes preferentes.

Concepto Los bienes preferentes son bienes privados que el sector público provee o subvenciona porque considera que, si cada individuo decidiera por sí solo, se consumirían en una cantidad inferior a la socialmente deseable. Este concepto es dinámico y depende del consenso social sobre la dignidad humana y el rol del Estado. Sin embargo, existe acuerdo generalizado en que sanidad, educación y vivienda deben considerarse bienes preferentes. En España, su carácter prioritario está respaldado por la Constitución.

Justificación La intervención pública en estos ámbitos se justifica por múltiples razones: fallos de mercado, búsqueda de eficiencia, criterios de equidad, motivaciones políticas y la necesidad de corregir decisiones individuales que podrían ser perjudiciales para el bienestar colectivo.

Sanidad

La sanidad se define como el conjunto de servicios destinados a preservar o recuperar la salud de la población, incluyendo atención médica y dispensación de medicamentos. Su provisión pública se justifica por razones de equidad (nadie debería quedar excluido por falta de recursos), paternalismo (protección de los menos precavidos), externalidades negativas (como epidemias o baja productividad) y problemas de selección adversa en mercados con información asimétrica.

Modelos de provisión Existen dos grandes modelos de provisión pública:

- Modelo de Seguridad Social (modelo Bismarck)[13]: Profesional, cubre a trabajadores y sus familias. Se financia mediante cotizaciones sociales proporcionales a la renta. Ventajas: mayor concienciación sobre el coste y contención del gasto. Ejemplos: Alemania, Suiza.
- Servicio Nacional de Salud (SNS, modelo Beveridge)[14]: Universal, cubre a todos los residentes legales. Se financia con impuestos generales. Ventajas: mayor equidad y

[13] El modelo debe su nombre al canciller alemán Otto von Bismarck, que implantó el primer sistema de seguro social obligatorio del mundo mediante la Ley del Seguro de Enfermedad de 1883, en vigor desde diciembre de 1884 y financiada con cotizaciones de empresarios y trabajadores. La medida respondía a una estrategia política para frenar el auge del movimiento socialista alemán, ofreciendo protección social desde el propio Estado.

[14] El modelo toma su nombre del Informe Beveridge (1942), que sentó las bases del Estado del Bienestar británico de posguerra y de su Servicio Nacional de Salud (NHS), creado en 1948 como sistema universal financiado con impuestos generales y no con cotizaciones ligadas al empleo.

redistribución. Ejemplos: España (desde la Ley General de Sanidad de 1986), Reino Unido, Suecia, Italia.

Modelos de producción La producción de servicios sanitarios puede ser pública o privada, incluso dentro de sistemas de provisión pública:

- Producción privada: El paciente acude a centros privados, pero el Estado cubre total o parcialmente el coste. Ventajas: libertad de elección, competencia entre profesionales, mayor visibilidad del coste. Ejemplo: Francia.
- Producción pública: El Estado construye infraestructuras, contrata personal y presta directamente el servicio. Ejemplo: España.

Problemas identificados

- Producción privada:
  - Incentivos perversos por demanda inducida (decidida por médicos).
  - Pérdida de control del gasto público.
  - Desabastecimiento en zonas rurales o poco rentables.
- Producción pública:
  - Ineficiencia por falta de competencia y presión del mercado.
  - Listas de espera, sobrecostes y calidad variable.
  - Asignación automática de centro y médico, aunque algunas regiones como Madrid han implementado el modelo de "área única"[15] para permitir elección.
- Problemas generales:
  - Envejecimiento poblacional: Aumenta la demanda y el gasto sanitario.
  - Demanda excesiva: Por percepción nula del coste en el momento de uso, que genera riesgo moral, porque el paciente asegurado consume más servicios de los que pagaría de su bolsillo.
- Copago: Introducido como posible solución para racionalizar el uso. En España, el sistema se reformó en 2012[16], afectando especialmente a pensionistas. Sin embargo, el copago puede ser regresivo, inefectivo (por la inelasticidad de la demanda) y contraproducente si retrasa tratamientos.

[15] Ley 6/2009, de 16 de noviembre, de Libertad de Elección en la Sanidad de la Comunidad de Madrid, desarrollada por el Decreto 51/2010, de 29 de julio, que permite elegir médico y hospital en toda la región, no solo en el área sanitaria de residencia.

[16] Real Decreto-ley 16/2012, de 20 de abril, que eliminó la gratuidad farmacéutica para pensionistas: copago del 10 %, con un tope mensual de 8 € para rentas anuales inferiores a 18.000 € (18 € y 60 € para tramos de renta superiores).

 © 2026 Manuel Cabado García

Gasto farmacéutico España destina cerca del 19 % del gasto sanitario a medicamentos, por encima del 15 % de Alemania y de la media de la OCDE (16 %). Para reducirlo se han adoptado medidas como:

- Centralización de compras.
- Prescripción por principio activo (genéricos).
- Ajuste del número de unidades por envase.
- Reformas en el copago farmacéutico desde 2012.

Educación

La intervención pública en educación se justifica por múltiples razones:

- Externalidades positivas: las externalidades positivas que genera (como mayor productividad, cohesión social o innovación) provocan que, en ausencia de intervención, se infraproduzca[17].
- Mercados de capital imperfectos, dificultan que individuos con alto potencial puedan financiar su formación.
- Equidad: desde una perspectiva de equidad, la Teoría del Capital Humano de Becker (1964) señala que la educación incrementa la productividad y los salarios. Si solo los más ricos acceden a ella, la desigualdad se acentúa.

Modelos de provisión

- En España coexisten la provisión pública y privada. Dentro de la pública, la producción puede ser también privada, como ocurre en los colegios concertados.
- Educación obligatoria: La enseñanza es obligatoria desde los 6 hasta los 16 años, y gratuita desde los 3 años. En el curso 2016/2017, el 68 % de los alumnos asistía a centros públicos, el 26 % a concertados y el 6 % a privados. España es uno de los países europeos con menor proporción de alumnos en escuelas puramente públicas.
- Educación universitaria: El 87 % de los estudiantes cursa estudios en universidades públicas, donde se paga una matrícula que cubre aproximadamente el 20 % del coste. Las privadas concentran más alumnos en áreas como ciencias de la salud.

Retos y propuestas

[17] La conexión formal entre capital humano/educación y crecimiento de largo plazo la formaliza Paul Romer, "Endogenous Technological Change", *Journal of Political Economy*, 98(5), 1990, pp. S71-S102 — Nobel de Economía 2018 por integrar la innovación tecnológica en el análisis del crecimiento a largo plazo.

- Primaria y secundaria: El debate gira en torno a si el Estado debe apostar por centros públicos o concertados. Aunque el gasto por alumno es menor en los concertados, esto se explica por factores como mayor número de alumnos por clase, menores salarios docentes y menor presencia de estudiantes con necesidades especiales. Además, la escuela pública garantiza cobertura territorial, incluso en zonas poco rentables.
    - La solución pasa por mantener ambos modelos, pero con reformas:
    - Introducir más competencia en la pública (por ejemplo, rankings ajustados por contexto socioeconómico).
    - Reforzar el control estatal en la concertada (evitar selección de alumnos o cuotas excluyentes).
- Universitaria:
    - Combatir la regresividad de las tasas con precios ajustados a la renta y más becas.
    - Mejorar eficiencia mediante préstamos, tasas más altas en segundas matrículas y contabilidad de costes.
    - Fomentar la colaboración con el sector privado para adaptar la oferta formativa al mercado laboral.
    - Abordar la sobrecualificación incentivando la formación profesional, clave para cubrir vacantes en sectores técnicos y especializados.

Vivienda

El acceso a una vivienda digna es un pilar fundamental del Estado del Bienestar, justificado por razones de equidad y por las externalidades positivas que genera, como la reducción de la delincuencia o la mejora de la cohesión social.

Oferta y demanda El sector público puede intervenir para facilitar el acceso a la vivienda desde dos vías:

- Por el lado de la demanda: mediante incentivos fiscales a la compra o ayudas directas al arrendatario. Sin embargo, los incentivos a la propiedad pueden ser contraproducentes si la oferta es inelástica, ya que benefician principalmente a promotores y encarecen el mercado. Además, fomentan la inmovilidad laboral, agravando el paro friccional.
- Por el lado de la oferta: a través de la construcción de viviendas de protección oficial, el aumento del suelo urbanizable o incentivos fiscales a los arrendadores para dinamizar el mercado del alquiler.

 © 2026 Manuel Cabado García

Además, las políticas públicas deben considerar la eficiencia energética y funcional de los edificios, promoviendo construcciones sostenibles que reduzcan el impacto ambiental y los costes a largo plazo.

Una estrategia equilibrada entre propiedad y alquiler, junto con una planificación urbana inclusiva, es clave para garantizar el derecho a la vivienda en contextos de creciente desigualdad y presión demográfica.

Ejemplos en España

- El Plan 2022–2025 introdujo el Bono Alquiler Joven, impulsó la rehabilitación energética mediante fondos europeos y fomentó la colaboración con entidades sociales para ampliar el parque de vivienda asequible. Este plan refleja una apuesta por equilibrar propiedad y alquiler, mejorar la eficiencia energética y garantizar el acceso a vivienda digna en todo el territorio.
- Como proyecto, el Plan Estatal de Vivienda 2026–2030 triplicará la inversión pública hasta 7.000 millones, consolidará un parque público de vivienda asequible con calificación permanente, reforzará la rehabilitación en zonas rurales y creará una base de datos transparente de precios.

Otros bienes preferentes

Además de sanidad, educación y vivienda, existen otros bienes que, por su impacto en la equidad y el bienestar colectivo, pueden considerarse preferentes en el marco del Estado del Bienestar.

- Seguridad ciudadana es un bien esencial que garantiza la integridad física y la convivencia pacífica. Su provisión pública responde a la necesidad de proteger a todos los ciudadanos, especialmente a los más vulnerables, frente a riesgos que no podrían afrontar individualmente.
- Justicia cumple una función redistributiva y de garantía de derechos. El acceso universal a la tutela judicial efectiva, consagrado en el artículo 24 de la Constitución Española, evita que la resolución de conflictos dependa del poder adquisitivo, reforzando la cohesión social y la confianza institucional.
- Medioambiente se ha consolidado como bien preferente por sus externalidades positivas sobre la salud, el desarrollo sostenible y la calidad de vida. La protección ambiental no solo beneficia a las generaciones presentes, sino que preserva recursos para las futuras, lo que justifica la intervención pública.
- Conectividad digital, en el contexto actual, se ha convertido en un requisito básico para el acceso a la información, la educación, el empleo y la participación demo-

crática. La brecha digital reproduce desigualdades sociales, por lo que garantizar el acceso universal a internet y a competencias digitales es ya una prioridad política.

- Cultura es un vehículo de desarrollo personal, integración social y participación ciudadana. Su acceso equitativo permite combatir la exclusión, fomentar el pensamiento crítico y preservar la identidad colectiva.

19.5. Impacto redistributivo ⏳6'

El sector público desempeña un papel clave en la redistribución de la renta, tanto a través de los ingresos como del gasto:

- Un sistema impositivo progresivo —donde quienes tienen más contribuyen proporcionalmente más— potencia la capacidad redistributiva del Estado.
- Por otro lado, el gasto público orientado a servicios sociales, transferencias y políticas inclusivas puede reducir las brechas económicas.

La eficacia de esta redistribución se refleja en indicadores como el índice de Gini, que mide el grado de desigualdad: cuanto más bajo es su valor, menor es la desigualdad que queda tras la intervención pública.

Índice de Gini El Índice de Gini es una medida de desigualdad en la distribución de la renta o riqueza dentro de una población. Se basa en la curva de Lorenz, que representa gráficamente el porcentaje acumulado de renta que posee cada porcentaje acumulado de población, ordenada de menor a mayor renta. El índice se calcula como el área entre la curva de Lorenz y la línea de perfecta igualdad, dividida por el área total bajo esa línea. Su valor oscila entre 0 y 1: un Gini de 0 indica igualdad perfecta (todos tienen la misma renta), mientras que un Gini de 1 representa desigualdad extrema (una sola persona concentra toda la renta). Aunque no mide pobreza directamente, es útil para comparar niveles de desigualdad entre países o a lo largo del tiempo, como recoge la Figura 19.3.

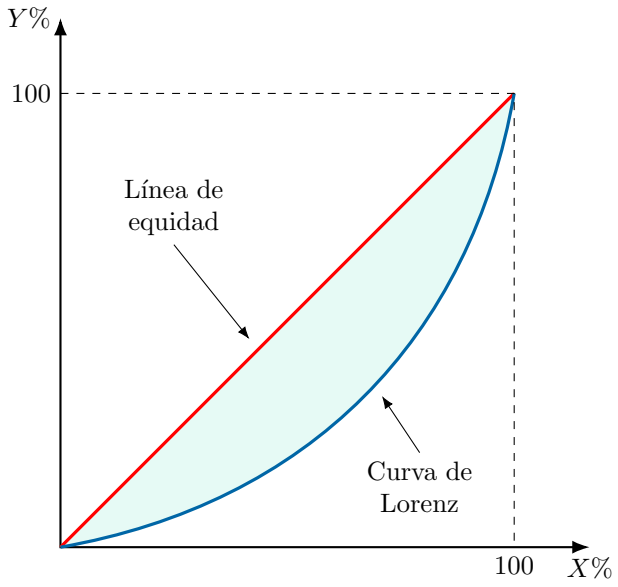
A pesar de la relevancia y el peso del gasto público en los presupuestos fiscales, su análisis desde una perspectiva distributiva es relativamente reciente y aún limitado. Tradicionalmente, la atención se ha centrado en la incidencia distributiva de los ingresos públicos, debido en parte a las dificultades metodológicas para definir y medir el impacto del gasto.

Enfoques de gasto En este contexto, De Wulf (1981) presentó cuatro enfoques principales centrados en el gasto exclusivamente:

1. Incidencia impacto: asigna el gasto público a quienes lo perciben directamente.

 © 2026 Manuel Cabado García

Figura 19.3: Índice de Gini



2. Incidencia normativa o legal: analiza en beneficio de quién se realizan los gastos, según su diseño institucional.
3. Incidencia sobre precios relativos: estudia cómo el gasto público afecta los precios de los factores y productos, alterando la distribución indirectamente.
4. Incidencia del beneficio: evalúa el efecto del gasto sobre el bienestar de los beneficiarios, en términos de utilidad percibida.

Identificación de beneficiarios El enfoque más sencillo consiste en identificar a los beneficiarios directos y asumir que solo ellos mejoran su bienestar. Sin embargo, este método presenta dos limitaciones:

1. Equilibrio general: no considera que el GP también beneficia a quienes participan en su provisión (profesores, personal sanitario, funcionarios, etc.).
2. Externalidades: muchos gastos generan efectos positivos sobre terceros, como ocurre con la sanidad o la educación pública.

Beneficio del gasto entre individuos Para distribuir el beneficio del gasto entre unidades individuales, se distingue entre:

- Gastos divisibles: bienes privados provistos por el sector público, como la sanidad o las transferencias monetarias. El beneficio se imputa directamente al consumidor.
- Gastos no divisibles: bienes públicos que benefician al conjunto de la sociedad, como la defensa, la justicia o las infraestructuras, sin un beneficiario identificable.

Evidencia empírica para España

El Décimo Informe del Observatorio de FEDEA (López Laborda et al., 2026) cuantifica el impacto distributivo de impuestos y prestaciones sobre los hogares españoles en 2023, con especial énfasis en el efecto del gasto público. La renta de mercado ampliada creció un 8,7 %, aumentando ligeramente su desigualdad (Gini) de 0,5601 a 0,5636[18]. Sobre ese horizonte, el estudio analiza prestaciones monetarias, sanidad y educación, y la carga fiscal.

Prestaciones monetarias Las prestaciones monetarias (pensiones, desempleo, invalidez, asistencia social y ayudas diversas) aumentaron un 6,4 % en 2023, aunque su peso relativo en la renta bruta cayó al 20,3 % (0,3 puntos menos que en 2022):

- Su efecto redistributivo, medido por el índice de Reynolds y Smolensky (1977), corrigió un 23,1 % de la desigualdad previa, frente al 23,9 % en 2022.
- Índice de Gini tras prestaciones monetarias: 0,4333
- Las pensiones de jubilación aportaron el 15,3 % de esa corrección, seguidas de supervivencia (4,0 %), desempleo (2,1 %) e invalidez (1,7 %).

Sistema fiscal El sistema fiscal, con un tipo medio efectivo del 35,5 % de la renta bruta, redujo la desigualdad en un 1,55 %, frente al 2,65 % de 2022.

- El IRPF, con un tipo medio del 10,4 %, sigue siendo la principal herramienta redistributiva, aunque su capacidad cayó del 4,88 % al 4,42 %.
- Índice de Gini tras impuestos: 0,4266
- El Impuesto sobre Sociedades, imputado solo a accionistas residentes, elevó su tipo medio del 0,9 % al 1,1 % de la renta bruta, mientras que patrimonio y especiales apenas influyeron.
- Las cotizaciones sociales aumentaron la desigualdad en 1,86 % (frente al 0,85 % de 2022).

Gasto en especie El gasto en especie de sanidad y educación completó la redistribución: su peso en la renta bruta se situó en el 8,7 % (sanidad) y el 5,2 % (educación), y redujo la desigualdad un 10,0 % (8,5 % sanidad, 3,0 % educación).

[18] El índice de Gini se expresa a veces en escala 0–100 (como en las estadísticas oficiales del INE) y otras en escala decimal 0–1 (como en el informe de FEDEA), sin que ello suponga ninguna contradicción: son la misma medida en dos escalas distintas.

El informe de FEDEA descompone además el proceso redistributivo paso a paso, partiendo de una renta de mercado con Gini 0,5636 (56,36), que se reduce a 0,4333 tras prestaciones monetarias, a 0,4266 tras impuestos y a 0,3838 con sanidad y educación, mostrando cómo cada intervención pública contribuye a reducir la desigualdad paso a paso.

 © 2026 Manuel Cabado García

- Sanidad pública benefició al 100 % de hogares, con subsidios medios efectivos del 31,6 % para el primer quintil y apenas 1,3 % para el tramo más alto.
- En educación, los hogares del primer quintil recibieron subsidios del 12,1 %, frente al 0,9 % del último centil.
- Índice de Gini tras gasto en especie: 0,3838

Total combinado Combinando impuestos, prestaciones monetarias y en especie, la intervención pública disminuye la desigualdad de la renta de mercado en un 31,9 %.

- Del total de esa corrección, el 72,4 % procede de las prestaciones monetarias, el 23,8 % de sanidad y educación, y solo el 3,7 % de la imposición.
- La redistribución neta muestra que el 60 % de hogares de menores ingresos (tres primeros quintiles) son beneficiarios netos, con subsidios efectivos que van del 81,5 % de su renta bruta en el primer quintil al 14,9 % en el tercero.
- Los dos quintiles superiores son contribuyentes netos: del –4,6 % para el cuarto quintil al –23,1 % para los centiles 91-99, con el centil más rico (el 1 % superior) en –19,7 %.
- Índice de Gini final (tras todas intervenciones): 0,3838

Índice de Gini ajustado por ciclo vital Una aproximación preliminar desde la perspectiva de ciclo vital, que ajusta el índice de Gini por edad del sustentador, revela menor desigualdad inicial:

- Gini ajustado inicial: 0,5074
- Gini ajustado final (tras intervención pública): 0,3793
- La corrección global baja del 31,9 % al 25,3 %.

Otros gastos Incluyendo en el análisis otros gastos no monetarios del sector público (defensa, medio ambiente, servicios generales, vivienda, asuntos económicos y protección social):

- Impacto redistributivo adicional: –9,1 puntos porcentuales
- Corrección total estimada: 41 %
- Índice de Gini final estimado con todos gastos públicos: ≈0,3325

Enfoque alternativo: pensiones como salario diferido Como novedad, el Décimo Informe presenta un enfoque alternativo al tratamiento habitual de las prestaciones contributivas: las pensiones y parte de las prestaciones por desempleo pasan a considerarse

retribuciones salariales diferidas, integradas en la renta de mercado en lugar de en las prestaciones públicas, mientras que las cotizaciones sociales se tratan como ahorro forzoso y dejan de computar como tributo.

- Bajo este enfoque, el Gini de la renta de mercado ampliada baja de 0,5636 a 0,4782.
- El efecto redistributivo conjunto de impuestos y prestaciones pasa del 31,9 % (enfoque tradicional) al 19,7 %.
- Las prestaciones monetarias explican el 41,6 % de la reducción, los impuestos el 13 %, y sanidad y educación el 45,4 %.
- Si se añaden los gastos públicos no clasificados (ver apartado anterior), la corrección se eleva al 30,4 %.

Comparación internacional A nivel internacional, según las cuentas distributivas nacionales, España corrige el 32,5 % de la desigualdad primaria, situándose por debajo de la media del 40 % observada en Europa occidental. Esta diferencia se explica, en parte, por la menor capacidad redistributiva del sistema fiscal español, que apenas aporta el 2,8 % de la corrección total, frente a más del 10 % en otros países europeos. En cambio, más del 90 % del efecto redistributivo en España proviene de prestaciones monetarias, sanidad, educación y otros gastos públicos. Esta estructura sitúa a España en una posición intermedia dentro del modelo de bienestar europeo, con margen de mejora fiscal.

19.6. Conclusiones ⏳30"

Como conclusión, existen dos visiones sobre el papel del sector público en la economía:

1. La visión optimista sostiene que el Estado, mediante el gasto público, garantiza servicios esenciales, reduce desigualdades y promueve la movilidad social, actuando como herramienta clave para la cohesión y el desarrollo económico.
2. La visión pesimista advierte que una mala asignación de recursos puede generar ineficiencias, favorecer intereses particulares y limitar el crecimiento, especialmente si no se evalúa el impacto redistributivo de cada partida.

Por ello, el foco debe situarse en la calidad del gasto: no basta con gastar más, sino con gastar mejor. La redistribución efectiva exige transparencia, rendición de cuentas y políticas públicas orientadas a reducir brechas sociales. En definitiva, el verdadero reto no es cuánto gasta el Estado, sino cómo lo hace y a quién beneficia.

 © 2026 Manuel Cabado García

Apéndice

19.A.1 Modelo de generaciones solapadas

El modelo de generaciones solapadas (Overlapping Generations, OLG) mencionado en el cuerpo del tema tiene dos versiones principales. Samuelson (1958) desarrolló la versión de intercambio puro, sin capital ni producción, centrada en el papel del dinero como depósito de valor intergeneracional. Diamond (1965) extendió el modelo incorporando capital físico y producción. Es esta segunda versión la que realmente sustenta la comparación entre reparto y capitalización desarrollada en el cuerpo del tema.

El modelo de Diamond (1965) En el modelo de Diamond, los individuos viven dos periodos, juventud y vejez, y en cada periodo conviven ambas generaciones. Los jóvenes trabajan, consumen una parte de su salario y ahorran el resto. Los viejos consumen la totalidad del ahorro acumulado más su rendimiento, sin dejar herencia, ya que saben que no vivirán un tercer periodo. Bajo utilidad logarítmica, el ahorro óptimo resulta una proporción constante del salario, independiente del tipo de interés, siendo β el factor de descuento intertemporal:

$$s_t = \frac{\beta}{1+\beta} \cdot y_t \tag{19.4}$$

El equilibrio del mercado de fondos prestables iguala el ahorro agregado de los jóvenes con el stock de capital del periodo siguiente, lo que da lugar a la ecuación fundamental de acumulación de capital del modelo:

$$k_{t+1} = \frac{1}{1+n} \cdot \frac{\beta}{1+\beta} \cdot [f(k_t) - f'(k_t) \cdot k_t] \tag{19.5}$$

donde k_t es el capital por trabajador, n la tasa de crecimiento poblacional y $f(\cdot)$ la función de producción neoclásica por trabajador. El estado estacionario resultante, aunque único y estable en los casos habituales, no tiene por qué ser eficiente en el sentido de Pareto.

Eficiencia dinámica A diferencia de los modelos de horizonte infinito, el equilibrio competitivo del modelo de generaciones solapadas puede no ser eficiente incluso sin ningún fallo de mercado, porque el primer teorema del bienestar supone un número finito de agentes, y en este modelo se suceden infinitas generaciones. Cuando el tipo de interés de equilibrio es inferior a la tasa de crecimiento de la población, la economía sufre sobreacumulación de capital, situación conocida como ineficiencia dinámica: todas las generaciones podrían aumentar su consumo reduciendo el ahorro agregado. Esta es precisamente la posibilidad que se discute en el apartado “Concepto de eficiencia dinámica” del cuerpo del tema, donde Abel, Mankiw, Summers y Zeckhauser (1989) concluyeron que las economías desarrolladas no suelen encontrarse en ese régimen, aunque Blanchard (2019) y Geerolf

(2018) lo cuestionaron con datos más recientes.

Aplicación a los sistemas de pensiones El modelo de Diamond permite formalizar la elección entre reparto y capitalización ya vista en el cuerpo del tema. En un sistema de capitalización, el ahorro forzoso sustituye exactamente al ahorro voluntario, por lo que el stock de capital de la economía no varía. En un sistema de reparto, la rentabilidad implícita de las cotizaciones no es el tipo de interés de mercado, sino la tasa de crecimiento conjunta de la población y de los salarios. El sistema de reparto mejora el bienestar de todas las generaciones frente a la capitalización cuando esa tasa conjunta supera al tipo de interés de mercado —la condición de rentabilidad ya vista en el cuerpo del tema, en el apartado "¿Cuál elegir?", es esta misma desigualdad expresada en tasas de crecimiento continuas— . Esta condición es conocida en la literatura como condición de Aaron (1966), quien formalizó por primera vez bajo qué circunstancias demográficas un sistema de reparto resulta preferible a uno de capitalización.

19.A.2 Cuentas nocionales

Las cuentas nocionales de contribución definida (*Notional Defined Contribution*, NDC), implantadas en países como Suecia o Italia y mencionadas en el cuerpo del tema, simulan cuentas de capitalización individual sin acumular fondos reales: las cotizaciones se contabilizan en una cuenta nocional que se revaloriza cada año a un tipo de interés nocional, pero se siguen financiando colectivamente, como en un sistema de reparto puro.

La OCDE (2005) compara formalmente las cuentas nocionales con los otros dos grandes diseños de pensiones contributivas, prestación definida y sistema de puntos, y muestra que los tres comparten una estructura algebraica equivalente. La pensión nocional del individuo puede escribirse:

$$NA = \sum_{i=0}^{R} w_i \cdot \frac{c}{A} \cdot (1 + r_N)^{R-i} \tag{19.6}$$

donde w_i es el salario de cada año de cotización, c el tipo de cotización, r_N el tipo de interés nocional aplicado al capital acumulado y A el factor de anualización, que traduce el capital nocional acumulado en el momento de la jubilación en una pensión periódica ajustada a la esperanza de vida.

Esta expresión es formalmente equivalente a la de un sistema de prestación definida o de puntos: en los tres casos, la tasa de acumulación efectiva de derechos puede expresarse como el cociente entre dos parámetros de diseño, de modo que decisiones aparentemente distintas —revalorizar el salario computable, fijar el valor de un punto o elegir el tipo de interés nocional— tienen efectos equivalentes sobre la pensión final.

El elemento clave para la sostenibilidad del sistema es la elección del tipo de interés

 © 2026 Manuel Cabado García

nocional r_N. Cuando se fija igual a la tasa de crecimiento de los salarios y del empleo —la misma magnitud que aparece en la condición de Aaron del apartado anterior—, el sistema replica de forma automática el rendimiento implícito de un reparto puro, ajustando la pensión a la evolución demográfica sin necesidad de reformas discrecionales periódicas. Es, en ese sentido, una forma de reparto con contabilidad individual que traslada explícitamente el riesgo demográfico al propio diseño del sistema. Cichon (1999) resume esta naturaleza híbrida con una expresión ya clásica: las cuentas nocionales son "vino viejo en botellas nuevas", el mismo reparto de siempre, mejor comunicado gracias a la ficción de una cuenta individual.

El diseño de cuentas nocionales se ha implantado en varios países desde mediados de los años noventa. Suecia fue pionera, legislando la reforma en 1994 e implementándola en 1999. Italia lo hizo con la reforma Dini de 1995, Letonia en 1996, y Polonia, legislada en 1996, la aplicó desde 1999. Noruega se sumó más tarde, legislando en 2009 e implementando en 2011 una versión modificada del modelo sueco con ajuste automático por esperanza de vida. Fuera de Europa, Mongolia y Kirguistán han adoptado también elementos de cuentas nocionales, a diferencia de Kazajistán, que optó por un sistema de capitalización individual plena.

Un desarrollo posterior de este diseño es el Mecanismo de Ajuste Automático (*Automatic Balance Mechanism*, ABM), estudiado formalmente por Vidal-Meliá, Boado-Penas y Settergren (2009): cuando el balance actuarial del sistema —cociente entre los activos, que incluyen el fondo de reserva y el valor de las cotizaciones futuras, y los pasivos por pensiones— cae por debajo de 1, el tipo de interés nocional se reduce automáticamente hasta restaurar el equilibrio, sin necesidad de una reforma legislativa expresa. El mecanismo sueco se activó por primera vez en 2010, como consecuencia de la crisis financiera de 2008, reduciendo las pensiones nominales en 2010, 2011 y de nuevo en 2014 —la primera vez que un sistema público de pensiones se ajustó a la baja de forma automática, sin una decisión política discrecional—.

19.A.3 Inmigración y sostenibilidad del Estado del Bienestar

En el cuerpo del tema el envejecimiento aparece como la principal amenaza para el sistema de reparto, y la inmigración solo asoma como un riesgo, por el "efecto llamada" de la renta básica universal. Este apartado estudia si la inmigración puede aliviar el problema demográfico de las pensiones y en qué condiciones, con el instrumental de los apartados 19.A.1 y 19.A.2.

La inmigración en la condición de Aaron En el modelo de Diamond la población trabajadora solo crece por la natalidad, a la tasa n. Si en cada periodo llega además un flujo neto de inmigrantes jóvenes igual a una proporción m de la generación joven del

periodo anterior, la tasa de inmigración, la población trabajadora crece a la tasa $n + m$. La condición de rentabilidad del reparto del apartado "¿Cuál elegir?", con la tasa de crecimiento de los salarios, $\frac{\dot{w}}{w}$, la de los jubilados, $\frac{\dot{J}}{J}$, y el tipo de interés, r, queda:

$$\frac{\dot{w}}{w} + n + m - \frac{\dot{J}}{J} > r \tag{19.7}$$

La inmigración amplía así la base de cotizantes y eleva el rendimiento implícito del reparto sin necesidad de subir el tipo de cotización, τ.[19]

Este efecto no es gratuito en equilibrio general. Si los inmigrantes llegan sin capital, la ecuación de acumulación del apartado 19.A.1 se convierte en:

$$k_{t+1} = \frac{1}{1+n+m} \cdot \frac{\beta}{1+\beta} \cdot [f(k_t) - f'(k_t) \cdot k_t] \tag{19.8}$$

de modo que el capital por trabajador, k, disminuye, el salario baja y el tipo de interés, r, sube. Parte de la ventaja se pierde, porque un tipo de interés más alto encarece la deuda pública y un salario más bajo reduce la recaudación (Storesletten, 2000, p. 301). Además, los inmigrantes que hoy suman en m se jubilarán unas décadas después y sumarán entonces en $\frac{\dot{J}}{J}$.

El saldo fiscal neto según la edad de llegada y la cualificación Storesletten (2000) cuantificó esta idea con un modelo de generaciones solapadas calibrado para Estados Unidos. Calculó el valor actual neto, VAN, de los impuestos que pagaría un inmigrante adicional menos el gasto público que generaría, incluido el de sus descendientes. Gráficamente, en un plano de edad de llegada y VAN como el de la Figura 19.4, se representan los perfiles que obtuvo para inmigrantes de cualificación alta, media y baja. Así:

- Todos los perfiles tienen forma de joroba, con un máximo entre los 35 y los 44 años de edad de llegada (Storesletten, 2000, p. 317).[20]
- El máximo, 177.000 dólares de 1993, unas siete veces el producto nacional bruto per cápita, corresponde a un inmigrante muy cualificado que llega con 40 a 44 años (Storesletten, 2000, p. 302).
- El inmigrante poco cualificado tiene un VAN negativo a cualquier edad (Storesletten, 2000, p. 317). El inmigrante representativo de cada grupo aportaba 96.000, −2.000 y −36.000 dólares según su cualificación fuera alta, media o baja, frente a −88.000 dólares de un recién nacido nativo (Storesletten, 2000, p. 302).

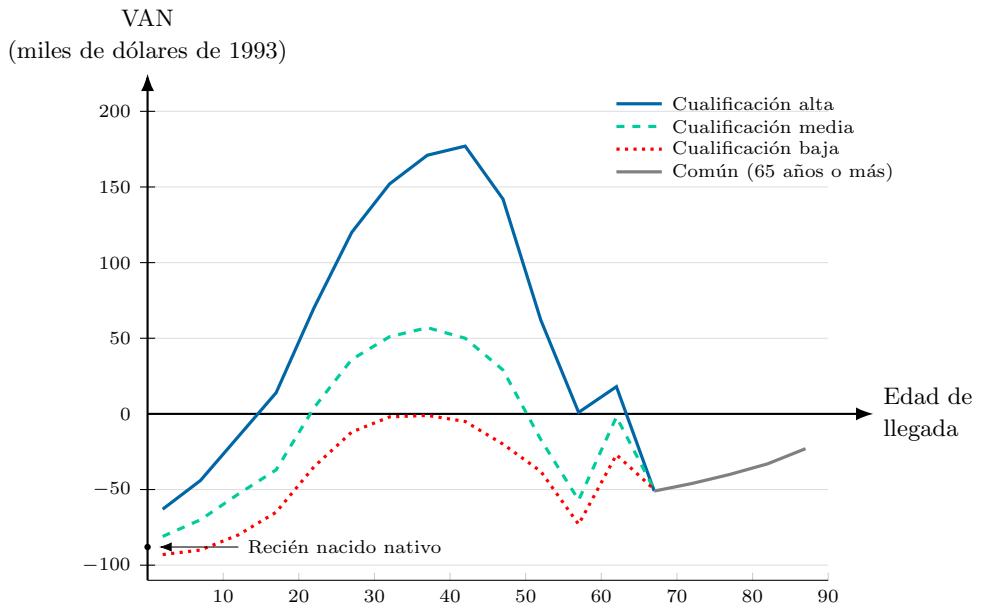
Su conclusión fue que una política selectiva bastaría para equilibrar el presupuesto sin reforma fiscal. Por ejemplo, bastaría con admitir cada año 1,6 millones de inmigrantes

[19] La intuición es que quien llega con treinta años cotiza desde el primer día sin que el sistema haya financiado su crianza ni su educación.

[20] La intuición es que quien llega antes ofrece más años de cotización pero también tendrá más hijos a cargo del presupuesto, y quien llega más tarde se acerca demasiado a la jubilación.

 © 2026 Manuel Cabado García

Figura 19.4: Valor actual neto fiscal de un inmigrante adicional según la edad de llegada y la cualificación



muy cualificados de 40 a 44 años, lo que elevaba la inmigración anual del 0,44 % al 0,62 % de la población. En cambio, más inmigración con la composición media de entonces no lo conseguiría (Storesletten, 2000, p. 301).[21]

La evidencia empírica rebaja el orden de magnitud. La OCDE estimó para 25 países y el periodo 2006-2018 que, en todos ellos, los inmigrantes aportaron en impuestos y cotizaciones más de lo que el Estado gastó en su protección social, sanidad y educación (Damas de Matos, 2021, p. 112). Imputados solo esos gastos individualizables, su contribución neta media fue del 1,56 % del PIB. Imputada además su parte del gasto en bienes públicos puros, como la defensa, pasó a un −0,16 % del PIB, en un periodo en que el saldo de los nativos también fue negativo por el déficit público (Damas de Matos, 2021, p. 120). Cuando la posición de los inmigrantes en edad central de trabajar era menos favorable que la de los nativos, se debía a sus menores contribuciones y no a un mayor gasto público en ellos. El gasto era apenas un 4 % mayor, pero sus contribuciones eran un 17 % menores (Damas de Matos, 2021, pp. 113 y 130).

Para el Reino Unido, Dustmann y Frattini (2014) concluyeron que los inmigrantes llegados desde 2000 habían aportado más de lo que habían recibido, en especial los procedentes del Espacio Económico Europeo. Más recientemente, la *Office for Budget Responsibility* (OBR), el organismo fiscal independiente británico, comparó perfiles ilustrativos de toda la vida (Office for Budget Responsibility, 2024, pp. 107-108):

- Un trabajador inmigrante que llega con 25 años y cobra el salario medio es contri-

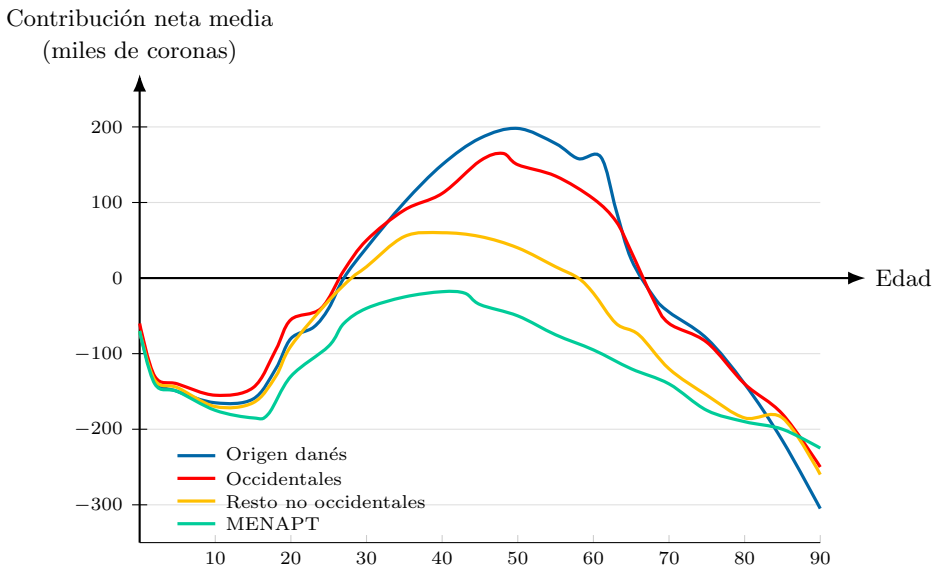
[21] La alternativa sin cambiar la política migratoria era subir 4,4 puntos el impuesto sobre la renta (Storesletten, 2000, p. 302).

buyente neto desde su segundo año. Un residente medio de su misma edad no lo es hasta pasados los 40.

- Con un salario un 30 % superior a la media, el inmigrante seguiría siendo beneficioso para las finanzas públicas aunque viviera hasta los 100 años.
- Con la mitad del salario medio, pasa a ser menos beneficioso que el residente medio a principios de la cuarentena.[22]

La evidencia danesa: el saldo fiscal a lo largo de la vida El Ministerio de Finanzas danés calculó para 2018, con los registros administrativos de Danmarks Statistik, cuánto pagó cada residente en impuestos y cotizaciones y cuánto recibió en transferencias y en servicios públicos como la sanidad, la educación o el cuidado infantil (Finansministeriet, 2021, p. 7). La diferencia es su contribución neta a las finanzas públicas. Gráficamente, en un plano de edad y contribución neta media, en miles de coronas, como el de la Figura 19.5, se representan los perfiles de cuatro grupos de origen: el danés, el occidental, el de Oriente Medio, Norte de África, Pakistán y Turquía (MENAPT) y el resto de países no occidentales.[23] Así:

Figura 19.5: Contribución neta media a las finanzas públicas por edad y origen, Dinamarca, 2018



- Todos los perfiles tienen forma de joroba. La contribución es negativa al principio y

[22] Aun así, la OBR concluyó que cambiar el salario medio de los inmigrantes o la duración de su estancia no altera de forma fundamental la dinámica de la deuda británica a largo plazo, que sigue siendo insostenible al final de su horizonte de cincuenta años (Office for Budget Responsibility, 2024, p. 113).

[23] Cada grupo de origen extranjero incluye a los inmigrantes y a sus descendientes nacidos en Dinamarca. La figura es una versión simplificada de la Figura 2.7 de Finansministeriet (2021, p. 13). A partir de los 70 años el original muestra una media móvil de tres años.

 © 2026 Manuel Cabado García

al final de la vida, porque niños y mayores usan más los servicios públicos y apenas participan en el mercado de trabajo (Finansministeriet, 2021, pp. 12-13).

- En la infancia los perfiles casi coinciden. Las diferencias aparecen en la edad de trabajar (Finansministeriet, 2021, p. 13).[24]
- El grupo de origen danés alcanza un máximo de unas 200.000 coronas hacia los 50 años y el occidental se queda algo por debajo. El resto de no occidentales apenas supera las 50.000 coronas y el grupo MENAPT no llega a ser contribuyente neto a ninguna edad.

Las medias por grupo esconden un efecto de composición. Los inmigrantes suelen llegar ya en edad de trabajar y muchos se marchan antes de jubilarse, de modo que, al corregir por la estructura de edades, su contribución media empeora (Finansministeriet, 2021, p. 13). Sin esa corrección, la contribución media por persona en 2018 fue de 39.000 coronas para los inmigrantes occidentales y de −85.000 para los procedentes de países MENAPT, frente a 8.000 para la población de origen danés (Finansministeriet, 2021, p. 14).[25] Entre los 25 y los 64 años, el 74 % de la población de origen danés fue contribuyente neta, frente al 53 % de los inmigrantes del resto de países no occidentales y al 35 % de los procedentes de países MENAPT (Finansministeriet, 2021, pp. 5 y 14).

El propio informe identificó la variable decisiva, que es el empleo y no el origen. Sea cual sea el motivo de residencia, quien trabaja contribuye en media positivamente, incluidos los llegados como refugiados, y quienes llegan por motivos laborales aportan por persona bastante más que la población en su conjunto (Finansministeriet, 2021, pp. 16-17). Si los inmigrantes no occidentales hubieran tenido la misma tasa de empleo que los daneses de su edad, esta habría subido 28 puntos para los procedentes de países MENAPT y 17 para los del resto. El saldo de los primeros habría mejorado de −13.000 a −3.000 millones de coronas y el de los segundos habría pasado de −1.000 a +5.000 millones (Finansministeriet, 2021, p. 19).[26]

Un estudio reciente para los Países Bajos, con contabilidad generacional sobre microdatos de toda la población, llegó a resultados parecidos (van de Beek et al., 2024). Los inmigrantes laborales que llegan entre los 20 y los 50 años aportan más de 100.000 euros netos a lo largo de su vida. Los llegados por asilo o por reagrupación familiar suponen, en cambio, un saldo negativo de unos 400.000 y 200.000 euros, respectivamente. El nivel educativo resultó decisivo y, de nuevo, el saldo negativo se debía más a las bajas contribuciones que al gasto (van de Beek et al., 2024, p. 40).[27]

[24] La intuición es que un niño cuesta lo mismo en la escuela o en el centro de salud sea cual sea el origen de sus padres, mientras que lo que aporta un adulto depende de si trabaja y de cuánto gana.

[25] Al tipo central de Dinamarca en el mecanismo de tipos de cambio europeo, unas 7,46 coronas por euro, equivalen a unos 5.200, −11.400 y 1.100 euros.

[26] El ministerio advirtió de que el cálculo sobrestima el efecto, porque supone que los nuevos ocupados ganarían lo mismo que los inmigrantes que ya trabajan (Finansministeriet, 2021, p. 20).

[27] Es un documento de trabajo sin revisión por pares. Sus autores concluyeron también que la segunda

¿Solución o alivio transitorio? Los inmigrantes también envejecen Naciones Unidas planteó la pregunta en su informe *Replacement Migration* (United Nations, 2000). Mantener constante la población en edad de trabajar de la Unión Europea entre 2000 y 2050 exigiría unos 79 millones de inmigrantes, 1,6 millones al año. Mantener constante la relación entre la población en edad de trabajar y los mayores de 65 años exigiría 674 millones, unos 13 millones al año (United Nations, 2000, pp. 2-3).[28] Craveiro et al. (2019) repitieron el ejercicio para 2015-2060 y reafirmaron sus conclusiones: el volumen necesario para frenar la caída de la población es del orden de los flujos recientes, pero detener el envejecimiento sigue siendo "un objetivo poco realista" (Craveiro et al., 2019, p. 1324).

Los datos recientes de Naciones Unidas confirman el doble filo. En los países de renta alta, la inmigración neta aportó entre 2000 y 2020 más al crecimiento de la población, 80,5 millones, que el saldo vegetativo, 66,2 millones, y en las próximas décadas será su única fuente de crecimiento (United Nations, 2022, p. 20). Sin embargo, la propia organización advierte de que la migración internacional, por sí sola, no puede compensar a largo plazo el declive ni el envejecimiento de la población y no debe verse como una "solución" a la transición demográfica (United Nations, 2024, p. 23).[29]

La razón es que los inmigrantes acaban, como señaló McFadden, en el numerador de la tasa de dependencia (McFadden, 2001, p. 65). En términos de la condición anterior, una oleada puntual eleva m durante su vida laboral, pero devuelve el efecto en $\frac{j}{J}$ cuando se jubila. Mantener m constante exige flujos absolutos crecientes. Börsch-Supan, Leite y Rausch (2019, p. 7) lo ilustraron para Alemania: ni 500.000 inmigrantes netos al año frenarían el aumento de su tasa de dependencia, para lo que harían falta unos 1,5 millones anuales durante quince años.

Evidencia para España España pasó de 0,35 millones de residentes extranjeros en 1995, el 1 % de la población, a 4,1 millones en 2006, en torno al 10 % (Izquierdo et al., 2010, p. 410). Estos autores concluyeron que ese flujo mejoró las cuentas de la Seguridad Social, pero que ningún escenario migratorio plausible evitaría un aumento significativo del gasto en pensiones (Izquierdo et al., 2010, p. 411). Para 2006-2018, la OCDE estimó una contribución neta de los inmigrantes del 1,70 % del PIB imputando solo los gastos individualizables y del −0,15 % imputando también los bienes públicos puros, cifras muy próximas a la media de la OCDE (Damas de Matos, 2021, p. 120). En cambio, igualar su tasa de empleo con la de los nativos de igual sexo y nivel educativo apenas mejoraría

generación tiende a reproducir la posición fiscal de sus padres, por lo que el problema no se resuelve solo con el relevo generacional (van de Beek et al., 2024, p. 40).

[28] La Unión Europea del informe era la de quince miembros. El propio informe concluyó que mantener esa relación solo con inmigración "parece fuera de alcance" y que, en la mayoría de los casos, bastaría con elevar el límite superior de la edad de trabajar hasta unos 75 años (United Nations, 2000, p. 4).

[29] El *World Social Report 2023* precisó el mecanismo: como los inmigrantes son sobre todo adultos jóvenes, retrasan el envejecimiento unos pocos años sin alterar la tendencia de fondo, y si se quedan y envejecen en el país de destino acaban engrosando las cohortes de mayor edad (United Nations, 2023, p. 32).

 © 2026 Manuel Cabado García

el saldo en un 0,09 % del PIB, frente al 0,37 % de Dinamarca o el 0,69 % de Bélgica (Damas de Matos, 2021, p. 133).

Los estudios y datos oficiales más recientes apuntan en la misma dirección:

- Mantener constante la relación entre población en edad de trabajar y mayores exigiría unos 2 millones de inmigrantes netos al año entre 2015 y 2060, frente a los 273.000 anuales observados entre 1996 y 2015 (Craveiro et al., 2019, p. 1336).
- El Banco de España calculó que mantener constante la tasa de dependencia durante los próximos treinta años exigiría una población nacida en el extranjero en edad de trabajar tres veces mayor que la proyectada por el INE (Banco de España, 2024, p. 135).
- El INE proyecta que los nacidos en España pasen del 79,8 % al 59,6 % de la población en cincuenta años. Sin migración, la población caería a 31,7 millones en 2076, frente a los 53,0 millones del escenario central (INE, 2026, pp. 1 y 5).
- La AIReF estimó que unos flujos migratorios un 15 % mayores o menores que los de su escenario base solo alterarían el gasto en pensiones de 2050 en una décima de PIB, aunque la deuda variaría en unos 6 puntos (AIReF, 2025, p. 168).[30]
- EsadeEcPol estimó que pasar de un saldo migratorio nulo a 330.000 personas al año, la media de 2000-2024, reduce el déficit simulado en 1,7 puntos del PIB, pero que elevarlo a 550.000, la media de 2021-2024, solo lo recorta siete décimas más. Concluyó que "la ventaja demográfica de la inmigración es transitoria" (Almunia y García-Guzmán, 2026, p. 27).[31]
- Funcas concluyó que la inmigración "debe entenderse como un factor de amortiguación, no como una solución permanente" y recordó que en 2025 el 22 % de los inmigrantes residentes en España tenía ya 55 años o más (Cebolla Boado y Miyar Busto, 2026, pp. 9-10).

Implicaciones de política La inmigración es un alivio transitorio y no una solución permanente. Su efecto fiscal depende más de la composición que del volumen. Por ello, la política migratoria puede orientarse hacia trabajadores con la edad y la cualificación que maximizan el saldo neto (Storesletten, 2000, p. 302), anticipando las necesidades que dejarán las jubilaciones (Banco de España, 2024, p. 209). Como el saldo desfavorable procede de menores ingresos y no de más prestaciones, y los casos danés y neerlandés

[30] La propia AIReF subrayó el calendario. Los inmigrantes que llegan ahora cotizan durante las décadas de 2030 y 2040, cuando se jubila el *baby boom*, y alcanzan su etapa de aportación negativa cuando el envejecimiento nativo ya se ha estabilizado. Lo denominó "dividendo fiscal de la inmigración" (AIReF, 2025, p. 176).

[31] El mismo estudio estimó que, entre los 30 y los 54 años, el saldo fiscal neto medio es de 15.900 euros anuales para una persona con estudios superiores, 6.300 con bachillerato o FP de grado medio y 2.500 sin superar la ESO, y que la población nacida en el extranjero presenta saldos menores que la nativa incluso a igual nivel educativo (Almunia y García-Guzmán, 2026, resumen ejecutivo y p. 25).

muestran que el empleo y la educación pesan más que el origen, la palanca principal es la integración laboral, lo que matiza el "efecto llamada" del cuerpo del tema. Por último, la inmigración solo compra tiempo para reformas como el retraso de la edad de jubilación o un tipo nocional ligado a la masa salarial, como en el apartado 19.A.2, que absorbe automáticamente tanto la subida transitoria de $\frac{\dot{L}}{L}$ como su posterior reversión.

 © 2026 Manuel Cabado García

Bibliografía

Aaron, H. (1966). The social insurance paradox. *Canadian Journal of Economics and Political Science 32*(3), 371–374.

Abel, A. B., N. G. Mankiw, L. H. Summers y R. J. Zeckhauser (1989). Assessing dynamic efficiency: Theory and evidence. *The Review of Economic Studies 56*(1), 1–19.

AIReF (2025, marzo). Opinión sobre la sostenibilidad de las administraciones públicas a largo plazo: Demografía y cambio climático. Opinión 2/25, AIReF, Madrid.

Almunia, M. y P. García-Guzmán (2026, junio). Las transferencias intergeneracionales y el reto demográfico. Esadeecpol insight, EsadeEcPol, Madrid.

Banco de España (2024). Informe anual 2023. Informe anual, Banco de España, Madrid.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research / Columbia University Press.

Beveridge, W. (1942). Social insurance and allied services. Informe técnico Cmd. 6404, HMSO.

Blanchard, O. (2019). Public debt and low interest rates. *American Economic Review 109*(4), 1197–1229.

Börsch-Supan, A., D. N. Leite y J. Rausch (2019). Demographic changes, migration and economic growth in the euro area. En *Twenty Years of European Economic and Monetary Union. ECB Forum on Central Banking, 17–19 June 2019, Sintra, Portugal*, pp. 193–230. Fráncfort del Meno: European Central Bank.

Brogueira de Sousa, J., J. Díaz-Saavedra y R. Marimon (2019, mayo). Introducing an Austrian Backpack in Spain. Working paper, European University Institute, ADEMU Project.

Cebolla Boado, H. y M. Miyar Busto (2026). *Los límites de la inmigración para el ajuste demográfico en España*. núm. 113 de Estudios de la Fundación. Serie Economía y Sociedad. Madrid: Funcas.

Cichon, M. (1999). Notional defined-contribution schemes: Old wine in new bottles? *International Social Security Review 52*(4), 87–105.

Craveiro, D., I. Tiago de Oliveira, M. Sousa Gomes, J. Malheiros, M. J. G. Moreira y J. Peixoto (2019). Back to replacement migration: A new European perspective applying the prospective-age concept. *Demographic Research 40*(45), 1323–1344.

Damas de Matos, A. (2021). The fiscal impact of immigration in OECD countries since the mid-2000s. En *International Migration Outlook 2021*, cap. 4, pp. 111–162. París: OECD Publishing.

de Kerchove, A.-M., J. J. Gabszewicz y L.-A. Gérard-Varet (1995). Unemployment benefits versus employment subsidies: A welfare appraisal. CORE Discussion Paper 1995041, CORE, Université catholique de Louvain.

De Wulf, L. (1981). Incidence of budgetary outlays: Where do we go from here? *Public Finance / Finances Publiques 36*(1), 55–76.

Diamond, P. A. (1965). National debt in a neoclassical growth model. *The American Economic Review 55*(5), 1126–1150.

Dustmann, C. y T. Frattini (2014). The fiscal effects of immigration to the UK. *The Economic Journal 124*(580), F593–F643.

Finansministeriet (2021, octubre). Økonomisk analyse: Indvandreres nettobidrag til de offentlige finanser i 2018. Informe técnico, Finansministeriet, Copenhague.

Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.

Geerolf, F. (2018). Reassessing dynamic efficiency. Working paper, UCLA.

INE (2026, junio). Proyecciones de población. años 2026–2076. Nota de prensa, INE, Madrid.

Izquierdo, M., J. F. Jimeno y J. A. Rojas (2010). On the aggregate effects of immigration in Spain. *SERIEs: Journal of the Spanish Economic Association 1*(4), 409–432.

López Laborda, J., C. Marín González y J. Onrubia (2026). Observatorio sobre el reparto de los impuestos y las prestaciones entre los hogares españoles. décimo informe – 2022 y 2023. Informe técnico EEE 2026-13, Fundación de Estudios de Economía Aplicada (FEDEA).

McFadden, D. (2001). Comment [on lee and tuljapurkar, "population forecasting for fiscal planning"]. En A. J. Auerbach y R. D. Lee (eds.), *Demographic Change and Fiscal Policy*, pp. 58–68. Cambridge: Cambridge University Press.

OECD (2005). *Pensions at a Glance: Public Policies across OECD Countries*. Paris: OECD Publishing.

Office for Budget Responsibility (2024, septiembre). Fiscal risks and sustainability. Command Paper CP 1142, Office for Budget Responsibility, Londres.

 © 2026 Manuel Cabado García

Reynolds, M. y E. Smolensky (1977). *Public Expenditures, Taxes, and the Distribution of Income: The United States, 1950, 1961, 1970*. New York: Academic Press.

Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy 66*(6), 467–482.

Storesletten, K. (2000). Sustaining fiscal policy through immigration. *Journal of Political Economy 108*(2), 300–323.

Thaler, R. H. y C. R. Sunstein (2003). Libertarian paternalism. *American Economic Review 93*(2), 175–179.

Tobin, J., J. A. Pechman y P. M. Mieszkowski (1967). Is a negative income tax practical? *Yale Law Journal 77*(1), 1–27.

United Nations (2000, marzo). Replacement migration: Is it a solution to declining and ageing populations? Informe técnico ESA/P/WP.160, United Nations, Department of Economic and Social Affairs, Population Division, Nueva York.

United Nations (2022). World population prospects 2022: Summary of results. Informe técnico UN DESA/POP/2022/TR/NO. 3, United Nations, Department of Economic and Social Affairs, Population Division, Nueva York.

United Nations (2023). *World Social Report 2023: Leaving No One Behind in an Ageing World*. Nueva York: United Nations, Department of Economic and Social Affairs.

United Nations (2024). World population prospects 2024: Summary of results. Informe técnico UN DESA/POP/2024/TR/NO. 9, United Nations, Department of Economic and Social Affairs, Population Division, Nueva York.

van de Beek, J., J. Hartog, G. Kreffer y H. Roodenburg (2024, diciembre). The long-term fiscal impact of immigrants in the Netherlands, differentiated by motive, source region and generation. IZA Discussion Paper 17569, IZA Institute of Labor Economics, Bonn.

Vidal-Meliá, C., M. d. C. Boado-Penas y O. Settergren (2009). Automatic balance mechanisms in pay-as-you-go pension systems. *The Geneva Papers on Risk and Insurance – Issues and Practice 34*(2), 287–317.

 © 2026 Manuel Cabado García

Tema 20

Teoría de la imposición: traslación e incidencia de los impuestos. Enfoques de equilibrio parcial y general. Efecto renta y efecto sustitución de los impuestos. Concepto y medición del exceso de gravamen.

20.1. Introducción ⏳ 1'

La Economía del Sector Público estudia cómo la autoridad pública interviene en el mercado mediante ingresos y gastos para cumplir con funciones de asignación, distribución y estabilidad. Para financiar estas acciones, el Estado recurre a impuestos, definidos como transferencias coactivas del sector privado al sector público, exigidas en virtud del principio de autoridad y sin contraprestación directa.

Pero un impuesto cuesta más que lo que recauda. Al alterar los precios relativos genera una pérdida de eficiencia que nadie recibe y, además, quien lo soporta realmente no tiene por qué ser quien lo ingresa en Hacienda. De ahí surgen dos preguntas: ¿cuánto cuesta, en eficiencia, recaudar un euro? y ¿quién paga realmente ese euro?

La presente exposición responde a ambas preguntas a través del siguiente esquema:

1. Exceso de gravamen: concepto y medición, a partir de los efectos renta y sustitución.
2. Incidencia y traslación de impuestos, en equilibrio parcial y general.

20.2. Exceso de gravamen: concepto y medición ⏳ 10'30"

Comenzando con el exceso de gravamen, todo impuesto disminuye el bienestar de los individuos y afecta a su comportamiento al producir un:

- Efecto renta: pérdida de poder adquisitivo que supone pagar el impuesto, cuyo equivalente monetario es la recaudación. Todos los impuestos que dan lugar a un mismo volumen de recaudación tienen el mismo efecto renta.
- Efecto sustitución: cambio en las decisiones de los individuos, que sustituyen lo gravado por lo no gravado porque el impuesto altera los precios relativos.[1] Así, los agentes tomarán sus decisiones influidos por criterios fiscales, lo cual genera una caída del bienestar que, al contrario de lo que ocurre con el efecto renta, no pasa a formar parte de la recaudación. Esta pérdida de bienestar social derivada del efecto sustitución es lo que se conoce como exceso de gravamen (EG).

A las distorsiones consecuencia del efecto sustitución se las denomina coste de eficiencia o exceso de gravamen de los impuestos, debido a que como consecuencia de los impuestos distorsionantes se alteran los precios relativos introduciendo la denominada "cuña fiscal"[2], entre los precios pagados por los consumidores y los percibidos por los productores, por lo que estos dejan de ser señales correctas de la escasez de bienes en los mercados, no cumpliéndose la condición de eficiencia paretiana[3].

Esta distinción entre efecto renta y efecto sustitución es importante porque, dada la inevitabilidad del efecto renta, el objetivo del diseño del sistema tributario, desde el punto de vista de la eficiencia, será minimizar el exceso de gravamen.

Concretamente, el único impuesto que no es distorsionante es el de suma fija, ya que no altera los precios relativos, tan solo generando un efecto renta puro[4][5][6].

No obstante, hay que señalar que los impuestos de suma fija, a cambio de ser eficientes, son altamente inequitativos, ya que al gravar por igual a todos los individuos con

[1] La intuición es que el efecto renta es lo que el contribuyente pierde porque el dinero pasa a Hacienda, mientras que el efecto sustitución es lo que pierde, además, por cambiar sus decisiones para pagar menos.

[2] Concretamente, es la diferencia entre el precio pagado por el consumidor y el precio recibido por el productor.

[3] Existe, no obstante, un tipo de impuestos que alteran los precios relativos y, sin embargo, mejoran la eficiencia: los impuestos pigouvianos, que gravan una actividad con externalidades negativas en la cuantía del daño marginal, de modo que corrigen una distorsión previa y pueden aumentar el bienestar social en lugar de disminuirlo.

[4] Si no hubiera fallos de mercado (que requieran impuestos distorsionantes, pigouvianos) ni objetivos de equidad distributiva, un impuesto de capitación (suma fija) sería un impuesto óptimo ya que es neutral.

[5] Ejemplo: impuesto por tener una determinada edad.

[6] Que los impuestos de suma fija no generen efecto sustitución y, por lo tanto, no alteren las decisiones en el margen (es decir, mantengan intactas las CPO), no quiere decir que no tengan ningún efecto sobre las cantidades de equilibrio, pues el efecto renta disminuirá la demanda de los bienes (si éstos son normales).

 © 2026 Manuel Cabado García

independencia de su nivel de renta, funcionan de forma regresiva[7]. De esta manera, se suelen usar impuestos distorsionantes, por lo que se estudiará primero cómo actúan sobre el comportamiento de los agentes y después se medirá su impacto en la economía.

Efecto renta y efecto sustitución según el impuesto

El peso relativo de los efectos renta y sustitución depende de la figura impositiva. Considerando un individuo que elige entre consumir, disfrutar de ocio (o trabajar) y ahorrar (Albi et al., 2018):

- Impuesto de suma fija: solo genera efecto renta, porque no altera ningún precio relativo. Por eso no distorsiona y su exceso de gravamen es nulo.
- Impuesto selectivo sobre un bien: encarece ese bien frente a los demás y frente al ocio, y el consumidor lo sustituye por otros bienes y por ocio. Genera efecto renta y efecto sustitución, y el coste de la distorsión es mayor cuanto más elástica es la demanda compensada del bien.
- Impuesto general sobre el consumo: no altera el precio relativo entre bienes, de modo que entre ellos solo hay efecto renta. Pero, como el ocio no está gravado, encarece el consumo frente al ocio y desincentiva la oferta de trabajo. Equivale a un impuesto sobre los salarios si el trabajo es la única fuente de renta y no hay ahorro, y su coste es menor cuanto más inelástica es la oferta de trabajo compensada. Con oferta de trabajo fija sería eficiente (Auerbach, 1985, p. 109).
- Impuesto sobre los salarios o sobre la renta: el efecto sustitución reduce la oferta de trabajo, porque el ocio resulta más barato, y el efecto renta la aumenta, porque hay que trabajar más para mantener el consumo. El resultado es indeterminado y debe resolverse con evidencia empírica, pero la oferta de trabajo siempre es menor que con un impuesto de suma fija de igual recaudación. La evidencia sugiere que, para los trabajadores principales, el efecto sobre la oferta de trabajo es pequeño, pero que puede reducir de forma significativa la oferta de trabajo y la participación de las mujeres (Albi et al., 2018).
- Impuesto sobre los rendimientos del ahorro: reduce el tipo de interés que recibe el ahorrador. El efecto sustitución abarata el consumo presente y reduce el ahorro, mientras que el efecto renta reduce el consumo presente y aumenta el ahorro. El resultado es también indeterminado.

En todos los casos solo el efecto sustitución genera exceso de gravamen. Por eso, para una recaudación dada, el objetivo del diseño impositivo es minimizarlo, y a continuación se estudia cómo medirlo.

[7] El poll tax en R.U. (1990) le costó el puesto a Margaret Thatcher (posteriormente fue derogado). Consistía en que cada municipio dividía su presupuesto de gasto entre todos los contribuyentes y cobraba a cada uno de ellos el resultado del cociente (es decir, la misma cantidad a todos).

Medición del exceso de gravamen

Para medir el exceso de gravamen, en la teoría de la imposición se utiliza el equivalente monetario de la disminución de bienestar que no se traduce en recaudación (Auerbach, 1985). Alternativamente, puede definirse como el exceso de carga impositiva en relación con la que habría producido un impuesto de suma fija de cuantía igual a la recaudación necesaria.

La técnica de medición se basa en la variación equivalente, que puede calcularse con la demanda compensada (hicksiana) y permite evaluar en términos monetarios el cambio de bienestar para un nivel de utilidad dado. A continuación, se verá cómo se puede otorgar un valor monetario al exceso de gravamen en equilibrio parcial, a través de los métodos de la variación equivalente y de la variación de excedentes[8].

Para ello se supondrá:

- Una economía donde existen dos bienes, X e Y, donde el bien Y es numerario, es decir $P_Y = 1$.
- El precio competitivo de la demanda de una unidad es igual al valor marginal para el consumidor.
- El precio competitivo de oferta de una unidad representa el valor marginal en producción.
- Solo se analiza la eficiencia, es decir, la distribución del impuesto va a ser irrelevante.

Estos supuestos van a permitir relacionar precios y utilidad pudiendo cuantificar en unidades monetarias los cambios en el bienestar.

Medición

La idea intuitiva que explica la aparición del exceso de gravamen vincula el coste de ineficiencia al efecto sustitución.

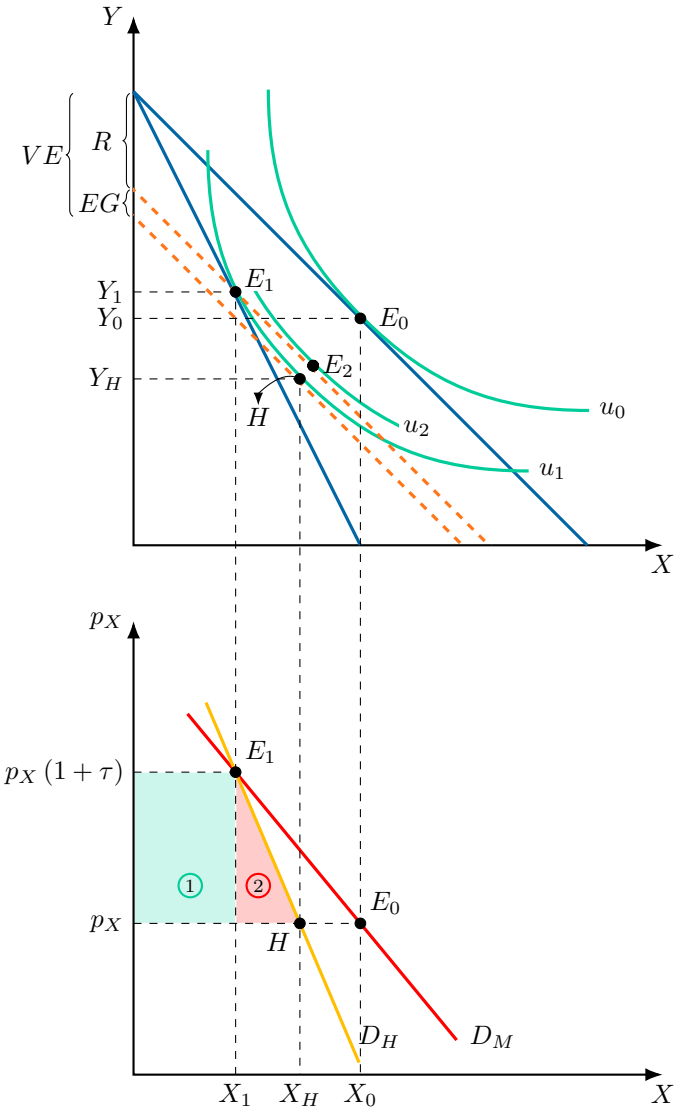
Gráficamente, en un plano de cantidades X e Y, como el de la Figura 20.1, se puede representar el equilibrio de partida de la economía antes de introducir el impuesto en el punto E_0, que representa el punto tangente entre la recta de balance –con pendiente P_X– y la curva de indiferencia más alejada del origen, u_0. Si se introduce un impuesto sobre el consumo de X, el precio de X, P_X, será mayor y la pendiente de la recta de balance cambiará, ahora será mayor y se llegará a un equilibrio en E_1, tangente a la curva de indiferencia u_1, dando lugar a una pérdida de bienestar.

Para medir monetariamente el cambio de bienestar producido se recurrirá al método de la variación equivalente, que mide a cuánta renta estaría dispuesto a renunciar el consumidor para evitar el impuesto. Para ello, se traslada la antigua recta presupuestaria

[8] Idea: se aplican conceptos de la dualidad en la teoría de la demanda a la medición del EG.

 © 2026 Manuel Cabado García

Figura 20.1: Exceso de gravamen



hasta la nueva utilidad hasta que sea tangente, en el punto H, obteniendo la variación equivalente, VE. Así, el consumidor estaría dispuesto a renunciar a VE unidades de renta para evitar el impuesto, representando el cambio de bienestar debido al impuesto en términos monetarios.

Sin embargo, el problema es que esa cantidad refleja una caída de bienestar que tiene dos orígenes[9]:

1. Efecto renta: dicho efecto, ER, mide la recaudación, R, del Gobierno, y será aquella cantidad de renta que hay que restar a la renta inicial para adquirir la cesta E_1.

[9]Nótese también que este análisis gráfico muestra que la cuantía del exceso de gravamen dependerá de forma importante de la curvatura de las curvas de indiferencia, esto es, de la relación entre los bienes: cuanto más sustitutivos sean (es decir, curvas de indiferencia más rectas), mayor será el efecto sustitución y, por lo tanto, mayor será el exceso de gravamen. Por lo tanto, el exceso de gravamen se reduce en bienes que tengan pocos sustitutivos (por ejemplo, la gasolina).

Gráficamente, se desplaza la recta presupuestaria inicial hasta que corte a E_1.
Si el impuesto, en lugar de ser distorsionante, hubiera sido de suma fija, la restricción presupuestaria se habría desplazado paralelamente hacia abajo hasta pasar por E_1 (desplazamiento de la recta presupuestaria inicial hasta que cruza la cesta final (Slutsky)). Así, la cuantificación monetaria de la recaudación (esto es, la pérdida de bienestar asociada al efecto renta) es ER (si Y es numerario, y $ER \cdot P_Y$ si no lo es)

2. Efecto sustitución o exceso de gravamen: en esta nueva restricción presupuestaria el consumidor podría situarse en un punto como E_2 que le reporta mayor utilidad que E_1, $u_2 > u_1$, es decir, con un impuesto de suma fija que detrajese directamente del individuo la cantidad que se deseaba recaudar, éste hubiese disfrutado de un mayor nivel de bienestar. Sin embargo, la alteración de los precios relativos provoca una pérdida de bienestar adicional que no se traduce en una mayor recaudación. Esa pérdida es el efecto sustitución, ES, o exceso de gravamen, EG.

$$ES = VE - ER = EG \tag{20.1}$$

El problema del método de la variación equivalente es que exige conocer las curvas de indiferencia, lo cual es complicado. Por ello, en la práctica, es un método difícilmente utilizable.

Por otro lado, la reducción del bienestar y el exceso de gravamen se pueden medir a través de las curvas de demanda (excedentes), es decir, en términos de la demanda compensada (hicksiana) asociada a la utilidad u_1. Se trasladarían todos los puntos a un plano adyacente y, uniendo E_1 con E_0 se obtendría la demanda ordinaria –que relaciona cantidades demandadas con precios, para un nivel de renta dado– y uniendo E_1 con H se obtendría la demanda compensada hicksiana –que relaciona cantidades demandadas con precios, manteniendo el nivel de utilidad constante, concretamente en u_1–, dando lugar a dos áreas. La suma de ambas sería el cambio en el bienestar, el área 1 la recaudación y el área 2 el exceso de gravamen.

Analíticamente, el exceso de gravamen, que corresponde al triángulo del área 2, puede aproximarse como[10]

$$EG = -\frac{1}{2} \cdot \varepsilon_{p,h} \cdot (p_X \cdot X_H) \cdot \tau^2 \tag{20.2}$$

donde $\varepsilon_{p,h} < 0$ es la elasticidad-precio de la demanda compensada de X, $p_X \cdot X_H$ el gasto en X neto de impuesto y τ el tipo impositivo ad valorem sobre X. El signo menos hace positivo el resultado porque $\varepsilon_{p,h}$ es negativa.

Las implicaciones de política económica serían que para reducir el exceso de gravamen

[10]Si se supone por simplicidad que la curva de demanda hicksiana es lineal, lo que es una buena aproximación para variaciones pequeñas del precio, el área que corresponde al EG en el gráfico es un triángulo y podría determinarse como la mitad de la base por la altura

 © 2026 Manuel Cabado García

(es decir, ineficiencia) de un sistema impositivo, es necesario[11][12]:

1. Gravar bienes cuya demanda sea inelástica (es decir, bienes con pocos sustitutivos).
2. Gravar bienes que no tengan un peso importante en la cesta de los individuos.
3. Establecer un tipo lo más bajo posible y, sobre todo, constante (suavización impositiva). De esta forma, el déficit público será contracíclico y el papel de la deuda será absorber los shocks.

El problema del método de los excedentes es que la demanda compensada no suele ser observable. Por ello, una alternativa imperfecta pero mucho más sencilla será medir el excedente del consumidor utilizando la demanda ordinaria (o demanda marshalliana)[13].

Costes del sistema fiscal sobre el bienestar

Por otro lado, para analizar los costes sobre el bienestar del sistema fiscal a nivel agregado hay que tener en cuenta las relaciones de sustituibilidad y complementariedad entre los bienes, es decir, no será posible sumar los excesos de gravamen de forma individual de cada uno de los impuestos sin tener en cuenta los otros impuestos.

El motivo está relacionado con la teoría del second best, según la cual, cuando existe una distorsión que no puede eliminarse, acercarse a las demás condiciones de eficiencia no garantiza más bienestar. Aplicada a los impuestos, si ya existe un impuesto distorsionante no es posible asegurar a priori que otro impuesto aumente el coste de eficiencia del sistema, e incluso puede reducirlo. Este resultado se explica por la interdependencia de mercados.

Para su análisis, se supondrán:

[11] Ejemplo para España: Albi et al. (2018) calculan, con datos de 2013, que el exceso de gravamen de los impuestos sobre el tabaco (impuesto especial + IVA) es del 30,3 % de la recaudación. Es decir, que por cada euro recaudado mediante impuestos sobre el tabaco, hay una pérdida social de 0,30 € (sin contar la reducción de externalidades negativas y la reducción del gasto en sanidad).

[12] Por lo que el exceso de gravamen será mayor cuanto:

- Mayor sea la sustituibilidad entre bienes, es decir, mayor sea, en valor absoluto, la elasticidad-precio de la demanda compensada del bien X, $\varepsilon_{p,h}$.
- Mayor sea el gasto neto de impuesto en el bien X, $p_X \cdot X_H$.
- Mayor sea el tipo impositivo, τ^2. El hecho de que el tipo esté elevado al cuadrado implica que el bienestar se reduce más que proporcionalmente al aumentar el tipo impositivo. Dupuit ya advirtió en 1844 que la pérdida es proporcional al cuadrado del impuesto (Salanié, 2003, p. 54).

[13] Este análisis se basa en los supuestos de la Teoría cardinal de la utilidad, según la cual:

- El individuo puede observar la utilidad que le reporta el consumo de cada uno de los bienes.
- La UMg del dinero es constante, de forma que puede usarse el dinero como unidad de medida.
- La utilidad es aditiva, separable y cuasilineal respecto a la renta.

Únicamente deberá tenerse en cuenta que, al hacerlo, se estará incurriendo en un error de medición que hará que, si el bien es normal, se sobreestime el exceso de gravamen (Albi et al., 2018). Dicho error será mayor cuanto mayor sea el efecto renta (que es el efecto adicional que capta la demanda ordinaria frente a la compensada). En el ejemplo concreto, el error de medición será igual al área E_0E_1H.

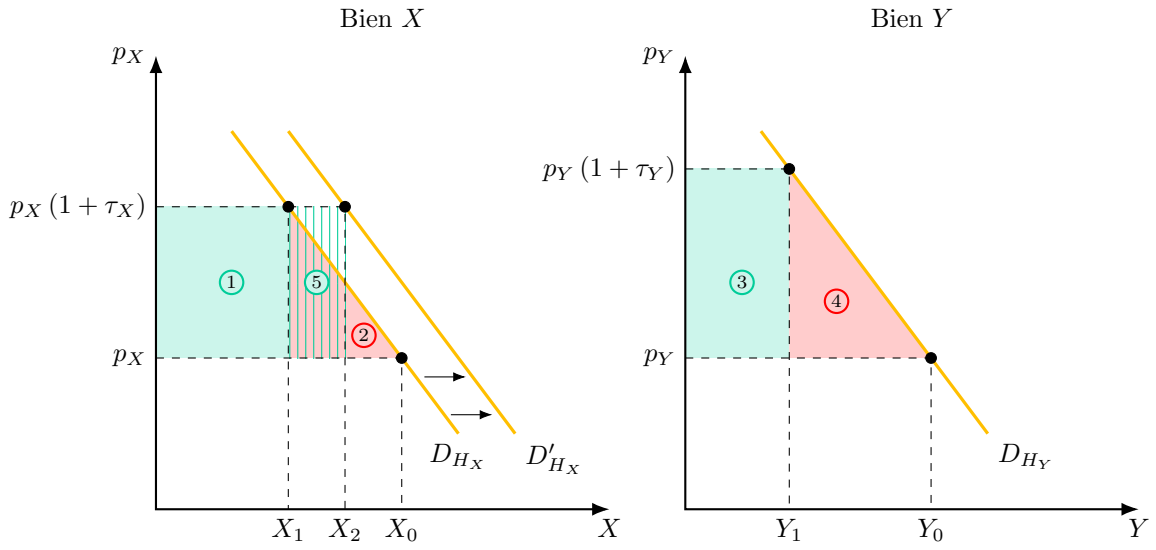
1. Una economía con dos bienes que son sustitutivos producidos en condiciones en las que los costes marginales son constantes[14], de forma que la oferta es perfectamente elástica y el impuesto recae sobre los consumidores y
2. Por simplicidad, se va a suponer que los efectos rentas son de una magnitud despreciable, por lo que se van a analizar demandas compensadas.

Gráficamente, en un plano P_x y x, como el de la Figura 20.2, se representa la demanda compensada de x, D_{H_x} y de forma análoga con el bien y. El bien x se someterá a un impuesto, aumentando el precio, $p_X(1+\tau_X)$, y generando dos áreas, 1 y 2, donde el exceso de gravamen será igual al área 2.

Para contrarrestar la disminución en las ventas de x, se decidirá gravar también el bien y que es sustitutivo de x a un tipo τ_y, que no necesariamente tiene que ser igual a τ_x, generando dos áreas donde el exceso de gravamen quedará determinado por el área 4.

Sin embargo, como son sustitutivos la disminución en la producción de y aumentará la demanda de x, por lo que la demanda compensada de x se desplazará hacia la derecha. La recaudación del impuesto τ_x aumentará en el rectángulo 5 sin disminuir el bienestar, recaudación que puede devolverse al contribuyente, aumentando así su utilidad y pudiendo compensar la pérdida en términos de bienestar producida por el exceso de gravamen, producida a su vez por el impuesto distorsionante establecido en el mercado de y.

Figura 20.2: Exceso de gravamen con más de un impuesto



Por tanto, el efecto total del nuevo impuesto sobre el exceso de gravamen del sistema fiscal es indeterminado. Es decir, aunque la introducción del nuevo impuesto genera una

[14]De hecho, una hipótesis de trabajo habitual en los estudios econométricos sobre los costes de ineficiencia asociados a los impuestos es que la oferta es muy elástica, que puede ser una buena aproximación para un horizonte temporal de medio y largo plazo.

 © 2026 Manuel Cabado García

pérdida de bienestar (es decir, exceso de gravamen) en ese mercado, esta pérdida podría verse más que compensada por el efecto positivo sobre el otro mercado.

En definitiva, el exceso de gravamen del sistema fiscal no es la mera suma de los excesos de gravamen de los impuestos individuales considerados por separado, pues existen relaciones de complementariedad y sustituibilidad entre los bienes. Si hay impuestos preexistentes la introducción de un nuevo impuesto puede aumentar la eficiencia del sistema.

En suma, el exceso de gravamen del sistema fiscal dependerá del tipo de bienes:

- Bienes independientes: suma de los excesos de gravamen en cada mercado.
- Bienes sustitutivos: menor que la suma de los excesos de gravamen en cada mercado (pues aumenta la recaudación en el otro bien).
- Bienes complementarios: mayor que la suma de los excesos de gravamen en cada mercado (pues disminuye la recaudación en el otro bien).

Por otro lado, para obtener el exceso de gravamen del sistema fiscal sería necesario comparar el exceso de gravamen de cada mercado y el incremento de la recaudación, por lo que ahora el exceso de gravamen sería el sumatorio de n bienes de consumo, multiplicado al cambio en la demanda compensada de i al variar el precio de j o efecto sustitución entre los bienes i y j, S_{ij}, por los precios y por los tipos impositivos[15][16]:

$$EG = -\frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} S_{ij} \cdot p_i \cdot p_j \cdot t_i \cdot t_j \tag{20.3}$$

Los términos con $i \neq j$ recogen las relaciones entre bienes: S_{ij} es positivo si i y j son sustitutivos, lo que reduce el exceso de gravamen del sistema, y negativo si son complementarios, lo que lo aumenta, como muestran los casos anteriores.

Sin embargo, los impuestos no solo afectan a la eficiencia sino que también tienen implicaciones en términos de equidad ya que, habitualmente, el diseño del sistema fiscal combina impuestos regresivos y progresivos.

Por ello, para conocer el verdadero efecto de los impuestos sobre la distribución de la renta, a continuación se estudia la traslación y la incidencia de los impuestos, primero mediante un análisis de equilibrio parcial –considerando un único mercado– y, después, mediante un enfoque de equilibrio general –considerando las interrelaciones entre sectores–.

[15] Esta expresión es la aproximación de segundo orden de Harberger (1964), tal como la recoge Auerbach (1985, p. 73).

[16] Ejemplo para España: Alonso-Carrera y Manzano (2003) calculan, con un modelo de equilibrio general dinámico, que el exceso de gravamen del conjunto del sistema impositivo español se sitúa entre el 7,8 % y el 9,6 % del PIB en estado estacionario.

20.3. Incidencia y traslación de impuestos ⏳ 8'

Una vez analizado el exceso de gravamen, se va a analizar tanto la incidencia como la traslación de los impuestos. La incidencia hace referencia a sobre quién recae el impuesto, es decir, quién soporta su carga. Se diferencia entre:

1. Incidencia legal, es decir, quién debe pagar el impuesto según la ley.
2. Incidencia económica, es decir, quién soporta realmente la carga del impuesto a través de los cambios de precios que provoca.

Generalmente, ambas incidencias no coinciden por lo que el impuesto es objeto de traslación impositiva[17] que puede ser total o parcial. Para conocer el efecto de los impuestos en la distribución de la renta hay que estudiar, por tanto, cómo la traslación separa la incidencia económica de la legal. Además, se puede analizar desde un punto de vista de equilibrio parcial o general.

Equilibrio parcial

Comenzando por un análisis de equilibrio parcial, es decir, solamente teniendo en cuenta el mercado que es objeto del impuesto, este es un análisis solamente válido para mercados pequeños. Además, se asume que se va a analizar la incidencia absoluta sin tener en cuenta el destino de la recaudación.

Competencia perfecta

Asumiendo un mercado en competencia perfecta[18], gráficamente, en un plano de cantidades de X y p_X como el de la Figura 20.3 el equilibrio E_0 se va a determinar entre la intersección entre la oferta, S_x, y la demanda, D_x, de X. Se va a introducir un impuesto unitario[19] sobre el consumo, τ, por ejemplo, sobre los hidrocarburos, por lo que el

[17] La traslación impositiva puede explicar fenómenos como por ejemplo:

- Que una reducción de las cotizaciones sociales a cargo de las empresas no abarate el coste laboral en la misma cuantía, porque parte de la rebaja acaba en salarios más altos.
  Ocurre cuando la oferta de trabajo es menos elástica al salario que la demanda, de modo que los trabajadores soportan la mayor parte de las cotizaciones, sea quien sea el obligado legal. La evidencia empírica por lo general apoya la idea de que las empresas trasladan sus cotizaciones sociales a los trabajadores vía menores salarios.
- Que el aumento del IVA produzca un repunte de la inflación.
- Que las ayudas públicas al alquiler y a la compra de vivienda no se traduzcan en una disminución significativa del precio neto que pagan los beneficiarios, porque, con una oferta poco elástica, buena parte de la ayuda se traslada a precios de mercado más altos.

[18] Definición de mercado competitivo. Un mercado (de bienes, de factores, etc.) es competitivo si se dan: i) muchos compradores y muchos vendedores, ii) producto homogéneo, iii) información perfecta, iv) libertad de entrada y de salida, etc.

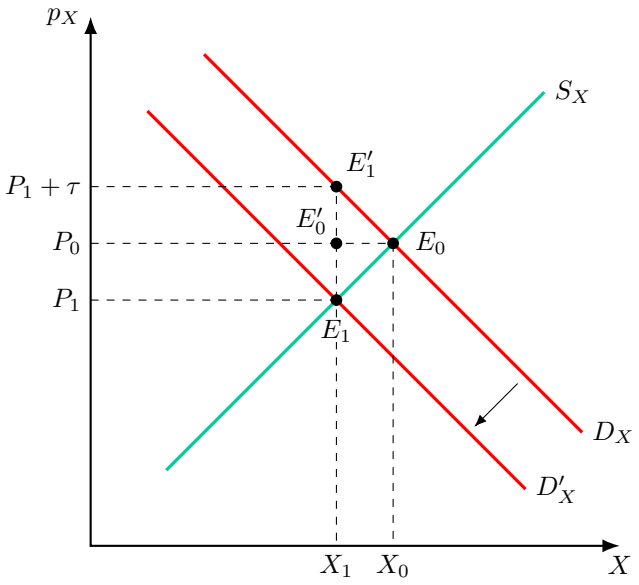
[19] No obstante, un impuesto específico/unitario y un impuesto ad-valorem son equivalentes: si en vez de un impuesto específico se estableciera un impuesto ad-valorem que permitiera obtener la misma re-

 © 2026 Manuel Cabado García

consumidor tendrá que pagar el precio que recibe el productor más el impuesto, $P_1 + \tau$, para consumir X. Así, la demanda tal como la perciben los productores se desplaza hacia abajo en la cuantía del impuesto, D'_X, porque por cada unidad solo reciben lo que paga el consumidor menos τ.

Si el productor se mantuviera en un precio P_0 como el inicial, existirá un exceso de oferta, lo que obliga a los productores a disminuir los precios hasta P_1. El impuesto recae legalmente sobre el consumidor, pero este puede trasladar parte de la carga al productor mediante la caída de la demanda. En el nuevo equilibrio E_1, el consumidor paga $P_1 + \tau$ (punto E'_1) y el productor recibe P_1. El segmento E'_1E_1 es la cuña impositiva τ: el tramo $E'_1E'_0$, de valor $P_1 + \tau - P_0$, lo soporta el consumidor, y el tramo E'_0E_1, de valor $P_0 - P_1$, lo traslada al productor.

Figura 20.3: Incidencia de un impuesto unitario al consumo en equilibrio parcial bajo competencia perfecta



Asimismo, se observa que:

- El triángulo $E'_1E_0E_1$ representa la pérdida de eficiencia o exceso de gravamen.
- La recaudación impositiva sería el rectángulo entre los nuevos precios y los nuevos equilibrios, $P_1 + \tau, P_1, E_1, E'_1$.
- El precio pagado por el consumidor es superior al inicial, $P_1 + \tau > P_0$, lo que reduce el excedente del consumidor.
- El precio percibido por el productor es inferior al inicial, $P_1 < P_0$, lo que reduce el excedente del productor.

caudación, los efectos serían exactamente los mismos.

Así, si se sustituye el impuesto específico por uno ad valorem que recaude lo mismo el desplazamiento de las curvas no habría sido paralelo sino que habría pivotado, pero el resto de conclusiones se mantendrían igual.

En este contexto, como afirma la ley de Dalton (Dalton, 1936, p. 73)[20] la incidencia económica de un impuesto en un mercado competitivo se reparte entre oferentes y demandantes de acuerdo con la elasticidad-precio de la oferta y de la demanda, con independencia de sobre quién recaiga la incidencia legal, de manera que el impuesto recaerá más sobre el agente cuya elasticidad-precio sea relativamente menor en valor absoluto[21][22].

En suma[23]:

1. La incidencia legal no tiene ningún efecto sobre la incidencia económica.
2. La incidencia económica depende de las elasticidades-precio relativas

Por otro lado, un impuesto ad valorem cambiaría la pendiente de la demanda pero el resultado sería similar, mientras que un impuesto sobre los beneficios puros, es decir, sobre los ingresos menos todos los costes, incluida la rentabilidad normal del capital, no provocaría un exceso de gravamen, ya que no cambia los precios relativos y sus efectos serían similares a un aumento de los costes fijos[24].

Monopolio

En un contexto de monopolio[25], a pesar del poder de mercado de la empresa el grado de traslación también dependerá de las elasticidades. El grado de traslación dependerá de las propiedades de la función de costes y de demanda.

Gráficamente, en un plano de precios y cantidades de x como el de la Figura 20.4, se representa la demanda que será lineal, D, el ingreso marginal, IMg, y el coste marginal

[20] Musgrave (1959, p. 292) advirtió de que esta interpretación de la regla solo es válida en un contexto de equilibrio parcial, porque se calcula con los demás precios constantes.

[21] La intuición es que la elasticidad mide la facilidad para escapar del impuesto. Si los consumidores pueden pasarse a otros bienes en cuanto sube el precio, los productores no pueden repercutir el impuesto y soportan la mayor parte de la carga. Si, en cambio, los productores no pueden reducir la cantidad que ofrecen, tienen que aceptar un precio neto menor y la carga recae sobre ellos.

[22] Casos extremos:

- Demanda perfectamente inelástica: el impuesto recaerá en su totalidad sobre los consumidores. Gráficamente la demanda sería vertical.
- Demanda perfectamente elástica: el impuesto recaerá en su totalidad sobre los productores. Gráficamente la demanda sería horizontal.
- Oferta perfectamente inelástica: el impuesto recaerá en su totalidad sobre los productores. Gráficamente la oferta sería vertical.
- Oferta perfectamente elástica: el impuesto recaerá en su totalidad sobre los consumidores. Gráficamente la oferta sería horizontal.

[23] Sin embargo, que un impuesto sobre la demanda no incida sobre ella no significa que no se vea afectada de ninguna manera por el impuesto. Solo significa que no se ve afectado el precio que paga el consumidor. Sin embargo, la cantidad se reduce, por lo que habría compradores que antes compraban el bien que ahora no lo compran.

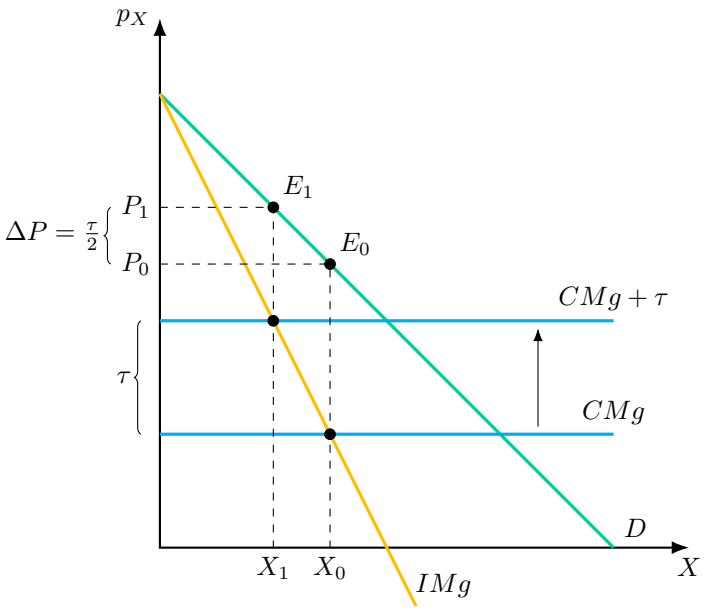
[24] Si el impuesto grava el beneficio contable, que incluye la rentabilidad normal del capital, deja de ser neutral. A corto plazo las empresas no pueden cerrar ni ajustar el tamaño de planta, por lo que se reducen sus beneficios pero se mantiene la oferta. A largo plazo, en cambio, algunas empresas salen del mercado, la cantidad ofrecida se reduce y el precio sube, de modo que el impuesto se traslada a los consumidores.

[25] Un mercado (de bienes, de factores, etc.) es monopolista si se dan: i) muchos compradores, pero un solo vendedor, ii) producto homogéneo y iii) barreras a la entrada.

 © 2026 Manuel Cabado García

constante, CMg. Las condiciones de primer orden establecen que el monopolio deberá de producir hasta que $IMg = CMg$. Si se establece un impuesto específico sobre la producción aumentará los costes marginales, desplazando la curva de costes marginales hacia arriba. La carga del impuesto se reparte al 50 % entre consumidores y monopolista ya que el precio aumenta en una cuantía igual a la mitad del impuesto (Stiglitz y Rosengard, 2015)[26].[27]

Figura 20.4: Incidencia de un impuesto unitario a la producción en equilibrio parcial bajo monopolio



Por otro lado:

- Si la función de demanda es isoelástica –es decir, de elasticidad constante, de modo que la variación porcentual de la cantidad demandada es siempre proporcional a la del precio[28]– y los costes marginales son constantes entonces se produce una sobretraslación, es decir, el monopolista traslada a los consumidores más del 100 % de la carga del impuesto. Esto se observa en la Figura 20.5.

[26] El proceso es el siguiente:

- Partiendo de una demanda lineal, $p = a - b \cdot x$, los precios y output están relacionados linealmente.
- Donde los ingresos totales son $p \cdot x$, es decir, $a \cdot x - b \cdot x^2$, y el ingreso marginal $a - 2 \cdot b \cdot x$.
- Se iguala el ingreso marginal al coste marginal más el impuesto, $a - 2 \cdot b \cdot x = CMg + \tau$ o $2 \cdot p = a + CMg + \tau$
- Por tanto $p = \frac{a+CMg+\tau}{2}$, es decir, si el impuesto τ incrementa en 2€ entonces el precio aumenta en 1€.

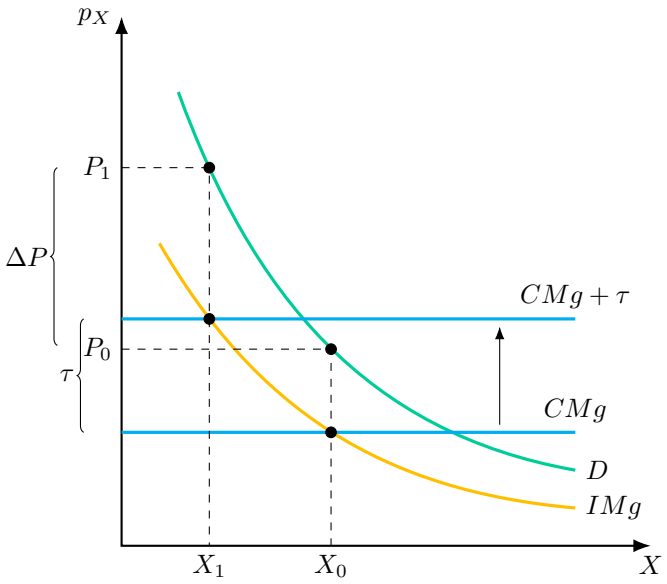
[27] Con el mismo coste marginal constante, la traslación depende de la forma de la demanda: la mitad con demanda lineal, el 100 % con demanda exponencial y más del 100 % con elasticidad constante (Weyl y Fabinger, 2013, pp. 540 y 548).

[28] ¿Por qué una demanda lineal no es isoelástica? Una demanda lineal no es isoelástica porque su elasticidad precio de la demanda cambia en cada punto de la curva.

La elasticidad precio de la demanda, ε_P se calcula como el cambio porcentual en la cantidad dividido

- Si la función de demanda es isoelástica y los costes marginales son lineales y crecientes, la sobretraslación del caso anterior se atenúa: cuanto más empinada sea la curva de coste marginal, menor es la traslación, que puede llegar a ser parcial.

Figura 20.5: Incidencia de un impuesto pero con monopolio bajo demanda isoelástica



Asimismo:

- Si el CMg es vertical, la incidencia económica de un impuesto recae íntegramente sobre los productores. Esto es equivalente a lo que ocurría en un mercado competitivo cuando la oferta era vertical.
- Cuanto más plana sea la pendiente de la curva de CMg, mayor será la incidencia económica del impuesto sobre la demanda. De nuevo, se trata de un razonamiento similar al de mercados competitivos.

Sin embargo, al contrario que en competencia, en el caso del monopolio el impuesto advalorem, para una misma cantidad de equilibrio, siempre logrará una mayor recaudación

---

por el cambio porcentual en el precio. Matemáticamente, esto es

$$\varepsilon_P = \frac{dQ}{dP} \cdot \frac{P}{Q}$$

Para una curva de demanda lineal, la pendiente, $\frac{dQ}{dP}$,es constante. Sin embargo, el cociente $\frac{P}{Q}$ cambia continuamente a lo largo de la curva. Esto es:

- A precios altos y cantidades bajas, el cociente $\frac{P}{Q}$ es grande, por lo que la demanda es elástica.
- A precios bajos y cantidades altas, el cociente $\frac{P}{Q}$ es pequeño, por lo que la demanda es inelástica.

Dado que el valor de la elasticidad cambia, una curva de demanda lineal no puede ser isoelástica. Una curva de demanda isoelástica (de la forma $Q = A \cdot P^{-\varepsilon}$) es aquella en la que la elasticidad es constante en todos sus puntos.

 © 2026 Manuel Cabado García

que un impuesto específico. Dicho de otra manera: si se compara un impuesto ad-valorem con otro específico que den lugar a la misma recaudación, el impuesto ad-valorem siempre dará lugar a una menor reducción de la cantidad de equilibrio, es decir, el monopolista producirá más con un impuesto ad-valorem.

Esto ocurre porque el impuesto ad-valorem reduce el IMg en una cantidad menor al impuesto, mientras que el impuesto específico reduce el IMg justo en la cuantía del impuesto. Por lo tanto, el impuesto ad-valorem da lugar a un nivel de producción de equilibrio mayor.

En suma, los resultados anteriores muestran qué ocurre con las principales reglas de la incidencia impositiva cuando se pasa de un mercado competitivo a un mercado monopolista:

1. La incidencia legal no tiene ningún efecto sobre la incidencia económica, como ocurría en el caso de la competencia perfecta.
2. La equivalencia entre impuestos específicos e impuestos ad-valorem que se daba en mercado de competencia perfecta no se da en mercados monopolísticos.
3. La incidencia económica se seguirá repartiendo entre consumidor y productor en la mayoría de los casos, pues el poder de monopolio no permite al productor trasladar completamente la carga del impuesto al consumidor. No obstante, con demanda isoelástica y coste marginal constante, como se ha visto, la traslación es incluso mayor que la cuantía del impuesto.

Oligopolio

En el caso de un oligopolio no existe una regla general de incidencia: el resultado depende de la conducta de las empresas y de la curvatura de la demanda (Weyl y Fabinger, 2013). Con costes marginales constantes:

- Si las empresas compiten en precios con un bien homogéneo –modelo de Bertrand–, el precio de equilibrio coincide con el coste marginal más el impuesto, de modo que este se traslada íntegramente a los consumidores.
- Si compiten en cantidades –modelo de Cournot– con demanda lineal y n empresas, la traslación es $\frac{n}{n+1}$: el 50 % con una sola empresa y tendiendo al 100 % cuando n crece.
- Si las empresas forman un cartel, actúan como un monopolio conjunto y la traslación es la del monopolio (el 50 % con demanda lineal).

Equilibrio general

No obstante, el análisis anterior tiene una serie de limitaciones ya que ignora las interdependencias entre mercados y no contempla el uso que se le da a la recaudación, por lo

que se va a analizar el impacto de los impuestos desde un marco de equilibrio general, es decir, la introducción de un impuesto también afecta al resto de mercados (Kotlikoff y Summers, 1987).

Para ello, se utilizará el modelo bisectorial de Harberger (1962), expuesto también en Atkinson y Stiglitz (2015), que se basa en los siguientes supuestos:

1. Dos bienes, x e y, producidos en competencia perfecta.
2. Oferta fija de capital, $\bar{K}$, y de trabajo, $\bar{L}$, y perfecta movilidad entre las industrias productoras de x y de y. Además, se remunerarán según el valor de su productividad marginal, que será igual entre sectores. Siendo p_x y p_y los precios de los bienes y PMg^x y PMg^y las productividades marginales en cada sector:

$$w = p_x \cdot PMg_L^x = p_y \cdot PMg_L^y$$
$$r = p_x \cdot PMg_K^x = p_y \cdot PMg_K^y$$

3. La tecnología presenta rendimientos constantes de escala y las funciones serán homotéticas, aunque la intensidad entre bienes puede diferir. Si, por ejemplo, x es más intensivo en capital que y entonces la relación capital-trabajo en x será superior que en y

$$\left(\frac{K}{L}\right)_x > \left(\frac{K}{L}\right)_y$$

4. Los consumidores tienen preferencias idénticas y homotéticas, es decir, ante variación en la renta no cambia el patrón de consumo. Además, se supone que el gobierno devuelve lo recaudado como una transferencia de suma fija. Gracias a estos supuestos, la ratio de consumo de bienes solo depende de los precios relativos y no de la renta de la economía.

Análisis sin impuestos

Las condiciones de equilibrio general competitivo, donde los consumidores maximizan su utilidad sujeta a la restricción presupuestaria y los productores maximizan sus beneficios, ambos en ausencia de impuestos son:

1. La condición de eficiencia en el consumo

$$RMS_{x,y} = \frac{P_x}{P_y} \tag{20.4}$$

2. La condición de eficiencia en la producción

$$RMST_{K,L}^x = RMST_{K,L}^y = \frac{w}{r} \tag{20.5}$$

 © 2026 Manuel Cabado García

3. La condición de eficiencia global

$$RMT_{x,y} = RMS_{x,y} \tag{20.6}$$

A continuación, se estudian los efectos de distintas figuras impositivas.

Impuesto general sobre el consumo

Un impuesto general sobre el consumo (es decir, mismo impuesto sobre X y sobre Y) no altera las CPO, ya que no afecta al precio relativo de los bienes, y por tanto tampoco al de los factores productivos.

$$\frac{P_x\,(1+\tau_c)}{P_y\,(1+\tau_c)} = \frac{P_x}{P_y} \tag{20.7}$$

Así, no se alterará la relación[29] de cantidades consumidas y producidas de ambos bienes. Sus consecuencias, por tanto, se limitan a la renta real: como la oferta de capital y de trabajo es fija, no cambian las cantidades producidas ni la retribución relativa de los factores, y el impuesto reduce en la misma proporción la retribución real de ambos, de modo que lo soportan los consumidores en proporción a su gasto. Solo si la oferta de factores respondiera a su retribución caerían también el empleo y la producción.

Impuesto específico sobre el consumo de x

El impuesto provoca un cambio de los precios relativos de los bienes. En este caso se obtiene de la resolución del problema de optimización del consumidor la siguiente CPO:

$$RMS_{xy} = \frac{P_x\,(1+\tau_x)}{P_y} \tag{20.8}$$

El impuesto abre una cuña entre el precio bruto soportado por la demanda y el percibido por la oferta. Como consecuencia, el equilibrio final no será una asignación eficiente porque la RMS será $(1+\tau_x)$ veces mayor que la RMT:

$$RMS_{xy} = \frac{P_x\,(1+\tau_x)}{P_y} > RMT_{x,y} = \frac{P_x}{P_y} \tag{20.9}$$

Por tanto:

- Como el bien x se encarece en relación al bien y, se reduce la demanda (y producción) del bien x y aumenta la demanda (y producción) de y. Esto se conoce como "efecto producción".
- Del sector x se libera más capital que trabajo, pues el bien x es intensivo en capital, mientras que el sector y absorbe más trabajo que capital, pues el bien y es intensivo

[29] Se subraya que se trata de la relación, no de las cantidades totales.

en trabajo. Esto hace que se produzca un exceso de oferta de capital y un exceso de demanda de trabajo, que se corregirá mediante una alteración del precio relativo de los factores productivos, es decir, aumentará la relación remuneración trabajo-capital, $\uparrow \frac{w}{r}$, en ambos sectores.

- Así, un impuesto selectivo sobre un bien reduce la retribución relativa del factor usado intensivamente en su producción (no obstante, si la oferta de factores no fuese fija las variaciones en la retribución serían amortiguadas por variaciones en su oferta).

Como ocurría con un impuesto general sobre todo el consumo, un impuesto selectivo sobre el consumo de un bien afecta tanto a los consumidores como a los propietarios de los factores productivos, aunque de una manera algo diferente:

- En cuanto a los propietarios de los factores productivos, la rentabilidad nominal (y real) del capital disminuye, y el salario nominal aumenta (con un impuesto general sobre todo el consumo, las retribuciones nominales –y reales– de todos los factores productivos disminuían). Además, el impuesto será soportado en mayor medida por los propietarios del factor intensivo en el bien gravado (en este caso, capital).
- En cuanto a los consumidores, hay que tener en cuenta que cuando se consideran todos los mercados de bienes, la demanda de los consumidores es más inelástica que la demanda de un bien concreto (porque para un bien concreto hay sustitutivos cercanos) y, por tanto, es probable que la incidencia impositiva sobre los consumidores resultante de un impuesto selectivo sea menor que la que produciría un impuesto general.

Impuesto general sobre un factor productivo

Si la oferta total de dicho factor productivo en la economía es fija –es decir, inelástica– (como se supone en este modelo), no se producirían cambios en las cantidades de equilibrio. Además, la incidencia del impuesto recaerá íntegramente sobre el factor gravado.

Si se relajase este supuesto, las cantidades de equilibrio podrían cambiar, y el impuesto también tendría efectos sobre otros factores.

Impuesto específico sobre un factor productivo

Si el sector público introduce un impuesto sobre las rentas del K empleado en el sector x abre una brecha entre el coste del capital para las empresas de x, $r \cdot (1 + \tau_{kx})$, y el que soportan las de y, r. El rendimiento neto r que perciben los propietarios del K es el mismo en ambos sectores, pues el capital es móvil.

En una economía competitiva sin impuestos el productor remunerará a los factores

 © 2026 Manuel Cabado García

productivos según su productividad marginal. En equilibrio, se cumple:

$$RMST^x_{K,L} = RMST^y_{K,L} = \frac{w}{r}$$

que es la condición que garantiza una asignación eficiente de K y L en ambos sectores.

El impuesto, sin embargo, hace que en equilibrio:

$$RMST^x_{K,L} = \frac{w}{r\,(1+\tau_{kx})} \neq RMST^y_{K,L} = \frac{w}{r} \tag{20.10}$$

De la introducción del impuesto se derivan dos efectos importantes sobre la producción:

1. Efecto sustitución factorial (cómo se produce): el impuesto encarece el capital usado en x respecto al trabajo usado en x, por lo que los empresarios sustituyen K por L, aumentando la demanda de trabajo y disminuyendo la de capital, lo que hará disminuir el precio relativo del K, aumentando la relación remuneración trabajo-capital, $\uparrow \frac{w}{r}$.
2. Efecto producción (cuánto se produce): además de a cómo se produce, el impuesto también afecta a cuánto se produce: el impuesto hace aumentar el coste unitario de x al aumentar el coste de uno de sus inputs. Para evitar pérdidas, el sector x aumentará p_x, aumentando su precio relativo.
   Así, el impuesto encarece el bien x respecto al bien y, por lo que se reduce la demanda de x (y de factores productivos en ese sector) y aumenta la demanda de y (y de factores productivos en ese sector). La caída en la producción de x, además, llevará a que varíe el precio relativo de los factores productivos, aumentando la relación remuneración trabajo-capital, $\uparrow \frac{w}{r}$, si el sector x es relativamente intensivo en K.
   Es el mismo efecto que se ha visto antes con un impuesto selectivo sobre el consumo de x, por lo que se produce un exceso de demanda de trabajo y un exceso de oferta de capital.

Por tanto:

- En este caso en que el impuesto recae sobre el factor en el que es intensivo ese sector (capital en x), ambos efectos van en la misma dirección, por lo que los efectos serán los mismos que con un impuesto sobre el consumo de x, pero más intensos: aumenta mucho la demanda de trabajo, y disminuye mucho la demanda de capital, por lo que aumenta mucho el salario nominal y disminuye mucho la rentabilidad nominal del capital.[30]

[30] Con funciones Cobb-Douglas, el capital soporta toda la carga. Harberger obtuvo ese resultado para EE.UU., pero los modelos de equilibrio general computable posteriores muestran que no es robusto (Salanié, 2003, p. 32).

- Si el impuesto recae sobre el factor en el que no es intensivo ese sector (por ejemplo, trabajo en x), los efectos producción y sustitución operarían en sentido contrario y su efecto total sería ambiguo.

Equivalencias impositivas

Los resultados de todo este análisis de equilibrio general se pueden sintetizar en las siguientes equivalencias impositivas del (20.1), donde z es la renta total, τ^z el impuesto proporcional sobre ella y la suma de dos impuestos (por ejemplo, $\tau^K + \tau^L$) indica que se aplican a la vez y al mismo tipo:

- Establecer un impuesto general sobre el consumo, τ^c, es igual que establecer impuestos selectivos iguales sobre el consumo de ambos bienes, $\tau_x + \tau_y$, e igual también que establecer impuestos generales iguales sobre los factores productivos, $\tau^K + \tau^L$.
- Establecer un impuesto selectivo sobre el consumo de x, τ_x, es igual que establecer impuestos selectivos iguales sobre los factores utilizados en dicho sector, $\tau_x^K + \tau_x^L$, y lo mismo aplica para el bien y.
- Establecer un impuesto general sobre un factor productivo capital, τ^K, es igual que establecer impuestos selectivos iguales sobre dicho factor, $\tau_x^K + \tau_y^K$, y lo mismo aplica para el factor trabajo.

Cuadro 20.1: Equivalencias impositivas

$$\begin{array}{ccccccc}
\tau_x^K & + & \tau_x^L & = & \tau_x & & \\
+ & & + & & + & & \\
\tau_y^K & + & \tau_y^L & = & \tau_y & & \\
= & & = & & = & & \\
\tau^K & + & \tau^L & = & \tau^c & = & \tau^z
\end{array}$$

Levantamiento de supuestos

Finalmente, cabe señalar que es posible relajar los supuestos del modelo:

1. n sectores: la ampliación a n sectores no aporta implicaciones adicionales.
2. Si la oferta de factores no es totalmente inelástica el impuesto específico sobre el rendimiento de un factor se traslada parcialmente al otro factor.
3. Si uno de los factores es inmóvil, el efecto será incrementar la carga de los impuestos que recaen sobre factores inmóviles. Por ejemplo, si el capital es inmóvil, un impuesto específico sobre el capital de x será soportado por los propietarios del capital en x. Por otra parte, el factor inmóvil también se puede beneficiar de un impuesto:

 © 2026 Manuel Cabado García

por ejemplo de impuesto selectivo sobre el consumo de x aumentará la demanda y producción de y, si este bien es intensivo en trabajo y el trabajo es móvil mientras que el capital es inmóvil, a medida que se incorpore más trabajo al sector la productividad del capital aumentará, y con ello su remuneración.

4. Heterogeneidad de preferencias, de modo que un impuesto que aumente el precio relativo de x –como un impuesto específico sobre el consumo de x o sobre el rendimiento del capital de x– recaerá con más intensidad sobre aquellos individuos en cuya cesta de consumo x tenga un peso mayor.
5. Variable tiempo: en el análisis llevado a cabo hasta ahora se ha prescindido de la dimensión temporal, por lo que en esta última parte del tema se amplía el análisis a un contexto dinámico que permite evaluar tres cuestiones relacionadas con la incidencia impositiva como son la capitalización impositiva, la incidencia impositiva intergeneracional y la incidencia sobre el ahorro y el crecimiento.
6. Traslación e incidencia regional.

Capitalización impositiva

Si se adopta un enfoque dinámico, puede introducirse el concepto de capitalización impositiva. Se dice que existe capitalización impositiva cuando un impuesto que grava la renta de un activo a lo largo del tiempo se refleja en una reducción del precio de mercado de dicho activo hoy.

Sin impuestos, el precio de mercado de un activo en un mercado competitivo (por ejemplo, la tierra) será igual al valor presente de los rendimientos esperados a lo largo de la vida útil del activo:

$$P = \sum_{t=1}^{\infty} \frac{R_t^e}{(1+r)^t} \tag{20.11}$$

Al introducir un impuesto ad valorem, el rendimiento neto esperado será $R_t^e\,(1-\tau)$ y el precio de mercado será:

$$P = \sum_{t=1}^{\infty} \frac{R_t^e\,(1-\tau)}{(1+r)^t} \tag{20.12}$$

donde R_t^e son las rentas de la tierra en cada período, τ es el impuesto proporcional sobre dichas rentas, y r es el tipo de descuento (que se supone igual al tipo de interés real).

Es decir, se habrá producido una reducción en el precio de mercado del activo. El impuesto incide por tanto sobre quienes sean los propietarios del activo en el momento en que se introduce el impuesto.

Como se ve, el impuesto (aún futuro) produce una reducción del precio actual del activo/bien gravado (hasta el punto de que, si la oferta de dicho bien/activo es totalmente inelástica, la capitalización será total, y quien adquiera el bien gravado después de haberse

anunciado el impuesto no soportará carga alguna).[31]

Este resultado matiza por tanto las conclusiones de los modelos estáticos, ya que implica que un impuesto sobre la renta de un factor de oferta fija incide sobre los propietarios que ya lo fueran cuando se estableció el impuesto, pero no sobre aquellos otros que adquieran el factor con el impuesto ya en vigor, ya que estos últimos adquirirán el activo a un precio más bajo que descuenta los impuestos futuros y no soportarán carga alguna.

Traslación intergeneracional

Al ampliar el análisis de la incidencia impositiva a un contexto dinámico no interesa ya solo conocer la distribución de la carga de los impuestos entre capitalistas y trabajadores, sino que también hay que analizar su incidencia intergeneracional, es decir, ¿qué generación soporta la carga del impuesto?

Para ello se supone un modelo de generaciones solapadas, OLG, donde conviven dos generaciones:

1. Jóvenes que trabajan a cambio de un salario que destinan al consumo y al ahorro.
2. Jubilados que no trabajan y consumen lo que ahorraron cuando eran jóvenes más el interés acumulado.

Si en un momento se introduce un impuesto sobre la renta del K cuya recaudación se transfiere a los jubilados, ocurre lo siguiente:

- Los jubilados no se verán afectados, ya que no pueden alterar su decisión de ahorro (la tomaron en el periodo anterior) y a corto plazo el impuesto es neutral: la rentabilidad que pierden los jubilados por la introducción del impuesto se compensa con la transferencia recibida.
- Los individuos jóvenes, sin embargo, aunque saben que su riqueza vital no ha cambiado (ya que en el futuro recibirán como transferencia los impuestos que hoy recaigan sobre su ahorro), encuentran más atractivo consumir hoy que hacerlo mañana, ya que el rendimiento del ahorro se habrá reducido por el impuesto.

Así, ahorrarán menos y consumirán más, dejando un menor stock de K a la generación de jóvenes del período siguiente. La reducción del K por trabajador en el futuro dará lugar a una caída de la PMg_L, y con ello del salario futuro.

En definitiva, el impuesto será trasladado a futuras generaciones[32].

[31] A medio y largo plazo, las ofertas de factores tienden a variar, de modo que la capitalización es solo parcial en la mayoría de los casos (Albi et al., 2018, apartado 3.2).

[32] El efecto habría sido distinto si la recaudación se hubiese transferido a los jóvenes, ya que estos habrían ahorrado una parte de la transferencia, aumentando el stock de K futuro. En este caso mientras que la generación de jóvenes vería aumentar su bienestar la de jubilados resultaría perjudicada.

 © 2026 Manuel Cabado García

Incidencia sobre el ahorro y equivalencia ricardiana

Completando el análisis dinámico, quedan dos cuestiones relacionadas con la incidencia en el tiempo: la incidencia de los impuestos sobre el ahorro y el crecimiento y el papel de la equivalencia ricardiana en la incidencia intergeneracional. Así:

1. La incidencia impositiva sobre el ahorro y el crecimiento fue analizada por Feldstein (1974) mediante el modelo de Solow con tasa de ahorro variable. Si el impuesto sobre los rendimientos del ahorro reduce la tasa de ahorro (porque domina el efecto sustitución), se reduce la acumulación y el stock de capital y, con ello, la renta per cápita de equilibrio –en estado estacionario–. Feldstein concluyó que buena parte de la carga se traslada así a los trabajadores, cuyo salario cae al reducirse la relación capital-trabajo.
2. La incidencia intergeneracional de la deuda pública, que se analiza partiendo de la equivalencia ricardiana. Esta establece que, en un contexto de certidumbre, con agentes con expectativas racionales y vida infinita, impuestos de suma fija y mercados de crédito perfectos, financiar el gasto público mediante deuda equivale a financiarlo con un aumento de impuestos futuro, que los agentes anticipan ahorrando hoy lo necesario para pagarlo. Por tanto, bajo estos supuestos la deuda no traslada carga alguna a las generaciones futuras. Si alguno de ellos falla, sí se traslada parte de la carga.[33]

Incidencia en el espacio

El modelo de Harberger ha sido utilizado por McLure (1969) y otros autores en el estudio de la traslación de impuestos entre regiones o países.

En efecto, aunque dicho modelo está referido a una economía cerrada, designando por x e y a la producción agregada de dos regiones, los resultados del modelo son plenamente aplicables:

- Un impuesto sobre la producción de las dos regiones al mismo tipo de gravamen será soportado por los residentes de ambas regiones en proporción a sus niveles de renta o gasto.
- Si la región x introduce un impuesto sobre su producción, la demanda de los agentes se recompondrá en favor de los productos de la región y. En la vertiente de los usos de renta, el impuesto tenderá a ser soportado por aquellos individuos que mayor preferencia tienen por x, con independencia de la región en la que residan.

Por su parte, la incidencia sobre las fuentes de renta dependerá en su signo de las intensidades factoriales:

[33] Buchanan (1958) define la carga de la deuda como la disminución de bienestar de las generaciones futuras por usar deuda en lugar de impuestos (Albi et al., 2018).

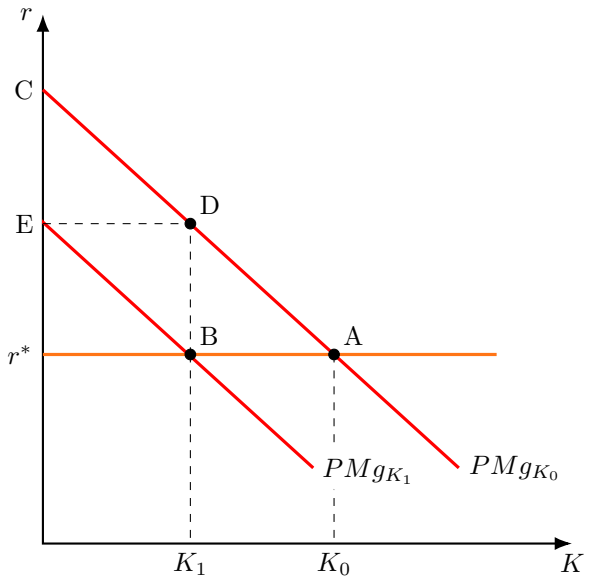
- Si la producción de la región x es intensiva en capital, los propietarios del capital de ambas regiones resultarán relativamente perjudicados por el impuesto, y viceversa.
- Si la región x decide gravar los beneficios del capital instalado en ella, la incidencia de este impuesto dependerá del grado de movilidad del capital.
    - Si éste es un factor fijo, los propietarios del capital en la región x soportarán plenamente el impuesto.
    - Si fuese plenamente móvil, la dirección de los cambios en los precios de los factores de producción dependerá del signo y el orden de magnitud de los efectos producción y sustitución factorial.

En suma, la política impositiva de una región o un país puede incidir sobre la renta y el bienestar de los individuos que residen en otras regiones o países, lo que se conoce como exportación impositiva. Su grado será función de la sustituibilidad en la demanda y en la producción –sustituibilidad de los factores–, de las intensidades factoriales, del grado de movilidad de los factores de producción y del tamaño relativo de las economías.

La importancia de la movilidad factorial y del tamaño se puede ilustrar si se considera el caso de un país pequeño que introduce un impuesto sobre las rentas del capital.

Al tratarse de un país pequeño, si se supone que el capital es perfectamente móvil, el tipo de interés tendrá que igualarse al del resto del mundo, r^*. Gráficamente, en un plano de tipo de interés y capital como el de la Figura 20.6, el equilibrio se encuentra inicialmente en el punto A, en el que la demanda de capital, dada por su productividad marginal, PMg_{K_0}, se iguala a la oferta, perfectamente elástica al tipo de interés mundial con un stock de capital K_0.

Figura 20.6: Impuesto al capital en un país pequeño



 © 2026 Manuel Cabado García

Si se introduce un impuesto sobre las rentas del capital, τ^K, se reduce la rentabilidad neta del capital y se produce una salida de capitales al exterior, de modo que el capital invertido en el país cae hasta K_1. La remuneración total del capital no se ve afectada, ya que el capital K_1 invertido en el país recibe un rendimiento neto igual al tipo de interés mundial, r^*, y el capital que salió al exterior, es decir, la diferencia entre K_1 y K_0, no está gravado y recibe r^*.

No obstante, se produce una pérdida de eficiencia (el exceso de gravamen, igual al triángulo ABD) y una reducción de la remuneración del trabajo igual a la recaudación del gobierno (el rectángulo con vértices en r^*, B, D y E) más el exceso de gravamen.

Por lo tanto, bajo estos supuestos, el impuesto sobre el capital es totalmente soportado por el trabajo. Sin embargo, si el país pudiese gravar los beneficios del capital invertido en el exterior, el capital soportaría íntegramente el impuesto y el trabajo no se vería afectado.[34]

Sin embargo, si el impuesto lo introduce un país grande,

- Se producirá una salida de capital desde ese país, al disminuir el rendimiento neto, hacia el resto del mundo, reduciéndose la productividad marginal del capital en el resto del mundo, de modo que el tipo de interés mundial se reduce.
- En el país grande la salida de capital hará que se reduzca la ratio capital-trabajo y que, por tanto, disminuya la productividad del trabajo y, consecuentemente, también lo haga el salario.
- Por el contrario, en el resto del mundo, la entrada de capital hará que aumente la ratio capital-trabajo y que, por tanto, se eleven la productividad del trabajo y el salario.
- Es decir, el impuesto lo soportan el trabajo en el país grande y el capital en el país grande y en el resto del mundo, mientras que el trabajo en el resto del mundo se ve beneficiado.

20.4. Conclusiones ⏳ 30"

En definitiva, los impuestos distorsionantes generan un coste de eficiencia, el exceso de gravamen, que depende de la elasticidad de la demanda compensada, del gasto en el bien y del cuadrado del tipo impositivo, y que en un sistema con varios impuestos no es la suma de los de cada uno por separado.

Por otro lado, la incidencia legal no coincide, en general, con la económica. En equilibrio parcial, la carga se reparte según las elasticidades relativas de la oferta y la demanda.

[34] Es la distinción entre un impuesto de fuente, sobre el capital situado en el país, y uno de residencia, sobre el capital de sus residentes dondequiera que esté, equivalente a un impuesto de suma fija sobre la riqueza que soportan los propietarios nacionales (Kotlikoff y Summers, 1987, p. 1067).

En equilibrio general, depende además de las intensidades factoriales y de la movilidad de los factores, y también del momento y del lugar en que se produce el ajuste, a través de la capitalización, las generaciones y las regiones.

Por ello, el diseño de un sistema impositivo debe tener en cuenta a la vez su coste de eficiencia y quién soporta realmente la carga, dos criterios que el Libro Blanco para la Reforma Tributaria incluye entre sus principios (efectos sobre la eficiencia y efectos redistributivos). Y, dado que este análisis no considera el destino de la recaudación, debe complementarse con el estudio del gasto público.

Finalmente, en la práctica los efectos de los impuestos sobre las rentas de los hogares se estiman con modelos de microsimulación fiscal, como EUROMOD, el modelo de impuestos y prestaciones de la Unión Europea (Sutherland y Figari, 2013), que simulan con microdatos las obligaciones fiscales y los derechos a prestaciones de cada hogar y complementan así los modelos teóricos de equilibrio parcial y general.

 © 2026 Manuel Cabado García

Apéndice

20.A.1. La incidencia de los impuestos

Este apartado amplía el análisis de la incidencia del cuerpo del tema. Como allí se indicó, quien está obligado por ley a pagar un impuesto puede no ser quien soporta finalmente su carga, y lo que interesa es conocer cómo los impuestos alteran la distribución de la renta.

Conceptos básicos de incidencia impositiva Siguiendo a Pantaleoni (1882), la traslación de un tributo se describe con los conceptos siguientes (Braña Pino, 2010):

- Percusión: el tributo recae sobre el sujeto pasivo, obligado por ley a liquidarlo.
- Traslación: el contribuyente obligado al pago transfiere el gravamen a otras personas, hacia delante, elevando el precio de lo que vende, o hacia atrás, rebajando lo que paga por lo que compra, como los salarios o los insumos. Puede ser total o parcial.
- Incidencia: impacto del tributo sobre quien lo soporta en última instancia, sin poder transferirlo. Puede manifestarse como difusión, capitalización o remoción.
- Difusión: reducción voluntaria del consumo para evitar parte de la carga del tributo.
- Capitalización (o amortización): disminución del valor de capital de las rentas gravadas.
- Remoción (o transformación): aumento del esfuerzo productivo para paliar el impacto del tributo.

Teoría moderna de la incidencia Seligman (1921, pp. 1–218) revisó las aportaciones de Pantaleoni y dio lugar a la teoría moderna de la incidencia, que Musgrave sintetizó en tres tipos (Musgrave, 1959; Braña Pino, 2010):

1. Incidencia absoluta: cambios en la distribución de la renta que produce la introducción o el aumento de un tributo, manteniendo constantes el gasto público y los demás tributos.[35]
2. Incidencia diferencial: cambios en la distribución de la renta que causa sustituir un tributo por otro de igual recaudación en términos reales, manteniendo constante el gasto público.
3. Incidencia de presupuesto equilibrado: efectos distributivos de un aumento de la tributación acompañado de un aumento del gasto, de modo que el presupuesto permanece equilibrado.

[35] Musgrave (1959, p. 211, n. 7) sustituyó la denominación "incidencia absoluta" por "incidencia específica", por considerarla engañosa.

Tipos de incidencia según el proceso de ajuste La traslación supone una cadena de acontecimientos, desde el impacto hasta el último destinatario del tributo. Con la notación de Braña (2010), se resume en cuatro situaciones de la distribución de la renta:

1. No existe el tributo (D_1).
2. Lo soportan íntegramente quienes están legalmente obligados a pagarlo (D_2).
3. Lo soportan las personas a las que la ley pretende gravar, estén o no obligadas a pagarlo directamente (D_3).
4. Situación real al final del proceso de ajuste, en la que el tributo lo soporta un número determinado de contribuyentes (D_4).

A partir de ellas se definen tres tipos de incidencia:

1. Incidencia de impacto ($D_2 - D_1$): resultado directo del tributo sobre el contribuyente legal. Es la que en el cuerpo del tema se denominó incidencia legal.
2. Incidencia pretendida ($D_3 - D_1$): la que la ley buscaba lograr. Musgrave la denominó incidencia de intención legislativa.
3. Incidencia efectiva ($D_4 - D_1$): resultado final del proceso de ajuste.

Musgrave distinguió dos contenidos del término traslación (Musgrave, 1959, pp. 230–231):

1. Traslación en sentido estricto: diferencia entre la incidencia efectiva y la de impacto, es decir, entre la situación final y la del contribuyente legal:

$$(D_4 - D_1) - (D_2 - D_1) = D_4 - D_2$$

2. Índice de frustración (Braña la denomina "incidencia fiscal"): diferencia entre la incidencia efectiva y la pretendida, es decir, el grado de fracaso de la política fiscal para lograr sus objetivos distributivos:

$$(D_4 - D_1) - (D_3 - D_1) = D_4 - D_3$$

Medición de la incidencia impositiva Para medir el efecto distributivo de los tributos se usan índices basados en el de Gini. Siendo G_{at} y G_{dt} los índices de Gini de la renta antes y después de impuestos, y C_T el índice de concentración de las cuotas, calculado ordenando a los contribuyentes de menor a mayor renta antes de impuestos:

- Índice de Musgrave: compara los índices de Gini de la renta después y antes de impuestos (Braña Pino, 2010):

$$I_m = (1 - G_{dt}) \,/\, (1 - G_{at})$$

 © 2026 Manuel Cabado García

- Índice de Reynolds-Smolensky: mide el efecto redistributivo como la reducción absoluta del índice de Gini que produce el impuesto (Albi et al., 2018):

$$\Pi^{RS} = G_{at} - G_{dt}$$

- Índice de Kakwani: mide la progresividad de un impuesto como la diferencia entre el índice de concentración de sus cuotas y el índice de Gini de la renta antes de impuestos, de modo que es positivo si el impuesto es progresivo, negativo si es regresivo y nulo si es proporcional (Albi et al., 2018):

$$\Pi^{K} = C_T - G_{at}$$

Ambos índices se relacionan mediante

$$\Pi^{RS} = \frac{t}{1-t} \cdot \Pi^{K} - D$$

donde t es el tipo medio y D el efecto reordenación, es decir, el cambio de posición de los individuos tras el impuesto (Braña Pino, 2010).[36]

Ejemplo: incidencia en España en 1990 Calonge[37] estimó los efectos redistributivos del sistema fiscal español en 1990 con dos hipótesis de incidencia (Braña Pino, 2010):

- Hipótesis I: el impuesto de sociedades se traslada a partes iguales entre capital, consumo y activos reales. Las cuotas empresariales a la Seguridad Social se trasladan en dos tercios a los trabajadores y un tercio al consumo.
- Hipótesis II, más progresiva: dos tercios del impuesto de sociedades recaen sobre los propietarios del capital y un tercio sobre el consumo. Las cuotas a la Seguridad Social las soportan íntegramente los trabajadores.

Con la hipótesis I, el sistema en su conjunto resulta casi proporcional: el índice de Kakwani del total de impuestos es 0,017 y el de redistribución 0,014, con un índice de Gini de la renta antes de impuestos de 0,385. El IRPF es el impuesto más progresivo (Kakwani de 0,303), mientras que el IVA es regresivo (−0,140) (Braña Pino, 2010).

[36] Algunos autores calculan el índice de Reynolds-Smolensky con el índice de concentración de la renta después de impuestos en lugar de su Gini. En ese caso no recoge la reordenación y se cumple exactamente $\Pi^{RS} = \frac{t}{1-t} \cdot \Pi^{K}$.

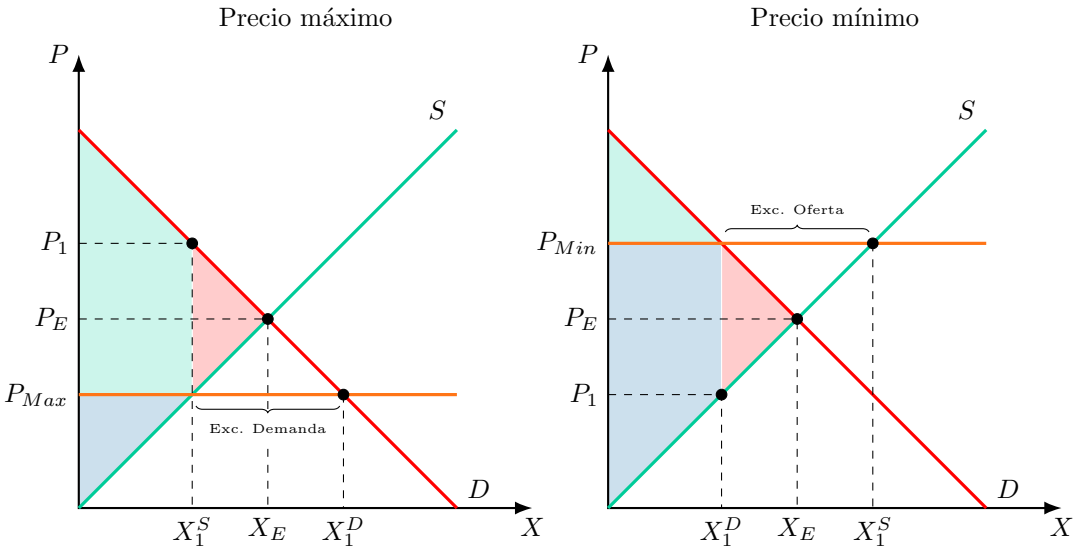
[37] Calonge, S., «Efectos redistributivos de impuestos y prestaciones en España y sus CC. AA.», Universidad de Barcelona (Centre de Recerca en Economia del Benestar), documento de trabajo, sin fecha.

20.A.2. Pérdida de eficiencia provocada por un precio máximo y un precio mínimo

En el cuerpo del tema se estudió el exceso de gravamen de los impuestos. Una pérdida de eficiencia análoga aparece cuando el sector público interviene directamente en el precio fijando un máximo o un mínimo. Los precios máximos y mínimos son instrumentos habituales de intervención, como los topes al precio de la vivienda o los precios mínimos que buscan evitar que los precios bajen más ante una crisis. Ambas medidas afectan a la eficiencia, como se aprecia en la Figura 20.7, donde se supone que el racionamiento es eficiente, es decir, que compran quienes más valoran el bien y venden las empresas de menor coste:

- Precio máximo, fijado por debajo del de equilibrio: la cantidad intercambiada es la que los oferentes venden a ese precio y aparece un exceso de demanda. El excedente del productor es el triángulo del origen, el del consumidor el trapecio superior y la pérdida de eficiencia el triángulo adyacente a ambos.
- Precio mínimo, fijado por encima del de equilibrio: la cantidad intercambiada es la que los consumidores demandan a ese precio y aparece un exceso de oferta. El excedente del productor es el trapecio, el del consumidor el triángulo superior y la pérdida de eficiencia el triángulo adyacente a ambos.

Figura 20.7: Pérdida de eficiencia por precio máximo y mínimo



20.A.3. Impuesto sobre los beneficios del monopolio

En el cuerpo del tema se señaló que un impuesto sobre los beneficios no genera exceso de gravamen en competencia perfecta. En monopolio ocurre lo mismo si el impuesto recae

 © 2026 Manuel Cabado García

sobre el beneficio puro, es decir, sobre los ingresos totales menos los costes totales, incluida una rentabilidad normal del capital (Stiglitz y Rosengard, 2015, p. 720). En ese caso, la cantidad y el precio no se ven afectados, porque el impuesto no cambia la condición de maximización del beneficio.

La condición de maximización del beneficio del monopolista es que el ingreso marginal (IMg) sea igual al coste marginal (CMg), siendo el beneficio (π):

$$\pi = p(x) \cdot x - C(x)$$

- Con un impuesto sobre el beneficio a un tipo constante τ, el beneficio neto del monopolista es:

$$\pi_{neto} = (1-\tau) \cdot [p(x) \cdot x - C(x)] = (1-\tau) \cdot \pi$$

- Para maximizarlo, el monopolista deriva respecto a la cantidad (x) y lo iguala a cero:

$$\frac{d\pi_{neto}}{dx} = (1-\tau) \cdot \frac{d\pi}{dx} = 0$$

- Como $(1-\tau)$ es una constante positiva, la condición se cumple solo si $\frac{d\pi}{dx} = 0$, que es la misma condición de maximización del beneficio antes de impuestos:

$$\frac{d\pi}{dx} = \frac{d[p(x) \cdot x]}{dx} - \frac{dC(x)}{dx} = 0$$

es decir, $IMg = CMg$.

Dado que el impuesto no altera la condición de maximización, la cantidad (x) y el precio (p) que el monopolista elige son los mismos que sin impuesto. El impuesto reduce el beneficio neto, pero no modifica su comportamiento: equivale a un impuesto de suma fija sobre el monopolista, por lo que no genera exceso de gravamen y su incidencia recae íntegramente sobre él (Stiglitz y Rosengard, 2015, p. 720).

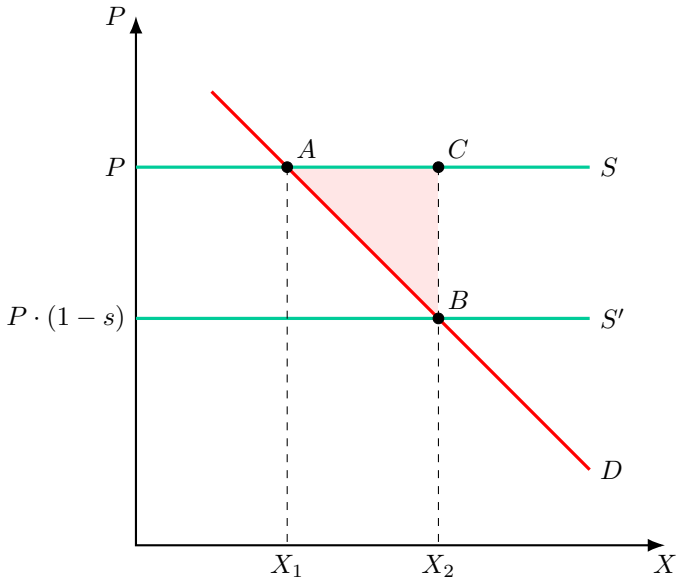
20.A.4. Exceso de gravamen de un subsidio

El cuerpo del tema se ha centrado en el exceso de gravamen de los impuestos, pero los subsidios también generan ineficiencias: funcionan como impuestos de tipo negativo y su exceso de gravamen se mide con la misma expresión (Albi et al., 2018).

Se supone, de nuevo, un mercado con una curva de demanda compensada con pendiente negativa y con una curva de oferta totalmente elástica.

Al introducir un subsidio, la curva de oferta se desplaza hacia abajo en la cuantía del subsidio, por lo que aumenta la cantidad de equilibrio.

Figura 20.8: Exceso de gravamen de un subsidio



Al ser la oferta totalmente elástica, la incidencia del subsidio se concentra íntegramente en los consumidores (de la misma forma que un impuesto también se concentraría íntegramente en ellos), por lo que su excedente aumenta en el trapecio de vértices P, A, B y $P\,(1-s)$, y el excedente de los productores no varía.

Sin embargo, el aumento del excedente de los consumidores es menor que el desembolso que hace el Gobierno por la subvención (el rectángulo de vértices P, C, B y $P\,(1-s)$), pues parte de dicho desembolso se pierde en forma de ineficiencia (el triángulo ABC).

Dicha ineficiencia se debe a que existe un tramo de producción $(X_2 - X_1)$ para el que la valoración privada del bien, dada por la demanda, es menor que su coste de producción, dado por la oferta. Es un equilibrio ineficiente en el sentido de Pareto: la producción podría reducirse en esa cantidad y todos los agentes saldrían beneficiados, con el consiguiente aumento del bienestar social.

20.A.5. Traslación del impuesto según la forma de la demanda

El cuerpo del tema indica que el reparto de la carga depende de las elasticidades en competencia perfecta y de la forma de la demanda y del coste marginal en monopolio. Se derivan aquí esos resultados.

Competencia perfecta Con un impuesto unitario t y p el precio que recibe el productor, el equilibrio cumple $X_D\,(p+t) = X_S\,(p)$ (Albi et al., 2018, cap. 3). Derivando respecto a t:

$$X'_D\left(\frac{dp}{dt}+1\right) = X'_S\,\frac{dp}{dt}$$

 © 2026 Manuel Cabado García

$$\frac{dp}{dt} = \frac{X'_D}{X'_S - X'_D} = \frac{\varepsilon_D}{\varepsilon_S - \varepsilon_D}$$

donde la segunda igualdad resulta de multiplicar numerador y denominador por p/X, y $\varepsilon_D < 0$ y ε_S son las elasticidades de demanda y de oferta. El precio que paga el consumidor, $p+t$, varía en $\frac{\varepsilon_S}{\varepsilon_S - \varepsilon_D}$. Por tanto, el consumidor soporta la proporción $\frac{\varepsilon_S}{\varepsilon_S + |\varepsilon_D|}$ del impuesto y el productor la proporción $\frac{|\varepsilon_D|}{\varepsilon_S + |\varepsilon_D|}$: la carga recae más sobre el lado menos elástico. Si la demanda es inelástica o la oferta infinitamente elástica, el consumidor soporta todo el impuesto. Si la oferta es inelástica o la demanda infinitamente elástica, lo soporta el productor.

Monopolio Con demanda inversa $p(x)$, coste $C(x)$ e impuesto unitario t, el monopolista maximiza $\pi = p(x)\,x - C(x) - tx$. La condición de primer orden es $p(x) + xp'(x) = C'(x) + t$, es decir, $IMg = CMg + t$. Derivando respecto a t:

$$[2p' + xp'' - C''] \frac{dx}{dt} = 1$$

y como $dp/dt = p'\,dx/dt$, el grado de traslación al precio es

$$\rho = \frac{dp}{dt} = \frac{p'}{2p' + xp'' - C''}$$

Este mismo grado de traslación lo expresan en función de elasticidades Weyl y Fabinger (2013, pp. 540 y 548). Con la expresión anterior se resuelven tres casos:

- Demanda lineal y coste marginal constante: si $p = a - bx$ y $C'' = 0$, entonces $p' = -b$ y $p'' = 0$, por lo que $\rho = \frac{1}{2}$: el precio sube la mitad del impuesto.
- Demanda de elasticidad constante y coste marginal constante: con $\eta > 1$ el valor absoluto de la elasticidad, $p' = -\frac{p}{\eta x}$ y $xp'' = \left(1 + \frac{1}{\eta}\right)\frac{p}{\eta x}$, de modo que $2p' + xp'' = -\frac{p}{\eta x}\left(1 - \frac{1}{\eta}\right)$. Entonces $\rho = \frac{\eta}{\eta - 1}$, mayor que 1: hay sobretraslación (con $\eta = 2$, el precio sube el doble del impuesto) (Stiglitz y Rosengard, 2015, pp. 554-555).
- Demanda de elasticidad constante y coste marginal creciente: con $C'' > 0$,

$$\rho = \frac{1}{\left(1 - \frac{1}{\eta}\right) + \frac{C''\eta x}{p}}$$

  La traslación disminuye al aumentar la pendiente del coste marginal y tiende a 0 cuando es vertical, caso en que el productor soporta todo el impuesto.

Ejemplo numérico (elaboración propia) Con $\eta = 2$, demanda $x = 100p^{-2}$ y coste marginal $1 + C''x$, la traslación vale 2,00 si $C'' = 0$, 1,53 si $C'' = 0{,}01$, 1,18 si $C'' = 0{,}05$, 0,88 si $C'' = 0{,}5$ y 0,76 si $C'' = 5$. Pasa así de sobretraslación a traslación parcial.

20.A.6. Coste marginal de los fondos públicos y evidencia para España

El exceso de gravamen medio del cuerpo del tema mide el coste de eficiencia del sistema en su conjunto. Para decidir si recaudar un euro más compensa, interesa en cambio el coste marginal.

Definiciones

- Exceso de gravamen medio: coste de bienestar total del sistema, por ejemplo en porcentaje del PIB.
- Exceso de gravamen marginal: aumento del coste de bienestar al recaudar un euro adicional.
- Coste marginal de los fondos públicos: el euro adicional recaudado más el coste de bienestar que genera, es decir, lo que cuesta realmente, en términos de bienestar, obtener un euro adicional de recaudación (Díaz Caro y Onrubia, 2015).

Evidencia para España

- Alonso-Carrera y Manzano (2003) estiman un exceso de gravamen medio del conjunto del sistema de entre el 7,8 % y el 9,6 % del PIB en estado estacionario.
- Sancho (2004) estima, con un modelo computable de equilibrio general, un coste marginal en bienestar del sistema de unos 0,5 euros por euro adicional recaudado.
- Díaz Caro y Onrubia (2015) estiman, con una elasticidad compensada media de la renta gravable de 0,415 y una elasticidad renta de $-0{,}20$, que el IRPF dual de 2007 elevó el coste marginal de bienestar del primer tramo de 0,08 a 0,33 euros por euro recaudado respecto del IRPF de 2006. En el resto de tramos el aumento fue mucho menor.

Los dos primeros resultados miden magnitudes distintas, media y marginal, y el tercero se refiere solo a un impuesto.

20.A.7. Exceso de gravamen con oferta de pendiente positiva

En el cuerpo del tema, el exceso de gravamen se mide con costes marginales constantes, es decir, con una oferta perfectamente elástica en la que el precio que cobra el productor no cambia. Aquí se relaja ese supuesto. Se mantiene la demanda compensada, sin efecto renta, y se parte de un mercado sin impuestos previos, con precio p y cantidad Q. Sean $\varepsilon_D < 0$ la elasticidad-precio de la demanda compensada, $\varepsilon_S > 0$ la de la oferta y τ el tipo ad valorem del impuesto (Salanié, 2003, cap. 2).

 © 2026 Manuel Cabado García

El impuesto abre una cuña entre el precio que paga el consumidor, p_c, y el que recibe el productor, p_p, de modo que $p_c - p_p = \tau p$. La cantidad intercambiada cae lo que indican ambas curvas:

$$\frac{\Delta Q}{Q} = \varepsilon_D \frac{\Delta p_c}{p} = \varepsilon_S \frac{\Delta p_p}{p}$$

Como $\Delta p_c/p - \Delta p_p/p = \tau$, resulta:

$$\frac{\Delta Q}{Q} = \frac{\varepsilon_D \varepsilon_S}{\varepsilon_S - \varepsilon_D}\,\tau$$

El exceso de gravamen es el triángulo cuya base es la cuña, τp, y cuya altura es la cantidad perdida, $|\Delta Q|$. Su área es:

$$EG = \frac{1}{2}\,\tau p\,|\Delta Q| = -\frac{1}{2}\,\frac{\varepsilon_D \varepsilon_S}{\varepsilon_S - \varepsilon_D}\,pQ\,\tau^2$$

La fracción de elasticidades es menor en valor absoluto que cada una de ellas, y el exceso de gravamen crece con la elasticidad de la demanda y con la de la oferta. Si la oferta es perfectamente elástica ($\varepsilon_S \to \infty$), la fracción tiende a ε_D y se recupera la fórmula del cuerpo, $EG = -\frac{1}{2}\varepsilon_D\,pQ\,\tau^2$. Si la oferta es perfectamente inelástica ($\varepsilon_S \to 0$), el exceso de gravamen desaparece, porque el impuesto no altera la cantidad.

Ejemplo ilustrativo Con un IVA del 20 % y una demanda de elasticidad unitaria, $\varepsilon_D = -1$, si la oferta también es unitaria el exceso de gravamen es de $0{,}01\,pQ$, alrededor del 5 % de la recaudación. Con oferta perfectamente elástica sube a $0{,}02\,pQ$, alrededor del 10 % de la recaudación (Salanié, 2003, cap. 2). El cociente entre el exceso de gravamen y la recaudación es proporcional al tipo, por lo que los tipos altos resultan comparativamente más costosos.

20.A.8. Incidencia en el modelo de Harberger: derivación algebraica

El cuerpo presenta el modelo de Harberger con gráficos y de forma cualitativa. Aquí se deriva el resultado con los supuestos de Albi et al. (2018, cap. 3) y el procedimiento de Salanié (2003, cap. 1): dos bienes producidos en competencia perfecta, rendimientos constantes a escala, factores con oferta fija y perfectamente móviles entre sectores, preferencias idénticas y homotéticas y recaudación devuelta como transferencia de suma fija. Se trabaja con variaciones relativas, $\hat{z} = dz/z$, y el subíndice X designa el sector gravado y Y el otro.

Notación

- λ_{KX} es la cuota del sector X en el uso total de capital, y $\lambda_{KY} = 1 - \lambda_{KX}$. Del mismo modo, λ_{LX} y λ_{LY} para el trabajo.
- s_{KX} es la participación del capital en la renta del sector X, y $s_{LX} = 1 - s_{KX}$. Del mismo modo, s_{KY} y s_{LY}.
- σ_X y σ_Y son las elasticidades de sustitución entre capital y trabajo en cada sector, y $\sigma_D > 0$ la elasticidad de sustitución en la demanda entre X e Y.
- $d\tau_{KX}$ es el impuesto sobre el capital empleado en X, y $d\tau_X$ el impuesto sobre el consumo de X. La variable que se resuelve es $\omega = \hat{w} - \hat{r}$, la variación del salario relativa a la del rendimiento del capital.

Las cinco condiciones del modelo

1. Con factores totalmente empleados, $\lambda_{KX}\hat{K}_X + \lambda_{KY}\hat{K}_Y = 0$ y $\lambda_{LX}\hat{L}_X + \lambda_{LY}\hat{L}_Y = 0$.
2. El impuesto encarece el capital en X, por lo que $\hat{K}_X - \hat{L}_X = \sigma_X\,(\omega - d\tau_{KX})$. En el sector Y, $\hat{K}_Y - \hat{L}_Y = \sigma_Y\,\omega$.
3. La producción de cada sector es la media ponderada de sus factores: $\hat{X} = s_{KX}\hat{K}_X + s_{LX}\hat{L}_X$, y $\hat{Y}$ del mismo modo.
4. La demanda relativa responde a los precios relativos, con el impuesto al consumo incluido: $\hat{X} - \hat{Y} = -\sigma_D\,(\hat{p}_X - \hat{p}_Y + d\tau_X)$.
5. Los precios de productor igualan los costes unitarios: $\hat{p}_X - \hat{p}_Y = (s_{LX} - s_{LY})\,\omega + s_{KX}\,d\tau_{KX}$.

Resolviendo el sistema se obtiene:

$$\Delta\,\omega = (\sigma_X a_X - \sigma_D \lambda^* s_{KX})\,d\tau_{KX} - \sigma_D \lambda^*\,d\tau_X$$

con $a_X = s_{KX}\lambda_{LX} + s_{LX}\lambda_{KX}$, $a_Y = s_{KY}\lambda_{LY} + s_{LY}\lambda_{KY}$, $\lambda^* = \lambda_{LX}\lambda_{KY} - \lambda_{LY}\lambda_{KX}$ y $\Delta = \sigma_X a_X + \sigma_Y a_Y + \sigma_D \lambda^*\,(s_{LX} - s_{LY})$. El denominador Δ es siempre positivo. El signo de λ^* depende de las intensidades factoriales, ya que $\lambda^* < 0$ si y solo si X es más intensivo en capital que Y.

La fórmula separa dos efectos. El término $\sigma_X a_X$ es el efecto sustitución factorial: el impuesto sobre el capital en X induce a emplear más trabajo, lo que reduce la renta relativa del capital. Los términos con $\sigma_D\lambda^*$ son el efecto producción: el impuesto encarece X, cae su demanda y se resiente el factor en el que X es más intensivo.

- Si X es intensivo en capital ($\lambda^* < 0$), ambos efectos van en la misma dirección. Entonces $\omega > 0$ y el capital soporta una carga superior a su participación en la renta.
- Si X es intensivo en trabajo ($\lambda^* > 0$), el efecto producción contrarresta al efecto

 © 2026 Manuel Cabado García

sustitución. Con σ_X pequeño, ω puede ser negativo, de modo que un impuesto sobre el capital en X lo soporte relativamente más el trabajo.

- Con solo un impuesto sobre el consumo de X ($d\tau_{KX} = 0$), $\omega = -\sigma_D \lambda^* d\tau_X / \Delta$. El capital pierde renta relativa si y solo si X es intensivo en capital.

El caso Cobb-Douglas Con $\sigma_X = \sigma_Y = \sigma_D = 1$, se cumple $\Delta = 1$ y $\omega = \lambda_{KX}\, d\tau_{KX}$. Tomando el salario como numerario, $\hat{w} = 0$, resulta $d(rK) = -rK_X\, d\tau_{KX}$, es decir, la pérdida de renta del capital es igual a la recaudación y los propietarios del capital soportan todo el impuesto. Esta es la conclusión de Harberger para Estados Unidos, que los modelos de equilibrio general computable posteriores no confirman con tanta rotundidad (Salanié, 2003, cap. 1).

20.A.9. Por qué no se utilizan impuestos de suma fija

Un impuesto de suma fija es aquel que no depende de ninguna decisión del contribuyente, ni de consumo ni de oferta de factores. Es neutral, no genera efecto sustitución y su exceso de gravamen es nulo. Además, todo impuesto neutral es equivalente a uno de suma fija, por lo que el exceso de gravamen se mide siempre respecto a él (Albi et al., 2018, cap. 1, epígrafe 3.2, y cap. 2). Si es tan eficiente, cabe preguntarse por qué no se usa.

Razones de su escasa utilización

- Equidad: un impuesto de suma fija suele ser inequitativo porque no atiende a la capacidad de pago. Los impuestos sobre la edad, la altura o el color de ojos son de suma fija, y el caso extremo es el de capitación, en el que todos pagan lo mismo. El único ejemplo reciente es el Poll Tax que el gobierno de Thatcher introdujo en 1990 para financiar los municipios, con una cuantía igual por persona dentro de cada municipio. Las protestas obligaron a retirarlo (Albi et al., 2018, cap. 1, epígrafe 3.2) y contribuyeron a la caída de Thatcher (Salanié, 2003, pp. 9-10).
- Información: para ser equitativo, el impuesto debería depender de la capacidad de pago, que el Sector Público no observa. Debe apoyarse en indicadores imperfectos, como los salarios o los patrones de consumo (Albi et al., 2018, cap. 2, epígrafe 3.2), y un impuesto que depende de ellos deja de ser de suma fija. Incluso el Poll Tax no lo era estrictamente, porque los contribuyentes podían mudarse a otro municipio (Salanié, 2003, pp. 9-10).
- Evasión: un impuesto de suma fija deja de ser neutral si existen posibilidades de evasión (Albi et al., 2018, cap. 2, epígrafe 3.1).

Como consecuencia, el Sector Público debe financiarse con impuestos distorsionantes. El problema se traslada entonces a la elección del sistema que minimiza el coste de eficiencia

para una recaudación dada, a partir de la información disponible, que es el problema del impuesto óptimo.

 © 2026 Manuel Cabado García

Bibliografía

Albi, E., J. M. González-Páramo, R. M. Urbanos y I. Zubiri (2018). *Economía Pública II: Teoría de la imposición. Sistema impositivo. Otros ingresos públicos. Economía Pública Internacional*. Barcelona: Ariel.

Alonso-Carrera, J. y B. Manzano (2003). Análisis dinámico del coste de bienestar del sistema impositivo español. Una exploración cuantitativa. *Hacienda Pública Española / Revista de Economía Pública 167*(4), 9–31.

Atkinson, A. B. y J. E. Stiglitz (2015). *Lectures on Public Economics* (Second ed.). Princeton University Press.

Auerbach, A. J. (1985). The Theory of Excess Burden and Optimal Taxation. En A. J. Auerbach y M. Feldstein (eds.), *Handbook of Public Economics*, vol. 1, cap. 2, pp. 61–127. North-Holland.

Braña Pino, F. J. (2010). Tema 13. La incidencia de los impuestos. Hacienda Pública I, Grupo III. Material docente, Facultad de Derecho, Universidad de Salamanca. Curso 2010-2011.

Buchanan, J. M. (1958). *Public Principles of Public Debt: A Defense and Restatement*. Homewood, IL: Richard D. Irwin.

Dalton, H. (1936). *Principles of Public Finance* (Ninth ed.). London: Routledge and Kegan Paul.

Díaz Caro, C. y J. Onrubia (2015). Elasticidades de la renta gravable y costes de eficiencia en el IRPF dual. Estudios sobre la Economía Española 2015/02, FEDEA.

Feldstein, M. S. (1974). Incidence of a Capital Income Tax in a Growing Economy with Variable Savings Rates. *The Review of Economic Studies 41*(4), 505–513.

Harberger, A. C. (1962). The Incidence of the Corporation Income Tax. *Journal of Political Economy 70*(3), 215–240.

Harberger, A. C. (1964). Taxation, Resource Allocation, and Welfare. En *The Role of Direct and Indirect Taxes in the Federal Revenue System*, pp. 25–80. Princeton University Press.

Kotlikoff, L. J. y L. H. Summers (1987). Tax Incidence. En A. J. Auerbach y M. Feldstein (eds.), *Handbook of Public Economics*, vol. 2, cap. 16, pp. 1043–1092. North-Holland.

McLure Jr., C. E. (1969). The Inter-regional Incidence of General Regional Taxes. *Public Finance 24*(3), 457–485.

Musgrave, R. A. (1959). *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill.

Pantaleoni, M. (1882). *Teoria della traslazione dei tributi: definizione, dinamica e ubiquità della traslazione*. Roma: A. Paolini.

Salanié, B. (2003). *The Economics of Taxation*. Cambridge, MA: MIT Press.

Sancho, F. (2004). Una estimación del coste marginal en bienestar del sistema impositivo en España. *Hacienda Pública Española / Revista de Economía Pública 169*(2), 117–132.

Seligman, E. R. A. (1921). *The Shifting and Incidence of Taxation* (Fourth ed.). New York: Columbia University Press.

Stiglitz, J. E. y J. K. Rosengard (2015). *Economics of the Public Sector* (Fourth ed.). WW Norton & Company.

Sutherland, H. y F. Figari (2013). EUROMOD: the European Union tax-benefit microsimulation model. *International Journal of Microsimulation 6*(1), 4–26.

Weyl, E. G. y M. Fabinger (2013). Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition. *Journal of Political Economy 121*(3), 528–583.

 © 2026 Manuel Cabado García

Tema 21

Imposición y oferta. Efectos incentivos de los impuestos. Medidas de progresividad impositiva. El compromiso entre eficiencia y equidad de la política tributaria.

21.1. Introducción ⏳ 1'

La Economía del Sector Público estudia cómo el Estado interviene en el mercado mediante ingresos y gastos para cumplir funciones de asignación, distribución y estabilidad. Los impuestos, como transferencias obligatorias del sector privado, son su principal herramienta.

Sin embargo, su impacto va más allá de la recaudación: afectan directamente a la oferta agregada, la inversión y el comportamiento de los agentes económicos. Por ejemplo, el llamado "impuesto al sol" en España desincentivó el autoconsumo energético.

Asimismo, el diseño progresivo del sistema fiscal influye en la equidad, al determinar cómo se distribuye la carga tributaria entre distintos niveles de renta. Este análisis es esencial para entender el compromiso entre eficiencia y equidad: un sistema impositivo eficiente minimiza distorsiones económicas, mientras que uno equitativo reparte la carga de forma justa. Encontrar el equilibrio entre ambos objetivos es el reto central de toda política fiscal moderna.

Por consiguiente, el objetivo de la presente exposición es analizar cómo el sistema impositivo cumple con las funciones del sector público y en qué medida es eficiente y equitativo a través del siguiente esquema:

1. Principios impositivos.
2. Eficiencia económica.

3. Equidad y medidas de progresividad.
4. Compromiso entre eficiencia y equidad.

21.2. Principios impositivos ⏳ 1'

Comenzando con los principios impositivos, bajo la premisa de que un impuesto es un instrumento que se utiliza para la maximización del bienestar de la sociedad, existe un acuerdo sobre las condiciones que debe reunir un buen sistema tributario. Las más importantes son:

1. Eficiencia económica, esto es, hasta qué punto interfiere o no en el proceso de libre elección de los contribuyentes y les fuerza –o no– a tomar decisiones inadecuadas desde el punto de vista de la asignación de recursos.
2. Justicia fiscal o equidad, los impuestos han de ser, idealmente, equitativos. La distribución de la carga se basa en diferentes principios. De entre ellos, los que han tenido una mayor importancia en la literatura hacendística son el beneficio, la capacidad de pago, la utilidad y el consumo.
3. Sencillez administrativa, implica que el sistema fiscal debe ser lo más sencillo posible, para lo cual se aconseja que haya pocos impuestos con base muy amplia, reglas sencillas de valoración, y pocos casos particulares.
4. Flexibilidad, implica que los impuestos deben adaptarse al contexto económico, facilitando así su tarea de estabilizadores automáticos de la economía.
5. Suficiencia, implica que la recaudación debe ser apropiada para cubrir los gastos planeados del sector público.

No obstante, a menudo estos principios entran en conflicto entre sí, por lo que el diseño de un sistema fiscal es un ejercicio de compromiso entre principios que pueden llegar a ser contradictorios.

21.3. Eficiencia económica ⏳ 10'30"

Centrando el análisis en la eficiencia económica, los impuestos afectan a la conducta económica de los agentes sobre los que recaen. Atendiendo a los efectos en términos de eficiencia, se distingue entre:

- Impuestos distorsionantes, aquellos cuya cuantía puede reducir el contribuyente cambiando su conducta, de modo que alteran los precios relativos a los que se enfrenta y, con ellos, sus decisiones.

 © 2026 Manuel Cabado García

- Impuestos no distorsionantes, aquellos cuya cuantía no depende de ninguna decisión del contribuyente, de modo que no alteran los precios relativos y solo generan un efecto renta. Este es el caso de los impuestos de suma fija o tanto alzado, esto es, una detracción de renta de cuantía fija a los contribuyentes.[1]

Desde el punto de vista de la economía del bienestar, los impuestos distorsionantes son ineficientes ya que al afectar a los precios relativos afectarán a las decisiones de consumo y producción. El resultado es una alteración de las condiciones de eficiencia y una generación de exceso de gravamen.

Sin embargo, normalmente se utilizan los impuestos distorsionantes porque los impuestos de suma fija no tienen en cuenta aspectos de equidad, lo que podría ser socialmente inaceptable[2]. Por ello, el objetivo será minimizar el impacto en la eficiencia de los impuestos distorsionantes a la hora de cumplir las funciones del sector público.

Reglas de imposición eficiente

Cuando la eficiencia es el único criterio de valoración, se define como sistema impositivo óptimo aquel que logra la recaudación requerida minimizando el exceso de gravamen total, entendiendo este como la pérdida de bienestar que causan los impuestos distorsionantes por encima de lo que recaudan.

Matemáticamente es un problema de minimización del exceso de gravamen, EG, definido como, $-1/2$[3] por el sumatorio de n bienes de consumo, multiplicado por el cambio en la demanda compensada de i al variar el precio de j o efecto sustitución entre los bienes i y j, S_{ij}, por los precios del bien i y del bien j y por los tipos impositivos de dichos bienes:

[1] La intuición es que un impuesto de suma fija también cambia la conducta, porque empobrece al contribuyente, pero no le ofrece ninguna forma de pagar menos. Con un impuesto distorsionante, en cambio, trabajar, consumir o ahorrar menos de lo gravado reduce la factura fiscal.

[2] Según Albi et al. (2018), el único ejemplo en la historia reciente de un impuesto de tanto alzado es, precisamente, el de un impuesto de capitación –impuesto de tanto alzado en el que todos los contribuyentes pagan la misma cantidad– introducido por el gobierno conservador de Margaret Thatcher en el Reino Unido en 1990 como instrumento de financiación de los municipios: el denominado Poll Tax.

Este impuesto básicamente consistía en que cada municipio dividiera su presupuesto de gasto entre todos los contribuyentes y cobrara a cada uno de ellos un impuesto igual a la cantidad resultante. Sin embargo, este impuesto se encontró con los problemas de equidad que, debido a los problemas de información, tienen en la práctica los impuestos de tanto alzado. Como el Poll Tax hacía pagar a todos –ricos o pobres– lo mismo, iba en contra de las normas de equidad dominantes y suscitó numerosas protestas que, en última instancia, forzaron a que fuera revocado. La inconsistencia con la equidad, interpretada como que el que tenga más pague impuestos más altos, es, probablemente, la causa del escaso uso en la práctica de los impuestos de tanto alzado.

[3] El signo negativo se debe a que el efecto sustitución propio, S_{ii}, es negativo, de modo que el sumatorio es negativo y el exceso de gravamen resulta positivo. El factor $1/2$ procede de aproximar linealmente la pérdida de bienestar por el área de un triángulo, lo que equivale a suponer demandas compensadas lineales (Albi et al., 2018).

$$\text{mín} \quad EG = -\frac{1}{2} \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} \cdot p_i \cdot p_j \cdot t_i \cdot t_j \tag{21.1}$$

sujeto a la recaudación requerida, $\bar{T}$

$$\bar{T} = \sum_{i=1}^{n} t_i \cdot p_i \cdot X_i \tag{21.2}$$

La resolución de este problema permite establecer una serie de reglas de imposición eficiente:

1. Regla de elección excluyente, cuando se pueda elegir entre dos impuestos, el mejor será aquel que grave el bien con la demanda compensada más inelástica. Esto se debe a que el exceso de gravamen producido por la introducción de un impuesto distorsionante será menor cuanto más inelástica sea la demanda.
2. Regla de igualación de distorsiones marginales –equimarginalidad–: si no se puede excluir entre bienes, es deseable gravarlos todos, con tipos que igualen el exceso de gravamen marginal por unidad monetaria recaudada en todos los impuestos.[4] La excepción se dará con un bien con una demanda totalmente inelástica y que proporcione una base suficientemente amplia.
3. Regla de la elasticidad inversa –planteada por Pigou y formalizada por Ramsey (1927)[5]–: si los efectos renta y los efectos sustitución cruzados son nulos, es decir, si los bienes son independientes, los tipos impositivos que minimizan el exceso de gravamen son inversamente proporcionales a las respectivas elasticidades-precio de las demandas, que en ese caso coinciden con las de las demandas compensadas. Por lo tanto, se deben gravar los bienes cuyas demandas sean más inelásticas[6].
4. Regla de Corlett-Hague (Corlett y Hague, 1953): como el ocio no puede gravarse, deben gravarse más los bienes complementarios del ocio y menos los sustitutivos del

[4] La intuición es que, como el exceso de gravamen crece con el cuadrado del tipo, si un impuesto distorsiona más que otro por el último euro recaudado, conviene recaudar un euro menos con él y uno más con el otro (Albi et al., 2018).

[5] La pregunta de cuál es el patrón de impuestos sobre los bienes que menos distorsiona, sin poder recurrir a impuestos de suma fija, se la planteó a Ramsey su profesor, A. C. Pigou, y Ramsey la resolvió en 1927 (Stiglitz y Rosengard, 2015, pp. 621-622). El año 1947 que figura en algunas referencias corresponde a la 3.ª edición de *A Study in Public Finance*, de Pigou, y no a la fecha de la pregunta. Por eso la regla se atribuye hoy a Ramsey, aunque algunos textos la llaman regla de Pigou (Albi et al., 2018).

[6] Diferencia con la regla de elección excluyente. Aunque en un primer momento esta regla parece ser la misma que la de la elección excluyente, esto no es así:

- La regla de elección excluyente solo tiene en cuenta la elasticidad y aconseja, por tanto, gravar un único bien
- La regla de la elasticidad inversa considera que, además de la elasticidad, el tipo impositivo también influye en el exceso de gravamen, y de forma más que proporcional, ya que este crece con el cuadrado del tipo. Por tanto, la regla de la elasticidad inversa aconseja gravar muchos bienes cuya demanda sea inelástica, de modo que el tipo sea reducido.

 © 2026 Manuel Cabado García

ocio, para reducir la distorsión sobre la oferta de trabajo.[7]

5. Regla de Ramsey-Diamond-Mirrlees (Ramsey, 1927; Diamond y Mirrlees, 1971), también llamada regla de Ramsey o de la reducción equiproporcional, estos autores generalizan la regla de la elasticidad inversa para que se pueda aplicar a mercados generales en los que existan relaciones de complementariedad y sustituibilidad. Llegan a la conclusión de que un sistema impositivo es óptimo si, tras su aplicación, disminuyen todas las demandas compensadas de forma equiproporcionada[8]. La regla de la elasticidad inversa es el caso particular en el que los bienes son independientes.

Efectos desincentivo de los impuestos sobre la oferta

Una vez analizadas las reglas de imposición eficiente, se van a analizar los efectos desincentivo de los impuestos sobre la oferta.

Oferta de trabajo

Comenzando con la oferta de trabajo, esta se va a obtener a través de un problema maximizador de la utilidad, u, que le reporta el nivel de consumo, C, y el nivel de ocio, O:

$$\text{máx} \quad u\left(C, O\right) \tag{21.3}$$

sujeto a

$$C = w \cdot L + I - T \tag{21.4}$$

donde I representa las rentas no salariales, T el impuesto de suma fija y $L = \bar{L} - O$ las horas trabajadas, siendo $\bar{L}$ el tiempo total disponible.

Gráficamente, en un plano de consumo, C, y ocio, O, como el de la Figura 21.1a, se representa la restricción presupuestaria[9] y la curva de indiferencia tangente, dando lugar a un equilibrio como E_0.

Impuesto de suma fija El establecimiento de un impuesto de suma fija va a producir un efecto renta puro, desplazando la restricción presupuestaria paralelamente hacia el interior en la cuantía T, sin variar su pendiente, $-w$, ni el tiempo total disponible, $\bar{L}$. Se llega así a un equilibrio como E_1 en el que, si el ocio es un bien normal, disminuye el nivel de ocio y aumentan las horas trabajadas, por lo que el impuesto incentiva el trabajo.

[7] La intuición es que gravar lo que se consume junto con el tiempo libre, como los servicios turísticos, es una forma indirecta de gravar el ocio, que el Estado no puede gravar directamente (Albi et al., 2018).

[8] La intuición es que esta regla indica que las distorsiones creadas por los impuestos no deberían modificar sustancialmente las proporciones en las que se consumen los distintos bienes.

[9] Nótese que en el extremo derecho de la recta, donde el ocio es máximo, el agente no trabaja y consume solo sus rentas no salariales: $C = I$ antes del impuesto y $C = I - T$ tras el impuesto de suma fija.

Impuesto proporcional Por otro lado, si se establece un impuesto proporcional, τ_L, sobre las rentas salariales como el de la Figura 21.1b, la restricción (21.4) cambiaría:

$$C = (1 - \tau_L) \cdot w \cdot L + I \tag{21.5}$$

Su pendiente pasa de $-w$ a $-(1 - \tau_L) \cdot w$, y la restricción pivota en torno al punto de no trabajar, $\left(\bar{L}, I\right)$, generando dos efectos:

1. Efecto renta, ya que el impuesto reduce la renta disponible y, si el ocio es un bien normal, disminuye su consumo, como el de los demás bienes normales, por lo que aumenta la oferta de trabajo.
2. Efecto sustitución –en el punto S, que se obtiene de trasladar la nueva restricción a la antigua curva de indiferencia–, ya que al disminuir el salario neto, disminuye el coste de oportunidad del ocio, lo que aumentará el ocio y disminuirá la oferta de trabajo.

Por lo tanto, dependiendo de cuál de los dos efectos predomine el efecto será incentivador o desincentivador sobre la oferta de trabajo[10].

Impuesto progresivo Por último, si el impuesto es progresivo sobre las rentas salariales como el de la Figura 21.1c, el tipo marginal crece con la renta. A modo de ejemplo, se consideran tres tramos con tipos $\tau_{Pr,1} < \tau_{Pr,2} < \tau_{Pr,3}$, de modo que la restricción presupuestaria es

$$C = w \cdot L + I - T(w \cdot L), \quad \text{con } T'(w \cdot L) = \tau_{Pr,k} \text{ en el tramo } k \tag{21.6}$$

donde $T(\cdot)$ es la función del impuesto. La restricción deja de ser una recta y pasa a tener en cada tramo una pendiente $-(1 - \tau_{Pr,k}) \cdot w$, cada vez más plana, de modo que cada hora adicional se grava más que la anterior (Rosen y Gayer, 2010, p. 418). El número de tramos es solo ilustrativo, ya que el razonamiento vale igual para más o para menos. En el caso dibujado, el equilibrio pasa de E_0 a E_1 con menos consumo y más ocio, por lo que se reduce la oferta de trabajo. Frente a un impuesto proporcional que recaude lo mismo, el efecto sustitución se refuerza, pero también se refuerza el efecto renta, por lo que el efecto general es ambiguo. No obstante, salvo que la oferta de trabajo tenga una pendiente muy negativa, lo más probable es que se reduzca, y la pérdida irrecuperable de eficiencia aumenta con el tipo marginal (Stiglitz y Rosengard, 2015).

[10] La evidencia empírica sugiere que la elasticidad de la oferta de trabajo de los asalariados principales, típicamente los hombres, es baja, en torno a 0,21-0,25, y considerablemente menor que la de los trabajadores secundarios, típicamente las mujeres (Albi et al., 2018, a partir de Blundell y MaCurdy, 1999).

 © 2026 Manuel Cabado García

Limitaciones del análisis Sin embargo, el análisis anterior tiene una serie de limitaciones (Atkinson y Stiglitz, 2015):

1. El modelo supone que se pueden elegir horas trabajadas libremente, cuando normalmente existen restricciones de las horas que se pueden trabajar.
2. Puede que el ocio sea un bien inferior y no normal para algunos individuos, en cuyo caso el efecto renta y el efecto sustitución actuarían en el mismo sentido.[11]
3. Según algunos autores, el análisis sería más adecuado en términos de esfuerzo o desutilidad.
4. Es un modelo estático, cuando las decisiones de trabajo se adoptan para un horizonte temporal.
5. No considera la incertidumbre ni las imperfecciones en el mercado de trabajo.

Imposición y ahorro

Por otro lado, parte de la renta puede ahorrarse, por lo que se va a analizar qué efectos tiene la introducción de un impuesto sobre las rentas del ahorro en un modelo de ciclo vital de dos periodos, base de los modelos de generaciones solapadas (Albi et al., 2018), para un agente que vive dos periodos y que solo obtiene ingresos salariales fijos en el periodo 1, no recibiendo ni dejando herencias.

Las decisiones de ahorro son un fenómeno intertemporal, por lo que se va a plantear un problema maximizador de la utilidad que le reporta el consumo tanto en el momento presente, C_1, como en el momento futuro, C_2, ponderado por un factor de descuento, β, menor que 1 que indica el grado de impaciencia en el tiempo, por lo que la tasa de descuento, ρ, es superior a 0:

$$\text{máx} \quad u(C_1) + \beta \cdot u(C_2), \quad \beta = \frac{1}{1+\rho} < 1 \tag{21.7}$$

sujeto a una restricción presupuestaria, donde el consumo en el presente, C_1, sumado al consumo en el futuro descontado al tipo de interés presente, $\frac{C_2}{1+r}$, es igual a la renta

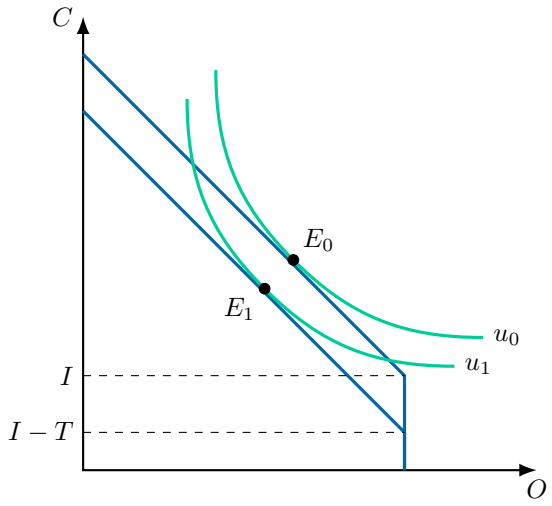
[11] Un bien es inferior cuando su consumo disminuye al aumentar la renta. Un ocio inferior implica, por tanto, que al subir la renta la persona desea menos ocio y trabaja más horas (Borjas, 2016, cap. 2). Es un caso poco habitual, pero puede darse en situaciones como estas:

- Trabajo con utilidad propia: cuando trabajar resulta en sí gratificante, como puede ocurrir entre investigadores, emprendedores y autónomos (Stern, 2004; Benz y Frey, 2008; Hamilton, 2000), la persona puede querer dedicarle más tiempo a medida que aumenta su renta.
- Emulación: si importa la renta de los demás, la presión por mantener o mejorar la posición relativa puede llevar a trabajar más horas (Bowles y Park, 2005).

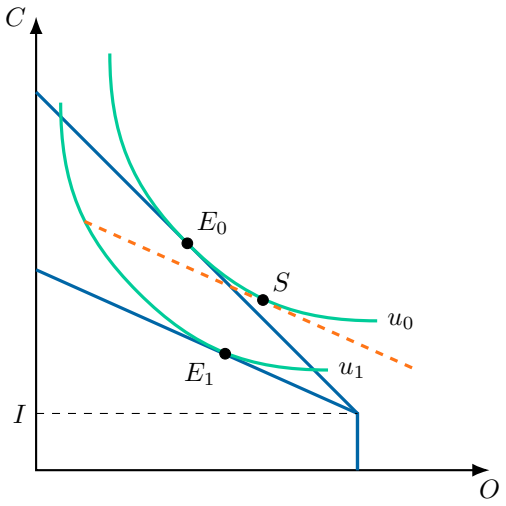
La evidencia sobre rentas no salariales, como los premios de lotería, indica en cambio que el ocio es un bien normal, con una propensión marginal a consumir ocio en torno a 0,1 por cada unidad de renta adicional (Imbens et al., 2001; Cesarini et al., 2017).

Figura 21.1: Efecto de impuestos sobre la oferta de trabajo

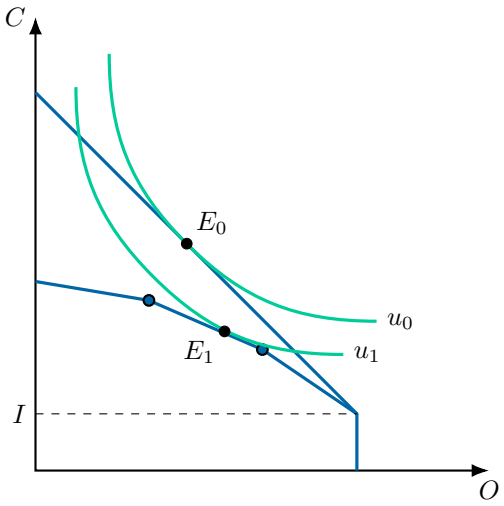
(a) Impuesto de suma fija



(b) Impuesto proporcional



(c) Impuesto progresivo



disponible, w

$$C_1 + \frac{C_2}{1+r} = w$$

donde r representa el tipo de interés.

Las CPO de este problema establecen la ecuación de Euler, según la cual el individuo debe consumir hasta que la relación marginal de sustitución entre ambos consumos, RMS_{C_1,C_2}, que es el cociente entre la utilidad marginal del consumo presente y la del consumo futuro, se iguale al cociente entre uno más el tipo de interés y uno más la tasa

 © 2026 Manuel Cabado García

de descuento:[12]

$$RMS_{C_1,C_2} = \frac{1+r}{1+\rho} \tag{21.8}$$

Gráficamente, en un plano consumo presente, C_1, y consumo futuro, C_2, como el de la Figura 21.2a, se parte de una situación de equilibrio, E_0, que se da en el punto de tangencia entre la restricción presupuestaria y la curva de indiferencia más alejada del origen, u_0.

Impuesto sobre el consumo Si se establece un impuesto sobre el consumo con el mismo tipo en ambos periodos, la restricción presupuestaria se desplaza paralelamente hacia el origen, ya que no altera el precio relativo del consumo futuro. Se produce, por tanto, solo un efecto renta, que reduce el consumo presente y el futuro hasta un nuevo equilibrio, E_1 (Atkinson y Stiglitz, 2015; Stiglitz y Rosengard, 2015). Si ambos consumos son bienes normales, el consumo presente cae menos que la renta, por lo que el ahorro disminuye.

Impuesto sobre el trabajo De la misma forma que en el análisis de la oferta de trabajo, un impuesto sobre las rentas salariales equivale a un impuesto a tanto alzado o de suma fija, dado que la oferta de trabajo es inelástica. Cuando no hay herencias, equivale también a un impuesto sobre el consumo (Stiglitz y Rosengard, 2015). Es decir, solo genera efecto renta y, por tanto, se reduce el consumo tanto en el primer periodo como en el segundo, y el ahorro disminuye.

Impuesto sobre el ahorro Por otro lado, si se establece un impuesto sobre los rendimientos del ahorro como el de la Figura 21.2b, la restricción presupuestaria pivotaría hacia abajo, estableciendo un nuevo equilibrio en E_1. De esta manera, se producen dos efectos sobre el consumo presente:

1. Efecto sustitución, ya que al disminuir los tipos de interés el consumo futuro se encarece aumentando el consumo presente y, por lo tanto, disminuyendo el nivel de ahorro.
2. Efecto renta, ya que al disminuir los tipos de interés si se supone que los individuos son acreedores, disminuirá su riqueza por lo que disminuirá el consumo tanto presente como futuro, aumentando los niveles de ahorro.

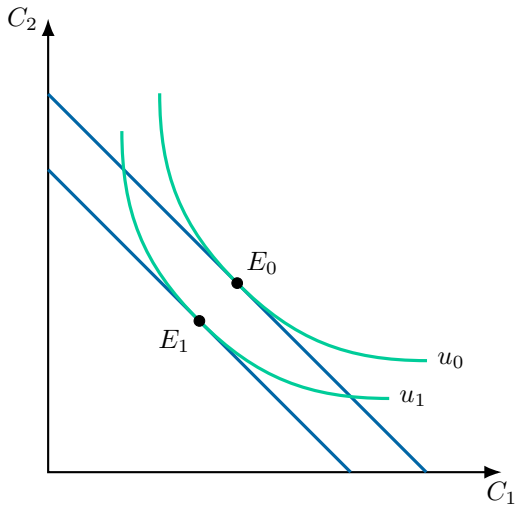
El efecto total sobre el ahorro es, por tanto, indeterminado, aunque el consumo futuro disminuirá[13].

[12] La intuición es que un euro consumido hoy debe proporcionar la misma utilidad que ahorrarlo y consumir $1+r$ euros mañana, descontados con el factor β. Por eso la condición puede escribirse $u'(C_1) = \beta \cdot (1+r) \cdot u'(C_2)$, y lo que se iguala al precio relativo del consumo presente en términos de consumo futuro, $1+r$, es la relación marginal de sustitución que ya incluye el descuento.

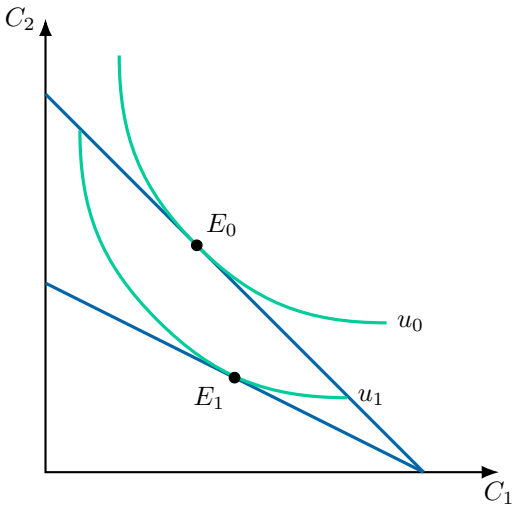
[13] La mayoría de las estimaciones internacionales de la elasticidad del ahorro respecto al tipo de interés se sitúan entre 0 y 0,4, con predominio de valores cercanos a cero, y la renta disponible parece un

Figura 21.2: Efectos de impuestos sobre el ahorro

(a) Impuesto al consumo



(b) Impuesto sobre los rendimientos del ahorro



Impuestos y deuda pública Finalmente, cabe señalar que los impuestos afectan al ahorro pero también a la deuda pública. Así, tal y como analiza la equivalencia ricardiana (Barro, 1974), una reducción de los impuestos en el período presente financiada con un aumento de la deuda pública, que dará lugar a mayores impuestos futuros para pagar esa deuda, no altera las decisiones de consumo y ahorro si los agentes prevén esos impuestos futuros y están conectados con las generaciones futuras por una cadena de transferencias intergeneracionales –como herencias o donaciones–, y si, además, los mercados de capital son perfectos y los impuestos son de suma fija.

Imposición y asunción de riesgos

Por otro lado, el grado óptimo de asunción de riesgos también puede verse distorsionado al introducir un impuesto. Normalmente un individuo puede invertir su ahorro en diferentes activos, que ofrecen diferentes tasas de rendimiento y diferentes niveles de riesgo.

Para analizarlo se utilizará una versión simplificada del modelo de cartera de Tobin (Tobin, 1958).

- En este modelo, un individuo racional buscará la maximización de rendimientos con el mínimo riesgo.
- Teniendo en cuenta que los individuos pueden dividir su riqueza en forma de demanda de bonos y demanda de dinero, la rentabilidad esperada de la cartera, R_T, si una proporción B de la riqueza se invierte en bonos, es:

$$R_T = B \cdot (g + r) \tag{21.9}$$

determinante más importante del ahorro que el tipo de interés (Albi et al., 2018).

 © 2026 Manuel Cabado García

donde el rendimiento se compone del tipo de interés, r, que ofrecen los bonos y la ganancia esperada de capital, g, que se puede obtener por la variación del precio de los bonos. El dinero ni rinde ni tiene riesgo.

- Por lo tanto un individuo se enfrentará al siguiente problema de maximización de la utilidad que dependerá positivamente del rendimiento, R_T, y negativamente del riesgo de la cartera, σ_T:

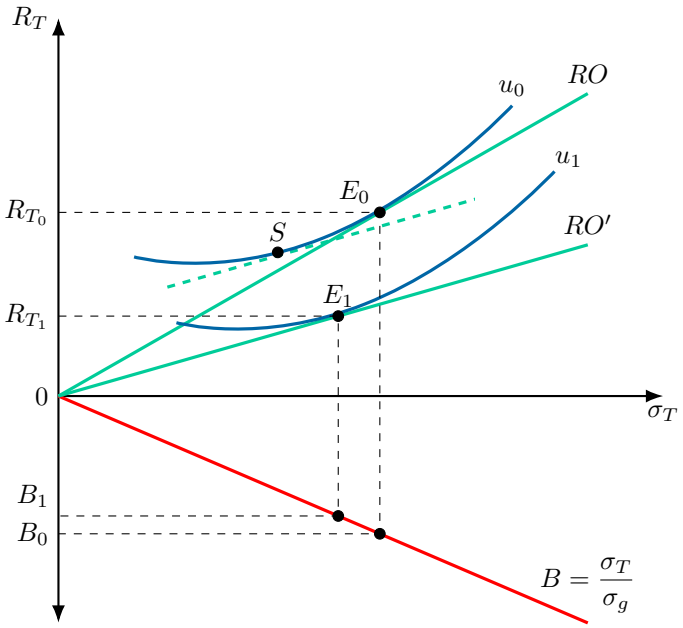
$$\text{máx} \quad u\left(R_T, \sigma_T\right) \tag{21.10}$$

sujeto a la restricción que relaciona rendimiento con riesgo, que es proporcional al de las ganancias o pérdidas de capital de los bonos, σ_g, es decir, $\sigma_T = B \cdot \sigma_g$, de modo que:

$$R_T = \frac{\sigma_T}{\sigma_g} \cdot (g + r) \tag{21.11}$$

Gráficamente, en un plano de riesgo en cartera y rendimiento en cartera como el de la Figura 21.3 en primer lugar la recta de oportunidades de inversión, RO, y, en segundo lugar, las curvas de indiferencia, u_0, que serán convexas porque se asume que los individuos son aversos al riesgo, representando un mayor bienestar cuando se desplacen hacia arriba y hacia la izquierda ya que el rendimiento es un "bien" y el riesgo es un "mal". El punto de equilibrio vendrá determinado por el punto de tangencia de la recta de oportunidades de inversión y la curva de indiferencia más alejada del eje de abscisas en E_0. Por otro lado, se utilizará la inversa del riesgo de mercado para analizar cómo distribuyen los individuos su riqueza entre demanda de bonos y demanda de dinero.

Figura 21.3: Impuestos sobre el rendimiento en cartera



En este contexto, si se introduce un impuesto sobre los rendimientos que no permite

deducir las pérdidas, la recta de oportunidades de inversión pivotará hacia la derecha con una menor pendiente, estableciendo un nuevo equilibrio en el que la rentabilidad será menor. Para analizar el efecto sobre el riesgo hay que considerar dos efectos:

1. Efecto sustitución, al disminuir los rendimientos el dinero será más barato, aumentando la demanda de dinero y disminuyendo la demanda de bonos, por lo tanto con un riesgo menor asociado.
2. Efecto renta, al disminuir los rendimientos, suponiendo que los individuos son acreedores, disminuirá su riqueza o sus recursos, disminuyendo la demanda de dinero y aumentando la demanda de bonos, es decir, con un mayor riesgo asociado.

Por lo tanto, el efecto sobre el riesgo dependerá de cuál de los dos efectos predomine.

En cambio, si existe derecho a deducción de pérdidas, el Estado se convierte en un socio que comparte tanto las ganancias como las pérdidas. El impuesto reduce entonces el rendimiento y el riesgo que soporta el inversor en la misma proporción, de modo que la recta de oportunidades de inversión no cambia de pendiente y desaparece el efecto sustitución. Queda solo el efecto renta, que lleva a asumir más riesgo para recuperar la renta perdida. Por tanto, cuando el dinero no rinde, como en este modelo, el riesgo total asumido aumenta (Domar y Musgrave, 1944). Si el activo sin riesgo tuviera rendimiento, el resultado sería indeterminado (Albi et al., 2018).

Imposición e inversión de las empresas

Hasta ahora se han analizado las decisiones de los hogares. Las empresas también ven alteradas las suyas, ya que el impuesto sobre el beneficio de las sociedades afecta a su inversión. Una empresa invierte en capital hasta que el valor del producto marginal del capital iguala su coste de uso, es decir, el tipo de interés más la depreciación (Hall y Jorgenson, 1967). Si el impuesto grava el beneficio y permite deducir todos los costes, incluido el de financiación, es neutral y no altera la decisión de inversión. Sin embargo, en la práctica no se deduce el coste de los fondos propios, por lo que sube el coste de uso del capital y se reduce la inversión (Rosen y Gayer, 2010).

Imposición y cumplimiento fiscal

Por último, los impuestos también afectan al cumplimiento fiscal. Los contribuyentes pueden reducir su carga por vías ilegales, la evasión, o legales, la elusión. En la evasión, el contribuyente decide cuánta renta declara, y la decisión depende de la probabilidad de que lo detecten y de la penalización. Aumentar cualquiera de las dos reduce la evasión, mientras que el efecto del tipo impositivo es ambiguo (Allingham y Sandmo, 1972). Además, la evasión es muy baja en la renta sobre la que informan terceros, como los salarios, y mayor en la renta autodeclarada (Kleven et al., 2011).

 © 2026 Manuel Cabado García

21.4. Equidad ⏳ 4'

Por otro lado, un buen sistema impositivo debe cumplir ciertos criterios de justicia o equidad. El problema es que el concepto de justicia no es compartido generalmente. La equidad se puede interpretar de varias formas:

1. Aplicando el criterio de beneficio, según el cual cada individuo debe pagar impuestos en función del beneficio que obtiene de los bienes públicos. En su versión más conocida, la de Lindahl, cada individuo paga por unidad de bien público un precio fiscal igual a su beneficio marginal, y la suma de esos precios cubre el coste marginal del bien. El problema de este método es que el sector público depende de que los individuos revelen sus preferencias y que, además, ignora la redistribución.
2. Aplicando el principio de capacidad de pago, que tiene a su vez dos interpretaciones:
   a) Equidad horizontal, quienes tienen la misma capacidad de pago deben pagar los mismos impuestos.
   b) Equidad vertical, quienes tienen mayor capacidad de pago deben pagar más impuestos, en una proporción que depende de la concepción de justicia que se adopte.[14]
3. Aplicando el principio de utilidad, que reemplaza el nivel de renta por el nivel de utilidad. Resuelve algún problema teórico, pero es difícil o imposible cuantificar la utilidad y compararla entre individuos (Pascual y Sarabia, 2003).
4. Aplicando el principio del consumo, según el cual solo se grava la parte de la renta que se consume y no la que se ahorra, ya que así se impone lo que el individuo obtiene de la sociedad y no lo que aporta. Pero tendería a pagar más, en proporción, quien consume una parte mayor de su renta, de modo que los estratos de renta baja, que no pueden ahorrar, pagarían un porcentaje mayor que los ricos (Pascual y Sarabia, 2003).

Centrando la atención en el criterio de equidad vertical, será preciso determinar el grado de progresividad del sistema tributario.

- Un sistema tributario será progresivo si el tipo medio crece al hacerlo su base imponible.
- No obstante, aunque existe unanimidad en torno a la definición, algunos autores asocian el criterio de progresividad a un impacto redistributivo, de ahí que existan diferentes indicadores de progresividad.

[14] La intuición es que la equidad vertical no exige por sí sola un impuesto progresivo, porque un impuesto proporcional ya hace pagar más a quien más gana. Que deba pagar más que proporcionalmente es un juicio de valor adicional (Albi et al., 2018).

A continuación, se va a distinguir entre índices de progresión local o estructural, que solo van a tener en cuenta la estructura impositiva, e índices de concentración, que también tendrán en cuenta aspectos redistributivos.

Índices de progresión local o estructural

El enfoque de la progresividad local define una medida de progresividad, $p\left(T\left(y\right)\right)$, en cada punto, y, de la función de impuestos, $T\left(\cdot\right)$. Entonces, una función es más progresiva que otra si es más progresiva en todos los puntos, esto es, $T_1\left(y\right)$ es más progresivo que $T_2\left(y\right)$ si:

$$p\left(T_1\left(y\right)\right) > p\left(T_2\left(y\right)\right), \quad \forall y$$

Musgrave y Thin (1948) propusieron cuatro medidas alternativas de progresividad local:

1. Progresión del tipo medio, $PTME$, evalúa la progresividad de una función en un punto por la pendiente del tipo medio, tme, que es igual a la diferencia entre el tipo marginal, tmg, y el tipo medio dividida por la renta, y:

$$PTME = \frac{d\left(tme\left(y\right)\right)}{dy} = \frac{tmg\left(y\right) - tme\left(y\right)}{y} \tag{21.12}$$

   por lo que una función es progresiva si, para todo y, $PTME$ es mayor que cero, igual a cero sería proporcional e inferior a cero sería regresivo. Así, un aumento de $PTME$ implica un aumento de la progresividad según esta definición.

2. Progresión del tipo marginal, $PTMA$, mide el crecimiento del tipo marginal en cada punto de la función de impuestos

$$PTMA = \frac{d\left(tmg\left(y\right)\right)}{dy} \tag{21.13}$$

   Sin embargo, este índice ha recibido fuertes críticas porque, por ejemplo, si un impuesto tiene tipos marginales constantes –por ejemplo, un impuesto lineal con tipo único y mínimo exento–, esta medida indica que el impuesto es proporcional cuando no lo es. Por esta razón no se considera un buen índice de progresividad.

3. Progresión de la carga o elasticidad impositiva –elasticidad de los impuestos pagados en relación a la renta antes de impuestos–, PC, mide, en cada punto, en qué porcentaje aumentan los impuestos pagados cuando la renta antes de impuestos aumenta en un 1 %, y es igual al cociente entre el tipo marginal y el tipo medio:

$$PC = \frac{tmg\left(y\right)}{tme\left(y\right)} \tag{21.14}$$

 © 2026 Manuel Cabado García

por lo que una función es progresiva si, para todo y, PC es mayor que 1, igual a 1 sería proporcional e inferior a 1 sería regresivo. Así, un aumento de PC implica un aumento de la progresividad según esta definición.

4. Progresión residual –elasticidad de la renta después de impuestos en relación a la renta antes de impuestos–, PRR, mide, en cada punto, en qué porcentaje aumenta la renta después de impuestos cuando la renta antes de impuestos aumenta en un 1 %:

$$PRR = \frac{1 - tmg}{1 - tme} \tag{21.15}$$

por lo que si, para todo y, PRR es inferior a 1, el impuesto es progresivo –igual a 1 sería proporcional y superior a 1 regresivo–. Esto significaría que un aumento de un 1 % en la renta antes de impuestos implica un aumento de menos del 1 % en la renta después de impuestos. Así, una reducción de PRR conlleva un aumento de la progresividad.[15]

Índices de concentración

Por otro lado, se pueden utilizar los índices de concentración:

1. Índice de Gini, que mide la desigualdad de una distribución como su distancia respecto a un reparto igualitario. Se obtiene a partir de la curva de Lorenz (Albi et al., 2018), que relaciona el porcentaje acumulado de la población, X %, ordenada de menor a mayor renta, con el porcentaje acumulado de la renta que le corresponde, Y %, como se observa en la Figura 21.4. El índice es el área comprendida entre la recta de perfecta igualdad y la curva de Lorenz dividida por el área total bajo esa recta. Si el índice de Gini es 0, la igualdad será absoluta y si es 1, un solo individuo dispondrá de todos los recursos.
2. Coeficiente de concentración, que, a diferencia del índice de Gini, ordena a los individuos por su renta antes de impuestos y mide cuánto se concentra otra magnitud –las cuotas del impuesto, T, o la renta después de impuestos, $X - T$– en los de mayor renta. Si $L_T(p)$ es la curva de concentración de las cuotas, es decir, el porcentaje acumulado del impuesto pagado por el porcentaje p de la población más pobre, el coeficiente es (Pascual y Sarabia, 2003):

$$C_T = 1 - 2\int_0^1 L_T(p)\, dp$$

[15] De las cuatro medidas locales, la PRR es la única consistente con la idea de redistribución como reducción de la desigualdad tras impuestos. En concreto, dadas dos funciones de impuestos, T_1 y T_2, la distribución de renta después de impuestos que produce T_1 domina en el sentido de Lorenz a la que produce T_2, cualquiera que sea la distribución antes de impuestos, si y solo si la PRR de T_1 es menor que la de T_2 para todo nivel de renta. Además, cada medida local induce solo una ordenación parcial, distinta de las demás, por lo que ninguna sirve por sí sola para clasificar todas las funciones de impuestos (Albi et al., 2018).

3. Índices agregados, que comparan la distribución de la renta antes de impuestos con la distribución después de impuestos o con el reparto de las cuotas

   a) Índice de Musgrave y Thin (1948), se obtiene a partir del índice de Gini antes de impuestos G_X, y el índice de Gini después de impuestos G_{X-T}:

   $$I_{MT} = \frac{1 - G_{X-T}}{1 - G_X} = 1 + \frac{G_X - G_{X-T}}{1 - G_X} \tag{21.16}$$

   si el impuesto es progresivo entonces $I_{MT} > 1$.

   b) Índice de Pechman y Okner (1974), se basa en la medición de la progresividad de un impuesto a partir de la reducción porcentual de la desigualdad medida a través del índice de Gini, y viene dado por

   $$I_{PO} = \frac{G_X - G_{X-T}}{G_X}$$

   por lo que el impuesto corrige la desigualdad si $I_{PO} > 0$ y la empeora si $I_{PO} < 0$ (Albi et al., 2018).

   c) Índice de Reynolds-Smolensky (1977), que mide el efecto redistributivo de un impuesto como la reducción absoluta del índice de Gini que produce, es decir, la diferencia entre el Gini antes de impuestos, G_X, y el Gini después de impuestos, G_{X-T} (Albi et al., 2018):

   $$I_{RS} = G_X - G_{X-T} \tag{21.17}$$

   por lo que el impuesto será redistributivo si $I_{RS} > 0$.[16]

   d) Índice de Kakwani (1977), que se basa en la desviación respecto de la proporcionalidad que conlleva el reparto de las cargas impositivas:

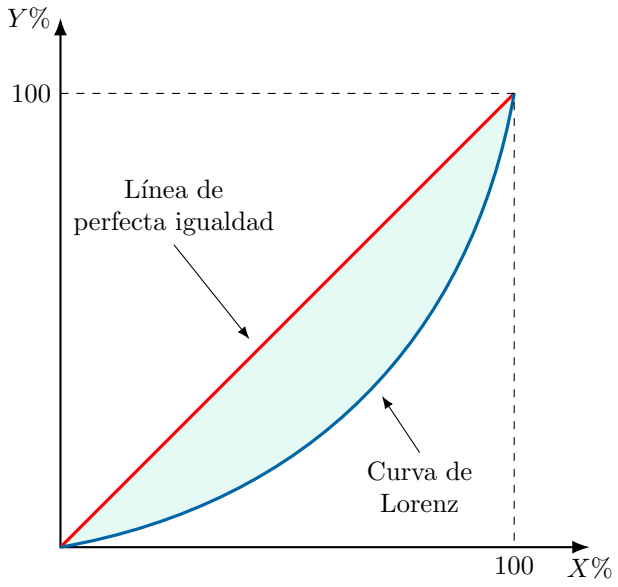
   $$K = C_T - G_X$$

   Cuanto más se aproxime K, si es positivo, a su valor máximo, $1 - G_X$, más progresivo será el impuesto, mientras que los valores negativos indican regresividad.

[16] Algunos autores calculan este índice con el índice de concentración de la renta después de impuestos, C_{X-T}, que ordena a los individuos por su renta antes de impuestos, en lugar de con su Gini: $I_{RS} = G_X - C_{X-T}$ (Pascual y Sarabia, 2003). Ambas versiones coinciden si el impuesto no altera la ordenación de los individuos. Con C_{X-T} se cumple exactamente $I_{RS} = \frac{t}{1-t} \cdot K$, siendo t el tipo medio y K el índice de Kakwani, mientras que con el Gini hay que restar además el efecto reordenación.

 © 2026 Manuel Cabado García

Figura 21.4: Índice de Gini y curva de Lorenz



21.5. Compromiso entre eficiencia y equidad 3'

Por último, haciendo referencia al compromiso entre la eficiencia y la equidad, en rigor, el análisis de la eficiencia es plenamente aplicable solo si las cuestiones distributivas son irrelevantes. Esto solo ocurriría si la economía estuviera formada por un solo individuo o múltiples individuos idénticos.

Sin embargo, si se introducen criterios de igualdad habrá que plantear un problema maximizador del bienestar:

$$\text{máx} \quad W(u_1, \ldots, u_n) \tag{21.18}$$

donde $W(\cdot)$ es una función de bienestar social, FBS, de tipo Bergson, que agrega las utilidades individuales, u_i, e incorpora juicios de valor sobre la distribución de recursos en la sociedad. La maximización está sujeta a la siguiente restricción

$$T(x, p, \tau) = \bar{T} \tag{21.19}$$

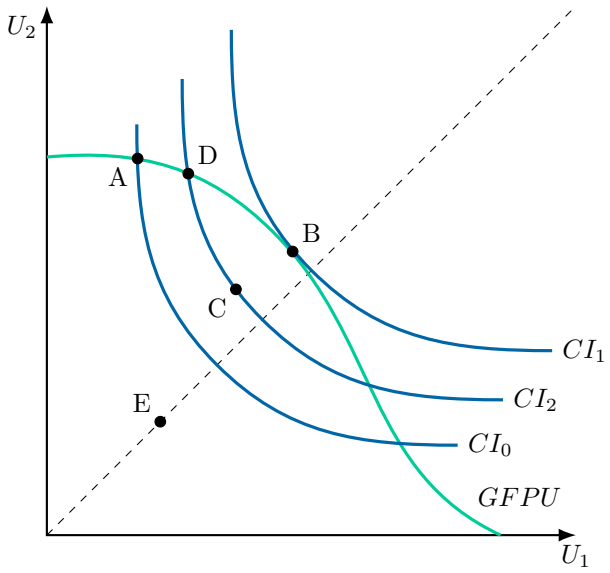
donde $T(\cdot)$ es una función impositiva que relaciona los vectores de precios, p, y demandas, x, con el conjunto de instrumentos impositivos de los que dispone la autoridad fiscal, τ, y que permite obtener el nivel de recaudación, $\bar{T}$. El sistema impositivo óptimo es el que, obteniendo esa recaudación, maximiza el bienestar social, aunque para ello tenga que aceptar un mayor coste de eficiencia a cambio de una distribución más igualitaria.

En un sistema ideal, redistribuir no debería acarrear necesariamente costes de eficiencia. Como afirma el segundo teorema de la economía del bienestar, bajo supuestos de competencia perfecta se podría alcanzar la meta redistributiva sin costes de eficiencia

mediante transferencias de suma fija.

Análisis gráfico Gráficamente, en un plano de utilidades de dos individuos 1 y 2, U_1 y U_2, como el de la Figura 21.5, se representa la Gran Frontera de Posibilidades de Utilidad, $GFPU$,[17] para dos individuos, que refleja la utilidad máxima que puede alcanzar uno dada la utilidad del otro, y las Curvas de Indiferencia Social, CI, asociadas a las Funciones de Bienestar Social. Además, se representa la bisectriz, que contiene los puntos en los que ambos individuos alcanzan la misma utilidad.

Figura 21.5: Compromiso entre eficiencia y equidad



De esta forma:

- Inicialmente, dado el conjunto de impuestos, se parte de un equilibrio como A, sobre la curva de indiferencia original, CI_0, que es eficiente dado que se sitúa sobre la $GFPU$ pero no maximiza el bienestar social, cuyo máximo se alcanza en el punto B, sobre la curva ideal, CI_1, el denominado óptimo social, y que es socialmente preferido.
- Para llegar al punto B, se aplicará el Segundo Teorema del Bienestar, por lo que se podrá alcanzar dicho óptimo social B mediante impuestos de suma fija sin resentir la eficiencia.

[17] La Frontera de Posibilidades de Utilidad, FPU, recoge las combinaciones de utilidad de los individuos que resultan de las asignaciones eficientes para un punto de producción dado. La gran FPU es la envolvente de todas ellas para todos los puntos de producción posibles (Boadway y Bruce, 1984, p. 79). Se usa aquí porque el compromiso entre eficiencia y equidad se plantea para el conjunto de la economía y no para una única asignación. Por el segundo teorema del bienestar, con impuestos de suma fija se puede alcanzar cualquier punto de la GFPU. Con impuestos distorsionantes la economía queda en el interior, como el punto C de la figura. Otros textos llaman FPU a esta misma frontera (Albi et al., 2018; Atkinson y Stiglitz, 2015).

 © 2026 Manuel Cabado García

- Sin embargo, el uso de estos impuestos es limitado debido a problemas de información, por lo que se utilizarán impuestos distorsionantes que permitirán alcanzar la asignación C que es socialmente preferida a A –se encuentra en la curva a la que se llega, CI_2, más alejada del origen que CI_0– pero ineficiente –dado que se sitúa por debajo de la $GFPU$, siendo D un punto eficiente asociado a la misma CI_2–.
- De esta forma existe un "trade-off" entre objetivos de eficiencia y equidad que la política impositiva no puede ignorar.
- Además, podría llegar un punto en el que si el gobierno quiere continuar con la redistribución puede que los beneficios sociales en equidad no superen a los costes en términos de eficiencia económica y la economía se acabe situando en E, que sería un punto de perfecta igualdad entre los individuos 1 y 2.

21.6. Conclusiones ⏳ 30"

Como conclusión, la política tributaria enfrenta una tensión estructural entre eficiencia económica y equidad distributiva. Un sistema eficiente debe minimizar las distorsiones que los impuestos generan sobre la oferta de bienes y servicios, la inversión y el comportamiento de los agentes económicos. Sin embargo, esta búsqueda no puede ignorar la equidad, que se persigue mediante la progresividad fiscal: un diseño tributario que asigna una carga creciente con la capacidad de pago.

A ello se suma el reto de combatir la evasión y elusión fiscal, que erosionan la base imponible y pueden debilitar la progresividad del sistema. Para abordar estos desafíos, se requiere un enfoque dual: positivo, que analice los efectos reales de los impuestos, y normativo, que proponga reformas orientadas a corregir fallos de mercado y mejorar el bienestar social.

Solo si se logra un buen equilibrio entre eficiencia, equidad y cumplimiento fiscal, se puede construir un sistema de impuestos que funcione bien y sea justo para todos.

Apéndice

21.A.1. Problemas de la imposición óptima

Las reglas de imposición óptima presentan algunos problemas, ya que pueden entrar en contradicción o no tener en cuenta la equidad.

Las reglas pueden entrar en contradicción entre ellas.

- Pan:
  - Las reglas de Ramsey y de la elasticidad inversa aconsejan gravarlo mucho porque su demanda es bastante inelástica.
  - La regla de Corlett-Hague aconseja gravarlo poco porque no es un bien complementario del ocio.
- Hoteles, restaurantes:
  - Las reglas de Ramsey y de la elasticidad inversa aconsejan gravarlos poco porque sus demandas son bastante elásticas.
  - La regla de Corlett-Hague aconseja gravarlos mucho porque son bienes complementarios del ocio.
- No obstante, en ciertos casos las reglas coinciden. Es el caso del alcohol o el tabaco:
  - Las reglas de Ramsey y de la elasticidad inversa aconsejan gravarlos mucho porque sus demandas son muy inelásticas.
  - La regla de Corlett-Hague aconseja gravarlos mucho porque son bienes complementarios del ocio.

Las reglas no tienen en cuenta la equidad:

- Si, por ejemplo, se comparan el pan y el caviar, según las reglas de la imposición óptima el pan debe gravarse mucho porque su demanda es inelástica, mientras que el caviar debe gravarse poco porque su demanda es muy elástica.
- Esta recomendación favorece la eficiencia, pero no tiene en cuenta la función redistribuidora que debe perseguir todo sistema fiscal.

Por lo tanto, un sistema fiscal, para ser considerado óptimo en su conjunto, debería tener en cuenta tanto la eficiencia como la equidad.

21.A.2. Impuesto sobre el valor del suelo

El impuesto sobre el valor del suelo –Land Value Tax, LVT– grava el valor del suelo y no el de lo construido sobre él. Su principal defensor histórico fue Henry George, quien en

 © 2026 Manuel Cabado García

Progreso y miseria (1879) propuso sustituir los impuestos sobre el trabajo y el capital por un único impuesto sobre el valor del suelo, bajo la idea de que la renta del propietario no responde a ningún esfuerzo productivo (Gruber, 2016, p. 736). Economistas de diferentes escuelas lo han defendido por su eficiencia, como Adam Smith[18], Milton Friedman, que lo calificó de "el menos malo de los impuestos" (Fox, 2017), o Paul Krugman, quien considera que los modelos de economía urbana respaldan este impuesto al menos para financiar el crecimiento de las ciudades, aunque no recaudaría lo bastante para sostener un Estado del bienestar moderno (Moore, 2009).

La razón es que la oferta de suelo es fija, de modo que el impuesto no la reduce ni altera las decisiones de los agentes, y la carga recae sobre el propietario (Stiglitz y Rosengard, 2015). Además, como no depende de lo que se construya, cuanto más se aprovecha el suelo, por ejemplo con más viviendas, menor es la carga en términos relativos. Eso incentiva a ofrecer más vivienda, lo que puede contribuir a reducir los precios (Arenas, 2024; Fox, 2017).

El impuesto grava el valor potencial del suelo y no los incrementos que resultan de las mejoras del propietario. La hipótesis es que ese valor se debe a circunstancias ajenas a la propiedad, habitualmente a la inversión pública, como la peatonalización de una calle o una estación de metro. Por tanto, redistribuye a la sociedad las externalidades positivas de las infraestructuras y los servicios públicos que revalorizan el suelo (Arenas, 2024). En modelos de bienes públicos locales, las rentas del suelo igualan el gasto público óptimo, resultado conocido como teorema de Henry George (Atkinson y Stiglitz, 2015).

También puede ser progresivo, porque los propietarios suelen tener mayor capacidad económica, y es difícil de eludir por deslocalización, ya que el suelo no se puede trasladar a otra jurisdicción (Arenas, 2024). Además, su justificación moral, la idea de que el suelo es de todos y de que sus frutos deben repercutir en la sociedad, puede ayudar a justificarlo políticamente (Arenas, 2024).

La experiencia de Pittsburgh apunta en esa dirección, aunque con matices. En 1979-1980 pasó a gravar el suelo a un tipo más de cinco veces superior al de las construcciones, y en los años ochenta la edificación aumentó muy por encima de la de otras ciudades de la región. Según Oates y Schwab (1997), el principal impulso fue la escasez de espacio comercial, y la mayor imposición sobre el suelo tuvo un papel de apoyo importante.

Como inconvenientes destacan:

- La estimación del valor potencial del suelo, es decir, la base imponible, ya que es difícil separar el valor del suelo del de las mejoras (Stiglitz y Rosengard, 2015).

[18] Este es un pasaje de su obra *La riqueza de las naciones* (libro V, capítulo II), sobre las rentas del suelo y de las casas:

"Though a part of this revenue should be taken from him in order to defray the expences of the state, no discouragement will thereby be given to any sort of industry. The annual produce of the land and labour of the society, the real wealth and revenue of the great body of the people, might be the same after such a tax as before."

- Los contribuyentes pueden encontrarse sin liquidez y verse forzados a descapitalizarse para pagar el impuesto. Sin embargo, pueden imaginarse soluciones a este problema, por ejemplo diferir el pago hasta su venta (Arenas, 2024).
- La introducción del impuesto genera perdedores claros: como se capitaliza en el valor del suelo, reduce de inmediato el valor de las propiedades y perjudica a quienes eran propietarios al anunciarse, especialmente a quienes las adquirieron hace poco (Stiglitz y Rosengard, 2015; Arenas, 2024).
- Es políticamente difícil de implantar, porque los propietarios son hoy una mayoría con alta propensión a votar y los impuestos sobre la propiedad son muy impopulares, como muestra la Proposición 13 de California (Fox, 2017). Friedman atribuía su impopularidad a que es el único impuesto que obliga a pagar de una vez una cantidad grande, y proponía retenerlo en origen (Fox, 2017). Pittsburgh abandonó su experimento de gravar más el suelo que las construcciones a comienzos de los años 2000 (Smith, 2017).

A pesar de estos inconvenientes, sus ventajas explican que se proponga como instrumento de política fiscal y de vivienda, aunque su viabilidad depende de resolver la valoración del suelo y la transición.

21.A.3. La inflación y el impuesto de la renta

Como indica Albi et al. (2018), la existencia de procesos inflacionistas origina problemas en el ámbito del impuesto sobre la renta personal. Los efectos de la inflación sobre el impuesto sobre la renta se transmiten por tres vías esencialmente: la base imponible, los elementos determinantes de la progresividad real, y los desfases y retardos recaudatorios.

La inflación y la base imponible

Las distorsiones de la inflación en la determinación de la base son de dos tipos:

1. Una reducción en términos reales de las cantidades exentas, y de las deducciones fijas.
2. Una falsa determinación de las rentas reales, sobre todo del capital. Las rentas del capital tienden a menudo a estar fijadas en términos monetarios y esto les perjudica, ya que con la inflación su valor real disminuye pero su tributación se mantiene. Igualmente, la inflación origina ganancias de capital puramente nominales.

Inflación y progresividad efectiva del impuesto

Por otro lado, los impuestos progresivos sobre la renta personal son vulnerables a la inflación. En presencia de procesos de inflación la estructura progresiva del impuesto

 © 2026 Manuel Cabado García

sobre la renta genera distorsiones en la distribución de la carga tributaria y altera la progresividad real del impuesto.

Los impuestos progresivos y los incrementos de precios originan aumentos impositivos que no se corresponden con aumentos reales en la capacidad de pago de los contribuyentes, si no se establecen correcciones. Este efecto se produce por dos vías:

1. La inflación, como ya se ha indicado, reduce el valor real de las cantidades exentas, desgravaciones y deducciones de cuantía fija que el impuesto introduce con objeto de reforzar la progresividad y ajustar la carga tributaria a las circunstancias personales o familiares de los contribuyentes. Como resultado, resulta necesario un ajuste anual o una indiciación de estas bonificaciones si se desea que éstas mantengan sus objetivos.
2. La escala impositiva se establece habitualmente en términos nominales, asignando tipos impositivos mayores a rentas nominales crecientes. Cuando el impuesto es progresivo, incrementos puramente nominales en la renta enfrentan a los contribuyentes a tipos marginales superiores.
   Es frecuente que, ante aumentos en el nivel general de precios, los salarios nominales crezcan para compensar así la pérdida de capacidad adquisitiva de los trabajadores. Así, aquellos contribuyentes cuya renta permanece constante en términos reales a lo largo del tiempo, pero que ven aumentar su salario nominal, se desplazan en la tarifa hacia tramos de renta superiores, que están sujetos a tipos de gravamen más elevados.
   Una posibilidad para resolver este problema consiste en deflactar automática y anualmente la tarifa, o bien establecer una indiciación automática de la base imponible con un índice de inflación. La otra posibilidad consiste en practicar políticas discrecionales de ajuste. Esta última es, sin embargo, la opción más habitual. Como resultado, en aquellos ejercicios en los que no se producen ajustes en la tarifa, el gobierno incrementa la recaudación sin que se hayan producido aumentos en la capacidad de pago real de los contribuyentes.

El efecto doble de la inflación sobre el impuesto progresivo de la renta personal, a través de las exenciones y deducciones nominales y de la tarifa nominal, en ausencia de ajustes permanentes, se denomina «progresividad en frío» o «rémora fiscal».

Retrasos y adelantos recaudatorios

Por último, los procesos de inflación y deflación muy agudos afectan a la recaudación del sector público en términos reales y, en consecuencia, a la carga tributaria real soportada por los contribuyentes[19].

[19] Por ejemplo, en períodos de inflación muy intensa los pagos de impuestos en el ejercicio siguiente al que tuvieron lugar los ingresos reducen la recaudación real del gobierno. Lo contrario ocurriría en

Con objeto de mantener el valor real de la recaudación cuando los precios varían a gran velocidad existen tres tipos de mecanismos esencialmente:

1. Sistemas de retenciones en el origen de la renta: en el momento en que se genera la renta, el pagador retiene un porcentaje que ha de ingresar en Hacienda en concepto de «anticipo» de la factura impositiva del perceptor.
2. Sistemas de pagos fraccionados frecuentes: similar al sistema de retenciones, consiste en que es el propio contribuyente quien, periódicamente, debe ingresar en Hacienda ciertas cantidades en concepto de adelanto de sus obligaciones tributarias.
3. Indiciación de la deuda tributaria con índices de inflación: esta última posibilidad consistiría en ajustar la factura impositiva de acuerdo con la evolución de los precios, de manera que se mantuviera constante el valor real de la recaudación.

Sin embargo, incluso cuando se implanta alguno de estos mecanismos puede existir discriminación entre los contribuyentes en presencia de procesos inflacionistas muy intensos. Por ejemplo, cuando, por problemas de control, no se generaliza el sistema de retención en origen a todas las modalidades de renta, contribuyentes con idéntica capacidad de pago que otros, pero que obtienen su renta sin retenciones, pueden posponer los pagos impositivos, reduciendo así su carga tributaria real. Esta discriminación distorsiona la equidad vertical y horizontal del impuesto.

21.A.4. Impuesto progresivo al consumo

En un estudio del NIESR de Reino Unido (Bhattacharjee et al., 2022) se trata el impuesto progresivo al consumo. La imposición directa sobre el consumo se distingue de un impuesto indirecto sobre el gasto corriente, como el Impuesto sobre el Valor Añadido (IVA), porque grava el consumo total de un individuo a lo largo de un año, en lugar de gravar el punto de gasto cada vez que se realiza una compra.

Según estos autores, un impuesto progresivo sobre el consumo, adecuadamente diseñado, mejora el bienestar de tres maneras.

1. En primer lugar, ayuda a los hogares a suavizar su consumo a lo largo del ciclo vital. La razón es que el consumo tiende a aumentar de forma constante con la edad hasta mediados de los 40 y luego disminuye bruscamente más adelante en la vida, pero las personas experimentan una mayor satisfacción a lo largo de la vida cuando pueden realizar un consumo más estable durante toda la vida adulta.
2. En segundo lugar, un impuesto progresivo sobre el consumo permite a los hogares tomar decisiones más flexibles sobre el equilibrio entre trabajo y ocio, al incentivar el empleo para todos los hogares, no solo los de renta baja, sino también los de renta más alta, ya que su trabajo no se grava, sino solo lo que consumen.

períodos de deflación muy intensa.

 © 2026 Manuel Cabado García

3. En tercer lugar, un impuesto progresivo al consumo permite a los hogares acumular riqueza porque genera incentivos para aumentar el ahorro de todos los grupos de renta.

No obstante, también apuntan que la introducción de un impuesto directo progresivo sobre el consumo plantea una serie de cuestiones de transición, como, por ejemplo, si se aplicaría de la misma manera a los bienes y servicios importados, cómo evitar la doble imposición de las pensiones y los bienes registrados y cómo abordar la fiscalidad diferenciada del capital –por ejemplo, el capital de las empresas frente al de la vivienda y el capital heredado–.

Una forma de abordar estos problemas de transición es introducir gradualmente un impuesto directo progresivo sobre el consumo, que suavizaría la carga que recaería sobre las personas que mantienen un alto nivel de consumo a partir del capital heredado. Ello se debe a que el paso de un sistema fiscal a otro modificará el valor de los activos, imponiendo una carga a la riqueza acumulada antes de la reforma y a la generación que posea el capital en ese momento. Otra forma es permitir exenciones, especialmente umbrales libres de impuestos.

Aunque las ventajas de un impuesto directo y progresivo sobre el consumo son evidentes, también es cierto que un sistema de este tipo tiene inconvenientes, como una larga transición, la dificultad de distinguir los distintos tipos de ingresos y gastos –por ejemplo, las herencias– y cuestiones relacionadas con los movimientos transfronterizos de bienes y servicios.

21.A.5. Curva de Laffer

A mediados de la década de los setenta, Arthur Laffer sugirió que cuando los tipos impositivos eran suficientemente elevados, introducían tales desincentivos a las ofertas factoriales que aumentarlos reduciría la renta del país y disminuiría la recaudación –porque los nuevos tipos más elevados se aplicarían a niveles de renta mucho más bajos–. Una reducción de los tipos impositivos, por su parte, generaría el efecto contrario: incentivaría las ofertas factoriales, lo que estimularía un fuerte crecimiento de la renta y esto, a su vez, produciría una recaudación más elevada.

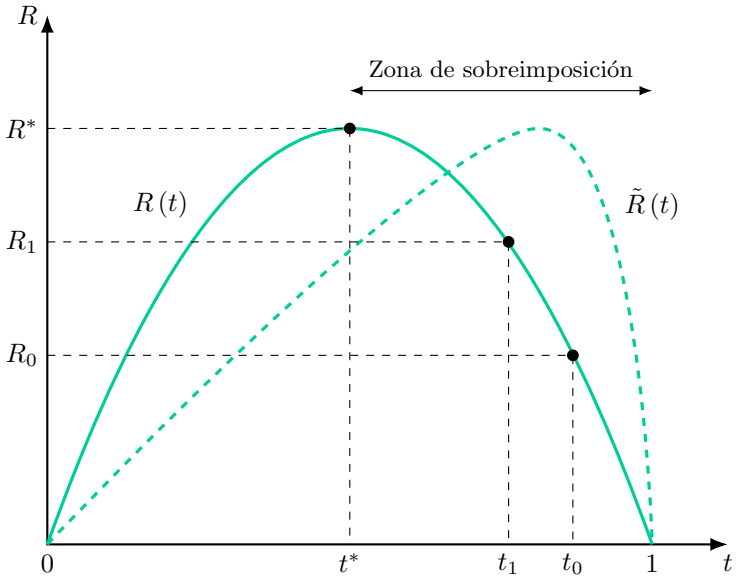
El razonamiento de Laffer parte de dos observaciones sencillas:

1. Cuando el tipo impositivo es cero, la recaudación es nula.
2. Cuando el tipo impositivo es igual a uno la recaudación también es nula, ya que si una persona ve que todo lo que gane lo va a tener que pagar en impuestos optará por no trabajar, y la base del impuesto desaparecerá.

Como se observa en la Figura 21.6, se denomina $R(t)$ a la recaudación como función del tipo impositivo. Dado que la función nace y muere en cero, lo más plausible es que, comenzando en $t = 0$, la recaudación aumente a medida que el tipo crece hasta alcanzar

un máximo en t^*. A partir de ese punto, la recaudación comienza a disminuir hasta hacerse cero cuando $t = 1$. De esta forma, tal y como se representa en la figura, la función de recaudación tendrá forma de U invertida. Para tipos impositivos superiores a t^*, una reducción de tipos aumentará la recaudación. Por ejemplo, una reducción del tipo impositivo de t_0 a t_1 aumentará la recaudación de R_0 a R_1.

Figura 21.6: La curva de Laffer



En la práctica, el razonamiento de Laffer presenta varios problemas.

- En primer lugar, una reducción de tipos no cambia inmediatamente el comportamiento de los individuos, ni genera de forma inmediata aumentos de la renta nacional. En el mejor de los casos, será un resultado de medio o largo plazo y en el intermedio el sector público deberá soportar una reducción de ingresos.
- Esta reducción, a su vez, forzará al sector público a reducir el gasto público o aumentar el endeudamiento, lo que redundará en una reducción de la demanda agregada que, en última instancia, puede retrasar o incluso eliminar los posibles aumentos de base imponible derivados del recorte fiscal.
- Finalmente, determinar cuál es el tipo a partir del cual comienza la sobreimposición es complicado. Con todo, para que una reducción de tipos aumente la recaudación, la elasticidad de la base imponible respecto al tipo impositivo tiene que ser, en valor absoluto, mayor que uno, lo que no parece ocurrir con las elasticidades de la oferta de trabajo y del ahorro vistas en este tema.[20] Además, los análisis cuantitativos más

[20] La recaudación es $R(t) = t \cdot B(t)$, siendo $B(t)$ la base imponible, de modo que $R'(t) = B(t) \cdot (1 - \varepsilon)$, donde $\varepsilon = -\dfrac{t}{B} \cdot \dfrac{dB}{dt}$ es la elasticidad de la base respecto al tipo. La recaudación cae al subir el tipo si $\varepsilon > 1$. Expresada respecto a la parte de la renta que queda tras el impuesto, $1 - t$, la condición equivale a una elasticidad superior a $(1 - t)/t$, que con un tipo del 50 % es 1 y con un tipo del 30 % es 2,3.

 © 2026 Manuel Cabado García

recientes indican que un recorte de tipos solo se autofinancia en parte: en Estados Unidos, el 32 % de un recorte de impuestos sobre el trabajo y el 51 % de uno sobre el capital, y el 54 % y el 79 %, respectivamente, en la UE-14 (Trabandt y Uhlig, 2011). La experiencia de Estados Unidos con los recortes de los años ochenta apunta en la misma dirección[21].

Todo lo anterior sugiere que es probable que la zona de sobreimposición, de existir, surja a niveles de presión fiscal próximos a la confiscación, como se observa en la curva discontinua, $\tilde{R}(t)$, pero no a los niveles de imposición que existen actualmente en los países más avanzados. Por ejemplo, el máximo de la curva de Laffer para los impuestos sobre el trabajo en Estados Unidos se sitúa en torno a un tipo medio del 58 % (Holter et al., 2019).

21.A.6. Razones para introducir un sistema fiscal progresivo

Como indican Pascual y Sarabia (2003), parece claro que el sistema fiscal puede ser: regresivo, proporcional o progresivo. Pero, ¿cuál de ellos es preferible? Para responder a esta cuestión, se analizarán brevemente las implicaciones de cada uno.

- Sistema fiscal regresivo. Puede aceptarse desde el punto de vista de la asignación óptima de los recursos aunque, atendiendo únicamente a la cuestión de la equidad, parece poco razonable proponer un sistema de este tipo. Con todo, cuando se apoya el principio del beneficio o del consumo, se tiende a apostar indirectamente por un sistema fiscal que produce efectos regresivos en la distribución de la renta.
- Sistema fiscal proporcional. Mantiene las diferencias y desigualdades económicas por lo que se produce un efecto distributivo cuando el sistema fiscal responde al principio de la capacidad de pago. Sin embargo, cabe conformarse con que el sistema fiscal sea proporcional, con el fin de evitar efectos negativos sobre la eficiencia económica, siempre que se asegure que el gasto público cumplirá la función de redistribución.
- Sistema fiscal progresivo. Reduce las desigualdades económicas y se justificaría si se sigue el principio de la utilidad al producirse una redistribución de las rentas, que la querrán evidentemente los que ganan con ella, y solo ellos.

Por un lado, la política fiscal progresiva es deseada por la presión de los individuos con renta baja, ya que saldrán ganando con las políticas redistributivas –ceden una proporción de su renta bruta para financiar el gasto público que es inferior a la que pagan los individuos con rentas elevadas–. Por lo tanto, puede haber una mayoría que sale beneficiada con el sistema progresivo.

[21] Concretamente, durante el mandato de Ronald Reagan (1981-1989) se aplicaron fuertes recortes impositivos, inspirados en los argumentos de la economía de la oferta. Los ingresos federales cayeron del 18,7 % del PIB en 1981 al 16,5 % en 1983-1984, y el déficit pasó del 2,5 % al 5,7 % del PIB entre 1981 y 1983 (U.S. Office of Management and Budget, 2026).

Por otro lado, existirá un grupo de contribuyentes de rentas altas que, en un sistema fiscal progresivo, paga una proporción de su renta mayor que la presión fiscal media. No obstante, parte de ellos puede aceptar una política fiscal progresiva si sus funciones de utilidad son interdependientes. En el modelo de Hochman y Rodgers (1969), la utilidad de un individuo, U_a, depende de su propia renta, Y_a, y de la utilidad de los demás individuos, U_b, U_c, etcétera, es decir, $U_a = f(Y_a, U_b, U_c, \ldots)$. Es lo que se denomina "altruismo de cercanía".

Otros autores sostienen que las rentas altas aceptan redistribuir por temor a que las desigualdades sociales se agraven (Brennan, 1973; Pauly, 1973), o porque una mejor distribución puede favorecer el crecimiento futuro y beneficiar también a esas rentas. Además, el deseo de redistribuir puede surgir de una voluntad social, independiente del nivel de renta personal.

21.A.7. Suavización impositiva

Cuando se cumple la equivalencia ricardiana, el Gobierno no tiene motivos para elegir una determinada senda de déficit presupuestario. Sin embargo, como se ha visto, los impuestos pueden afectar a los precios relativos y, por lo tanto, ser distorsionantes, presentando así un efecto sustitución además del efecto renta propio de los impuestos de suma fija. Siendo así, la distribución temporal de los impuestos sí tiene importancia.

Barro (1979) desarrolló un modelo para averiguar cuál debe ser la senda óptima impositiva que minimice las distorsiones derivadas del sistema impositivo. En él, el Gobierno parte de una secuencia conocida y exógena de gasto público, G_t, y de renta, Y_t, de un tipo de interés constante, r, y de una deuda inicial, B_0. Los ingresos fiscales, T_t, son una proporción de la renta igual al tipo impositivo medio, τ_t:

$$\tau_t = \frac{T_t}{Y_t}$$

El coste de recaudar, es decir, el exceso de gravamen del periodo t, CT_t, es:

$$CT_t = Y_t \cdot f(\tau_t) \tag{21.20}$$

donde el coste aumenta proporcionalmente más que el tipo de gravamen: es decir, $f(\tau_t)$ es estrictamente convexa, con $f' > 0$ y $f'' > 0$.

En dicho modelo, el problema del Gobierno es diseñar una secuencia impositiva que minimice el valor actual del coste de recaudar, CT:

$$\text{mín} \quad CT = \sum_{t=1}^{\infty} \frac{Y_t \cdot f(\tau_t)}{(1+r)^t} \tag{21.21}$$

sujeto a la restricción intertemporal del Gobierno, que iguala el valor actual del gasto más

 © 2026 Manuel Cabado García

la deuda inicial al valor actual de los ingresos:

$$\sum_{t=1}^{\infty} \frac{G_t}{(1+r)^t} + B_0 = \sum_{t=1}^{\infty} \frac{T_t}{(1+r)^t} \tag{21.22}$$

y a la condición de transversalidad, que evita soluciones explosivas y esquemas tipo Ponzi, donde B_t es la deuda al final del periodo t, que evoluciona según $B_t = (1+r)\, B_{t-1} + G_t - T_t$:

$$\lim_{t\to\infty} \frac{B_t}{(1+r)^t} = 0 \tag{21.23}$$

La CPO exige que el coste marginal de recaudar sea el mismo en todos los periodos. Esto ocurre con independencia del tipo de interés, porque la recaudación y su coste se producen en el mismo momento. Como f es estrictamente convexa, f' es estrictamente creciente y a cada valor del coste marginal le corresponde un único tipo. Por tanto:

$$f'(\tau_t) = f'(\tau_{t+1}) \implies \tau_t = \tau_{t+1} \tag{21.24}$$

Es decir, lo óptimo es que el tipo se mantenga constante, cualquiera que sea la senda de renta y de gasto. Y es que, como el coste marginal de la distorsión es creciente con el tipo impositivo, un tipo uniforme minimiza el exceso de gravamen.

Esto implica una suavización impositiva, por lo que es preferible:

1. Mantener tipos bajos durante muchos años antes que tipos altos durante un solo año.
2. Mantener el tipo constante y dejar actuar a los estabilizadores automáticos. El déficit aparece cuando se espera un aumento temporal del gasto sobre la renta o una caída temporal de la renta, como en recesiones o guerras, y el superávit en el caso contrario. O lo que es lo mismo: el déficit público será contracíclico –es decir, ahorro público, procíclico–. En este contexto, el papel de la deuda será, pues, el de absorber los shocks para permitir la deseada suavización impositiva.

21.A.8. EUROMOD: medir la redistribución y la progresividad con microsimulación

El cuerpo del tema desarrolla las medidas de progresividad y el compromiso entre eficiencia y equidad. EUROMOD es la herramienta con la que se mide, en la práctica, el efecto de los impuestos y las prestaciones sobre la renta de los hogares europeos.

Qué simula EUROMOD es el modelo de microsimulación de impuestos y prestaciones de la Unión Europea. A partir de microdatos de hogares basados en EU-SILC y de las

reglas fiscales de cada Estado miembro, simula los impuestos sobre la renta, las cotizaciones sociales de empleados, autónomos y empleadores y las prestaciones, y desde 2024 también los impuestos sobre el consumo (Joint Research Centre, sf; Sutherland y Figari, 2013). Lo gestiona el Joint Research Centre de la Comisión Europea desde 2021.

Cómo calcula la renta disponible Con Y_o la renta original (de mercado más transferencias regulares), B las prestaciones, SIC las cotizaciones sociales de los hogares y T los impuestos:

$$Y_d = Y_o + B - SIC - T$$

Qué mide La redistribución se define como el índice de Gini de la renta inicial menos el de la renta disponible, y se descompone con el marco de Kakwani:

$$RE = L \cdot P - RR$$

donde L es el nivel (impuestos y cotizaciones menos prestaciones, sobre la renta disponible), P la progresividad (el índice de Kakwani de esos impuestos netos, es decir, su índice de concentración menos el Gini de la renta inicial) y RR el reordenamiento de los hogares tras los impuestos (Joint Research Centre, sf). Es la misma estructura que relaciona el índice de Reynolds-Smolensky con el de Kakwani: la redistribución es el nivel, que hace de $t/(1-t)$ con t el tipo medio, multiplicado por la progresividad de Kakwani, menos el efecto reordenación.

Qué no hace

- No mide la traslación: al ser un modelo estático que supone fijas las rentas brutas (Klevmarken, 2022), aplica las reglas fiscales a las rentas observadas, lo que se corresponde con la incidencia de impacto, no con la efectiva. Cualquier hipótesis de traslación, por ejemplo sobre las cotizaciones empresariales o el impuesto de sociedades, debe imponerla el analista.
- No capta el efecto sustitución: sin respuestas de comportamiento, recoge la carga directa del impuesto, el efecto renta, pero no el efecto sustitución ni, por tanto, el exceso de gravamen ni los efectos incentivos sobre la oferta de trabajo. Para estos hacen falta elasticidades estimadas fuera del modelo.
- Supone cumplimiento pleno: por defecto, no hay evasión ni cobro incompleto de prestaciones, que solo se modelan con extensiones calibradas cuando la validación macro lo exige.

Impuestos sobre el consumo Desde 2024, EUROMOD los incluye, de modo que puede medir también cuánto soportan los hogares de estos impuestos, y no solo de los

 © 2026 Manuel Cabado García

impuestos directos.

EUROMOD es, por tanto, el complemento empírico de la teoría: mide con microdatos la incidencia de impacto y la redistribución, mientras que los efectos de comportamiento y el compromiso entre eficiencia y equidad requieren los modelos de este tema.

21.A.9. Impuestos e inversión empresarial

Una empresa que usa capital, K, y trabajo, L, maximiza su beneficio, Π:

$$\Pi = p \cdot F(K, L) - w \cdot L - q \cdot (r + \delta) \cdot K$$

donde p es el precio del producto, F la función de producción, w el salario, q el precio de los bienes de capital, r el tipo de interés y δ la tasa de depreciación. La condición de primer orden para el capital, $p \cdot F_K = q \cdot (r + \delta)$, iguala el valor del producto marginal del capital a su coste de uso.

Con un impuesto sobre el beneficio de tipo u y sin crédito fiscal a la inversión, el coste de uso, c, pasa a ser:

$$c = q \cdot (r + \delta) \cdot \frac{1 - u \cdot z}{1 - u}$$

donde z es el valor actual de las deducciones por depreciación de cada euro invertido (Hall y Jorgenson, 1967). Si la inversión se amortiza de inmediato, $z = 1$ y el impuesto es neutral. Si no se deduce nada, $z = 0$, el coste de uso es $q \cdot (r + \delta) / (1 - u)$, sube y la inversión se reduce. Entre ambos extremos, cuanto más rápida es la amortización fiscal, mayor es z y menor el coste de uso.

Hay otra vía de neutralidad: si el impuesto permite deducir los costes de financiación y la inversión marginal se financia con deuda, no altera la decisión de inversión (Rosen y Gayer, 2010). Pero esta neutralidad depende de supuestos. Si la inversión se financia con acciones o beneficios retenidos, o hay restricciones de financiación o riesgo de quiebra, el efecto depende de la estructura financiera. Y supone que a la empresa le cuesta lo mismo financiarse con fondos propios que con ajenos. Con restricciones de liquidez, el impuesto reduce los fondos internos y, con ellos, la inversión (Rosen y Gayer, 2010).

21.A.10. Evasión y elusión fiscal

La evasión consiste en el impago voluntario de impuestos debidos y es ilícita, mientras que la elusión consiste en reducir la carga tributaria mediante disposiciones legales y es lícita (Albi et al., 2018; Slemrod y Yitzhaki, 2002).

Allingham y Sandmo (1972) modelaron la decisión de evadir como una elección bajo incertidumbre. Un contribuyente con renta real W declara una renta X, sobre la que paga un tipo impositivo θ. Con probabilidad p es inspeccionado y paga además un tipo de penalización π, mayor que θ, sobre la renta no declarada. Maximiza su utilidad esperada

y declara menos de lo debido si $p \cdot \pi < \theta$. Un aumento de la probabilidad de detección o de la penalización aumenta la renta declarada. El efecto del tipo impositivo es ambiguo, porque el efecto sustitución reduce la renta declarada y el efecto renta la aumenta.

La evidencia de un experimento de auditoría en Dinamarca matiza el modelo: la evasión es casi nula para la renta con información de terceros, como los salarios, y sustancial para la renta autodeclarada (Kleven et al., 2011).

21.A.11. Curva del gran Gatsby

La curva del gran Gatsby relaciona la desigualdad de la renta en un momento dado con la movilidad entre generaciones. Alan Krueger, presidente del Consejo de Asesores Económicos de Estados Unidos, la popularizó a partir de los trabajos de Corak y le dio el nombre del personaje de Fitzgerald (Stiglitz y Rosengard, 2015; Durlauf et al., 2022). En el eje horizontal se mide la desigualdad con el coeficiente de Gini, y en el vertical la persistencia con la elasticidad intergeneracional de la renta, IGE, que relaciona la renta permanente de padres e hijos. Cuanto mayor es la IGE, menor es la movilidad.

La relación es positiva: los países con más desigualdad tienden a tener menos movilidad. Por ejemplo, los países de la OCDE con más desigualdad en los años ochenta tenían más persistencia de la renta entre generaciones, y Krueger estimó que la IGE de Estados Unidos aumentaría de 0,47 a 0,56 en la próxima generación (Stiglitz y Rosengard, 2015).

Parte de esta relación es mecánica. Si el logaritmo de la renta de los hijos, y_{t+1}, depende del de los padres, y_t, según

$$y_{t+1} = \alpha + \beta \cdot y_t + \varepsilon_t$$

donde α es una constante, β la IGE y ε_t un término de error, entonces la varianza de la renta futura es

$$\operatorname{Var}(y_{t+1}) = \beta^2 \cdot \operatorname{Var}(y_t) + \operatorname{Var}(\varepsilon)$$

de modo que la persistencia y la desigualdad futura están ligadas matemáticamente. Según Berman, esta relación explica casi dos tercios de la variación de la IGE entre países y más del 90 % de su variación en el tiempo en Estados Unidos (Durlauf et al., 2022). Por eso, interpretar que la desigualdad genera persistencia exige tratar la ecuación como una relación estadística y no estructural.

Durlauf, Kourtellos y Tan (2022) distinguen cinco clases de teorías, que no compiten entre sí, para explicar el mecanismo:

- Inversión de las familias: los padres reparten su renta entre educar a los hijos y consumir, y las restricciones de crédito hacen que la renta de los padres limite esa inversión.
- Habilidades: las capacidades cognitivas y no cognitivas dependen de los insumos familiares durante la infancia, incluida la estructura familiar y el momento en que

 © 2026 Manuel Cabado García

se reciben.

- Influencias sociales: la segregación de ricos y pobres en barrios distintos, con educación financiada localmente, transmite el estatus socioeconómico entre generaciones.
- Economía política: la desigualdad determina el voto sobre la redistribución y, con él, la inversión pública en educación, como en el modelo del votante mediano.
- Aspiraciones: la desigualdad aumenta la dispersión de las aspiraciones, que se forman por influencias familiares y sociales.

Más allá de la comparación entre países, la curva se ha estudiado dentro de países y regiones con resultados desiguales. En China, por ejemplo, la relación entre la desigualdad provincial y la persistencia se debilita sustancialmente al estimarla en primeras diferencias (Durlauf et al., 2022).

La curva interesa porque sugiere que una sociedad puede perseguir a la vez la igualdad de resultados y la de oportunidades, y porque cuestiona que Estados Unidos acepte más desigualdad a cambio de más oportunidades. Como el estatus de los padres es algo de lo que los hijos no son responsables, la persistencia indica que parte de la desigualdad es injusta según el igualitarismo sensible a la responsabilidad (Durlauf et al., 2022). Enlaza así con el compromiso entre eficiencia y equidad del cuerpo del tema.

21.A.12. Deflactación de la tarifa

La progresividad en frío, vista en el apartado 21.A.3, es un aumento del tipo medio sin que se haya producido una subida de la renta en términos reales. Para evitarla se deflactan los tramos de la tarifa, y la discusión es si deben ajustarse con la inflación o con el crecimiento de la renta (Almunia, 2022).

Deflactar tiene límites. Quienes están por debajo del mínimo exento no se benefician, porque no pagan el impuesto, y en ese tramo de la distribución cualquier rebaja del IRPF es regresiva. Además, deflactar cuando la inflación es muy alta estimula la demanda y puede causar más inflación, de modo que un IRPF sin deflactar actúa como estabilizador automático de la inflación. A ello se suma el coste fiscal, con un déficit estructural de aproximadamente el 4 % del PIB y una deuda del 118 % (Almunia, 2022).

En la práctica, Estados Unidos deflacta anualmente con la inflación desde los años ochenta, algo menos de la mitad de los países de la Unión Europea lo hacen de forma automática, entre ellos Francia y los países nórdicos, y Alemania cada dos años. La conclusión del autor es que deflactar es necesario a largo plazo, pero que a corto plazo se puede hacer de varias maneras y hay que elegir bien el momento (Almunia, 2022).

Bibliografía

Albi, E., J. M. González-Páramo, R. M. Urbanos y I. Zubiri (2018). *Economía Pública II: Teoría de la imposición. Sistema impositivo. Otros ingresos públicos. Economía Pública Internacional*. Barcelona: Ariel.

Allingham, M. G. y A. Sandmo (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics 1*(3-4), 323–338.

Almunia, M. (2022, octubre). Pros y contras de ajustar el IRPF con la inflación. Nada es Gratis.

Arenas, A. (2024, febrero). ¿Llegó la hora del impuesto sobre el valor del suelo? Nada es Gratis.

Atkinson, A. B. y J. E. Stiglitz (2015). *Lectures on Public Economics* (Updated ed.). Princeton: Princeton University Press.

Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy 82*(6), 1095–1117.

Barro, R. J. (1979). On the Determination of the Public Debt. *Journal of Political Economy 87*(5, Part 1), 940–971.

Benz, M. y B. S. Frey (2008). Being Independent is a Great Thing: Subjective Evaluations of Self-Employment and Hierarchy. *Economica 75*(298), 362–383.

Bhattacharjee, A., A. Pabst, L. da Silva Marioni, A. Aitken, E. Lisauskaite y J. van de Ven (2022, octubre). Fairer Taxation? The Case For and Against More Progressive Consumption Taxes in the UK. Informe, National Institute of Economic and Social Research.

Boadway, R. W. y N. Bruce (1984). *Welfare Economics*. Oxford: Basil Blackwell.

Borjas, G. J. (2016). *Labor Economics* (7 ed.). New York: McGraw-Hill Education.

Bowles, S. y Y. Park (2005). Emulation, Inequality, and Work Hours: Was Thorsten Veblen Right? *The Economic Journal 115*(507), F397–F412.

Brennan, G. (1973). Pareto Desirable Redistribution: The Non-Altruistic Dimension. *Public Choice 14*(1), 43–67.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo y R. Östling (2017). The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries. *American Economic Review 107*(12), 3917–3946.

 © 2026 Manuel Cabado García

Corlett, W. J. y D. C. Hague (1953). Complementarity and the Excess Burden of Taxation. *The Review of Economic Studies 21*(1), 21–30.

Diamond, P. A. y J. A. Mirrlees (1971, junio). Optimal Taxation and Public Production II: Tax Rules. *American Economic Review 61*(3), 261–278.

Domar, E. D. y R. A. Musgrave (1944, mayo). Proportional Income Taxation and Risk-Taking. *The Quarterly Journal of Economics 58*(3), 388–422.

Durlauf, S. N., A. Kourtellos y C. M. Tan (2022, agosto). The Great Gatsby Curve. *Annual Review of Economics 14*, 571–605.

Fox, J. (2017, noviembre). Why Economists Love Property Taxes and You Don't. Bloomberg Opinion.

Gruber, J. (2016). *Public Finance and Public Policy* (5 ed.). New York: Worth Publishers.

Hall, R. E. y D. W. Jorgenson (1967). Tax Policy and Investment Behavior. *American Economic Review 57*(3), 391–414.

Hamilton, B. H. (2000). Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment. *Journal of Political Economy 108*(3), 604–631.

Hochman, H. M. y J. D. Rodgers (1969). Pareto Optimal Redistribution. *American Economic Review 59*(4), 542–557.

Holter, H. A., D. Krueger y S. Stepanchuk (2019, noviembre). How Do Tax Progressivity and Household Heterogeneity Affect Laffer Curves? *Quantitative Economics 10*(4), 1317–1356.

Imbens, G. W., D. B. Rubin y B. I. Sacerdote (2001). Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players. *American Economic Review 91*(4), 778–794.

Joint Research Centre (s.f.). EUROMOD: the tax-benefit microsimulation model for the European Union. Comisión Europea.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen y E. Saez (2011, mayo). Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–692.

Klevmarken, A. (2022). Modeling Behavioural Response in EUROMOD. *International Journal of Microsimulation 15*(1), 89–96.

Moore, M. S. (2009, octubre). This Land Is Your Land. Pacific Standard.

Musgrave, R. A. y T. Thin (1948). Income Tax Progression, 1929-48. *Journal of Political Economy 56*(6), 498–514.

Oates, W. E. y R. M. Schwab (1997, marzo). The Impact of Urban Land Taxation: The Pittsburgh Experience. *National Tax Journal 50*(1), 1–21.

Pascual, M. y J. M. Sarabia (2003). La Medición de la Equidad en la Implementación de los Sistemas Impositivos. Documento de Trabajo 21/03, Instituto de Estudios Fiscales.

Pauly, M. V. (1973, febrero). Income Redistribution as a Local Public Good. *Journal of Public Economics 2*(1), 35–58.

Ramsey, F. P. (1927, marzo). A Contribution to the Theory of Taxation. *The Economic Journal 37*(145), 47–61.

Rosen, H. S. y T. Gayer (2010). *Public Finance* (9 ed.). New York: McGraw-Hill/Irwin.

Slemrod, J. y S. Yitzhaki (2002). Tax Avoidance, Evasion, and Administration. En A. J. Auerbach y M. Feldstein (eds.), *Handbook of Public Economics*, vol. 3, cap. 22, pp. 1423–1470. Amsterdam: Elsevier.

Smith, N. (2017, octubre). Faster Growth Begins With a Land Tax in U.S. Cities. Bloomberg Opinion.

Stern, S. (2004). Do Scientists Pay to Be Scientists? *Management Science 50*(6), 835–853.

Stiglitz, J. E. y J. K. Rosengard (2015). *Economics of the Public Sector* (4 ed.). New York: W. W. Norton & Company.

Sutherland, H. y F. Figari (2013). EUROMOD: the European Union tax-benefit microsimulation model. *International Journal of Microsimulation 6*(1), 4–26.

Tobin, J. (1958, febrero). Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies 25*(2), 65–86.

Trabandt, M. y H. Uhlig (2011, mayo). The Laffer Curve Revisited. *Journal of Monetary Economics 58*(4), 305–327.

U.S. Office of Management and Budget (2026). Federal Receipts as Percent of Gross Domestic Product (FYFRGDA188S) y Federal Surplus or Deficit as Percent of Gross Domestic Product (FYFSGDA188S). FRED, Federal Reserve Bank of St. Louis.

 © 2026 Manuel Cabado García

Tema 22

La teoría del federalismo fiscal: la asignación óptima de las funciones entre los distintos niveles de gobierno para alcanzar el bienestar máximo. La financiación de las haciendas territoriales: impuestos, transferencias y deuda.

22.1. Introducción ⏳ 1'

La Economía del Sector Público estudia la intervención del sector público en una economía de mercado fundamentalmente a través de sus programas de ingresos y gastos públicos, con el objetivo de cumplir sus funciones asignativa, distributiva y de estabilización y crecimiento.

No obstante, el sector público no toma todas sus decisiones de manera centralizada, sino que, generalmente, en todos los países existe un gobierno central que diseña la política exterior o de defensa y un gobierno de carácter local o regional que tiene asignada la capacidad de decisión en algunas cuestiones públicas en lo que afecta exclusivamente a los residentes en su jurisdicción.

Además, en algunos países, los denominados fiscalmente federales, existe un conjunto de gobiernos intermedios de carácter regional interpuestos entre el gobierno central y los gobiernos locales, como las CCAA en España, los Länder en Alemania, los Estados en EEUU o los Cantones en Suiza.

Concretamente, según el Índice de Autoridad Regional[1], España se sitúa entre los países más descentralizados del mundo, junto con Alemania, Bosnia y Herzegovina e India.

La Teoría del Federalismo Fiscal es la parte de la Economía del Sector Público que se encarga de analizar cómo se adoptan las decisiones descentralizadas, tratando de resolver cuestiones como:

- ¿Cuál es la justificación económica de la descentralización?
- ¿Cómo se deben repartir las competencias?
- ¿Cuántos niveles de gobierno deben existir?
- ¿Cómo se deben financiar los niveles de gobierno?

Por consiguiente, el objetivo del presente tema es responder a las cuestiones anteriores a través del siguiente esquema:

1. Justificación económica de la descentralización, desde un punto de vista de eficiencia, equidad y estabilización.
2. Niveles de gobierno y competencias.
3. Financiación, mediante impuestos propios, transferencias y emisión de deuda.

22.2. Justificación económica de la descentralización ⏳ 7'

Comenzando por la justificación económica, la descentralización se justifica en la medida en que el reparto del poder decisorio entre gobierno central y gobiernos territoriales permita alcanzar mejor los objetivos de eficiencia[2], equidad y estabilización.

Eficiencia

Partiendo de la base de que a los gobiernos territoriales les interesa el bienestar de sus ciudadanos, la descentralización por motivos de eficiencia está muy ligada a la provisión de bienes públicos locales, es decir, aquellos cuyos beneficios solo son disfrutados por aquellos individuos que vivan cerca de donde se provean. Ejemplos de ellos son el alumbrado público, el servicio de bomberos o una presa, pero no la defensa o las relaciones exteriores, que benefician a todos los residentes del país.[3]

[1] Ver 22.A.5 (Índice de Autoridad Regional).

[2] En el modelo básico NOEM de Galí y Monacelli (para una pequeña economía abierta), la función estabilizadora de la economía doméstica es también la que maximiza el BE (ambas funciones se cumplen simultáneamente).

[3] En sentido amplio, el federalismo fiscal aplica el mismo análisis a servicios que prestan los gobiernos territoriales, como la sanidad o la educación, aunque son bienes privados provistos públicamente, porque son rivales y se puede excluir de ellos (Stiglitz y Rosengard, 2015).

 © 2026 Manuel Cabado García

Existe una serie de motivos por los cuales se justifica la descentralización:

- Según Oates (1972), la descentralización permite solventar mejor el problema de información imperfecta y asimétrica pues los gobiernos territoriales, al estar más cerca de los residentes en una jurisdicción, conocen mejor sus preferencias. En consecuencia, el Gobierno central tenderá a proveer un nivel uniforme de bienes públicos en todos los territorios, lo que puede resultar ineficiente si estos tienen preferencias diferentes.

  Por ejemplo, es más que probable que los habitantes de Andalucía tengan unas preferencias más intensas por el gasto público en agricultura que los habitantes de Madrid.
- Teorema de descentralización: esto lleva a enunciar el Teorema de Descentralización de Oates, que sostiene que el nivel de bienestar alcanzado cuando cada jurisdicción escoge su nivel de consumo público de forma independiente será siempre igual o superior (y, casi siempre, superior) al que resultaría de un nivel uniforme de consumo, siempre que no existan ahorros de coste por provisión centralizada ni externalidades interjurisdiccionales.

Solución: descentralización fiscal Si se permite que cada jurisdicción decida de manera autónoma sobre su nivel de bien público, podrá adaptarlo a las preferencias de sus residentes y acercarse a la condición de Samuelson de provisión eficiente de bienes públicos, según la cual un bien público debe proveerse hasta que la valoración social, es decir, la suma de las utilidades marginales de los residentes, $\sum UMg_i$, iguale al coste marginal social, CMg_S[4]:

$$\sum UMg_i = CMgS \tag{22.1}$$

por ejemplo, una provisión uniforme del bien público, G^*, puede resultar ineficiente por exceso para el territorio A y por defecto para el territorio B.

Problema de revelación de preferencias Aun con descentralización fiscal, puede ocurrir que no se cumpla la condición de Samuelson si el Gobierno territorial no es capaz de agregar las preferencias de los individuos de su jurisdicción (es decir, obtener $\sum UMg_i$). Esto puede ocurrir si hay problemas de revelación y de agregación de preferencias:

- Problema de revelación de preferencias. Los individuos, sabedores de que sus preferencias reveladas determinarán la renta que el sector público les detraerá, pueden preferir revelar una utilidad menor y actuar como free-riders, por lo que se produciría infraproducción del bien público.

[4]Nótese que la condición de Samuelson se plantea como $\sum UMg_i = CMg_S$ porque se ilustra en un contexto de equilibrio parcial (como hace Bowen). Formulada en un contexto de equilibrio general, la condición de Samuelson sería $\sum RMS_i^{G,X} = RMT^{G,X}$ (es decir, las valoraciones del bien público G en términos del bien privado X).

- Problema de agregación de preferencias. Aun cuando los individuos estén dispuestos a revelar sus auténticas preferencias por el bien público, ¿cómo agrega el Gobierno territorial las preferencias de los distintos individuos para obtener $\sum UMg_i$?

Una posible solución a los dos problemas anteriores sería permitir el voto, de manera que el nivel de bien/gasto público sea decidido por mayoría simple. El problema del sistema de mayoría simple es que no asegura el cumplimiento de la condición de eficiencia de Samuelson, pues el nivel de gasto público vendrá determinado por las preferencias del votante mediano.

Para evitar los problemas de revelación y agregación y, además, asegurar la eficiencia de la condición de Samuelson, la descentralización fiscal y la regla del voto deben ir acompañadas de la libre movilidad de individuos entre las jurisdicciones. Es lo que se conoce como voto con los pies de Tiebout (1956).

Sistema de voto con los pies

- Cada hacienda territorial decide su combinación de bienes públicos locales e impuestos de acuerdo a las preferencias de los ciudadanos de su jurisdicción, esto es, mediante voto con mayoría simple (es decir, elecciones). Pero ello supone que el bien se proveerá conforme a las preferencias del votante mediano, como se ha advertido, no necesariamente eficiente.
- Si a la descentralización fiscal y al voto se añade movilidad entre jurisdicciones (y sus costes son bajos), los individuos descontentos con la política fiscal de su territorio se desplazarán a la jurisdicción que ofrezca una combinación fiscal más cercana a sus preferencias personales (es decir, votarán "con los pies").
- Conforme se acentúan los movimientos de personas entre jurisdicciones, las preferencias dentro de cada jurisdicción se vuelven más similares, hasta el punto de que todos los votantes de una jurisdicción acaban siendo homogéneos y se convierten todos ellos en el mismo "votante mediano". Por tanto, llegará un punto en que el nivel de bien/gasto público de cada jurisdicción será eficiente, pues, al venir determinado por el votante mediano y ser todos los ciudadanos de dicho territorio iguales al votante mediano, la asignación será eficiente.

Así, habrá jurisdicciones que ofrezcan muchas ventajas en términos de bienes públicos locales a familias con hijos (parques, columpios, etc.), mientras que otras estarán muy centradas en las necesidades de la tercera edad (accesibilidad, servicios sociales, etc.).

Críticas al sistema de voto con los pies

1. Coste de movilidad. La movilidad entre jurisdicciones es costosa (por ejemplo, cambio de casa, mudanza, hacer nuevas amistades, etc.). Puede ser factible a nivel municipal pero no a nivel país/entre países.

 © 2026 Manuel Cabado García

2. Congestión. La emigración en masa a una jurisdicción puede ser ineficiente si produce externalidades negativas (por ejemplo, congestión de los bienes públicos si son impuros –como las carreteras o los parques–)[5].
3. Otros factores. La decisión de un individuo de vivir en una determinada jurisdicción no solo se basa en sus preferencias sobre bienes públicos e impuestos, sino también en otras variables (incluso más determinantes) como la familia, las amistades, las costumbres, etc.
4. La existencia de un número limitado de jurisdicciones, que impide cubrir todo el abanico de preferencias de los agentes[6].
5. Evidencia empírica limitada. El modelo, aunque sugerente, solo se ajusta parcialmente a la evidencia empírica, que es limitada y mixta.

Restricciones políticas La descentralización se puede justificar también porque pueden existir restricciones políticas por parte del gobierno central para proveer distintas cantidades de bienes en distintos territorios. Así, los gobiernos centrales se pueden enfrentar a distintas restricciones éticas o jurídicas que impedirían la provisión de niveles eficientes en el sentido de Pareto del bien público en cada uno de los territorios.

A pesar de esto, no es raro encontrarse con programas de provisión de bienes públicos a cargo de un gobierno central que establecen diferentes niveles de provisión en función del territorio.

Conclusión El federalismo fiscal conduce a ganancias de eficiencia asignativa.
No obstante, hay que recordar que la eficiencia no se manifiesta únicamente en su forma asignativa (es decir, distribución de recursos de la mejor manera posible), sino también en su vertiente productiva (es decir, menores costes de producción posibles): una excesiva descentralización de ciertos servicios públicos puede conducir a una pérdida de economías de escala que produzca pérdidas importantes de eficiencia productiva (por ejemplo, aeropuertos).

Críticas y recomendaciones Más recientemente esta corriente de opinión se ha visto debilitada y contrarrestada por argumentos pro centralización. Las teorías que demuestran

[5]Por ejemplo, la emigración de una jurisdicción A a otra B genera, por un lado, un aumento de las bases fiscales en la localidad de destino, pero también puede ocasionar una mayor congestión en los bienes y servicios públicos (y al contrario ocurrirá en la localidad de origen). Como, en general, será difícil internalizar los efectos externos causados por los movimientos migratorios, el resultado final puede ser ineficiente. Por otra parte, las decisiones de los gobiernos subcentrales en ocasiones afectan a otros gobiernos como, por ejemplo, en la instalación de una industria contaminante en un área limítrofe, y en tanto esos efectos no se manifiesten en el mecanismo de precios, habrá ineficiencia.

[6]Solo con diez bienes públicos locales y tres niveles de provisión (alto-medio-bajo), serían necesarias 3^{10} jurisdicciones para cubrir todo el abanico de posibles preferencias por la provisión pública. Obviamente, en la realidad, tanto el número de jurisdicciones como la diferenciación de ofertas públicas locales es mucho menor.

las ventajas de la descentralización van acompañadas de que no suelen verificarse en la práctica, lo que ha estimulado la realización de trabajos empíricos. Esta revisión de la teoría, que incorpora la teoría de la elección colectiva (incentivos de políticos y burócratas, problemas de información) y abandona la idea del planificador benevolente, se conoce como federalismo fiscal de segunda generación. Además, la toma en consideración de otros aspectos de la economía política ha venido a poner ciertos frenos a las recomendaciones indiscriminadas de descentralización:

- Las contribuciones de Rodden (2003) han sido importantes en este sentido. Subrayan los riesgos de indisciplina fiscal que van ligados a una creciente descentralización. Además, vienen a demostrar que la supuesta reducción de gasto público asociada a la descentralización de las tareas de gobierno solo se cumple cuando los gobiernos subcentrales tienen una notable dosis de autonomía tributaria, y se financian en gran medida con impuestos propios sobre los que poseen una amplia capacidad normativa.
- También son reseñables las aportaciones de Tanzi (1996) o Huther y Shah (1998) en el sentido de subrayar las desventajas en un marco de un desarrollo institucional débil o la apertura de mayores posibilidades de cultivar la corrupción en el nivel local.

La mayor o menor eficiencia en la prestación descentralizada de bienes públicos se relaciona casi siempre con el carácter, local o nacional, de dichos bienes. Así, la recomendación consiste en que la provisión de bienes públicos de carácter local se haga por el nivel descentralizado mientras que la correspondiente a los bienes nacionales se provea a un nivel central. En la práctica, no obstante, existen frecuentes interacciones entre los distintos niveles de gobierno. O dicho de otra manera, es habitual reconocer las ganancias de eficiencia en la prestación de determinados bienes y servicios por parte del nivel descentralizado, pero con ciertas limitaciones, lo que justifica la existencia de regulaciones por parte del nivel central que dotan de cierta uniformidad a la prestación de los bienes y servicios[7].

Equidad

Al igual que ocurre con la provisión de los bienes públicos locales, podría también descentralizarse la función de redistribución, de forma que cada jurisdicción decidiera unilateralmente cuánta redistribución desea. Esto, sin embargo, plantea distintos problemas:

1. En las haciendas territoriales cuyos individuos tengan, por lo general, una elevada renta, las preferencias de sus ciudadanos por la redistribución serán, en principio,

[7] Ver 22.A.1 (otros desarrollos de la teoría del federalismo fiscal) y 22.A.2 (hechos estilizados sobre la descentralización).

 © 2026 Manuel Cabado García

menores, por lo que las políticas de equidad serán escasas. Esto hará que los individuos más pobres emigren a otras jurisdicciones donde las preferencias por la redistribución sean mayores, es decir, jurisdicciones generalmente más pobres, por lo que el coste de la redistribución se traslada a las jurisdicciones con menos recursos.

2. Al mismo tiempo, los individuos de elevada renta de las jurisdicciones más pobres se desplazarán a los territorios que menos gastan en redistribución, lo que a su vez disminuirá la capacidad redistributiva de las jurisdicciones más pobres. Por lo tanto, la descentralización de la redistribución puede hacer que las distintas jurisdicciones tengan distinta capacidad redistributiva.
   Así, las regiones con individuos más pobres tendrían menor capacidad para llevar a cabo la redistribución, con lo que persistiría la desigualdad entre los residentes de distintas jurisdicciones. Dado que la concepción de equidad es de ámbito global, esto obligaría al gobierno central a intervenir para reducir la desigualdad entre jurisdicciones.

Resultado El federalismo fiscal no promueve la equidad, por lo que es preferible que el Gobierno central conserve poderes fiscales para asegurar una redistribución entre territorios. Es necesario el papel del Estado para imponer coactivamente la redistribución.

Problema Aunque es cierto que la redistribución ejecutada a nivel central es más efectiva que una descentralizada, también es cierto que el Gobierno central puede no ejercer dicha labor de manera correcta por tener incentivos perversos. En efecto, podría infravalorar a las regiones más despobladas y pobres, que son poco importantes en las elecciones, y priorizar el gasto en las regiones más pobladas y ricas.
Además, ha habido otras aportaciones que subrayan que también existe campo para la acción de los gobiernos territoriales con fines redistributivos.

En este sentido, la aportación de Pauly (1973) puede considerarse que inició este tipo de recomendaciones:

- Partía de la base de que la redistribución puede considerarse como un bien público local, del cual puede haber distintas demandas en los distintos territorios.
- Los individuos de mayor nivel de renta obtienen una mayor utilidad de los aumentos de renta que obtengan los individuos más desfavorecidos en su territorio. Es por ello que los ricos de una región preferirán una política redistributiva con acento local.
- Dado que los individuos de mayor nivel de renta obtienen utilidad de las transferencias realizadas a personas de renta más baja, la tarea de redistribución se considera óptima en el sentido de Pareto.
- Esta idea de que los objetivos de redistribución pueden ser alcanzados de manera más eficiente a través de políticas ejercidas por los gobiernos subcentrales se fortalece

bajo supuestos de movilidad imperfecta y de eficiencia en la toma de decisiones colectivas.

Conclusión En definitiva, en la Teoría del Federalismo Fiscal no existe consenso respecto a la deseabilidad de la descentralización por motivos de equidad, aunque la visión predominante (Musgrave, 1959; Oates, 1972) es, como se ha visto, que la función redistributiva debe ser llevada a cabo de forma centralizada[8].

Estabilización

Por último, la estabilización es una función de carácter macroeconómico, cuyo objetivo es atenuar las fluctuaciones cíclicas de la economía en torno a su senda potencial de crecimiento, tratando de lograr la estabilidad de precios y empleo.

Existe un amplio consenso en que el objetivo de estabilización debe estar centralizado por al menos tres razones:

1. Los gobiernos territoriales, al no emitir su propia moneda carecen de política monetaria, elemento importante en el diseño de las políticas de estabilización y que debe coordinarse con la política fiscal.
   Además, en el caso de la política monetaria, la descentralización generaría una mayor expansión monetaria, fomentando la inflación. La política monetaria y la emisión de moneda están siempre en el núcleo duro de las funciones de cualquier configuración estatal, se reservan al ámbito central.
2. La mayoría de variables macroeconómicas –nivel de precios, tipo de cambio, tipo de interés e incluso, en función de la movilidad, la tasa de paro– se determinan en el ámbito nacional o supranacional, por lo que las jurisdicciones son incapaces de influir. Esto convierte a las jurisdicciones de un sistema federal en lo que en Macroeconomía se denomina "economías pequeñas abiertas".
3. Una política fiscal aplicada localmente se difuminaría rápidamente entre jurisdicciones, afectando relativamente menos a quien la aplica y generando efectos externos sobre el resto de jurisdicciones[9].

Resultado El federalismo fiscal no facilita la tarea estabilizadora de la política económica, por lo que es preferible que el Gobierno central conserve poderes fiscales para asegurar un correcto control del ciclo.

[8] Ver 22.A.4 (ordinalidad y redistribución).

[9] Cuando el gasto público de una jurisdicción afecta al bienestar de los residentes en otras regiones, se dice que existe un efecto externo interjurisdiccional.

 © 2026 Manuel Cabado García

Aplicación en España

Se ha visto que el federalismo fiscal se justifica por motivos de eficiencia, pero no por motivos de equidad ni de estabilización. De ahí que los países descentralizados hayan tratado de buscar un equilibrio en el respeto de estos principios.

- En España, por ejemplo, se ha desarrollado un sistema fiscal fuertemente descentralizado para asegurar la eficiencia, es decir, que las decisiones políticas se corresponden con las preferencias de los ciudadanos de cada territorio (por ejemplo, la mayor descentralización se ha producido en los bienes públicos más sensibles para los ciudadanos, como son sanidad, educación y servicios sociales).
- Pero, al mismo tiempo, se han articulado fondos para asegurar la equidad entre los distintos territorios (por ejemplo, Fondo de Compensación Interterritorial, Fondo de Garantía, Fondos de Convergencia) y se han conservado las potestades fiscales más importantes para garantizarse poderes estabilizadores del ciclo (por ejemplo, el Estado conserva la potestad legislativa sobre los impuestos más importantes –IVA e IS– y sobre la mayor parte del IRPF, en el que las CCAA solo tienen margen sobre la tarifa autonómica, habiéndose desprendido únicamente de tributos con escasa capacidad recaudadora –Impuesto sobre el Patrimonio, sobre Sucesiones y Donaciones, sobre el Juego, etc.–)[10].

22.3. Tamaño de la jurisdicción y el reparto de competencias ⏳ 4'30"

Hasta ahora se ha justificado por qué es beneficiosa la descentralización fiscal: la función asignativa debe descentralizarse, pero no las funciones redistributiva y estabilizadora.

Esto plantea dos cuestiones: cuántos niveles de gobierno debe haber y qué gastos de asignación deben corresponder a cada nivel. Para responder a la primera se analiza el tamaño óptimo de las áreas descentralizadas, es decir, el tamaño de la población a cubrir por cada uno de los gobiernos territoriales. Esta cuestión puede responderse a través de una analogía introducida por Buchanan: la Teoría de los Clubes (Buchanan, 1965).

Hipótesis Esta teoría se basa en los siguientes supuestos:

1. Jurisdicción = club. Buchanan señala que una jurisdicción se puede asimilar a un club: se trata de un grupo de individuos que se unen para lograr un fin común: la provisión de un bien público local. Determinar el tamaño óptimo de la jurisdicción es por tanto equivalente a decidir el número óptimo de socios de un club.

[10] Ver 22.A.3 (retos de la financiación autonómica y local en España).

2. Mismas preferencias.
3. Misma renta. En estas circunstancias (si se cumple que las preferencias y la renta son iguales), el bien público puede financiarse mediante un impuesto de suma fija (de capitación[11]) dado que este será eficiente y equitativo.

Desarrollo teórico La idea básica es que al ampliar el tamaño de la jurisdicción (es decir, al ampliar el ámbito de aplicación de un bien o servicio público a un individuo más), se producen dos efectos contrapuestos para los beneficiarios:

1. (+) Beneficio marginal. La financiación del bien se reparte entre más individuos, por lo que el coste de financiación per cápita (es decir, lo que paga cada beneficiario) es menor. Se supone que la provisión del bien público está sujeta a fuertes costes fijos, por lo que un aumento de los usuarios aumenta los costes variables, pero disminuye los costes unitarios (es decir, disminuye el pago que cada beneficiario tiene que hacer para financiarlo).
2. (–) Coste marginal. Al atender a más usuarios, el bien público se congestiona, de modo que para mantener el mismo nivel de servicio hay que aumentar la dotación y el coste total de provisión crece.

Por consiguiente, el ámbito óptimo de aplicación (es decir, número de individuos, o nivel de descentralización) será aquel en el que el BMg se iguale al CMg, es decir, aquel en el que el ahorro de impuestos por incorporar a un individuo adicional iguala al coste marginal de proveerle el servicio.[12]

Gráficamente, en un plano de coste por habitante, C/N, y tamaño de la jurisdicción, N, como el de la Figura 22.1 se puede representar la curva de costes medios, CMe, que tendrá forma de U, es decir, presenta rendimientos a escala mixtos. Aquí, $C(N, G)$ es el coste de proveer el nivel G del bien público a una jurisdicción de N habitantes, de modo que el coste medio es $CMe = C/N$ y el coste marginal, $CMg = \partial C/\partial N$, es el de incorporar un habitante más. Por lo tanto, la teoría de costes ayudará a resolver este problema.

Así, el tamaño óptimo de la jurisdicción vendrá determinado por el mínimo de la curva de costes medios que, además, coincidirá con el punto de intersección con la curva de costes marginales, CMg, de provisión del bien público, es decir, en N^*:

- Si $N < N^*$ (como en N_1) el bien público está infrautilizado, y aumentar N tiene un CMg bajo y permite reducir el CMe (existen rendimientos a escala en la provisión del bien público).

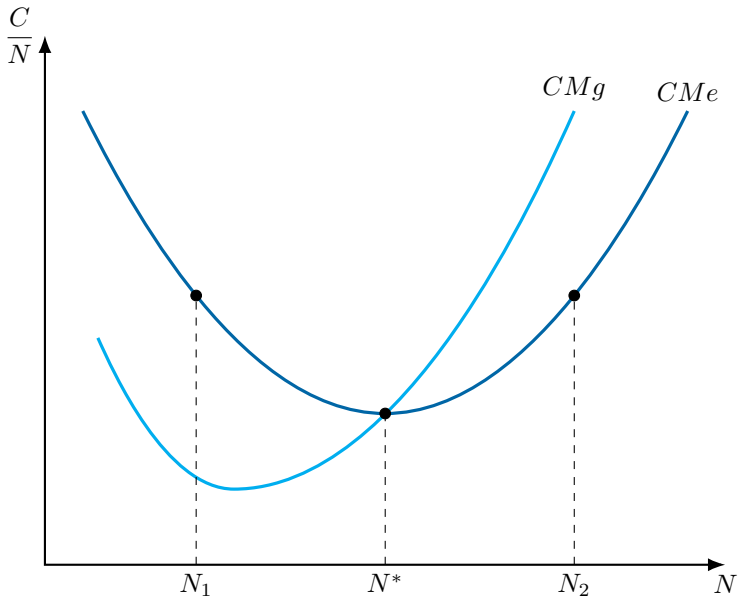
[11] En rigor, se trata de un impuesto según el beneficio que cada individuo obtiene del bien público, que equivale a un impuesto de capitación porque todos los individuos son idénticos (Albi et al., 2017).

[12] La intuición es que, mientras el bien está infrautilizado, cada nuevo residente cuesta menos de lo que paga, que es el coste medio, y su entrada rebaja el impuesto de los demás. Cuando la congestión hace que cueste más de lo que paga, conviene dejar de ampliar la jurisdicción.

 © 2026 Manuel Cabado García

- Si $N > N^*$ (como en N_2) el bien público se va congestionando, y para mantener la utilidad de los individuos constante serán necesarios desembolsos crecientes (por ejemplo, habrá que hacer más carreteras y aumentarán los costes de gestión).

Figura 22.1: Tamaño óptimo de la jurisdicción



Desarrollo analítico Se supone que la función de utilidad del agente representativo depende del consumo del bien privado, X, del consumo del bien público, G, y del número de individuos del club, N (es decir, del tamaño de la jurisdicción)

$$\max_{\{X,G,N\}} U(X,G,N) \tag{22.2}$$

sujeto a dos restricciones:

$$P_X \cdot X + \tau = Y \tag{22.3}$$

$$\tau \cdot N = CF + P_G \cdot G \tag{22.4}$$

donde τ es el impuesto de suma fija a pagar por cada beneficiario del bien o servicio público, CF son los costes fijos de provisión, P_X es el precio del bien privado (igual a su CMg) y P_G es el coste variable de cada unidad de bien público (CMg).

Resolviendo el problema[13], se obtiene que el tamaño de la región descentralizada (N) será mayor cuanto:

[13] El impuesto de suma fija será:

$$\tau = \frac{CF}{N} + \frac{P_G \cdot G}{N}$$

- Mayor sea la demanda del bien o servicio público, u'_G (utilidad marginal del bien público).
- Menor sea, en valor absoluto, la pérdida de utilidad por congestión que causa un miembro adicional, u'_N, que es negativa.
- Mayor sea el coste fijo de proveer el bien, CF.

Aplicación práctica El objetivo del modelo anterior es determinar un "club" o jurisdicción de individuos con las mismas preferencias sobre un bien o servicio público. Es decir, que habría tantos clubs/jurisdicciones como preferencias para cada bien público. Esto implica que un mismo individuo podría estar sujeto a diferentes jurisdicciones para cada bien público (por ejemplo, un individuo pertenecería a un club para la prestación del servicio de bomberos y a otro diferente para el servicio de recogida de basuras). Por lo tanto:

- Resulta obvio que la descentralización que es óptima en teoría resulta inviable en la práctica, porque los meros costes de gestión y coordinación de dicho sistema serían sin duda superiores a las ganancias de eficiencia[14].
- Así, la teoría de la descentralización óptima debe entenderse no como una teoría a aplicar a rajatabla, sino como una teoría que guíe la acción política.
- Además, tampoco puede olvidarse la existencia de razones históricas y políticas en el establecimiento de los límites jurisdiccionales.

Reparto de competencias Respecto a qué gastos asignativos corresponden a cada nivel, la descentralización se rige también por los dos principios anteriores: cada gobierno territorial debe decidir sobre las cuestiones que afectan exclusivamente a sus residentes, siempre que esto agote los rendimientos a escala en la provisión (Albi et al., 2017). De ello se derivan tres implicaciones:

- Incluso gobiernos de igual naturaleza, como los municipios, deberían tener competencias de gasto diferentes en función de su tamaño.
- Cuando la provisión individualizada suponga pérdidas considerables de rendimientos a escala, deberían crearse uniones entre jurisdicciones para proveer conjuntamente esos bienes y servicios.
- Si, por una descentralización imperfecta, el gasto local genera efectos externos, el gobierno central debería intervenir para internalizarlos, por ejemplo cofinanciando el gasto mediante una transferencia condicionada.

[14] De esta forma, el diseño de una descentralización óptima plantea un intercambio o trade-off entre las ganancias de eficiencia de crear múltiples niveles de gobierno local y los costes de gestión a que esto da lugar. El punto de equilibrio de este trade-off no se puede determinar en términos teóricos solamente, aunque es claro que elementos como la superficie del país y las diferencias económicas y culturales entre las diferentes zonas jugarán un papel importante a la hora de diseñar el grado "óptimo" de descentralización.

 © 2026 Manuel Cabado García

22.4. Financiación de las haciendas territoriales ⏳ 7'

Ya se han visto la justificación de la descentralización fiscal y el tamaño óptimo que deben adoptar las haciendas territoriales para proveer bienes y servicios públicos, además de otras cuestiones relacionadas. Ahora se analizan las diferentes formas en que dichas haciendas territoriales pueden financiar los bienes públicos.

Existen diferentes opciones, que pueden coexistir, como en el caso de España:

1. Impuestos propios.
2. Transferencias de niveles superiores de gobierno.
3. Emisión de deuda.

Impuestos propios

Comenzando por los impuestos, la financiación de las haciendas territoriales a través de tributos propios tiene como objetivo principal la corresponsabilidad fiscal, esto es, que las haciendas financien una parte sustantiva de su gasto mediante impuestos propios y visibles para los contribuyentes, pues así:

1. Los Gobiernos territoriales se vuelven más responsables en sus decisiones de gasto, tratando de ajustarlo a sus ingresos, claramente definidos. Es decir, permite limitar las expansiones ineficientes de gasto que se podrían dar si la financiación se realizase únicamente mediante transferencias.
2. Los ciudadanos de dicho territorio perciben mejor el verdadero coste de los bienes y servicios públicos que reciben, pues los financian en buena parte con sus propios impuestos.
3. Se rebajan las tensiones entre el Gobierno central y el territorial sobre las transferencias necesarias, pues con tributos propios las haciendas disponen, en teoría, de herramientas para cubrir sus gastos.

En el caso de la descentralización de los tributos, y a diferencia de lo que ocurre con el gasto, su objetivo principal no es la obtención de ganancias de eficiencia, sino corresponsabilizar al gobierno territorial en la obtención de los recursos. Así, la descentralización de los tributos se convierte en un ejercicio de compromiso entre:

- Conceder a los gobiernos territoriales discrecionalidad suficiente para que puedan determinar su nivel de ingresos.
- Limitar dicha discrecionalidad para no mermar la capacidad de redistribución y estabilización del gobierno central.

Tipos A su vez, la financiación mediante "tributos propios" puede adoptar diferentes formas de más a menos margen de actuación por parte de las haciendas territoriales[15]:

- Impuestos propios en sentido estricto, que son aquellos diseñados y gestionados (es decir, liquidación, recaudación, e inspección) por las haciendas territoriales. En el caso español, esto es posible siempre que los impuestos que se establezcan no invadan hechos imponibles ya gravados por el Estado, es decir, se trata básicamente de impuestos relacionados con el medioambiente.
- Impuestos cedidos, que son aquellos creados y regulados por el Estado, pero cuya recaudación y buena parte de la gestión se ceden a las haciendas territoriales, junto con una capacidad normativa limitada. En el caso español, se trata básicamente de impuestos sobre la riqueza (Patrimonio, Sucesiones y Donaciones, etc.).
- Impuestos compartidos, que son aquellos en los que el Estado diseña los elementos básicos del impuesto y los gestiona, pero deja cierta discrecionalidad de diseño a las haciendas territoriales. En el caso español, se trata del IRPF, para el que se prevé una tarifa autonómica. La tarifa autonómica del IRPF se corresponde con una participación autonómica del 50 % en el rendimiento de este impuesto.
- Impuestos participados (cesión parcial de la recaudación), que son aquellos regulados, gestionados y recaudados por el Estado, pero en los que este transfiere parte de la recaudación a las haciendas territoriales. En el caso español, se trata sobre todo de impuestos indirectos armonizados a nivel europeo (IVA, impuestos especiales, etc.).
    - 50 % de la recaudación líquida por IVA (según un índice de consumo).
    - 58 % de la recaudación líquida por impuestos especiales.
    - 100 % de la recaudación líquida por el Impuesto sobre la Electricidad.

Recargos Además de los tipos anteriores, en ocasiones no se descentraliza por completo un tributo, sino que se establece un sistema de recargos, en el que el gobierno central fija las normas para el cálculo de la base imponible y el esquema liquidatorio del tributo y los gobiernos territoriales establecen un recargo (aumento de los tipos) o una bonificación (disminución de tipos).

Competencia fiscal Otorgar autonomía a las haciendas territoriales en el diseño de los impuestos permite cumplir mejor con el principio de corresponsabilidad fiscal, pero puede dar lugar a competencia fiscal, es decir, al intento de cada hacienda territorial de atraer a su jurisdicción, con impuestos más bajos, empresas y factores productivos instalados en

[15] Esta clasificación es didáctica. En la terminología de la LOFCA, todos estos impuestos son tributos cedidos, total o parcialmente, con o sin capacidad normativa autonómica: el IRPF se cede al 50 % con capacidad normativa sobre la tarifa, y el IVA y los impuestos especiales se ceden parcialmente sin ella.

 © 2026 Manuel Cabado García

otras (política de tipo beggar-thy-neighbour). Si todas hacen lo mismo, ninguna consigue atraer factores, porque la base imponible de cada hacienda acaba siendo la misma, pero los tipos quedan por debajo de los necesarios para financiar el nivel adecuado de gasto público. Se trata de una situación del dilema del prisionero.

Cuadro 22.1: Dilema del prisionero en competencia fiscal

	C	NC
C	$2, 2$	$0, 3\checkmark$
NC	$3\checkmark, 0$	$1\checkmark, 1\checkmark$

Los gobiernos territoriales pueden cooperar (C,C) y establecer un tributo razonable que les permita una recaudación aceptable, o pueden no cooperar, reduciendo el tipo impositivo con el fin de aumentar la base imponible del tributo y con ello su recaudación. Como se ve, se trata de un juego de tipo dilema del prisionero que lleva a un equilibrio de Nash ineficiente (NC, NC). Ello hace necesario establecer mecanismos de cooperación regional que eviten un race to the bottom.

Eficiencia El principio de eficiencia establece que el tributo no debe generar movimientos de factores o actividades económicas. De acuerdo con este principio, un impuesto será descentralizable si una diferenciación de tipos entre jurisdicciones no genera movimientos significativos.
De ahí que los impuestos que en mayor o menor medida se suelen descentralizar sean aquellos que gravan actividades poco móviles: fallecimiento (Sucesiones), patrimonio, bienes inmuebles (IBI), rendimientos del trabajo (IRPF, ineficiencia reducida dada la limitada movilidad de los trabajadores), consumo (IVA[16]), etc., mientras que se ha mantenido la competencia estatal plena sobre otros impuestos que gravan actividades más móviles (por ejemplo, impuesto de sociedades, cuya descentralización generaría ineficiencias importantes por ser el capital muy móvil).

Transferencias

En ocasiones, la descentralización de tributos puede no ser suficiente para financiar la totalidad de los gastos de los gobiernos territoriales, por lo que se complementan sus ingresos mediante transferencias, que son una cantidad de dinero que un gobierno (donante) entrega a otro (beneficiario) para que financie sus gastos.

[16] El IVA: genera distorsiones si la imposición es:

- En destino: los consumidores se desplazan al lugar con menor gravamen.
- En origen: distorsiona las decisiones de producción/localización.

Ventajas Las transferencias funcionan como un mecanismo de equidad y redistribución. Las transferencias de nivelación cubren la diferencia entre la capacidad fiscal y la necesidad de gasto de cada jurisdicción. Corrigen dos tipos de desequilibrio: el vertical, entre niveles de gobierno (por ejemplo, el Estado conserva las fuentes tributarias pero las CCAA asumen competencias de gasto), y el horizontal, entre jurisdicciones del mismo nivel. Corregir el vertical no elimina el horizontal, que exige llevar recursos de las regiones ricas a las pobres. Además, también existen motivos de eficiencia que justifican la existencia de transferencias entre niveles de gobierno, en particular, transferencias de naturaleza compensatoria.

Desventajas

- No contribuyen a la corresponsabilidad fiscal, porque una parte del gasto la pagan contribuyentes de otros territorios, lo que puede llevar a un gasto público excesivo. El riesgo es mayor si el gobierno territorial espera que, cuando gaste o se endeude en exceso, el gobierno central acuda a rescatarlo, situación que se conoce como restricción presupuestaria blanda[17].
- Puede crear tensiones entre las haciendas territoriales y el Estado.

Tipos Las transferencias que realiza el Gobierno a las haciendas territoriales pueden ser de tres tipos[18]:

1. Transferencias incondicionales, que son aquellas que se conceden sin condiciones sobre su destino (por ejemplo, el Gobierno central transfiere 1.000 M€ al territorio, y éste es libre para decidir su destino).
2. Transferencias condicionales no compensatorias, que son aquellas que se conceden condicionadas a un uso particular, pero cuya cuantía no depende del nivel de gasto del beneficiario en dicho concepto. El donante transfiere una cuantía fija independientemente del gasto del beneficiario (por ejemplo, el Gobierno central transfiere 1.000 M€ al territorio para que éste lo dedique a infraestructuras).
3. Transferencias condicionales compensatorias (cofinanciación), que son aquellas que se conceden condicionadas a un uso particular, y, además, su cuantía depende del nivel de gasto de la hacienda territorial en dicho concepto (por ejemplo, el Gobierno central aporta un euro por cada euro que aporta el territorio, es decir, financia el 50 % de su gasto en infraestructuras).

Transferencias incondicionales

Las transferencias incondicionales son aquellas que se conceden sin condiciones sobre su destino, por lo que no hay una restricción del uso de los fondos. Para analizarlas, se partirá

[17] Ver 22.A.1 (la restricción presupuestaria blanda).
[18] Ver 22.A.6 (ejemplos de transferencias en España y en la UE).

 © 2026 Manuel Cabado García

de los siguientes supuestos:

- Existe una jurisdicción que provee un único bien público local, G, no sujeto a congestión y de coste unitario de producción, p_G, definido en términos de renta privada a la que se renuncia para financiar el gasto del bien público.
- La jurisdicción tiene una renta Y_0, que se supone fija, de modo que la renta disponible, Y_d, tras el pago de impuestos para financiar el bien será: $Y_d = Y_0 - p_G \cdot G$. Por lo tanto, en ausencia de transferencias, los impuestos serán suficientes para financiar G.
- La decisión entre usos públicos y privados se realiza a través de una función de utilidad que depende de la renta disponible y el gasto público[19]: $u\left(Y_d, G\right)$.

De esta forma, el nivel óptimo de provisión del bien público se obtendrá resolviendo el siguiente problema de maximización:

$$\text{máx} \quad u\left(Y_d, G\right) \tag{22.5}$$

sujeto a la restricción presupuestaria de los agentes de esta jurisdicción, según la cual la renta es igual a la renta disponible (que determina sus posibilidades de consumo privado) más la cuantía destinada al consumo del bien público local:

$$Y_0 = Y_d + p_G \cdot G \tag{22.6}$$

que ante la transferencia incondicional pasaría a ser

$$Y_0 + T = Y_d + p_G \cdot G \tag{22.7}$$

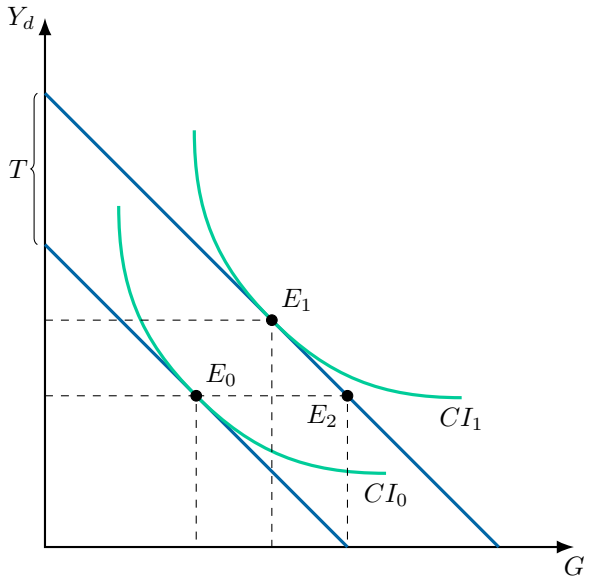
Gráficamente, en un plano Y_d y G como el de la Figura 22.2, se representa la curva de indiferencia CI_0 y la restricción, $\dfrac{Y_0}{p_G}$, siendo la solución óptima el punto de tangencia E_0 donde se alcanza el nivel óptimo de provisión del bien público. En este contexto:

- Si se establece una transferencia incondicional, T, para financiar el gasto público, se genera un efecto renta puro ya que no afecta a los precios relativos $\dfrac{p_G}{p_x}$, donde p_x es el precio del bien privado x, de modo que se incrementa el consumo del bien público y del bien privado si ambos bienes son normales. Además, el aumento del gasto público será inferior a la cuantía de la transferencia, ya que aumenta también la demanda del bien privado.
- Por consiguiente, el nuevo equilibrio sería E_1, donde se produce la tangencia entre

[19] Esta función de utilidad se puede interpretar bien como que los individuos son idénticos y es la del agente representativo, bien como las preferencias del votante mediano o bien como una función de bienestar social

la curva de indiferencia CI_1 y la restricción, $\dfrac{Y_0+T}{p_G}$, y no E_2 donde la transferencia se destinaría plenamente al gasto público[20].

Figura 22.2: Transferencia incondicional



Transferencias condicionales no compensatorias

Por otro lado, las transferencias condicionales no compensatorias son aquellas que se conceden condicionadas a un uso particular, pero cuya cuantía no depende del nivel de gasto de la hacienda territorial en dicho concepto.

Así, la restricción de condicionalidad podría interpretarse como el requisito de que el gasto final en el bien para el que se realiza la transferencia aumente, al menos, en la cuantía de lo transferido.

Gráficamente, como se observa en la Figura 22.3 se produce de nuevo un desplazamiento paralelo hacia afuera de la restricción presupuestaria de la hacienda territorial, con la diferencia de que ahora la nueva restricción se define por tramos: no son alcanzables aquellos puntos en los que el nuevo gasto público sea inferior a la transferencia, ya que esta se concede para dedicarse en su integridad a dicho fin. Con esto en cuenta el nuevo equilibrio sería el punto E_2, donde el gasto público aumentaría en la cuantía de la transferencia.

[20] No obstante, existen dos casos en los que la transferencia se traduce íntegramente en gasto público:

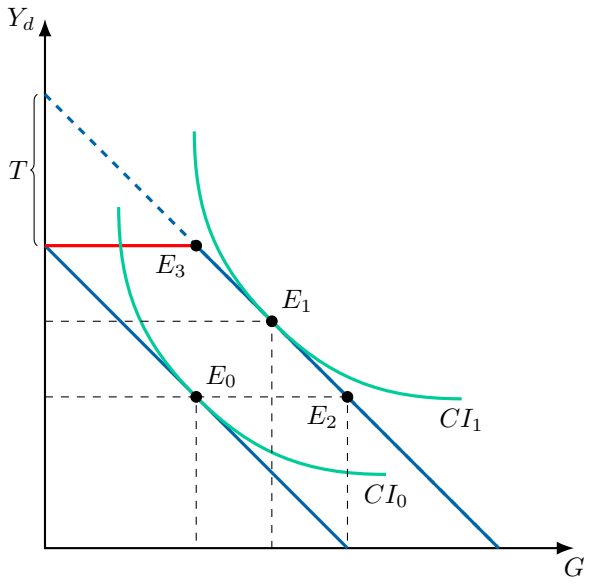
1. Bien privado neutro. Si el bien privado es neutro (por contraposición a "bien" y "mal"), entonces los individuos no lo valoran nada, y la curva de indiferencia será vertical. El equilibrio será una solución de esquina.
2. Bien privado independiente de la renta. Si el bien privado es independiente de la renta (por contraposición a "bien normal" y "bien inferior"), entonces la demanda de dicho bien no depende de su renta.

 © 2026 Manuel Cabado García

Sin embargo:

- Para que esto fuera viable el donante debería saber cuánto habría gastado el beneficiario en gasto público de no haberse realizado la transferencia. El problema es que en la mayoría de los casos relevantes no dispondrá de esta información[21].
- Siendo esto así, la restricción de condicionalidad solo puede interpretarse de forma que el gasto final, G_1, en el bien para el que se realiza la transferencia sea mayor que la cuantía de lo transferido, $p_G \cdot G_1 \geq T$.
- Esto se traduce en una extensión del tramo presupuestario por la línea de puntos, donde los puntos a la izquierda del punto E_3 no serían accesibles dada la restricción de condicionalidad.
- El nuevo equilibrio dependerá de las preferencias comunitarias por la provisión pública.[22]

Figura 22.3: Transferencia condicional no compensatoria



[21] Por ejemplo, un gobierno que realiza una transferencia para mejorar las escuelas públicas difícilmente sabrá cuánto se habría gastado de sus recursos propios el gobierno beneficiario de no haberse realizado la transferencia.

[22] Como se acaba de señalar, el nuevo equilibrio tras la transferencia condicional no compensatoria dependerá de cómo sean las preferencias comunitarias por la provisión pública, aunque, al igual que en el caso de la transferencia incondicional, el aumento del gasto será inferior al importe de la transferencia. Si la elasticidad renta del consumo privado es muy alta y la elasticidad renta de consumo de bienes públicos es muy baja, entonces el aumento del gasto es mayor que si la transferencia fuese incondicional: con una incondicional se situaría a la izquierda de E_3, pero con una condicional debe situarse al menos en E_3. Por el contrario, si las elasticidades renta del bien público y del bien privado son parecidas, el efecto será equivalente al de una transferencia incondicional, situándose en los dos casos en un punto como E_1, a la derecha de E_3.

Transferencias condicionales compensatorias o cofinanciación

Por su parte, las transferencias condicionales compensatorias o cofinanciación son aquellas que se conceden condicionadas a un uso particular, pero cuya cuantía depende del nivel de gasto de la hacienda territorial en dicho concepto. La transferencia compensatoria supone que el gobierno central aporta τ unidades monetarias por cada unidad monetaria que aporta el beneficiario, por lo que el coste unitario de provisión del bien público para el gobierno subcentral se reduce de p_G a $\dfrac{p_G}{1+\tau}$.

Ahora la restricción presupuestaria sería:

$$Y_d + \frac{p_G}{1+\tau} G = Y_0 \tag{22.8}$$

Gráficamente, en un plano Y_d y G como el de la Figura 22.4, la transferencia:

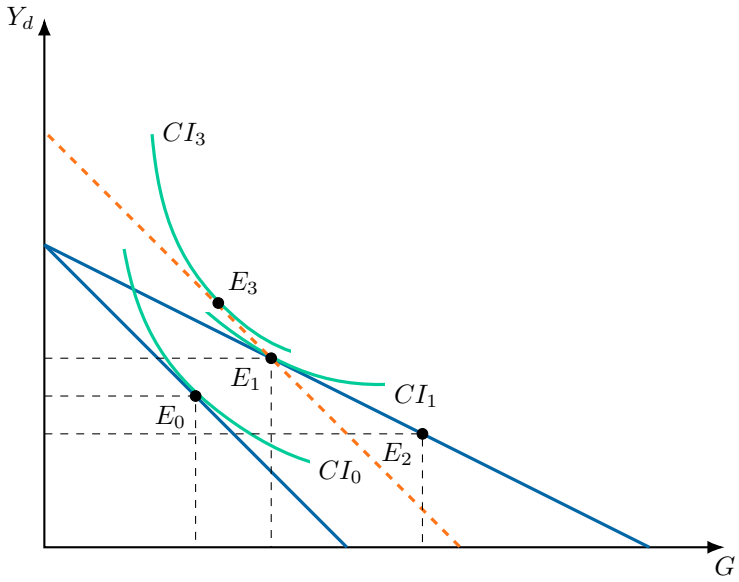
- Modifica la pendiente de la restricción, pasándose de un equilibrio inicial, E_0, a un equilibrio final en el punto E_1.
- También podría darse en el equilibrio E_2, pero la evidencia empírica señala que es más factible el punto E_1[23].
- Por otro lado, si se compara con una transferencia incondicional de idéntica cuantía, es decir, si se traza una línea paralela a la restricción inicial que corte al equilibrio E_1, se accede a una curva de indiferencia más alejada del origen, CI_3, en el punto E_3. Del punto E_3 se puede observar que tiene dos propiedades importantes en relación a E_1:
  - El nivel de provisión del bien público es menor en E_3 que en E_1, ya que la transferencia condicionada compensatoria de E_1 genera, además del efecto renta, un efecto sustitución. Este efecto sustitución, al abaratar relativamente el bien público, refuerza la expansión del gasto público por encima de la que se produciría con una transferencia incondicional.
    Por lo tanto, a igual coste para el donante, el aumento del gasto público de la hacienda territorial es mayor con cofinanciación que con una transferencia incondicional o condicional no compensatoria.
  - El nivel de utilidad de la jurisdicción es mayor en E_3, ya que la transferencia condicionada compensatoria de E_1 limita la elección al obligar a consumir más del bien público, reduciendo de esta forma la posibilidad de alcanzar niveles de utilidad mayores.

[23] En el punto E_2 la renta privada se vería reducida y el aumento del bien público estaría por encima de la transferencia. Esto dependerá de la elasticidad-precio de la demanda del bien público: si es menor a 1 entonces el equilibrio será E_1 (el gasto público aumenta menos que la transferencia) pero si es mayor a 1 entonces el equilibrio será E_2 (aumenta más).

No obstante, pese a que ambos resultados son factibles la evidencia empírica señala que el equilibrio E_1 es el más factible, siendo la elasticidad inferior a 1.

 © 2026 Manuel Cabado García

Figura 22.4: Transferencia condicional compensatoria



Conclusiones de la financiación territorial mediante transferencias

Las conclusiones respecto a las transferencias son claras:

- Si el Gobierno central quiere estimular el gasto público territorial, lo aconsejable es una transferencia condicional compensatoria (esto es, la cofinanciación).
  Si el gasto público local o regional produce efectos externos positivos sobre otras jurisdicciones, los gobiernos locales estarían infravalorando los beneficios de ese bien público por lo que lo infraproveen. En estos casos, una transferencia condicionada compensatoria, al estimular la provisión del bien, servirá para internalizar el efecto externo.
- Si el Gobierno central quiere maximizar el bienestar territorial, lo aconsejable es una transferencia incondicional.
  Las transferencias incondicionales están generalmente ligadas a objetivos de equidad interregional, ya que las regiones más pobres tendrían unos niveles de provisión de bienes públicos más bajos. Por consiguiente, las transferencias incondicionales permitirían garantizar unas provisiones mínimas aceptables.

Sistema español

El sistema español de financiación autonómica se basa, además de en los impuestos propios vistos anteriormente, en transferencias del Estado (que se nutren, a su vez, de parte de la recaudación de los "impuestos propios" de las CCAA). Dichas transferencias se articulan a través de fondos.

El más importante de ellos es el "Fondo de Garantía de Servicios Públicos Funda-

mentales", que busca la equidad, es decir, que todas las Comunidades puedan ofrecer un nivel similar en la prestación de los servicios públicos fundamentales: sanidad, educación y servicios sociales.

Sin embargo, se trata de un fondo que no condiciona sus transferencias al gasto en dichos bienes públicos. Es decir, se trata de transferencias no condicionales, por lo que, conforme al análisis teórico visto antes, puede decirse que, en el fondo, el sistema español busca maximizar el bienestar territorial más que estimular el gasto público.

A continuación se harán dos precisiones con respecto al análisis anterior de los efectos en el gasto público de la financiación vía transferencias:

1. Bien público como bien inferior.
2. Efecto adherencia o efecto papel matamoscas.

Bien público inferior

Los análisis anteriores se basaban en dos bienes, el privado y el público, que eran normales, de manera que cuando aumentaba la renta de los individuos, aumentaba el consumo de los bienes. Sin embargo, merece la pena plantearse la posibilidad de un bien o servicio público inferior, es decir, aquel cuya demanda por parte de la comunidad disminuye cuando aumenta su renta:

- Pensar en un bien público inferior no es un mero ejercicio teórico sin ninguna aplicabilidad práctica. La sanidad y la educación públicas, servicios provistos públicamente aunque no sean bienes públicos puros, pueden comportarse como bienes inferiores: cuando no disponen de renta suficiente, muchos ciudadanos acuden a los hospitales públicos y envían a sus hijos a colegios públicos.
- Pero cuando su renta aumenta lo suficiente, abandonan la sanidad y educación públicas y comienzan a utilizar la sanidad privada (para ahorrarse las esperas) y la educación privada (para proporcionar a sus hijos mayores oportunidades), por lo que la demanda del bien público se reduce con la renta.
- Pues bien, en el caso de bienes públicos inferiores, una transferencia del Gobierno central a las haciendas territoriales podría disminuir el consumo de bien público (disminuirá seguro en el caso de una transferencia incondicional –efecto renta negativo–, y podría disminuir en el caso de cofinanciación –efecto renta negativo pero efecto sustitución positivo–).

Implicaciones de política económica Por lo tanto, a la hora de diseñar sus objetivos de gasto público, el Gobierno central no solo debe analizar el tipo de transferencia que hará, sino también la naturaleza del bien público cuyo consumo quiere estimular.

 © 2026 Manuel Cabado García

El efecto adherencia o efecto papel matamoscas

Pese a que el modelo anterior permite analizar las transferencias sobre el gasto algunas de sus implicaciones no son consistentes con la evidencia empírica:

- Por ejemplo, según el modelo una transferencia incondicionada de T_0 es equivalente a un incremento de la renta privada de los residentes de la jurisdicción en la misma cuantía, provocando el mismo desplazamiento en la recta presupuestaria como se puede observar en (22.7).
- Por consiguiente, las transferencias incondicionales dan lugar al mismo aumento del gasto público que los aumentos de renta privada de la misma magnitud.
- No obstante, los estudios empíricos no corroboran dicha predicción de la teoría, sino que sugieren que una transferencia incondicional de una unidad monetaria da lugar a un aumento del gasto público mayor que un aumento de la misma magnitud en la renta privada de los residentes de la jurisdicción[24]. Este fenómeno se denomina efecto adherencia o efecto papel matamoscas –flypaper effect–, porque refleja que el dinero tiende a quedarse "pegado" en el primer sector, público o privado, en el que "impacta".

Albi et al. (2017) señalan que al efecto papel matamoscas se le han dado diversas explicaciones. La más frecuente consiste en suponer que las elecciones públicas se adoptan por votación mayoritaria, pero que los votantes desconocen la cuantía de las transferencias incondicionadas que reciben sus gobernantes.

Además, al igual que en la teoría de la burocracia de Niskanen (1971), el objetivo de los gobernantes es maximizar el tamaño de su presupuesto. Por ello, con el fin de aumentar su presupuesto y mantener los votos, los gobernantes pueden introducir una ilusión fiscal cuando reciben una transferencia T_0, consistente en hacer creer a los votantes que el coste de provisión del bien G es menor que su coste real. Por ejemplo, podrían reducir los impuestos en la cuantía de la transferencia T_0 y hacer creer a los votantes que el coste unitario de provisión del gasto público G es $(p_G \cdot G - T_0)/G$ en vez de p_G. Ante esta reducción del precio observado –reflejada en menores impuestos–, los ciudadanos votarán aumentos en el nivel de gasto.

Implicaciones de política económica Por lo tanto, a la hora de diseñar sus objetivos de gasto público, el Gobierno central debe considerar la posibilidad de que una transferencia incondicional a las haciendas territoriales no sea equivalente a una transferencia de

[24] La intuición es que el dinero que llega al ayuntamiento tiende a quedarse en el presupuesto en lugar de volver a los vecinos en forma de rebajas de impuestos. Según las estimaciones que recogen Rosen y Gayer (2010), cada unidad monetaria de transferencia genera unos 40 céntimos de gasto público, mientras que una unidad adicional de renta privada solo genera unos 10.

Albi et al. (2017) señalan los trabajos de Gramlich (1977) y Bastida et al. (2009), los cuales recogen evidencia empírica del efecto para el conjunto de los municipios españoles.

recursos al sector privado, por lo que, si su objetivo es maximizar el bienestar de los ciudadanos de esa jurisdicción, podría ser conveniente transferir dichos recursos directamente a los ciudadanos (aunque en la práctica, es muy difícil para el Gobierno central articular transferencias directas a los ciudadanos de un determinado territorio –por ejemplo, una bajada de impuestos afecta a todos los territorios–)[25].

Emisión de deuda

Por último, el gobierno central puede permitir que las haciendas territoriales emitan deuda para financiarse, lo que se justifica por dos motivos: financiar inversiones de largo alcance repartiendo su coste entre generaciones (por equidad intergeneracional, como recoge la regla de oro de las finanzas públicas), y servir de alternativa a los impuestos cuando la capacidad tributaria propia es limitada. No obstante, su uso debe ser limitado por dos motivos:

1. Erosión de la función estabilizadora del Gobierno central: la deuda no solo es un medio de financiación, sino también un instrumento fundamental de política económica por sus efectos sobre el tipo de interés, la demanda agregada, etc., pudiéndose poner en riesgo la política de estabilización del gobierno central[26].
2. Riesgo moral y free-rider: existen incentivos a endeudarse indefinidamente, llevando a las haciendas territoriales a aplicar políticas fiscales expansivas financiadas con deuda, pues existe un problema de riesgo moral y de free-rider[27]. La hacienda territorial tendrá incentivos para llevar políticas fiscales expansivas, pues se beneficiará del aumento de la demanda agregada de dicha política, pero sin tener que hacer frente a los costes asociados a esta[28].

En efecto, las regiones saben que la insostenibilidad de su política fiscal no se va a traducir directamente en un aumento del riesgo región, ya que con la unión monetaria que representa el país el riesgo región de cada jurisdicción se difumina en el riesgo moneda,

[25] Este "efecto pegamento" ha sido considerado por el Gobierno español a la hora de diseñar algunos fondos de apoyo a la financiación autonómica. Así, por ejemplo, el Fondo para la Financiación de los Pagos a Proveedores, puesto en marcha por el Gobierno central en 2012, que tiene como objetivo liquidar las deudas comerciales de las administraciones territoriales. Para ello, el Gobierno central paga directamente a la empresa a la que adeuda la hacienda territorial, por lo que se asegura que los recursos movilizados (unos 41.800 M€ entre comunidades autónomas y entidades locales) llegan en su integridad al tejido empresarial. La administración territorial devuelve después esos recursos mediante un préstamo sujeto a un plan de ajuste.

[26] Concretamente, en España durante la crisis inmobiliaria se vio cómo la deuda autonómica fue un determinante fundamental del déficit del país.

[27] Ver 22.A.1 (la restricción presupuestaria blanda).

[28] También en España se dio dicha situación, donde durante el boom inmobiliario ayuntamientos y CCAA se endeudaron en exceso aprovechándose del bajo riesgo país. Sin embargo, tras el estallido de la crisis en 2008 los mercados de deuda se fragmentaron y las primas de riesgo se dispararon, obligando al Estado Central a un "rescate" encubierto a través del Fondo de Liquidez Autonómica, FLA, sentando un precedente que puede acentuar en el futuro los problemas de riesgo moral.

 © 2026 Manuel Cabado García

común para todo el país (es decir, se considera que los bonos regionales, emitidos en la moneda común, son perfectamente sustitutivos).

Además, aun cuando eventualmente los mercados de deuda se fragmentasen y aumentara la prima de riesgo de la región, la hacienda territorial podría tener la esperanza de recibir ayuda financiera del Gobierno central ante el miedo de éste a que:

- Se produzca un contagio y se extienda el pánico financiero al resto del país, y
- Los gobiernos regionales culpen al Gobierno central de forma creíble al no disponer las haciendas territoriales de una capacidad tributaria propia plena.

Esto es precisamente lo que ha ocurrido en España: muchas CCAA y ayuntamientos se endeudaron en exceso aprovechándose del bajo riesgo país de España. Y cuando los mercados de deuda se fragmentaron y las primas de riesgo de algunas CCAA y Ayuntamientos aumentaron, se produjo un "rescate" encubierto del Estado a través del Fondo de Liquidez Autonómica (planteándose incluso la posibilidad de quitas), sentando un precedente que puede acentuar en el futuro los problemas de riesgo moral.

Soluciones Por lo tanto, las soluciones deben pasar por:

1. Limitar el endeudamiento de las haciendas territoriales. En España a través de la Ley Orgánica 2/2012 de Estabilidad Presupuestaria y Sostenibilidad financiera, el artículo 13 fija el límite de deuda de cada comunidad autónoma en el 13 % de su PIB regional.
2. Sujetar la emisión de deuda territorial a la autorización estatal. Artículo 135.3 de la Constitución, que exige autorización por ley para emitir deuda pública o contraer crédito, y artículo 14.3 de la LOFCA: "Para concertar operaciones de crédito en el extranjero y para la emisión de deuda o cualquier otra apelación de crédito público, las Comunidades Autónomas precisarán autorización del Estado"
3. Compromiso de no rescate de las haciendas territoriales por parte del Estado.

22.5. Conclusiones ⏳ 30"

Como conclusión, el federalismo fiscal no es un simple reparto de competencias, sino que obliga a reflexionar sobre su aplicación práctica en España.

Por un lado, la descentralización a las CCAA ha creado una situación única donde el poder de gasto y el control de los ingresos se superponen. Sin embargo, la cesión simultánea de soberanía fiscal a la UE, junto al control del BCE sobre la política monetaria, genera una tensión que dificulta la capacidad del Gobierno central para estabilizar la economía.

Por otro lado, como señala un informe de FEDEA de 2024, la existencia de una "restricción presupuestaria blanda" y la falta de transparencia en el sistema actual plantean

un interrogante crucial: ¿cómo se puede garantizar la equidad y la responsabilidad fiscal si los incentivos políticos no están alineados con la sostenibilidad financiera? La teoría es clara, pero la realidad demuestra que los sistemas de financiación deben ser lo suficientemente flexibles y transparentes para adaptarse a un entorno de gobernanza multinivel sin sacrificar la coherencia ni la estabilidad.

 © 2026 Manuel Cabado García

Apéndice

22.A.1. Otros desarrollos de la teoría del federalismo fiscal

A continuación se presentan otros desarrollos no vistos en el tema

Transparencia y rendición de cuentas

Otras aportaciones de la Teoría del Federalismo Fiscal se centran en la relación de agencia entre políticos (agentes) y electorado (principal). En este caso, el electorado debe tener un criterio para valorar la actuación del gobierno y decidir sobre su reelección.

La descentralización favorece además la competencia por comparación entre gobiernos (argumento a favor de la descentralización).[29] Al poder comparar los resultados de distintos gobiernos, los votantes obtienen información adicional que les permite valorar mejor la actuación del suyo de cara a la reelección, lo que aumenta los incentivos de los gobiernos a seguir el interés público (Boadway y Shah, 2009).

La provisión descentralizada abre la puerta a la competencia entre gobiernos locales. La teoría de la Hacienda pública subraya a menudo la carencia de incentivos para actuar de manera eficiente en la labor de gobierno, lo que sucede en mayor medida en el marco de un único gobierno que ejerce todas las competencias. Las ventajas de la competencia entre gobiernos son mayores cuanto mayor sea la capacidad de los individuos de irse a vivir a otra región. Breton (1996) situó esta competencia entre gobiernos como la principal razón para descentralizar el sector público. Brennan y Buchanan (1980) llevaron este argumento más lejos. Si el sector público no es un planificador benevolente, sino un Leviatán que trata de maximizar su tamaño, la competencia entre gobiernos y la movilidad de los contribuyentes limitan su capacidad para recaudar y, con ella, el nivel de gasto (Albi et al., 2017).

Asociado a las ventajas de competencia, se encuentran los incentivos a incrementar la innovación en el campo del suministro de los bienes locales. No siempre está claro cuál es el método más eficiente de prestar un determinado servicio público. Ante la duda, una buena receta es la de dejar ensayar métodos diferenciados en diferentes regiones, lo que permitirá comparar los resultados e ir convergiendo hacia la utilización del método más eficiente.

Reparto del riesgo entre las distintas jurisdicciones

Se trata de una cuestión relacionada con la función aseguradora de la política fiscal. Si las diferentes jurisdicciones se enfrentan a shocks aleatorios imperfectamente correlacionados y existe aversión al riesgo queda justificada la contratación de un seguro. Para ello existen varias soluciones:

[29] El término procede de la literatura de *yardstick competition*, formalizada por Salmon (1987) y Besley y Case (1995).

- El Gobierno central asume la función aseguradora, en cuyo caso aparecen problemas de riesgo moral.
- Constitución de rainy day funds: los gobiernos regionales constituyen provisiones en periodo de bonanza para hacer frente a tiempos difíciles. Es una solución menos habitual, aunque se utiliza en algunos países (por ejemplo, Finlandia).
- Creación de un sistema de reparto de riesgo entre jurisdicciones, basado en la diversificación del riesgo o risk pooling: cada región aporta a un fondo común que compensa a la que sufre un shock, de modo que el coste se reparte entre todas. Funciona mejor cuanto menos correlacionados estén los shocks, y si son independientes la varianza del coste que soporta cada una de las N regiones se divide por N.

Relación más estrecha entre costes y beneficios

El papel que juegan los precios en el mercado a la hora de estimular la eficiencia se puede trasladar a la prestación de servicios públicos: la eficiencia aumentará cuanto más estrecha sea la vinculación entre los servicios provistos y el pago correspondiente de impuestos. Esto es, cuando exista una estrecha correspondencia entre el pago de impuestos por parte de los ciudadanos y los beneficios que reciben en forma de servicios públicos. Parece claro que esta vinculación, o al menos la percepción de la misma, es mayor para los ciudadanos cuando la prestación es llevada a cabo por una administración local.

Ahora bien, esta adecuación entre costes y beneficios será mayor cuando el gasto correspondiente a la provisión de bienes públicos por parte de los gobiernos locales se haya financiado con recursos provenientes de impuestos financiados por los ciudadanos de ese territorio.

Si el gobierno local depende de transferencias del gobierno central para financiar la prestación, los incentivos se trasladan hacia la negociación entre gobiernos con el fin de obtener las máximas transferencias posibles. De esta forma se reducen los incentivos de los ciudadanos a controlar la acción de gobierno local a través del voto. Este razonamiento proporciona un motivo para hacer descansar la financiación de los gobiernos locales preferentemente en impuestos propios.

La restricción presupuestaria blanda

Se habla de restricción presupuestaria blanda cuando un gobierno subcentral espera que el gobierno central le preste ayuda financiera si gasta en exceso o se endeuda demasiado, de modo que no soporta todo el coste de sus decisiones. Es un problema de riesgo moral propio de los sistemas descentralizados (Boadway y Shah, 2009).

Surge porque los gobiernos subcentrales dependen de transferencias, que en parte se diseñan para que puedan cubrir sus necesidades de gasto, y porque prestan servicios esenciales como la sanidad o la educación. Si no pueden prestarlos con un mínimo de

 © 2026 Manuel Cabado García

calidad, el gobierno central se siente obligado a intervenir, a veces por mandato legal (en Alemania, por ejemplo, un municipio no puede quebrar). El incentivo a excederse es mayor cuanto más discrecionales son las transferencias. La capacidad de endeudarse añade una vía más, aunque el problema puede darse también sin deuda.

En España se manifiesta en la percepción de que parte de los costes de las CCAA puede trasladarse a la Administración Central mediante aportaciones adicionales de fondos o la condonación de la deuda acumulada (de la Fuente, 2024). El Fondo de Liquidez Autonómica es el ejemplo más reciente (ver "Emisión de deuda").

Anunciar que no habrá rescate no basta, porque ese compromiso es difícil de sostener. Las soluciones que se plantean son:

- Transferencias basadas en fórmulas objetivas y no en el déficit real de cada gobierno, que resultan más creíbles que las discrecionales.
- Que los gobiernos subcentrales se financien con deuda propia, sometida a la disciplina de los mercados (como en Canadá), lo que exige una autonomía real sobre su presupuesto.
- Cláusulas de no rescate y reglas fiscales. El Tratado de Maastricht prohíbe que el BCE y los bancos centrales nacionales financien a los gobiernos y que la UE o los demás Estados miembros asuman sus deudas (cláusula de no rescate), y el Pacto de Estabilidad refuerza esas disposiciones.

22.A.2. Hechos estilizados sobre la descentralización

A continuación se presentan algunos hechos estilizados sobre la descentralización.

La descentralización en el mundo A nivel mundial, existe una clara tendencia hacia la descentralización. Un ejemplo de ello es Latinoamérica, cuyo índice de centralización del gasto público pasó del 76 % en 1975 al 70 % en 1995. Además, parece haber una correlación positiva entre el nivel de descentralización y nivel de desarrollo de los países. En los países más desarrollados la centralización del gasto público suele ser sustancialmente inferior, rondando el 50 %. Los SP africanos suelen estar muy centralizados, con un grado de centralización superior al 80 %.

Nivel de descentralización por funciones del sector público Se pueden encontrar una serie de patrones en cuanto al grado de descentralización en la provisión del SP en la mayoría de países:

- Los servicios comunitarios y de vivienda suelen ser los más descentralizados, con un grado de descentralización del 71 %.

- Les siguen los servicios de educación y sanidad, con un grado de descentralización del 64 %.
- Los servicios menos descentralizados son los relacionados con la Seguridad Social y el Bienestar, con una media del 18 %.

Estos grados de descentralización son consistentes con el enfoque normativo, que establece que los servicios de distribución son alcanzados en mayor grado a nivel central.

Determinantes de la descentralización La literatura positiva de la descentralización ha encontrado varias fuerzas que promueven la descentralización:

- Oates (1972) encontró, a través de un análisis de sección cruzada a nivel nacional, que el tamaño del país y la renta per cápita son determinantes a la hora de explicar la descentralización.
- Panizza (1999), a partir de la estimación de un modelo teórico de un país lineal, encontró que la descentralización aumenta de nuevo con el tamaño del país y la renta per cápita, pero también con el nivel de democracia, la fragmentación etnolingüística y la diversidad cultural.
- Finalmente, la descentralización puede no responder a motivos de eficiencia, sino a motivos políticos: la descentralización se utiliza también para acallar deseos secesionistas. Italia (Lombardía) y Bélgica (Flandes) son buenos ejemplos de este tipo de descentralización.

22.A.3. Retos de la financiación autonómica y local en España

El apunte "Financiación autonómica y local" de Ángel de la Fuente (2024), publicado por Fedea en enero de 2024, recopila y resume los principales trabajos de Fedea sobre la materia.

Diagnóstico del sistema actual Se identifican tres grandes problemas en el sistema de financiación territorial:

1. Complejidad y falta de transparencia. El modelo es tan enrevesado que resulta incomprensible para el ciudadano medio. Las transferencias entre el Estado y las comunidades autónomas (CCAA) se canalizan con criterios variados y a veces contradictorios, lo que dificulta la rendición de cuentas.
2. Distribución desigual y arbitraria. El reparto de recursos entre CCAA no sigue criterios claros de equidad ni de capacidad fiscal. Sin contar las comunidades forales, y con datos de 2021, hay un abanico de unos 25 puntos porcentuales entre la mejor y la peor financiada por habitante ajustado. Las comunidades forales (País Vasco

 © 2026 Manuel Cabado García

y Navarra) reciben entre 50 y 100 puntos más que la mejor financiada del régimen común.

3. Déficit de responsabilidad fiscal. Las CCAA, sobre todo las de régimen común, no son plenamente responsables de sus decisiones de gasto, ya que parte de sus costes son asumidos por la Administración Central. Esto genera una "restricción presupuestaria blanda", que incentiva el gasto excesivo sin asumir el coste político de subir impuestos.

Propuestas de reforma Se propone avanzar hacia un sistema más simple, transparente, equitativo y responsable, mediante una reforma gradual y coordinada. Las principales recomendaciones son:

- Revisión de competencias y herramientas fiscales. Es necesario adaptar la división de competencias entre administraciones y revisar el régimen foral para que participe en los mecanismos de solidaridad interterritorial.
- Fondo complementario de nivelación. Como medida inmediata, se propone crear un fondo financiado por el Estado para reducir o eliminar la distancia que separa de la media del régimen común a las comunidades peor financiadas (Murcia, Valencia, Andalucía y Castilla-La Mancha). El coste estimado sería de unos 3.000 millones de euros.
- Mejoras técnicas. Se sugiere reformar el sistema de retenciones y entregas a cuenta para que reflejen más rápidamente las decisiones fiscales de las CCAA, mejorar el cálculo de la recaudación normativa y regular las escalas de referencia de los tributos cedidos.
- Mecanismos colectivos de responsabilidad fiscal. Para evitar desequilibrios horizontales, se plantea crear tramos autonómicos colegiados del IVA y los Impuestos Especiales, cuyos tipos serían fijados por las CCAA y repartidos según población ajustada. Esto permitiría aumentar los recursos sin beneficiar desproporcionadamente a unas regiones sobre otras.

Necesidad de pacto político Se subraya que dos cuestiones clave requieren consenso político:

- Nivel de servicios públicos deseado. Es necesario decidir cuánto se está dispuesto a pagar en impuestos para mantener o mejorar las prestaciones públicas.
- Grado de nivelación territorial. Se debe pactar qué parte de las diferencias fiscales entre territorios debe corregirse mediante transferencias. Aunque la nivelación debe ser alta para garantizar la cohesión, podría no llegar al 100 % si eso facilita el acuerdo con las comunidades más reacias a la redistribución.

Se advierte del riesgo de una deriva hacia un sistema de financiación "a la carta", negociado bilateralmente con algunas comunidades, lo que podría debilitar la capacidad del Estado para garantizar la cohesión social y territorial. Se reconoce que esta situación ya existe en el régimen foral, pero se insiste en que no debe extenderse a otras regiones.

22.A.4. Ordinalidad y redistribución

La ordinalidad y la nivelación son dos principios que se contraponen en el debate sobre la financiación autonómica.

Ordinalidad La ordinalidad es el principio según el cual la nivelación no debe alterar el orden de las comunidades autónomas por capacidad fiscal por habitante, de modo que una comunidad con más capacidad fiscal que otra antes de las transferencias debe seguir teniendo más financiación por habitante después, aunque la diferencia entre ellas se reduzca.[30] Sus defensores sostienen que incentiva la responsabilidad fiscal y la eficiencia: si una comunidad se esfuerza por mejorar su economía y recaudar más, se beneficia de ello, y se premia así la buena gestión de sus finanzas.

Nivelación La nivelación (o redistribución) es el proceso por el cual el sistema de financiación transfiere recursos de las CCAA con más capacidad fiscal a las que tienen menos, para que todas dispongan de una financiación similar por habitante ajustado, es decir, por unidad de necesidad de gasto, y puedan ofrecer servicios públicos básicos (como educación y sanidad) en condiciones de igualdad. Es fundamental para la cohesión social y territorial, porque el acceso a los servicios no depende entonces del lugar de residencia, y refuerza la solidaridad dentro del país. Sin embargo, una nivelación muy intensa puede desincentivar a las comunidades de renta alta a recaudar más, ya que gran parte de ese esfuerzo se destina a otras regiones, y puede generar riesgo moral en las receptoras, que sienten menos presión para gestionar eficientemente sus recursos o aumentar sus ingresos.

Críticas al principio de ordinalidad De la Torre y Fernández-Villaverde (2025) rechazan el principio de ordinalidad, tanto en su versión débil, la definida arriba, como en la fuerte, según la cual quien más aporta debe recibir más, y lo consideran una idea equivocada que se presenta como razonable y justa. Sus razones son las siguientes:

- Contradice la lógica redistributiva del Estado del bienestar, en el que cada persona contribuye según su capacidad y recibe según su necesidad. Por ejemplo, una comu-

[30] El principio figura en el artículo 206.5 del Estatuto de Autonomía de Cataluña de 2006, según el cual el Estado debe garantizar que los mecanismos de nivelación no alteren la posición de Cataluña en la ordenación de rentas per cápita anterior a la nivelación. La STC 31/2010 lo interpretó en el sentido de que la garantía solo opera si la alteración se debe exclusivamente a esos mecanismos.

 © 2026 Manuel Cabado García

nidad con más niños escolarizados debe recibir más financiación aunque eso altere el orden de renta entre comunidades.

- Exigir que las comunidades mantengan su posición relativa de renta tras impuestos y transferencias ignora las diferencias en necesidades de gasto, como la estructura demográfica. Un ejemplo ilustrativo es el de dos regiones que comparten un sistema común de pensiones y de sanidad: una más productiva y joven y otra menos productiva y más envejecida. Tras cotizaciones, impuestos y transferencias, la región menos productiva y más envejecida puede acabar con más recursos por habitante que la otra, lo que, en su opinión, muestra que el principio es absurdo. Según estos autores, regiones con ese perfil existen en España, como Asturias o Cantabria.
- Carece, según estos autores, de consenso académico y político.
- En su versión fuerte, un contribuyente de renta muy alta recibiría más del Estado que un jubilado sin ingresos.
- Reconocen como legítimo el argumento de que una redistribución excesiva desincentiva el esfuerzo, pero consideran que debe plantearse de forma abierta y no a través de la ordinalidad.
- Fernández-Villaverde resume su postura señalando que no es posible defender a la vez el Estado del bienestar y la ordinalidad territorial sin contradicción.

La tensión entre ambos principios Existe una tensión inherente entre la ordinalidad y la nivelación. Una nivelación total (que los recursos por habitante ajustado sean iguales en todas las comunidades para los mismos servicios) puede alterar el orden original de las comunidades. Por el contrario, respetar estrictamente la ordinalidad puede generar grandes desigualdades en la capacidad de gasto. El desafío en cualquier reforma es encontrar un equilibrio que promueva la solidaridad sin penalizar la eficiencia ni la buena gestión fiscal. En esta línea, Fedea considera que el grado de nivelación y el respeto de la ordenación original (la ordinalidad) ofrecen un margen de negociación que podría acomodar parcialmente las demandas de los territorios de renta alta dentro de un esquema razonable (de la Fuente, 2024).

22.A.5. El Índice de Autoridad Regional

El Índice de Autoridad Regional (RAI, por sus siglas en inglés) mide la autoridad política que ejercen las regiones o estados subnacionales dentro de un país. Lo ha desarrollado un equipo de académicos liderado por Liesbet Hooghe, Gary Marks y Arjan H. Schakel (Hooghe et al., 2016; Shair-Rosenfield et al., 2021). Su objetivo es ir más allá de la distinción tradicional entre Estados "federales" y "unitarios" y cuantificar de forma matizada el grado de descentralización en 96 países, con datos anuales de 1950 a 2018. Se considera

región cualquier nivel de gobierno situado entre el municipal y el nacional cuya población media supere los 150.000 habitantes.

Frente a los indicadores fiscales El índice nace como alternativa a las medidas fiscales habituales, como el peso del gasto o de los ingresos subnacionales, que tienen dos limitaciones. El volumen de gasto no indica quién decide: un gobierno regional puede gestionar mucho dinero con poca autonomía real. Además, dan una sola cifra para todos los niveles subnacionales y no recogen regímenes diferenciados, como los territorios forales del País Vasco y Navarra. Los datos lo confirman: el gasto subnacional explica solo en parte la variación del RAI ($R^2 = 0{,}39$) y el desequilibrio vertical, medido como transferencias sobre gasto subnacional, no guarda relación con él ($R^2 = 0{,}00$). Canadá, Suiza o Estados Unidos, muy descentralizados, tienen un desequilibrio comparable al de países centralizados como Bulgaria o Eslovaquia.

Medición de la autoridad regional El RAI evalúa la autoridad regional a través de diez dimensiones, que se agrupan en dos grandes categorías:

1. Autogobierno ("self-rule"). Mide la capacidad de una región para ejercer poder dentro de sus propias fronteras. Evalúa si la región tiene su propio parlamento y ejecutivo, el alcance de sus competencias políticas (como la sanidad o la educación), y su autonomía fiscal y de endeudamiento.
2. Participación en el gobierno central ("shared-rule"). Mide la influencia que una región ejerce en el gobierno central del país. Se analiza su participación en la aprobación de leyes nacionales, su control sobre el ejecutivo central, su poder en la distribución de los impuestos nacionales y su papel en las reformas constitucionales.

Cada dimensión puntúa en una escala corta (por ejemplo, de 0 a 3 la profundidad institucional y de 0 a 4 la autonomía fiscal), de modo que cada región o nivel regional obtiene hasta 18 puntos de autogobierno y 12 de gobierno compartido, un máximo de 30. La puntuación de un país suma la de todos sus niveles, ponderada por población, por lo que puede superar los 30 y favorece a los países con más niveles. El índice se actualiza cada año y registra más de 1.300 cambios entre 1950 y 2010.

Posición de España Con datos de 2018, España suma 35,6 puntos (22,6 de autogobierno y 13,0 de gobierno compartido)[31]. Solo la superan Alemania (37,7) y Bosnia y Herzegovina (36,3), y empata con India (35,6). Queda por delante de Bélgica (33,9), Estados Unidos (29,6), Canadá (27,8), Suiza (26,5), Italia (26,0), Francia (21,8) y Reino Unido (9,6), y es el país con mayor puntuación del mundo en gobierno compartido. Esta posición es notable porque España no es un Estado federal en el sentido clásico.

[31] Datos de RAI-Country y RAI-Region, versión 3.1 (abril de 2021), disponibles en la web de Gary Marks (UNC).

 © 2026 Manuel Cabado García

La trayectoria muestra la rapidez del proceso: el índice pasa de 3,4 en 1977 a 29,4 en 1983, cuando todas las comunidades autónomas ya tenían estatuto e instituciones propias, y se estabiliza en torno a 35 desde 2002. Los propios autores del índice describen esa etapa como una cascada de pactos territoriales iniciada por el País Vasco, Cataluña y Galicia, que el Estado extendió a todo el territorio con la fórmula del "café para todos" (Shair-Rosenfield et al., 2021).

La posición depende de cómo se agregue. Si se considera solo el nivel regional más autoritativo, que en España son las comunidades autónomas, la puntuación es de 23,7 sobre 30, la octava de 79 países, por detrás de Alemania (27,0), Bosnia y Herzegovina (26,7), Suiza (26,5), Bélgica (25,5), Australia (25,5), Argentina (25,0) y Estados Unidos (24,6). La diferencia se debe a que la cifra nacional suma también otros niveles, como las provincias, y las regiones con régimen diferenciado.

Perfil de las comunidades autónomas El perfil es desigual. Las comunidades puntúan al máximo en instituciones propias (3 sobre 3) y representación (4 sobre 4), y alto en competencias (3 sobre 4) y gobierno compartido, gracias a las conferencias sectoriales y a la Conferencia de Presidentes. En cambio, su autonomía fiscal es de 2,9 sobre 4, frente a unos 4 puntos en Estados Unidos, Canadá o Suiza (salvo el País Vasco y Navarra, que alcanzan el máximo por su régimen foral), y su autonomía de endeudamiento es de 1 sobre 3, por la autorización previa del Estado, frente a 3 en Estados Unidos, Canadá, Alemania o Suiza. La descentralización española es, por tanto, más de gasto y de participación que de ingresos y deuda, lo que ayuda a entender el déficit de corresponsabilidad fiscal y la restricción presupuestaria blanda.

22.A.6. Ejemplos de transferencias

A continuación se presentan ejemplos prácticos tanto de España como de la UE.

Transferencias incondicionales Son fondos que se otorgan sin restricciones sobre su uso. Los gobiernos receptores tienen total libertad para gastarlos en lo que consideren más adecuado para sus ciudadanos.

- Ejemplo en España. El Fondo de Suficiencia Global. Es una transferencia que el Estado hace a las Comunidades Autónomas para financiar sus competencias, y los recursos pueden ser utilizados de forma discrecional.

Transferencias condicionales no compensatorias También conocidas como subvenciones no compensatorias, estas transferencias están condicionadas a que el gobierno receptor gaste los fondos en un área específica, pero la cantidad de la transferencia no depende del gasto que la entidad receptora ya esté realizando en ese sector.

- Ejemplo en España. Los fondos de políticas activas de empleo del Servicio Público de Empleo Estatal (SEPE). Se reparten entre las comunidades autónomas en la Conferencia Sectorial de Empleo y Asuntos Laborales con criterios objetivos y deben destinarse a esas políticas, pero la cuantía de cada comunidad no depende de cuánto gaste con sus recursos propios.
- Ejemplo en la UE. Las subvenciones del Mecanismo de Recuperación y Resiliencia (NextGenerationEU). Cada Estado miembro tiene asignada una cuantía máxima de contribución financiera no reembolsable, que recibe a medida que cumple los hitos y objetivos de las reformas e inversiones acordadas con la Comisión Europea. La cantidad no se calcula como un porcentaje del gasto propio del Estado, sino que se vincula al cumplimiento de esos hitos y objetivos.

Transferencias condicionales compensatorias (cofinanciación) Son fondos que "igualan" o complementan una parte del gasto del gobierno receptor en un área específica. Para recibir el dinero, el gobierno subnacional debe aportar su propia parte del coste.

- Ejemplo en la UE. Los Fondos Estructurales y de Cohesión, como el FEDER (Fondo Europeo de Desarrollo Regional), el Fondo Social Europeo Plus (FSE+) y el Fondo de Cohesión. La UE cofinancia proyectos de desarrollo regional (como la construcción de un parque tecnológico o la modernización de pymes) o de transporte y medio ambiente (como una nueva línea de alta velocidad o una planta de gestión de residuos), con la condición de que el Estado o la comunidad autónoma también aporten un porcentaje del coste. La cantidad que recibe la entidad receptora aumenta a medida que aumenta su propio gasto en el proyecto, hasta un tope de cofinanciación (en el Fondo de Cohesión, el 85 % del coste).

 © 2026 Manuel Cabado García

Bibliografía

Albi, E., J. M. González-Páramo y I. Zubiri (2017). *Economía Pública I: Fundamentos. Presupuesto y Gastos*. Barcelona: Ariel.

Bastida, F., B. Benito y M.-D. Guillamón (2009). An Empirical Assessment of the Municipal Financial Situation in Spain. *International Public Management Journal 12*(4), 484–499.

Besley, T. y A. Case (1995). Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. *American Economic Review 85*(1), 25–45.

Boadway, R. y A. Shah (2009). *Fiscal Federalism: Principles and Practice of Multiorder Governance*. Cambridge: Cambridge University Press.

Brennan, G. y J. M. Buchanan (1980). *The Power to Tax: Analytical Foundations of a Fiscal Constitution*. Cambridge: Cambridge University Press.

Breton, A. (1996). *Competitive Governments: An Economic Theory of Politics and Public Finance*. New York: Cambridge University Press.

Buchanan, J. M. (1965). An Economic Theory of Clubs. *Economica 32*(125), 1–14.

de la Fuente, Á. (2024). Financiación autonómica y local. Apuntes 2024/04, FEDEA.

de la Torre, F. y J. Fernández-Villaverde (2025). *La factura del cupo catalán: privilegios territoriales frente a ciudadanía*. Madrid: La Esfera de los Libros.

Gramlich, E. (1977). Intergovernmental Grants: A Review of the Empirical Literature. En W. E. Oates (ed.), *The Political Economy of Fiscal Federalism*, pp. 219–239. Lexington, MA: Lexington Books.

Hooghe, L., G. Marks, A. H. Schakel, S. Chapman Osterkatz, S. Niedzwiecki y S. Shair-Rosenfield (2016). *Measuring Regional Authority: A Postfunctionalist Theory of Governance, Volume I*. Oxford: Oxford University Press.

Huther, J. y A. Shah (1998). Applying a Simple Measure of Good Governance to the Debate on Fiscal Decentralization. Policy Research Working Paper 1894, World Bank.

Musgrave, R. A. (1959). *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill.

Niskanen, W. A. (1971). *Bureaucracy and Representative Government*. Chicago: Aldine-Atherton.

Oates, W. E. (1972). *Fiscal Federalism*. New York: Harcourt Brace Jovanovich.

Panizza, U. (1999). On the Determinants of Fiscal Centralization: Theory and Evidence. *Journal of Public Economics 74*(1), 97–139.

Pauly, M. V. (1973). Income Redistribution as a Local Public Good. *Journal of Public Economics 2*(1), 35–58.

Rodden, J. (2003). Reviving Leviathan: Fiscal Federalism and the Growth of Government. *International Organization 57*(4), 695–729.

Rosen, H. S. y T. Gayer (2010). *Public Finance* (9 ed.). New York: McGraw-Hill/Irwin.

Salmon, P. (1987). Decentralisation as an Incentive Scheme. *Oxford Review of Economic Policy 3*(2), 24–43.

Shair-Rosenfield, S., A. H. Schakel, S. Niedzwiecki, G. Marks, L. Hooghe y S. Chapman-Osterkatz (2021). Language Difference and Regional Authority. *Regional & Federal Studies 31*(1), 73–97.

Stiglitz, J. E. y J. K. Rosengard (2015). *Economics of the Public Sector* (4 ed.). New York: W. W. Norton & Company.

Tanzi, V. (1996). Fiscal Federalism and Decentralization: A Review of Some Efficiency and Macroeconomic Aspects. En M. Bruno y B. Pleskovic (eds.), *Annual World Bank Conference on Development Economics 1995*, pp. 295–316. Washington, DC: World Bank.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy 64*(5), 416–424.

 © 2026 Manuel Cabado García

Glosario

A

Acción oculta Conducta de una parte que la otra no puede vigilar una vez firmado el acuerdo, como el esfuerzo de un empleado. Técnicamente, es la asimetría sobre decisiones que se toman después del contrato, no son observables ni verificables y alteran el resultado para la otra parte, lo que origina el riesgo moral.

Aditividad Supuesto de que, si dos planes de producción son posibles por separado, también lo es llevarlos a cabo a la vez. Técnicamente, es la propiedad del conjunto de producción por la que la suma de dos planes factibles también es factible. Equivale a la libre entrada en el agregado y es compatible con rendimientos constantes o crecientes.

Administraciones Públicas (sector S.13) Conjunto de organismos públicos que prestan servicios sin venderlos en el mercado y redistribuyen la renta. Técnicamente, en el Sistema Europeo de Cuentas son las unidades institucionales cuya función principal es producir bienes y servicios no destinados a la venta y redistribuir la renta, financiadas sobre todo con impuestos y cotizaciones sociales. Quedan fuera las empresas públicas de mercado, cuyas ventas cubren más de la mitad de sus costes de producción, y el déficit y la deuda públicos se miden siempre sobre este sector.

Agregación vertical de demandas (análisis de Bowen) Forma de obtener cuánto valora la sociedad un bien público sumando lo que cada persona pagaría por la misma unidad, en vez de sumar cantidades. Técnicamente, en el análisis de equilibrio parcial de Bowen la demanda colectiva es la suma vertical de las disposiciones marginales a pagar, porque todos consumen la misma cantidad, y la provisión eficiente se alcanza donde esa suma iguala al coste marginal.

Amante del riesgo Persona a la que le gusta arriesgar. Técnicamente, es el agente que prefiere jugar una lotería antes que recibir con certeza su valor esperado, lo que corresponde a una función de utilidad convexa.

Amenaza creíble Advertencia que a quien la hace le conviene cumplir llegado el momento. Técnicamente, una amenaza es creíble si la acción anunciada es óptima en el punto del juego en que habría que ejecutarla, y la racionalidad secuencial descarta los equilibrios basados en amenazas no creíbles, como la de mantener la producción ante la entrada de un rival.

Análisis coste-beneficio Método para decidir si un proyecto público merece la pena comparando todo lo que aporta a la sociedad con todo lo que le cuesta. Técnicamente, es la técnica de evaluación, basada en el criterio de Kaldor-Hicks, que acepta un proyecto si el valor presente de sus beneficios sociales supera al de sus costes sociales, aunque no recaigan sobre las mismas personas.

Análisis de frontera estocástica Método para medir cuánto se aleja cada empresa del coste mínimo posible. Técnicamente, estima econométricamente una frontera de costes con forma funcional dada (SFA) y separa, en la distancia de cada empresa a ella, la parte debida al azar de la debida a la ineficiencia.

Análisis envolvente de datos Forma de comparar empresas para ver cuáles aprovechan mejor sus recursos. Técnicamente, es un método no pa-

ramétrico (DEA) que mide la eficiencia de las empresas trazando, mediante optimización lineal, una frontera eficiente que envuelve las observaciones sin imponer una forma funcional a la tecnología.

Animal spirits Impulso espontáneo a actuar de los empresarios, más ligado al optimismo o al pesimismo que al cálculo. Técnicamente, según Keynes, este componente psicológico y volátil domina las decisiones de inversión y las hace impredecibles.

Arbitraje (discriminación de precios) Compra de un bien donde es barato para revenderlo donde es caro. Técnicamente, en la discriminación de precios es la reventa entre consumidores que pagan precios distintos, y su imposibilidad es condición necesaria para que el monopolista pueda discriminar.

Argumento de la industria naciente Justificación de proteger temporalmente una industria nacional frente a la competencia exterior hasta que sea competitiva. Técnicamente, se apoya en economías de escala externas o de aprendizaje, porque al crecer la producción nacional bajan los costes medios, de modo que la protección sería solo transitoria.

Arquitectura de la elección Forma en que se presentan las opciones ante quien decide, que nunca es neutral. Técnicamente, es el diseño del contexto de decisión (orden, formato, opción por defecto), y quien lo organiza, el arquitecto de la elección, influye en lo que se elige aunque no prohíba nada.

Asignación justa en el sentido de Foley Reparto en el que nadie se cambiaría por otro y que, además, no desperdicia nada. Técnicamente, es la asignación que, además de eficiente en sentido de Pareto, está libre de envidia, porque ningún agente prefiere la asignación que tiene el otro.

Ausencia de dictador Condición de que nadie decida por todos. Técnicamente, axioma de Arrow que exige que no exista ningún individuo cuya preferencia estricta entre dos alternativas se imponga siempre como preferencia social, con independencia de las preferencias de los demás.

Ausencia de distorsión en el extremo superior Resultado según el cual al más eficiente se le pide lo mismo que si se conociera su tipo. Técnicamente, en los modelos de selección adversa el contrato del tipo más eficiente coincide con el de información completa y toda la distorsión recae sobre los tipos menos eficientes.

Autoridad fiscal independiente Organismo técnico que vigila si el gobierno cumple sus propias reglas de gasto y déficit. Técnicamente, es un organismo de perfil técnico, ajeno al ciclo político y a los grupos de presión, encargado de supervisar el cumplimiento de las reglas fiscales y dar credibilidad a los planes presupuestarios.

Autoselección o filtro (screening) Forma de distinguir a quien no se puede observar dejándole elegir entre varias opciones pensadas para cada tipo. Técnicamente, es la solución a la selección adversa por la que la parte no informada diseña mecanismos, como un menú de contratos, para que cada agente revele su tipo con su elección.

Aversión a la ambigüedad Preferencia por apostar cuando se conocen las probabilidades antes que cuando se desconocen. Técnicamente, los agentes eligen loterías con probabilidades conocidas frente a otras con probabilidades desconocidas aunque no haya motivo para considerarlas peores, como mostró la paradoja de Ellsberg, lo que no puede representarse con ninguna asignación coherente de probabilidades subjetivas.

Aversión a las pérdidas Tendencia a sentir una pérdida con más intensidad que una ganancia de la misma cuantía. Técnicamente, en la teoría de las perspectivas se traduce en una función de valor más pronunciada en el tramo de pérdidas que en el de ganancias, y explica fenómenos como el efecto dotación o la rigidez a la baja de los salarios nominales.

Aversión al riesgo Preferencia por un resultado seguro frente a una apuesta con el mismo valor esperado. Técnicamente, el agente prefiere recibir con certeza el valor esperado de una lotería antes que jugarla, lo que corresponde a una función de utilidad cóncava.

Axioma de independencia Principio según el cual añadir lo mismo a dos opciones no debe cam-

 © 2026 Manuel Cabado García

biar cuál se prefiere. Técnicamente, si dos loterías se combinan con una misma tercera lotería y en la misma proporción, el orden de preferencia entre ellas se mantiene.

Axioma débil de preferencia revelada Exigencia de que el consumidor sea coherente al elegir entre dos cestas. Técnicamente, es el axioma según el cual, si el consumidor elige una cesta cuando otra también era asequible, solo elegirá la segunda en una situación en que la primera ya no sea asequible.

Axioma fuerte de preferencia revelada Exigencia de que las elecciones del consumidor sean coherentes también a través de cadenas de cestas, no solo entre dos. Técnicamente, si una cesta se revela preferida a otra, directamente o a través de una cadena de cestas, la segunda no puede revelarse nunca directamente preferida a la primera, lo que impone la transitividad de las preferencias reveladas.

B

Banda de inacción Intervalo de situaciones en las que a la empresa no le compensa ni contratar ni despedir. Técnicamente, con costes de contratación y despido lineales, es el tramo de empleo comprendido entre el nivel en que la productividad marginal iguala el salario más el coste anualizado de contratar y aquel en que iguala el salario menos el coste anualizado de despedir.

Barreras de entrada Obstáculos que dificultan que nuevas empresas entren en un mercado. Técnicamente, son impedimentos técnicos, legales o estratégicos a la entrada de nuevos competidores que permiten a las instaladas mantener beneficios extraordinarios a largo plazo. En la acepción de Stigler, son el coste de producción que debe soportar la empresa entrante pero no las ya instaladas.

Beneficio económico Lo que gana realmente la empresa una vez descontado lo que podrían haber rendido sus recursos en otro uso. Técnicamente, es el beneficio contable menos el coste de oportunidad de los recursos empleados. En el largo plazo competitivo es nulo aunque el beneficio contable sea positivo.

Beneficio extraordinario Ganancia que supera la necesaria para que la empresa siga en la actividad. Técnicamente, es un beneficio económico positivo, es decir, ingresos superiores a todos los costes, incluido el coste de oportunidad de los recursos propios, que en competencia perfecta solo puede darse a corto plazo.

Beneficio normal Rentabilidad mínima que el empresario necesita para mantener sus recursos en la actividad. Técnicamente, es la retribución del capital propio y del riesgo asumido igual a su coste de oportunidad, que la teoría económica cuenta como coste, de modo que obtenerla equivale a un beneficio económico nulo.

Bien de club Bien del que se puede excluir a quien no paga, pero que varias personas pueden usar a la vez sin perjudicarse mientras no se llene, como una piscina o una autopista de peaje. Técnicamente, es un bien excluible y no rival hasta el punto de congestión, por lo que puede financiarse con cuotas de acceso entre sus usuarios.

Bien de lujo Bien en el que se gasta proporcionalmente más a medida que se gana más. Técnicamente, es un bien normal cuyo consumo aumenta en mayor proporción que la renta, con elasticidad renta superior a la unidad.

Bien de primera necesidad Bien cuyo consumo crece con la renta, pero menos que ella. Técnicamente, es un bien normal cuyo consumo aumenta en menor proporción que la renta, con elasticidad renta entre cero y uno.

Bien Giffen Bien que, en contra de lo habitual, se compra más cuando sube su precio. Técnicamente, es un bien inferior cuyo efecto renta supera en valor absoluto al efecto sustitución, de modo que su demanda aumenta cuando sube su precio.

Bien inferior Bien que se compra menos cuando se gana más. Técnicamente, es el bien cuyo consumo disminuye cuando aumenta la renta, con elasticidad renta negativa.

Bien normal Bien que se compra más cuando se gana más. Técnicamente, es el bien cuyo consumo aumenta cuando aumenta la renta, con elasticidad renta positiva.

Bien preferente Bien privado cuyo consumo la sociedad considera tan valioso que el Estado lo

provee o lo subvenciona, como la sanidad o la educación. Técnicamente, es un bien rival y excluible cuya provisión pública se justifica por un juicio de valor, en la idea de que los individuos infravaloran sus beneficios.

Bien público Bien que, una vez provisto, disfrutan todos a la vez y del que no se puede dejar fuera a nadie, como la defensa nacional. Técnicamente, se caracteriza por la no rivalidad y la no exclusión en el consumo, por lo que el mercado, por sí mismo, tiende a proveerlo en cantidad insuficiente.

Bien público global Bien público que beneficia a todo el planeta, más allá de cualquier frontera. Técnicamente, es un bien no rival y no excluible cuyos beneficios se extienden a todas las naciones, clases sociales y generaciones, sin importar las fronteras.

Bien público impuro Bien que se parece a un bien público, pero no del todo. Técnicamente, es el que cumple solo en parte la no rivalidad o la no exclusión, como una carretera que se congestiona o un servicio del que se puede excluir con un coste.

Bien público inferior Servicio público que se usa menos cuando la renta aumenta, porque se sustituye por una alternativa privada. Técnicamente, es un bien provisto públicamente cuya demanda por parte de la comunidad tiene elasticidad renta negativa, de modo que una transferencia incondicional reduce su provisión.

Bien público intertemporal Bien público cuyos beneficios llegan también a las generaciones futuras, como la estabilidad del clima. Técnicamente, extiende la no rivalidad y la no exclusión en el tiempo, y tiende a proveerse de menos porque la generación presente soporta los costes mientras las futuras, que también se benefician, no intervienen en las decisiones.

Bien público puro Bien público en sentido estricto, que no se agota por mucho que se use y del que es imposible excluir a nadie. Técnicamente, es totalmente no rival, con un coste marginal nulo de un usuario adicional, y no excluible, a diferencia del bien público impuro.

Bien Veblen Bien que resulta más deseable precisamente por ser caro. Técnicamente, es un bien de lujo cuya demanda aumenta con el precio porque el precio elevado forma parte de su atractivo social, lo que se conoce como consumo conspicuo.

Bienes públicos locales Bienes públicos que solo benefician a quienes viven en un territorio concreto, como el alumbrado de una calle. Técnicamente, son no rivales y no excluibles dentro de un ámbito geográfico limitado, por lo que su provisión eficiente corresponde al gobierno de ese territorio.

Búsqueda de rentas Gasto de recursos en conseguir favores del Estado en lugar de en producir. Técnicamente, concepto de Tullock (1967) al que Krueger (1974) dio nombre, es la actividad de los grupos de presión que dedican recursos a influir en las decisiones públicas en favor de sus intereses privados, como protecciones o exenciones, en lugar de a actividades productivas, lo que despilfarra recursos y sesga las decisiones públicas.

C

Caja de Edgeworth Gráfico que muestra todas las formas posibles de repartir entre dos partes unas cantidades fijas de dos cosas. Técnicamente, es un instrumento gráfico cuyas dimensiones vienen dadas por las dotaciones totales de dos factores o de dos bienes y que representa su reparto entre dos empresas o dos consumidores.

Capacidad fiscal Recursos que un territorio podría obtener con sus propios impuestos. Técnicamente, es la recaudación que obtendría una jurisdicción si aplicara a sus bases imponibles un esfuerzo fiscal normal o de referencia, con independencia de los tipos que fije en realidad. Comparada con la necesidad de gasto, determina las transferencias de nivelación.

Capital constante En la teoría de Marx, parte del capital invertida en medios de producción, como máquinas y materias primas. Técnicamente, es trabajo muerto, ya realizado, que no genera plusvalía, pues para Marx la maquinaria solo es productiva por el trabajo que se emplea en ella.

Capital humano Conocimientos y habilidades de una persona que la hacen más productiva. Técnicamente, es un stock que se acumula invirtiendo

 © 2026 Manuel Cabado García

en educación y formación, rinde durante años y no puede venderse: el trabajador solo alquila sus servicios a las empresas.

Capital variable En la teoría de Marx, parte del capital que se paga a los trabajadores como salario. Técnicamente, equivale al valor necesario para reproducir la fuerza de trabajo y es la única parte del capital que genera plusvalía, porque el trabajo vivo produce más valor del que recibe.

Capitalización impositiva Pérdida de valor de un bien cuando se crea un impuesto sobre las rentas que dará en el futuro. Técnicamente, es la reducción del precio actual de un activo cuando se establece un impuesto que grava su renta futura, de modo que la carga recae sobre los propietarios en el momento de introducirse el impuesto. En la terminología de Pantaleoni se llama también amortización del impuesto.

Captura del regulador Situación en la que el regulador acaba sirviendo a la industria que regula en lugar de al interés general. Técnicamente, según Stigler (1971), los grupos regulados tienen mucho más interés e información sobre la regulación que el público y logran orientarla en su favor, algo que facilitan las asimetrías de información, las puertas giratorias y la escasez de recursos del regulador.

Cártel Acuerdo entre empresas para actuar juntas como si fueran un único monopolio. Técnicamente, es la forma de colusión explícita en la que los oligopolistas renuncian a su autonomía y delegan en una agencia central la fijación de precios y el reparto de cuotas para actuar como un monopolista.

Ciclo político Oscilación de la política económica y del gasto público ligada al calendario electoral. Técnicamente, en el modelo oportunista de Nordhaus (1975), ante votantes miopes con expectativas estáticas, el gobierno aplica políticas expansivas que culminan justo antes de las elecciones y las corrige después, aunque la evidencia apunta a ciclos en la inflación más que en la producción o el empleo.

Clasificación del gasto público Forma de ordenar el gasto público para analizarlo. Técnicamente, combina tres criterios, el orgánico (quién gasta), el económico (cómo se gasta, por capítulos) y el funcional (para qué se gasta), cuyo equivalente armonizado internacional es la clasificación COFOG de Naciones Unidas, que distingue diez funciones.

Cláusula de no rescate (territorial) Compromiso de que el gobierno central no asumirá las deudas de los gobiernos territoriales. Técnicamente, busca que estos soporten el coste de sus excesos y combatir así la restricción presupuestaria blanda, pero por sí solo es poco creíble y suele acompañarse de reglas fiscales y de transferencias basadas en fórmulas objetivas. En la Unión Europea, el Tratado de Maastricht incluyó una prohibición de este tipo.

Cliometría Estudio de la historia económica con las herramientas de la economía y la estadística. Técnicamente, es la aplicación de la teoría económica y del contraste estadístico riguroso a la historia económica, cuyo método característico es el análisis contrafactual.

Coeficiente de aversión absoluta al riesgo de Arrow-Pratt Medida de cuánto le disgusta a un agente arriesgar una cantidad fija de dinero. Técnicamente, es el cociente, cambiado de signo, entre la segunda y la primera derivada de la función de utilidad de Bernoulli, que mide su curvatura y permite aproximar la prima de riesgo, de modo que, cuanto mayor es, mayor es la prima de riesgo que exige el agente.

Coeficiente de aversión relativa al riesgo Medida de cuánto le disgusta a un agente arriesgar un porcentaje fijo de su riqueza. Técnicamente, es el coeficiente de aversión absoluta multiplicado por la riqueza, y equivale a la elasticidad de la utilidad marginal respecto a la riqueza cambiada de signo.

Coeficiente de concentración (o índice de concentración) Medida de cuánto se concentra una magnitud, como las cuotas de un impuesto o la renta después de impuestos, en los individuos de mayor renta. Técnicamente, se calcula como el índice de Gini, pero ordenando siempre a los individuos por su renta antes de impuestos, y es la base del índice de Kakwani.

Colusión tácita Coordinación entre empresas que se logra sin pactar nada, solo mediante la observación mutua. Técnicamente, es la cooperación que los oligopolistas sostienen a largo plazo sin acuerdo explícito, lo que dificulta probar su existencia.

Comercio intraindustrial Intercambio entre países de variedades distintas de un mismo tipo de producto. Técnicamente, el modelo de Krugman, basado en el de Dixit-Stiglitz, lo explica por la preferencia por la variedad y las economías de escala internas, y genera ganancias del comercio incluso entre países idénticos en dotaciones, tecnología y preferencias.

Comparación interpersonal de utilidad Juicio sobre si lo que gana una persona vale más o menos que lo que pierde otra. Técnicamente, consiste en medir y comparar en una escala común la utilidad de distintos individuos, algo que exige una utilidad cardinal y que el criterio de Pareto, basado en una utilidad ordinal, evita a propósito. Los criterios de compensación la llevan implícita, y la necesita toda función de bienestar social que pondere ganancias y pérdidas entre personas.

Compatibilidad de incentivos Propiedad de unas reglas de juego que hacen que a cada participante le convenga decir la verdad. Técnicamente, un mecanismo es compatible con los incentivos cuando revelar honestamente la información privada es una estrategia óptima para cada agente, lo que no asegura por sí solo la eficiencia en sentido de Pareto. Se dice entonces que el mecanismo es incentivo-compatible.

Compensación de Hicks Ajuste de la renta, tras un cambio de precios, que permite al consumidor mantener su utilidad inicial. Técnicamente, es el criterio de renta real constante en que se basa la demanda hicksiana.

Compensación de Slutsky Ajuste de la renta, tras un cambio de precios, que permite al consumidor seguir comprando su cesta inicial. Técnicamente, es el criterio de renta real constante que usa la teoría de la preferencia revelada, porque no requiere conocer las curvas de indiferencia.

Competencia fiscal Pugna entre territorios por atraer empresas y contribuyentes bajando impuestos. Técnicamente, cuando la base es móvil, cada gobierno subcentral tiene incentivos a reducir sus tipos para atraer factores de otras jurisdicciones, y el equilibrio de Nash resultante, con tipos bajos en todas, es inferior en el sentido de Pareto porque deja infrafinanciado el gasto público local. Su resultado extremo se conoce como carrera a la baja.

Competencia imperfecta Situación de un mercado en la que alguna empresa puede influir en el precio que cobra. Técnicamente, engloba toda estructura distinta de la competencia perfecta, como el monopolio, el oligopolio y la competencia monopolística. Su estudio nació de la crítica de Sraffa a la competencia perfecta marshalliana y lo desarrollaron Robinson y Chamberlin.

Competencia monopolística Mercado con muchos vendedores de productos parecidos pero no idénticos, en el que entrar y salir es fácil. Técnicamente, cada empresa tiene cierto poder sobre el precio de su propia variedad, pero la libre entrada lleva los beneficios extraordinarios a cero a largo plazo.

Competencia perfecta Mercado con tantas empresas y compradores que ninguno puede influir en el precio. Técnicamente, es la estructura de mercado con muchos compradores y vendedores precio-aceptantes, producto homogéneo, información perfecta, igual tecnología a largo plazo y libertad de entrada y salida. Es el modelo de referencia para comparar las demás estructuras.

Competencia por comparación (yardstick competition) Regulación que pone a competir, de forma indirecta, a monopolistas que operan en mercados separados. Técnicamente, la retribución de cada empresa se fija según los costes de otras empresas comparables, lo que le impide influir en su propia referencia y la incentiva a reducir costes. En federalismo fiscal, designa también la competencia entre gobiernos territoriales cuyos votantes juzgan a cada uno por los resultados de los demás.

Competencia por el mercado (subasta de Demsetz) Rivalidad entre empresas por conseguir el derecho a ser la única que presta un servicio. Técnicamente, propuesta por Demsetz

 © 2026 Manuel Cabado García

(1968), consiste en adjudicar mediante subasta la explotación en exclusiva de un monopolio natural durante un periodo, de modo que la competencia entre licitadores antes de la adjudicación y la repetición periódica de la subasta sustituyen a la regulación.

Complementariedad de costes Situación en la que producir más de un bien abarata producir los demás. Técnicamente, el coste incremental de cada bien disminuye al aumentar la producción de los otros bienes de la empresa. Junto con las economías de alcance, es una de las condiciones de existencia del monopolio natural multiproducto, porque con varios bienes las economías de escala ya no garantizan que los costes sean subaditivos.

Complementarios brutos Bienes que se consumen juntos, de modo que si uno se encarece se compra menos también del otro. Técnicamente, lo son cuando la elasticidad precio cruzada de la demanda marshalliana es negativa, por lo que el criterio recoge a la vez el efecto sustitución y el efecto renta, a diferencia de los complementarios netos.

Complementarios netos Bienes que el consumidor usa juntos, de modo que, si uno se encarece, compra menos del otro aunque se le compense la pérdida de renta. Técnicamente, son bienes cuyo efecto sustitución cruzado, medido con la demanda hicksiana, es negativo.

Complementarios perfectos Bienes que solo sirven usados juntos y en una proporción fija. Técnicamente, solo aumentan la utilidad si se consumen conjuntamente en esa proporción y generan curvas de indiferencia en forma de L, propias de la función de Leontief, con elasticidad de sustitución nula.

Complementos estratégicos Variables de decisión en las que a una empresa le conviene moverse en la misma dirección que su rival. Técnicamente, son aquellas cuyas funciones de reacción tienen pendiente positiva, como los precios cuando el producto está diferenciado.

Completitud Capacidad del consumidor para comparar siempre dos cestas cualesquiera. Técnicamente, es el axioma según el cual, ante dos cestas cualesquiera, el consumidor prefiere una a la otra o le son indiferentes, lo que garantiza que no quede ningún par de cestas sin ordenar.

Composición orgánica del capital Grado de mecanización de la producción en la teoría de Marx. Técnicamente, es el cociente entre el capital constante, invertido en medios de producción, y el capital variable, pagado en salarios, y su aumento reduce la tasa de beneficio si la tasa de explotación no cambia.

Compromiso entre eficiencia y equidad Dilema por el que repartir mejor la renta con impuestos tiene un coste en eficiencia. Técnicamente, es el conflicto por el que, al no poder usarse impuestos de suma fija, redistribuir con impuestos distorsionantes mejora la equidad a costa de alejarse de la frontera eficiente.

Condición de Aaron Regla que indica cuándo un sistema de pensiones de reparto rinde más que uno de capitalización. Técnicamente, el reparto es preferible cuando la suma de las tasas de crecimiento de la población y de los salarios supera al tipo de interés de mercado, de modo que la rentabilidad implícita de las cotizaciones excede a la del ahorro.

Condición de agregación de Cournot Restricción que exige que, cuando cambia el precio de un bien, el gasto total siga igualando a la renta. Técnicamente, la suma de las elasticidades de todos los bienes respecto a ese precio, ponderadas por su participación en el gasto, es igual a la participación de ese bien cambiada de signo.

Condición de agregación de Engel Condición que asegura que, cuando cambia la renta, el gasto total siga coincidiendo con ella. Técnicamente, es la restricción según la cual la suma de las elasticidades renta de todos los bienes, ponderadas por su participación en el gasto, es igual a uno.

Condición de eficiencia económica Regla según la cual la empresa combina los factores de modo que no le compensa sustituir uno por otro a los precios de mercado. Técnicamente, es la condición de equilibrio del productor por la que la relación marginal de sustitución técnica, tasa a la que la empresa puede intercambiar internamente los factores, se iguala al cociente de sus precios, tasa de intercambio en el mercado.

Condición de Gorman (forma polar de Gorman) Condición para que la demanda de un mercado dependa solo de la renta total y no de cómo se reparte entre los consumidores. Técnicamente, exige preferencias con la forma polar de Gorman, cuyas curvas de Engel son rectas paralelas, con la misma pendiente para todos los consumidores, de modo que una redistribución de la renta no altera la demanda agregada y esta puede tratarse como la de un consumidor representativo.

Condición de homogeneidad Restricción según la cual variaciones de precios y renta en la misma proporción no alteran la cantidad demandada, lo que implica ausencia de ilusión monetaria. Técnicamente, en términos de elasticidades, la suma de todas las elasticidades precio de un bien más su elasticidad renta es cero.

Condición de Hosios Condición para que el mercado genere el paro socialmente adecuado: cada parte debe llevarse una parte del excedente proporcional a lo que aporta a los emparejamientos. Técnicamente, el equilibrio del modelo DMP es eficiente si el poder de negociación de los trabajadores coincide con la elasticidad de la función de emparejamiento respecto al paro y la renta del parado no incluye transferencias como la prestación por desempleo (Hosios, 1990).

Condición de negatividad Condición de que encarecer un bien, compensando al consumidor para que no pierda bienestar, nunca le lleve a comprar más de él. Técnicamente, es la restricción según la cual el efecto sustitución de la demanda de un bien respecto a su propio precio es negativo o nulo.

Condición de Samuelson Regla que indica cuánto bien público conviene producir: hasta que lo que todos juntos valoran una unidad más iguale lo que cuesta. Técnicamente, es la condición de provisión eficiente de un bien público según la cual la suma de las relaciones marginales de sustitución de todos los individuos debe igualarse a la relación marginal de transformación.

Condición de simetría Propiedad de la demanda según la cual la sustituibilidad entre dos bienes, medida por el efecto sustitución, es recíproca. Técnicamente, el efecto de una subida del precio de un bien sobre la demanda compensada del otro es igual al efecto inverso.

Condición de Solow Regla que dice hasta dónde le conviene a la empresa subir el salario cuando pagar más hace trabajar mejor. Técnicamente, el salario de eficiencia es aquel en el que la elasticidad del esfuerzo respecto al salario es igual a uno, lo que minimiza el coste por unidad de esfuerzo (Solow, 1979).

Condicionalidad Exigencia de cumplir ciertas condiciones para recibir una ayuda, que evita que contar con ella invite a la imprudencia. Técnicamente, es una solución al riesgo moral de los gobiernos que sujeta la financiación de las instituciones multilaterales a requisitos estrictos de política económica, de modo que solo resulte atractiva en situaciones extremas.

Conjetura de Coase Idea de que quien vende en exclusiva un bien duradero acaba compitiendo consigo mismo y pierde su poder de mercado. Técnicamente, también llamada paradoja de Coase (1972), sostiene que, si los consumidores son pacientes y anticipan las rebajas futuras, el monopolista no puede discriminar en el tiempo y el precio tiende al coste marginal.

Conjunto de capacidad Abanico de vidas que una persona puede llevar de verdad con lo que tiene y en sus circunstancias. Técnicamente, en el enfoque de Sen, es el conjunto de combinaciones de funcionamientos que una persona puede alcanzar dados sus bienes y circunstancias, que mide su libertad real para elegir formas de vida valiosas.

Conjunto de posibilidades de producción Conjunto de todo lo que una empresa es capaz de producir con la tecnología disponible. Técnicamente, reúne todos los vectores de producción técnicamente posibles, en los que los componentes positivos son outputs (productos) y los negativos, inputs (factores).

Conocimiento común Lo que todos los jugadores saben y saben que los demás saben. Técnicamente, un hecho es de conocimiento común cuando todos lo conocen, todos saben que los demás lo conocen, todos saben que los demás saben que

 © 2026 Manuel Cabado García

lo conocen, y así indefinidamente. Es el supuesto habitual sobre las reglas del juego.

Constitución fiscal Reglas legales que atan las manos a los gobiernos para que el gasto y los impuestos no crezcan sin control. Técnicamente, es el mandato legal, a veces constitucional, que limita la discrecionalidad de la política fiscal para frenar el crecimiento desmedido del sector público.

Consumidor indiferente Consumidor al que le da igual comprar a una empresa o a otra porque obtiene con ambas la misma utilidad neta. Técnicamente, en el modelo de Hotelling es aquel para el que el precio más el coste de transporte coincide en las dos tiendas, y su posición reparte el mercado entre ambas empresas.

Contabilidad generacional Método que calcula cuánto pagará y recibirá del sector público cada generación a lo largo de su vida. Técnicamente, desarrollado por Auerbach, Gokhale y Kotlikoff en 1991, estima el valor actual de los impuestos netos de transferencias de cada cohorte para valorar la sostenibilidad y la equidad intergeneracional de la política fiscal.

Continuidad de las preferencias Supuesto de que las preferencias no dan saltos bruscos ante cambios pequeños en las cestas. Técnicamente, es el axioma que descarta saltos en las preferencias y que, junto con el preorden completo, garantiza la existencia de una función de utilidad continua que las representa.

Contrato de franquicia Contrato en el que el agente se queda con todo el resultado a cambio de pagar al principal una cantidad fija. Técnicamente, es el contrato óptimo cuando el principal es averso al riesgo y el agente neutral, en el que el agente asume todo el riesgo y paga al principal un canon constante, de modo que el principal queda plenamente asegurado.

Contrato de precio fijo Contrato en el que la empresa cobra una cantidad cerrada y se queda con todo lo que ahorre o pierda. Técnicamente, es el esquema de incentivos de mayor poder, en el que el regulador no cubre ninguna desviación de costes, lo que maximiza el incentivo al esfuerzo pero deja más renta a la empresa eficiente. El price cap se aproxima a él.

Contrato de reembolso de costes (cost plus) Contrato en el que el regulador paga a la empresa todos los costes en que incurra. Técnicamente, es el esquema de menor poder, que cubre cualquier desviación de costes, no deja rentas pero elimina el incentivo a reducirlos. La regulación por coste del servicio se aproxima a él y conviene cuando el riesgo de captura es alto.

Contrato incompleto Contrato que no puede prever todo lo que puede llegar a pasar. Técnicamente, no especifica las obligaciones de las partes en todos los estados de la naturaleza, porque algunos no pueden preverse o verificarse ante un tercero, de modo que pasa a ser decisivo quién decide en las situaciones no previstas.

Contratos implícitos Acuerdos no escritos por los que la empresa garantiza al trabajador una renta estable a lo largo del ciclo a cambio de un salario medio algo menor. Técnicamente, la teoría de Baily, Azariadis y Gordon explica la rigidez del salario real como un reparto eficiente del riesgo entre empresas neutrales al riesgo y trabajadores aversos a él, que se sostiene en la reputación de la empresa y no genera paro involuntario.

Contribución fiscal neta Diferencia entre lo que una persona paga al sector público y lo que recibe de él. Técnicamente, son los impuestos y cotizaciones que paga menos las transferencias y el gasto en servicios públicos que se le imputan, medidos en un año o, descontados, a lo largo de toda su vida.

Controversia de las dos Cambridge Polémica sobre si tiene sentido sumar todas las máquinas y equipos de una economía en una sola cifra de capital. Técnicamente, fue el debate de mediados del siglo XX entre los poskeynesianos del Cambridge británico y los neoclásicos del Cambridge estadounidense sobre la imposibilidad de medir el capital agregado con independencia de la tasa de beneficio.

Convexidad del conjunto de producción Propiedad según la cual mezclar dos técnicas factibles en cualquier proporción da también una técnica factible. Técnicamente, toda media ponderada de dos planes factibles pertenece al conjunto

de producción. Se cumple, por ejemplo, cuando hay a la vez aditividad y divisibilidad, y, si es posible la inacción, es compatible con rendimientos constantes o decrecientes, pero no crecientes.

Convexidad estricta de las preferencias Axioma que refleja la preferencia por la variedad, de modo que el consumidor prefiere las cestas equilibradas a las que tienen mucho de un bien y poco de otro. Técnicamente, cualquier combinación ponderada de dos cestas distintas e igual de valoradas se prefiere estrictamente a ambas, lo que hace estrictamente convexas las curvas de indiferencia, decreciente la relación marginal de sustitución y estrictamente cuasicóncava la función de utilidad.

Convexidad estricta del conjunto de producción Propiedad según la cual, al mezclar dos técnicas factibles distintas, se obtiene una técnica con la que podría producirse algo más de lo que indica la propia mezcla. Técnicamente, toda media ponderada de dos planes factibles distintos queda en el interior del conjunto de producción y no en su frontera, lo que implica rendimientos estrictamente decrecientes a escala.

Copago Parte del precio de un servicio público que paga directamente quien lo usa. Técnicamente, es un instrumento contra el riesgo moral: al hacer que el usuario soporte parte del coste en el momento del consumo, reduce la demanda de servicios cuyo precio percibido sería nulo.

Corresponsabilidad fiscal Principio de que quien decide gastar debe asumir también el coste político de recaudar. Técnicamente, exige que las haciendas territoriales financien una parte sustantiva de su gasto con impuestos propios y visibles para los contribuyentes, sobre los que tengan capacidad normativa.

Corto plazo Periodo en el que la empresa no tiene tiempo de cambiar algunos de sus recursos, como sus instalaciones. Técnicamente, es el horizonte temporal en el que al menos un factor, generalmente el capital, es fijo y no hay progreso técnico, de modo que la producción solo puede alterarse modificando los factores variables.

Coste contable Coste que recoge la contabilidad de la empresa a partir de los gastos efectivamente realizados. Técnicamente, se calcula con las convenciones que sirven para elaborar los estados financieros y cumplir las obligaciones fiscales, y no incluye el coste de oportunidad de los recursos propios.

Coste de agencia Lo que se pierde porque el principal no puede ver directamente lo que hace el agente. Técnicamente, es la diferencia entre el resultado con información simétrica y el alcanzable con riesgo moral, debida a que, para incentivar el esfuerzo, el agente averso al riesgo tiene que soportar parte de él.

Coste de oportunidad Lo que se deja de ganar al elegir una opción en lugar de la mejor alternativa disponible. Técnicamente, en la empresa es el rendimiento que obtendrían sus recursos en su mejor uso alternativo, e incluirlo es lo que distingue el coste económico del coste contable.

Coste de transporte (diferenciación espacial) Molestia o gasto que soporta el consumidor por comprar una variedad alejada de la que más le gusta. Técnicamente, en los modelos de diferenciación espacial es el coste en que incurre el consumidor al desplazarse hasta una variedad, y refleja la heterogeneidad de sus gustos.

Coste de uso del capital Lo que le cuesta a una empresa utilizar una unidad de capital durante un periodo. Técnicamente, es el precio del bien de capital multiplicado por la suma del tipo de interés y la tasa de depreciación, y la empresa invierte hasta que el valor del producto marginal del capital lo iguala. Un impuesto sobre el beneficio lo eleva salvo que permita deducir de inmediato la inversión o todos los costes, incluido el de financiación.

Coste económico Coste que incluye todo lo que la empresa sacrifica al producir, se pague o no. Técnicamente, añade a los gastos que registra la contabilidad el coste de oportunidad de los recursos propios, y es el concepto de coste que utiliza la teoría económica.

Coste fijo medio Parte del coste fijo que corresponde a cada unidad producida. Técnicamente, es el coste fijo dividido por la cantidad producida, y disminuye de forma continua al aumentar la producción porque el coste fijo se reparte entre

 © 2026 Manuel Cabado García

más unidades.

Coste incremental Coste adicional que supone para la empresa añadir un bien a su producción. Técnicamente, es la diferencia entre el coste de producir todos sus bienes y el de producirlos todos salvo ese, y sirve para comprobar si el precio de un bien está libre de subsidios cruzados.

Coste marginal Lo que cuesta producir una unidad más. Técnicamente, es la derivada del coste total respecto a la producción y, a corto plazo, es igual al precio del factor variable dividido por su productividad marginal.

Coste marginal de los fondos públicos (o coste sombra) Lo que cuesta de verdad a la sociedad obtener un euro más de recaudación. Técnicamente, es la suma del euro adicional recaudado y del exceso de gravamen marginal que provoca, y es el dato relevante para decidir si compensa un gasto público financiado con impuestos. En la regulación, como en el modelo de Laffont y Tirole, es lo que hace costoso dejar rentas a la empresa regulada y actúa como multiplicador en la regla de precios de tipo Ramsey-Boiteux.

Coste marginal social Lo que le cuesta de verdad a la sociedad producir una unidad más, incluido el daño a terceros. Técnicamente, es la suma del coste marginal privado y del coste marginal externo, o daño marginal causado a terceros, de producir una unidad adicional de un bien.

Coste total medio Lo que cuesta de media cada unidad producida, contando todos los costes. Técnicamente, es la suma del coste variable medio y el coste fijo medio. Tiene forma de U y el coste marginal lo corta en su mínimo.

Coste variable medio Lo que cuesta de media cada unidad producida sin contar los costes fijos. Técnicamente, es el coste variable por unidad producida, decreciente hasta que lo corta el coste marginal y creciente a partir de ese punto.

Costes de ajuste (del empleo) Costes que soporta la empresa por cambiar el tamaño de su plantilla, como contratar, formar o despedir. Técnicamente, convierten la demanda de trabajo en un problema dinámico. Si son cuadráticos, el empleo se ajusta de forma gradual, y si son lineales aparece una banda de inacción con ajustes bruscos.

Costes de menú Gastos y molestias que supone para una empresa cambiar sus precios. Técnicamente, son los costes en que incurre una empresa al cambiar sus precios de venta y que, según Mankiw, explican la rigidez nominal de los precios.

Costes de transacción Costes de ponerse de acuerdo con otra parte y de hacer cumplir lo pactado, aparte del precio de lo que se intercambia. Técnicamente, incluyen buscar a la contraparte, informarse, negociar, redactar y vigilar el acuerdo. Su ausencia es un supuesto del teorema de Coase, y su tamaño explica que la negociación privada no resuelva muchas externalidades.

Costes fijos Gastos que la empresa tiene que pagar aunque no produzca nada, como el alquiler del local. Técnicamente, son los costes que no dependen del volumen de producción y que solo existen a corto plazo, porque a largo plazo todos los factores son variables.

Costes hundidos Gastos ya hechos que no se pueden recuperar pase lo que pase. Técnicamente, son costes ya realizados que no pueden recuperarse de ninguna manera, por lo que son irrelevantes para las decisiones futuras y pueden actuar como barrera de entrada.

Costes variables Costes que cambian con el volumen de producción. Técnicamente, a corto plazo su curva suele tener forma de S invertida, porque crece primero a ritmo decreciente, mientras los rendimientos marginales del factor variable son crecientes, y después a ritmo creciente, cuando pasan a ser decrecientes.

Crisis de las tijeras Situación en la que los ingresos públicos se estancan mientras el gasto sigue creciendo, como las hojas de unas tijeras que se abren. Técnicamente, Tarschys llamó así en 1983 a la crisis de los años setenta, en la que la evasión, los costes de control y los efectos desincentivadores frenaban la recaudación mientras la crisis elevaba el gasto del Estado del Bienestar, lo que abrió el debate sobre el tamaño del sector público.

Criterio de Hicks Criterio de compensación que da por bueno un cambio si a quienes pierden no les compensaría pagar a quienes ganan para que

renuncien a él. Técnicamente, una asignación es preferible a otra si los perdedores no pueden compensar a los ganadores para que renuncien al cambio quedando ellos, aun así, mejor que si el cambio se produjera.

Criterio de Hurwicz Regla de decisión intermedia entre el optimismo y el pesimismo. Técnicamente, es un criterio de decisión bajo incertidumbre que pondera el mejor y el peor resultado de cada estrategia mediante un coeficiente de optimismo y elige la de mayor valor ponderado.

Criterio de Kaldor Criterio de compensación que da por bueno un cambio si lo que ganan unos basta para indemnizar a quienes pierden. Técnicamente, una asignación es preferible a otra si los ganadores del cambio pueden compensar a los perdedores y, aun así, seguir beneficiándose, aunque la compensación no llegue a pagarse.

Criterio de Laplace Regla de decisión que, a falta de información, trata todas las situaciones posibles como igual de probables. Técnicamente, es un criterio de decisión bajo incertidumbre que asigna la misma probabilidad a todos los estados de la naturaleza y elige la estrategia de mayor valor esperado.

Criterio de Little Criterio de compensación que, además de la eficiencia, tiene en cuenta el reparto. Técnicamente, solo da por bueno un cambio si cumple a la vez los criterios de Kaldor y de Hicks y, además, mejora la distribución de la renta.

Criterio de Pareto Regla para comparar situaciones económicas según la cual un cambio es una mejora si beneficia a alguien sin perjudicar a nadie. Técnicamente, ordena los estados mediante la superioridad paretiana y descansa en juicios de valor débiles: individualismo, unanimidad, independencia del proceso, no paternalismo y benevolencia.

Criterio de Pareto (axioma de Arrow) Condición de que la sociedad respete lo que todos prefieren. Técnicamente, es el axioma de Arrow, también llamado optimalidad paretiana débil, según el cual, si todos los individuos prefieren una alternativa a otra, la regla de elección colectiva debe ordenarlas socialmente igual. En su versión fuerte basta con que alguien la prefiera y nadie prefiera la contraria.

Criterio de Samuelson Criterio según el cual una situación económica es mejor que otra si, con el reparto adecuado, permitiría que todos estuvieran mejor que en cualquier reparto de la otra. Técnicamente, una situación es preferible a otra si su frontera de posibilidades de utilidad queda por completo fuera de la de esta, aunque los puntos concretos no sean comparables según Pareto.

Criterio de Scitovsky Regla que solo da por buena una mejora potencial si se mantiene al mirar el cambio en los dos sentidos. Técnicamente, exige que se cumplan a la vez los criterios de Kaldor y de Hicks (doble criterio), lo que evita la paradoja de Scitovsky.

Criterio del coste medio Regulación que obliga al monopolio a vender a un precio que solo cubra sus costes. Técnicamente, impone un precio igual al coste medio para asegurar beneficios nulos, a costa de reducir el incentivo del monopolista a rebajar costes.

Criterio maximax Regla de decisión optimista que apuesta por lo mejor que puede pasar. Técnicamente, es un criterio de decisión bajo incertidumbre que identifica el mejor resultado posible de cada estrategia y elige aquella cuyo mejor resultado es el más alto.

Criterio maximin Regla de decisión pesimista que se protege frente a lo peor que puede pasar. Técnicamente, es un criterio de decisión bajo incertidumbre, atribuido a Wald, que elige la estrategia cuyo peor resultado es el menos malo.

Criterio media-varianza Forma simplificada de comparar alternativas con riesgo atendiendo solo a su rendimiento medio y a su dispersión. Técnicamente, trata la media como un bien y la varianza como un mal y prescinde de los momentos de orden superior. Solo es exactamente óptimo con utilidad cuadrática o si los rendimientos siguen una distribución normal, y es la base de la teoría de selección de carteras de Markowitz.

Criterio minimax de Savage Regla de decisión que busca arrepentirse lo menos posible de lo elegido. Técnicamente, es un criterio de decisión bajo incertidumbre que construye una matriz de arrepentimiento o coste de oportunidad y elige

 © 2026 Manuel Cabado García

la estrategia que minimiza el máximo arrepentimiento.

Criterios de compensación Reglas para dar por buenos cambios que benefician a unos y perjudican a otros, siempre que los que ganan pudieran indemnizar a los que pierden. Técnicamente, son criterios que buscan una ordenación completa de los estados económicos admitiendo cambios en los que los ganadores podrían compensar potencialmente a los perdedores.

Crítica de Lucas Advertencia de que la gente cambia su comportamiento cuando cambian las reglas de la política económica. Técnicamente, es el argumento según el cual los modelos econométricos tradicionales no sirven para evaluar políticas, porque sus parámetros cambian cuando los agentes, al anticipar las nuevas reglas, modifican su comportamiento.

Cuasi-renta Ganancia que obtienen a corto plazo los factores que no pueden ajustarse, como una máquina ya instalada. Técnicamente, es la remuneración de los factores fijos a corto plazo, igual a los ingresos que exceden de los costes variables, que desaparece a largo plazo al no existir factores fijos.

Cuasioferta Oferta de una empresa cuyo coste depende de lo que produce toda la industria. Técnicamente, es la cantidad que la empresa ofrecería a cada precio dado un nivel de producción de la industria, y su suma da la oferta de mercado cuando hay economías o deseconomías externas.

Cuasirracionalidad Tendencia a no reoptimizar cuando hacerlo apenas compensa. Técnicamente, un agente cuasirracional mantiene sus decisiones ante una perturbación si la pérdida de no ajustarse es de segundo orden, es decir, muy pequeña en comparación con el tamaño de la perturbación.

Cuentas nocionales Sistema de pensiones que anota a cada trabajador lo que cotiza como si lo ahorrara, aunque ese dinero sirve para pagar las pensiones de hoy. Técnicamente, simula cuentas individuales de capitalización revalorizadas a un tipo de interés nocional, pero sin acumular fondos reales, pues se financia colectivamente como un sistema de reparto.

Cuña fiscal Brecha que abre un impuesto entre lo que paga el comprador y lo que cobra el vendedor. Técnicamente, es la diferencia entre el precio pagado por el consumidor y el precio recibido por el productor que introduce un impuesto distorsionante.

Curva de aprendizaje Representación de cómo se produce cada vez más rápido a medida que se gana práctica. Técnicamente, muestra cómo disminuyen las horas de trabajo necesarias por unidad producida a medida que aumenta la producción acumulada.

Curva de Armey Gráfico que muestra que el crecimiento económico es mayor con un sector público de tamaño intermedio que con uno muy pequeño o muy grande. Técnicamente, es la relación empírica en forma de U invertida entre el tamaño del sector público y el crecimiento económico, también llamada curva BARS, que estima el tamaño que maximiza el crecimiento.

Curva de Beveridge Relación que muestra que, cuando hay muchas vacantes, hay poco paro, y al revés. Técnicamente, es la relación decreciente y convexa entre la tasa de vacantes y la tasa de paro que resulta de igualar las entradas y salidas del paro. Un desplazamiento hacia fuera indica un emparejamiento menos eficiente o una destrucción de empleo más rápida.

Curva de colusión Conjunto de repartos de la producción entre los miembros de un cártel que agotan las ganancias de cooperar. Técnicamente, une los puntos de tangencia entre las curvas isobeneficio de las empresas, desde la cantidad de monopolio de una hasta la de la otra.

Curva de contrato Conjunto de acuerdos entre dos partes que no se pueden mejorar para una sin empeorar a la otra. Técnicamente, es el lugar geométrico de los puntos de tangencia entre las curvas de indiferencia de dos consumidores, o entre las isocuantas de dos bienes, dentro de una caja de Edgeworth, y en la negociación entre sindicato y empresa, entre las curvas de indiferencia del sindicato y las isobeneficio de la empresa, de modo que recoge todos los acuerdos eficientes en el sentido de Pareto.

Curva de costes a largo plazo Coste mínimo de producir cada cantidad cuando la empresa puede elegir el tamaño de su planta. Técnicamente, es la envolvente de las curvas de costes a corto plazo, que nunca quedan por debajo de ella, y es una función de planificación en la que la empresa elige el tamaño de planta.

Curva de creación de empleo Relación que indica cuántas vacantes compensa abrir para cada salario: cuanto más caro es el trabajo, menos. Técnicamente, iguala el coste esperado de cubrir una vacante con el valor de un puesto ocupado, y define una relación decreciente entre el salario y la tensión del mercado que hace el papel de la demanda de trabajo en el modelo DMP.

Curva de Engel Curva que muestra cuánto compra el consumidor de un bien según lo que gane. Técnicamente, es la relación entre el nivel de renta y la cantidad comprada de un bien, derivada de la curva de renta-consumo, que permite distinguir bienes normales e inferiores.

Curva de indiferencia Conjunto de combinaciones de bienes que le gustan al consumidor exactamente lo mismo. Técnicamente, es el lugar geométrico de las cestas que proporcionan al consumidor el mismo nivel de utilidad.

Curva de indiferencia social (bienestar social) Repartos del bienestar entre individuos que la sociedad considera igual de deseables. Técnicamente, es el conjunto de combinaciones de utilidades individuales a las que la función de bienestar social asigna el mismo valor.

Curva de Laffer Relación entre el tipo impositivo y lo que se recauda, que primero crece y después cae. Técnicamente, tiene forma de U invertida, con recaudación nula con tipos del 0 % y del 100 %, de modo que por encima del tipo que maximiza la recaudación bajar los tipos la aumenta, aunque la evidencia sitúa esa zona en tipos próximos a la confiscación, lejos de los habituales.

Curva de Lorenz Gráfico que muestra qué parte de la renta total tiene el 10 %, el 20 %, etc. más pobre de la población. Técnicamente, relaciona el porcentaje acumulado de población, ordenada de menor a mayor renta, con el porcentaje acumulado de renta que le corresponde, y coincide con la diagonal cuando hay igualdad perfecta.

Curva de no escaqueo Curva que indica cuánto hay que pagar, como mínimo, para que a los trabajadores no les compense escaquearse. Técnicamente, da para cada nivel de empleo el salario más bajo compatible con el esfuerzo, y es creciente porque, cuanto menor es el paro, menos cuesta ser despedido.

Curva de oferta a largo plazo de la industria Cantidad que ofrece un sector a cada precio cuando hay tiempo para que entren o salgan empresas. Técnicamente, es la relación entre precio y cantidad ofrecida por la industria con libre entrada y salida. Es horizontal con costes constantes, creciente con costes crecientes y decreciente con costes decrecientes.

Curva de oferta de la empresa competitiva Cantidad que la empresa precio-aceptante desea vender a cada precio. Técnicamente, a corto plazo coincide con el tramo creciente de su coste marginal por encima del mínimo del coste variable medio, y es nula para precios inferiores.

Curva de oferta de la industria Cantidad total que desean vender todas las empresas de un mercado a cada precio. Técnicamente, se obtiene sumando horizontalmente las ofertas individuales, lo que exige que el precio de los factores no cambie cuando la industria produce más.

Curva de oferta de Lucas Relación según la cual la producción solo se aparta de su nivel natural cuando los precios sorprenden. Técnicamente, los productores no distinguen si la subida de su precio es relativa o general y responden a ella aumentando la producción, por lo que la curva tiene pendiente positiva ante variaciones no anticipadas del nivel de precios y es vertical ante las anticipadas, incluso a corto plazo.

Curva de oferta de trabajo en S invertida Curva de oferta individual que primero baja, luego sube y al final puede volver a bajar a medida que aumenta el salario. Técnicamente, refleja que con salarios muy bajos domina el efecto renta, con salarios intermedios el de sustitución y con salarios altos de nuevo el de renta, forma que Frisch dedujo de sus estimaciones de la utilidad marginal de la renta.

 © 2026 Manuel Cabado García

Curva de Phillips Relación que muestra que, cuando baja el paro, suele subir la inflación. Técnicamente, es la relación negativa y no lineal entre inflación (originalmente, crecimiento de los salarios monetarios) y desempleo, que Samuelson y Solow interpretaron como una disyuntiva explotable por la política económica.

Curva de Phillips aumentada con expectativas Versión de la curva de Phillips que tiene en cuenta la inflación que la gente espera. Técnicamente, es la versión de Friedman en la que la inflación depende de la inflación esperada y de la brecha entre el paro y su tasa natural, de modo que la disyuntiva es solo temporal y la curva es vertical a largo plazo.

Curva de producto total Curva que muestra cuánto se produce según la cantidad que se emplea de un factor, con los demás fijos. Técnicamente, representa el nivel máximo de producción para cada cantidad del factor variable, dada la cantidad del factor fijo.

Curva de renta-consumo Recorrido que sigue la cesta elegida cuando aumenta la renta sin que cambien los precios. Técnicamente, une los puntos de tangencia entre las curvas de indiferencia y una familia de rectas de balance paralelas, y de ella se deriva la curva de Engel de cada bien.

Curva del gran Gatsby Relación positiva entre la desigualdad de la renta de un país y la persistencia de la renta entre padres e hijos. Técnicamente, enfrenta el índice de Gini con la elasticidad intergeneracional de la renta, de modo que los países más desiguales tienen menos movilidad. La popularizó Alan Krueger a partir de los trabajos de Corak.

Curva en U de la participación femenina Evolución de la participación laboral de la mujer casada, que primero cae y luego se recupera con el desarrollo. Técnicamente, Goldin mostró con más de dos siglos de datos de Estados Unidos que bajó con el paso de una sociedad agraria a una industrial y repuntó con el crecimiento de los servicios, impulsada por el cambio estructural y por las normas sociales sobre el papel de la mujer.

Curva isobeneficio Curva que une los repartos de la producción con los que una empresa gana lo mismo. Técnicamente, es el lugar geométrico de las combinaciones de producción de dos empresas que proporcionan el mismo beneficio a una de ellas.

Curva salarial (modelo DMP) Relación que muestra que el salario negociado sube cuando hay más vacantes por cada parado. Técnicamente, es el resultado de la negociación de Nash en el modelo DMP, una relación creciente entre el salario y la tensión del mercado que ocupa el lugar de la oferta de trabajo.

D

Debate del cálculo económico Discusión sobre si una economía planificada puede asignar los recursos tan bien como el mercado. Técnicamente, Mises sostuvo que sin precios de mercado para los medios de producción no es posible el cálculo económico racional, y Hayek añadió que la información necesaria está dispersa y es en buena parte tácita, de modo que solo el sistema de precios puede transmitirla. Lange respondió con el socialismo de mercado, y el debate ha vuelto con el big data y la inteligencia artificial.

Defensa de la competencia Política que vigila que las empresas compitan y sanciona a quien lo impide. Técnicamente, actúa ex post, castigando conductas anticompetitivas y controlando las concentraciones, y es la respuesta adecuada cuando el poder de mercado procede de conductas estratégicas o de barreras legales y la competencia sigue siendo posible.

Déficit estructural Déficit que tendría el sector público si la economía estuviera en su nivel normal de actividad. Técnicamente, es el saldo presupuestario corregido de los efectos del ciclo económico y de las medidas temporales, y es la magnitud que limitan el artículo 135 de la Constitución y la ley de estabilidad presupuestaria.

Deflactación de la tarifa Actualización de los tramos y de las cantidades fijas de un impuesto progresivo para que la inflación no eleve la carga real. Técnicamente, corrige la progresividad en frío ajustando la tarifa con la inflación o con el crecimiento de la renta, aunque no beneficia a quienes están por debajo del mínimo exento

y tiene un coste recaudatorio.

Demanda condicionada de factores Cantidad de cada factor que la empresa emplea para obtener una producción dada al menor coste. Técnicamente, también llamada hicksiana, depende de los precios de los factores y del nivel de producción y es la solución del problema dual del productor.

Demanda de mercado Cantidad total que desean comprar todos los consumidores a cada precio. Técnicamente, es la suma horizontal de las demandas individuales, de modo que la cantidad demandada a cada precio es la suma de las cantidades demandadas por todos los consumidores.

Demanda derivada Demanda de un factor que existe porque hay demanda del bien que ayuda a producir. Técnicamente, la demanda de trabajo se obtiene de la maximización del beneficio de la empresa y varía con el precio y la demanda del producto.

Demanda efectiva (Chamberlin) Demanda que la empresa tiene de verdad cuando sus rivales responden a sus cambios de precio. Técnicamente, es la curva de demanda que refleja las ventas de la empresa cuando todos los competidores ajustan sus precios y cada una mantiene su cuota de mercado.

Demanda hicksiana Demanda que se observaría si, tras cada cambio de precio, se ajustara la renta del consumidor para que conservara su nivel de utilidad. Técnicamente, también llamada compensada, resulta de minimizar el gasto para una utilidad dada, depende de los precios y de esa utilidad y solo recoge el efecto sustitución.

Demanda incondicionada de factores Demanda de un factor cuando la empresa decide a la vez cuánto producir y cómo hacerlo. Técnicamente, resulta de maximizar el beneficio, depende de los precios de los factores y del precio del producto, y equivale a la demanda condicionada evaluada en la producción que maximiza el beneficio.

Demanda inducida Consumo de servicios sanitarios que deciden los propios profesionales más que la necesidad del paciente. Técnicamente, como el médico conoce mejor que el paciente el tratamiento necesario y decide por él, en los sistemas de producción privada con financiación pública puede recomendar más servicios de los precisos, lo que dificulta controlar el gasto.

Demanda marshalliana Función que indica cuánto compra el consumidor de cada bien según los precios y su renta, también llamada demanda ordinaria. Técnicamente, resulta de maximizar la utilidad sujeta a la restricción presupuestaria, es homogénea de grado cero en precios y renta, por lo que no hay ilusión monetaria, y agota toda la renta disponible.

Demanda ordinaria de factores Cantidad de cada factor que la empresa emplea para obtener la máxima producción con un coste dado. Técnicamente, también llamada marshalliana, depende de los precios de los factores y del coste, es la solución del problema primal y es homogénea de grado cero en ambos.

Demanda planeada (Chamberlin) Demanda que la empresa espera tener al cambiar su precio, porque cree que sus rivales no van a reaccionar. Técnicamente, es la curva de demanda que la empresa cree afrontar suponiendo que los competidores no reaccionan a sus cambios de precio, y es más elástica que la efectiva.

Demanda quebrada Curva de demanda con un codo en el precio vigente, porque los rivales imitan las bajadas de precio pero no las subidas. Técnicamente, coincide con la demanda planeada por encima de ese precio y con la efectiva por debajo. Surge en el análisis de Chamberlin, pero fue Sweezy (1939) quien la convirtió en un modelo de oligopolio.

Demandas recíprocas Idea de que lo que cada país quiere comprar del otro determina a qué precio relativo comercian. Técnicamente, es la demanda de cada país de los productos importados, que determina a la vez su oferta de exportaciones y con la que Mill fijó la relación real de intercambio en el punto en que el valor de las exportaciones iguala al de las importaciones.

Democracia directa Sistema en el que los propios ciudadanos votan las decisiones públicas, como en un referéndum. Técnicamente, cada individuo dispone de un voto, y las decisiones presupuestarias se adoptan aplicando a sus preferencias una

 © 2026 Manuel Cabado García

regla de votación fijada de antemano.

Democracia representativa Sistema en el que los ciudadanos eligen a representantes que deciden por ellos. Técnicamente, las decisiones presupuestarias se votan de forma indirecta a través de los parlamentarios, lo que reduce los costes de negociación pero introduce en el proceso a políticos, funcionarios y grupos de presión con objetivos propios.

Derechos residuales de control Poder de decidir qué se hace con un bien en todo aquello que el contrato no ha previsto. Técnicamente, según Grossman y Hart (1986), son los derechos de control sobre un activo no asignados por contrato, cuya titularidad define la propiedad y aumenta los incentivos de su dueño a realizar inversiones específicas.

Desagradable aritmética monetarista Situación en la que un déficit público persistente acaba atando las manos del banco central frente a la inflación. Técnicamente, es la situación descrita por Sargent y Wallace en la que una política fiscal sistemáticamente expansiva impide a la política monetaria controlar la inflación, porque financiar el déficit con deuda acaba siendo más inflacionista que monetizarlo.

Descomposición de Oaxaca-Blinder Método que separa una brecha salarial en la parte debida a diferencias de características y la que no se explica por ellas. Técnicamente, descompone la diferencia de los salarios medios en logaritmos en un término de dotaciones, valorado a los rendimientos de un grupo, y otro de diferencias en los rendimientos, que se asocia, con cautela, a la discriminación.

Descremado del mercado (cream skimming) Entrada de un competidor solo en la parte más rentable de un mercado. Técnicamente, en un monopolio natural un entrante menos eficiente puede servir únicamente los segmentos rentables, lo que eleva el coste total de abastecer la demanda y puede hacer insostenible al incumbente. La regla del precio eficiente de los componentes impide que entren rivales menos eficientes.

Deseconomías de escala Desventaja de coste que sufre la empresa al crecer, porque cada unidad le sale más cara. Técnicamente, es la situación en la que el coste total crece en mayor proporción que la producción, de modo que el coste medio es creciente e inferior al coste marginal.

Desempleo friccional Paro que existe porque encontrar trabajo, o a quien contratar, lleva tiempo, aunque haya puestos para todos. Técnicamente, es el desempleo que generan las fricciones de búsqueda y emparejamiento entre parados y vacantes, y es compatible con un exceso de demanda de trabajo nulo.

Desempleo involuntario Paro de quienes desean trabajar al salario vigente, o incluso a uno menor, y no encuentran empleo. Técnicamente, para Keynes se debe a una demanda agregada insuficiente, por lo que no desaparece con rebajas salariales, mientras que en las teorías de la rigidez salarial surge porque el salario se fija por encima del que vaciaría el mercado, como con un sindicato monopolista, con salarios de eficiencia o en el modelo de insiders-outsiders.

Desempleo voluntario Paro de quienes, al salario vigente, prefieren no trabajar. Técnicamente, es el único desempleo posible en el modelo neoclásico, en el que el salario se ajusta hasta vaciar el mercado, de modo que solo está sin empleo quien tiene un salario de reserva superior al salario vigente.

Desequilibrio horizontal Diferencia de recursos entre jurisdicciones de un mismo nivel de gobierno para prestar servicios parecidos. Técnicamente, se debe a sus distintas capacidades fiscales y necesidades de gasto, no desaparece al corregir el desequilibrio vertical y exige llevar recursos de las regiones ricas a las pobres mediante la nivelación.

Desequilibrio vertical Desajuste entre lo que cada nivel de gobierno tiene que gastar y lo que puede recaudar. Técnicamente, aparece cuando el gobierno central conserva las principales fuentes tributarias y los gobiernos territoriales asumen buena parte de las competencias de gasto, y se corrige con transferencias entre niveles de gobierno.

Dicotomía clásica Idea de que la cantidad de dinero no influye en lo que realmente se produce y se consume, sino solo en los precios. Técnicamen-

te, es la separación entre variables reales y nominales propia de la macroeconomía clásica, según la cual las variables reales se determinan sin intervención de las nominales. De ella se deriva la neutralidad del dinero: los cambios en la cantidad de dinero solo afectan a los precios.

Diferenciación de productos Situación en la que las empresas venden versiones distintas de un mismo tipo de bien que los consumidores no consideran idénticas. Técnicamente, los productos son sustitutivos cercanos pero imperfectos por su localización, calidad, diseño o servicios asociados, lo que da a cada empresa una demanda con pendiente negativa y cierto poder sobre el precio de su variedad.

Diferenciación horizontal Diferencia entre productos sobre la que los consumidores no se ponen de acuerdo acerca de cuál es mejor, como el color de un jersey. Técnicamente, a igualdad de precio unos consumidores prefieren una variedad y otros otra, porque no hay unanimidad sobre la característica que distingue a los productos y sus gustos son heterogéneos.

Diferenciación máxima Tendencia de las empresas a alejarse todo lo posible unas de otras en el espacio de características para suavizar la competencia en precios. Técnicamente, es el resultado que d'Aspremont, Gabszewicz y Thisse (1979) obtuvieron en el modelo de Hotelling con costes de transporte cuadráticos, en el que el efecto competición supera al efecto demanda y las empresas se sitúan en los extremos.

Diferenciación vertical Diferencia entre productos que todos los consumidores ordenan de la misma forma, como un coche más o menos potente. Técnicamente, a igualdad de precio todos preferirían la mejor versión, y las de peor calidad sobreviven porque los consumidores difieren en su disposición a pagar por la calidad.

Diferencias salariales compensatorias Diferencias de salario que pagan las condiciones desagradables o peligrosas de un empleo. Técnicamente, entre empleos que exigen la misma cualificación, el mercado competitivo iguala las ventajas netas, no los salarios, por lo que los empleos con peores atributos pagan más, como recoge la función salarial hedónica de Rosen.

Difusión Reacción de quien, ante un impuesto, compra menos para pagar menos. Técnicamente, es la disminución voluntaria del consumo para eludir parte de la carga de un tributo.

Difusión del riesgo (risk spreading) Reparto del coste de un riesgo entre muchas personas, de modo que a cada una le corresponde una parte pequeña. Técnicamente, al dividir un riesgo dado entre muchos individuos el colectivo se comporta como si fuera neutral al riesgo aunque cada miembro sea averso, como ocurre con las acciones de una empresa o los créditos sindicados. A diferencia de la puesta en común del riesgo, no agrupa riesgos independientes, sino que fracciona uno solo.

Dilema del prisionero Situación en la que dos jugadores que persiguen cada uno su propio interés acaban peor que si hubieran cooperado. Técnicamente, la estrategia dominante de cada jugador es no cooperar, y el equilibrio de Nash resultante no es un óptimo de Pareto, como ocurre con la tentación de incumplir un acuerdo de cártel.

Discriminación de precios Venta del mismo producto a precios distintos a clientes distintos. Técnicamente, es la práctica del monopolista que cobra precios distintos según el consumidor o las unidades vendidas para aumentar su beneficio. Exige distinguir consumidores con distinta disposición a pagar y que no sea posible el arbitraje.

Discriminación de precios de cuarto grado Cobro de precios distintos según el momento de la compra, más caros al principio y más baratos después. Técnicamente, es la discriminación intertemporal que cobra precios altos al principio a los consumidores más impacientes y los reduce después para atraer a los más pacientes.

Discriminación de precios de primer grado Cobro a cada cliente del máximo que está dispuesto a pagar. Técnicamente, es la discriminación perfecta en la que cada unidad se vende al precio máximo que su comprador está dispuesto a pagar, de modo que el monopolista se apropia de todo el excedente y el resultado es eficiente.

Discriminación de precios de segundo grado Discriminación en la que el vendedor ofrece a

 © 2026 Manuel Cabado García

todos el mismo menú de precios y cada consumidor elige cuánto comprar. Técnicamente, es una fijación no lineal de precios: el precio por unidad varía con la cantidad adquirida, como en los descuentos por volumen o las tarifas en bloques.

Discriminación de precios de tercer grado Cobro de precios distintos a distintos grupos de clientes que se pueden identificar. Técnicamente, el monopolista segmenta el mercado por características observables de los consumidores y cobra a cada grupo un precio distinto, mayor cuanto menor sea su elasticidad.

Diseño de mecanismos Rama de la teoría económica que diseña las reglas de una interacción para que cada participante, buscando su propio interés, contribuya al resultado que se pretende. Técnicamente, es el enfoque de Hurwicz, Maskin y Myerson que diseña las reglas de un juego para que agentes con información privada, persiguiendo su propio interés, alcancen el resultado deseado.

Diversificación Reparto de una inversión entre activos distintos para no depender de uno solo. Técnicamente, al combinar activos cuyos rendimientos no están perfectamente correlacionados, la varianza de la cartera se reduce sin rebajar su rentabilidad esperada.

Divina coincidencia Caso favorable en el que el banco central no tiene que elegir entre controlar la inflación y estabilizar la actividad. Técnicamente, es el resultado de Blanchard y Galí según el cual, sin rigideces salariales, estabilizar la inflación equivale a estabilizar la brecha de producción, de modo que ambos objetivos del banco central no entran en conflicto.

Divisibilidad Supuesto de que cualquier forma de producir puede llevarse a cabo a menor tamaño. Técnicamente, es la propiedad del conjunto de producción por la que cualquier plan factible puede reducirse de escala tanto como se desee, lo que implica rendimientos constantes o decrecientes.

División del trabajo Reparto de las tareas de producción de modo que cada trabajador se dedique a una sola parte del proceso. Técnicamente, es la especialización de cada trabajador en una fase del proceso productivo, que Smith consideró el principal determinante de la productividad y que, según él, depende de la extensión del mercado y de la acumulación de capital.

Doble dividendo Idea de que un impuesto ecológico logra dos beneficios a la vez: menos contaminación y un sistema fiscal mejor. Técnicamente, Pearce sostuvo que la recaudación del impuesto corrector permite rebajar otros impuestos distorsionantes, de modo que se gana eficiencia además de internalizar la externalidad.

Doctrina del fondo de salarios Idea clásica de que los salarios se pagan con una cantidad de capital ya acumulada. Técnicamente, el salario medio es el cociente entre ese fondo, adelantado por los capitalistas antes de vender la producción, y el número de trabajadores.

Dominancia estocástica de primer orden Situación en la que una apuesta es mejor que otra para cualquiera que prefiera tener más dinero. Técnicamente, es la relación por la que una distribución es preferida a otra por cualquier agente que prefiera más riqueza a menos, sea cual sea su actitud ante el riesgo.

Dominancia estocástica de segundo orden Situación en la que, entre apuestas que dan lo mismo en promedio, una es mejor que otra para cualquiera a quien le disguste el riesgo. Técnicamente, es la relación por la que, entre loterías con igual valor esperado, una distribución es preferida a otra por cualquier agente averso al riesgo.

Dominio no restringido Exigencia de que el procedimiento de decisión sirva sean cuales sean las preferencias de la gente. Técnicamente, es el axioma de Arrow, también llamado universalidad, que exige que la regla de elección colectiva funcione para todas las configuraciones posibles de preferencias individuales.

Dualidad del mercado de trabajo División de los asalariados entre quienes tienen un empleo estable y bien protegido y quienes encadenan contratos temporales. Técnicamente, surge de la diferencia de costes de despido entre contratos indefinidos y temporales, que hace recaer el ajuste del empleo sobre los temporales, y es una forma

de la división entre insiders y outsiders.

E

Economía del bienestar Parte de la economía que juzga qué situaciones son mejores o peores para la sociedad en su conjunto. Técnicamente, es la rama de la teoría económica que valora, desde un punto de vista colectivo, la deseabilidad social de los estados alternativos en que puede encontrarse un sistema económico.

Economía del comportamiento Rama de la economía que estudia cómo deciden realmente las personas, con sus errores y atajos mentales. Técnicamente, incorpora al análisis de la elección los sesgos sistemáticos que observa la psicología, como la aversión a las pérdidas o la inercia, en lugar de suponer una racionalidad ilimitada.

Economías de aglomeración Ventajas que obtienen empresas y personas por estar cerca unas de otras. Técnicamente, son reducciones de costes derivadas de la concentración geográfica, que se dividen en economías de localización, dentro de un mismo sector, y de urbanización, entre sectores diversos.

Economías de alcance Ahorro que se consigue al fabricar varios productos en la misma empresa. Técnicamente, también llamadas economías de gama o de variedad, se dan cuando producir varios bienes en una sola empresa resulta más barato que producirlos por separado en varias.

Economías de aprendizaje Abaratamiento de la producción que se logra con la experiencia acumulada. Técnicamente, también llamadas economías de escala dinámicas, son reducciones de costes que se obtienen a medida que crece la producción acumulada gracias al aprendizaje de trabajadores, directivos y proveedores.

Economías de escala Ventaja de coste que obtiene la empresa al hacerse más grande, porque cada unidad le sale más barata. Técnicamente, es la situación en la que el coste total crece en menor proporción que la producción, de modo que el coste medio es decreciente y superior al coste marginal. Es un concepto de la teoría de costes, distinto de los rendimientos a escala.

Economías de escala externas Ventajas de coste que obtiene cada empresa por el crecimiento de su sector, no por el suyo propio. Técnicamente, pueden deberse a factores más baratos (pecuniarias) o a mejoras tecnológicas compartidas, y hacen que la oferta de la industria a largo plazo tenga pendiente negativa.

Economías de escala internas Ahorros de coste que consigue una empresa gracias a su propio crecimiento. Técnicamente, son reducciones del coste medio que resultan de la expansión de la producción de la propia empresa y quedan recogidas en la forma de su curva de costes.

Ecuación de Euler Regla que indica cómo repartir el consumo entre hoy y mañana: el último euro consumido hoy debe dar la misma utilidad que ahorrarlo y consumir mañana lo que rinda. Técnicamente, es la condición de primer orden del problema intertemporal del consumidor, según la cual la utilidad marginal del consumo presente iguala a la del consumo futuro multiplicada por uno más el tipo de interés y por el factor de descuento, que es la inversa de uno más la tasa de descuento. Equivale a igualar el cociente de utilidades marginales al cociente entre uno más el tipo de interés y uno más la tasa de descuento.

Ecuación de Mincer Ecuación que relaciona el salario con los años de estudio y de experiencia laboral. Técnicamente, explica el logaritmo del salario con la escolarización, la experiencia y su cuadrado, y su coeficiente de la escolarización se interpreta como el rendimiento de un año adicional de educación.

Ecuación de Slutsky Relación que separa las dos razones por las que cambia lo que se compra cuando varía un precio: el bien pasa a ser relativamente más caro o más barato y el consumidor se vuelve más pobre o más rico. Técnicamente, descompone el efecto total de la variación de un precio sobre la demanda marshalliana en un efecto sustitución, dado por la demanda hicksiana, y un efecto renta.

Efecto Averch-Johnson Tendencia de las empresas reguladas a invertir en capital más de lo necesario. Técnicamente, es la sobrecapitalización de la empresa sometida a regulación de la tasa de

 © 2026 Manuel Cabado García

rendimiento del capital, que elige una combinación ineficiente de factores con mayores costes y precios.

Efecto certeza Valor desproporcionado que se concede a lo que es completamente seguro. Técnicamente, en la teoría de las perspectivas, pasar de la certeza a una probabilidad algo inferior se percibe como una pérdida de valor mucho mayor que la que corresponde a esa diferencia de probabilidad, y es el mecanismo que explica la paradoja de Allais.

Efecto competición (modelo de Hotelling) Inconveniente de situarse cerca del rival, porque obliga a competir con más dureza en precios. Técnicamente, en el modelo de Hotelling con localización endógena, es la intensificación de la competencia en precios cuando una empresa se acerca a su rival y ambas se diferencian menos.

Efecto concentración Paso de tareas de las administraciones locales al gobierno central en tiempos de crisis. Técnicamente, es el traspaso de funciones de las administraciones locales a la central durante las perturbaciones sociales, por considerarse que esta actúa con más eficacia.

Efecto de red Aumento del valor que un bien tiene para cada usuario a medida que lo usan más personas. Técnicamente, también llamado externalidad de red o economía de escala del lado de la demanda, hace que la utilidad de cada usuario dependa del número de usuarios de productos compatibles. Genera una retroalimentación positiva que puede favorecer la concentración del mercado, y no debe confundirse con las economías de escala externas, que son reducciones de costes. Es directo cuando el valor depende del número de usuarios del mismo tipo, como en la telefonía, y cruzado cuando depende de los usuarios del otro lado de una plataforma.

Efecto demanda (modelo de Hotelling) Ventaja que obtiene una empresa al situarse cerca del centro, donde puede atraer a más clientes. Técnicamente, en el modelo de Hotelling con localización endógena, es el aumento del mercado de una empresa al acercarse al centro de la ciudad.

Efecto desplazamiento de Peacock y Wiseman Subida brusca del gasto público tras una crisis grave que después no se deshace. Técnicamente, es el aumento discontinuo del gasto público, descrito por Peacock y Wiseman, que se produce porque tras una conmoción social los contribuyentes aceptan niveles impositivos más altos y el gasto no vuelve después a su nivel inicial.

Efecto dotación Tendencia a valorar más un bien por el simple hecho de tenerlo. Técnicamente, consiste en exigir, por desprenderse de un bien que se posee, más dinero del que se estaría dispuesto a pagar por adquirirlo.

Efecto escala (demanda de factores) Parte de la respuesta de la demanda de un factor a su precio que se debe al cambio en la cantidad producida. Técnicamente, también llamado efecto producción, recoge que una subida del salario eleva el coste marginal, reduce la producción que maximiza el beneficio y, con ella, el trabajo demandado. Como refuerza al efecto sustitución, no existen factores Giffen.

Efecto expulsión (crowding-out) Reducción del gasto privado que provoca el aumento del gasto público. Técnicamente, el mayor gasto público, sobre todo si se financia con deuda, eleva los tipos de interés y desplaza la inversión privada, de modo que la política fiscal expansiva tiene menos efecto del previsto. La expulsión es total si la caída del gasto privado iguala el aumento del público.

Efecto externo interjurisdiccional Efecto que las decisiones de un gobierno territorial tienen sobre los habitantes de otros territorios. Técnicamente, es el efecto de esas decisiones, sobre todo de su gasto público, sobre el bienestar de los residentes de otras jurisdicciones, que el gobierno que lo causa no tiene en cuenta al decidir. Si es positivo, lleva a infraproveer el bien, lo que justifica una transferencia condicional compensatoria del gobierno central para internalizarlo.

Efecto herencia Aumento del ahorro de los padres que quieren compensar a sus hijos por lo que estos pagan para financiar las pensiones. Técnicamente, ante la transferencia intergeneracional que supone el sistema de reparto, los individuos con motivo herencia elevan su ahorro para aumentar

los legados y anular el efecto de esa transferencia sobre la renta de sus hijos.

Efecto inspección Toma de conciencia, provocada por una crisis, de necesidades sociales que antes pasaban inadvertidas. Técnicamente, en el análisis de Peacock y Wiseman, descubrimiento durante las conmociones sociales de ámbitos en los que la actuación pública es insuficiente, que refuerza la disposición a financiar un mayor gasto y contribuye a que este no vuelva a su nivel inicial.

Efecto jubilación Aumento del ahorro que se produce porque las pensiones adelantan la jubilación y hay que financiar más años sin salario. Técnicamente, al inducir una jubilación anticipada, el sistema alarga el periodo de desahorro y acorta el de acumulación, lo que eleva el ahorro durante la vida activa y compensa en parte el efecto sustitución de la riqueza.

Efecto papel matamoscas Tendencia del dinero que un gobierno recibe de otro a quedarse en el gasto público en lugar de volver a los ciudadanos. Técnicamente, una transferencia incondicional a una jurisdicción eleva su gasto público más que un aumento de igual cuantía de la renta privada de sus residentes. También llamado efecto adherencia (flypaper effect).

Efecto producción Caída de la producción de un bien que un impuesto encarece, que perjudica a los factores que más se usan en él. Técnicamente, es el cambio en la composición de la producción provocado por un impuesto que encarece relativamente un bien, lo que reduce su demanda y la de los factores empleados intensivamente en él.

Efecto renta Parte de la reacción de la demanda ante un cambio de precio que se debe a que el consumidor, con la misma renta monetaria, se ha vuelto más rico (si el precio baja) o más pobre (si sube). Técnicamente, es la variación de la demanda debida al cambio de la renta real, y su sentido depende de que el bien sea normal o inferior.

Efecto renta de un impuesto Empobrecimiento del contribuyente por el simple hecho de pagar el impuesto, que el Estado recupera como recaudación. Técnicamente, es la reducción de renta real que produciría un impuesto de suma fija de igual recaudación, por lo que es igual para todos los impuestos que recaudan lo mismo y no genera exceso de gravamen.

Efecto renta dotación Parte de la respuesta a una subida salarial que lleva a trabajar menos porque el tiempo disponible vale más y la persona es más rica. Técnicamente, es el efecto del aumento de la renta potencial, igual al valor de la dotación de tiempo, que, si el ocio es un bien normal, eleva su consumo y supera al efecto renta ordinario cuando el individuo trabaja.

Efecto reordenación Cambio de puestos en la clasificación por renta que provoca un impuesto. Técnicamente, aparece cuando algunos contribuyentes pasan a tener menos renta después de impuestos que otros que tenían menos que ellos antes. Hace que la reducción del índice de Gini sea menor que la que se deduciría solo del tipo medio y de la progresividad del impuesto medida por el índice de Kakwani.

Efecto sustitución Parte de la reacción de la demanda ante un cambio de precio que se debe a que el bien se ha vuelto más barato o más caro en comparación con los demás. Técnicamente, es la variación de la demanda que se produce sin que cambie la renta real del consumidor, y siempre lleva a comprar más del bien que se abarata y menos del que se encarece.

Efecto sustitución de la riqueza Tendencia a ahorrar menos por contar con una pensión pública en el futuro. Técnicamente, es la reducción del ahorro privado que se produce porque los individuos perciben sus cotizaciones como riqueza futura, lo que en un sistema de reparto puede disminuir el ahorro nacional.

Efecto sustitución de las pensiones Aumento del gasto en pensiones que se produce porque las pensiones nuevas son más altas que las que se extinguen. Técnicamente, al reemplazar las bajas por altas de mayor cuantía, la pensión media y el gasto crecen aunque no aumente el número de pensionistas. En España, las nuevas pensiones son en torno a un 30 % superiores a las que se extinguen.

Efecto sustitución de un impuesto Cambio de conducta del contribuyente para pagar menos im-

 © 2026 Manuel Cabado García

puestos, desplazando su gasto o su esfuerzo hacia lo que no está gravado. Técnicamente, es la reasignación que provoca la alteración de los precios relativos, que reduce el bienestar sin aumentar la recaudación y origina el exceso de gravamen.

Efecto sustitución en la oferta de trabajo Aumento de las horas trabajadas cuando sube el salario porque el ocio se vuelve más caro. Técnicamente, es la variación del ocio ante un cambio del salario real manteniendo constante la utilidad, que siempre reduce el ocio y eleva la oferta de trabajo.

Efecto sustitución factorial Tendencia de las empresas a usar menos el factor productivo gravado y más el que no lo está. Técnicamente, es la sustitución del factor gravado por el no gravado en un sector cuando un impuesto encarece uno de los factores, lo que altera la remuneración relativa de ambos.

Efecto trinquete (consumo) Resistencia de las familias a reducir su consumo cuando cae su renta. Técnicamente, en la teoría de la renta relativa de Duesenberry el consumo depende también de la renta más alta alcanzada en el pasado, por lo que ante una caída de la renta se recorta el ahorro antes que el nivel de vida y el consumo resulta rígido a la baja.

Efecto trinquete (regulación) Tendencia a esforzarse menos cuando hacerlo bien hoy hace que mañana se exija más. Técnicamente, es la ocultación de eficiencia o la reducción del esfuerzo de una empresa regulada que anticipa que sus buenos resultados endurecerán el contrato o el factor X de la siguiente revisión.

Efecto Wicksell Cambio en el valor del capital de una economía cuando varía el tipo de interés. Técnicamente, combina un efecto precio, la revalorización de los bienes de capital existentes, y un efecto real, el cambio de técnica, de modo que con varios bienes de capital la relación entre el tipo de interés y la cantidad de capital no tiene por qué ser monótona.

Eficiencia asignativa Se produce la cantidad de cada bien que la sociedad desea, ni más ni menos. Técnicamente, se da cuando el beneficio marginal de la última unidad para el consumidor coincide con su coste marginal de producción, lo que maximiza el excedente total. En la medición de la eficiencia de la empresa, uso de la combinación de factores que minimiza el coste dados sus precios.

Eficiencia dinámica (acumulación de capital) Situación en la que la economía no ahorra de más: ninguna generación podría consumir más sin que otra consumiera menos. Técnicamente, una senda es dinámicamente eficiente si el capital por trabajador no supera al de la regla de oro, lo que equivale a que el tipo de interés no sea inferior a la tasa de crecimiento de la economía.

Eficiencia económica Uso de los recursos que no desperdicia ninguno de ellos. Técnicamente, en economía del bienestar se identifica con el óptimo de Pareto y no garantiza por sí sola el máximo bienestar social. En la teoría de la empresa, designa la combinación de factores que, además de técnicamente eficiente, es la de menor coste dados sus precios.

Eficiencia marginal del capital Rentabilidad que el empresario espera obtener de un nuevo bien de capital. Técnicamente, para Keynes es el tipo de descuento que iguala el valor actual de los rendimientos esperados de ese bien con su coste de reposición, y se invierte mientras supere al tipo de interés.

Eficiencia productiva Se produce sin desperdiciar recursos, es decir, al menor coste posible. Técnicamente, en competencia perfecta a largo plazo se alcanza porque cada empresa opera en el mínimo de su coste medio.

Eficiencia técnica Situación en la que no se desperdician factores productivos. Técnicamente, una combinación de factores es técnicamente eficiente si con ella no puede obtenerse más producto ni el mismo producto con menos cantidad de algún factor, y la función de producción solo recoge combinaciones de este tipo.

Ejército industrial de reserva Conjunto de parados que el capitalismo mantiene siempre disponibles. Técnicamente, según Marx, es la reserva de parados que el capitalismo genera al sustituir trabajadores por máquinas y en las crisis, y su función es presionar los salarios a la baja, pues

cualquier ocupado puede ser reemplazado por un desempleado.

Elasticidad de escala Medida de cuánto crece la producción cuando se aumentan todos los factores a la vez. Técnicamente, es la variación porcentual de la producción ante una misma variación porcentual y proporcional de todos los factores. Coincide con la suma de las elasticidades producto y, si la función es homogénea, con su grado.

Elasticidad de Frisch Cuánto cambian las horas trabajadas ante un cambio pasajero o previsto del salario. Técnicamente, es la elasticidad de las horas respecto al salario manteniendo constante la utilidad marginal de la riqueza, siempre mayor que la hicksiana y, si el ocio es un bien normal, también que la marshalliana, y es clave en la teoría del ciclo real.

Elasticidad de sustitución Medida de la facilidad con que se sustituye un bien por otro en el consumo, o un factor por otro en la producción. Técnicamente, es la variación porcentual de la proporción en que se combinan ante una variación porcentual de la relación marginal de sustitución, o de la relación marginal de sustitución técnica. Es infinita para sustitutivos perfectos y nula para complementarios perfectos.

Elasticidad de sustitución entre variedades Grado de facilidad con que el consumidor cambia una variedad por otra cuando se abarata. Técnicamente, es la variación porcentual de la cantidad relativa demandada ante una variación del 1 % del precio relativo y, en el modelo de Dixit-Stiglitz, es igual a la inversa de la preferencia por la variedad.

Elasticidad de sustitución intertemporal Disposición a trasladar consumo de un periodo a otro cuando cambia lo que cuesta hacerlo. Técnicamente, es la variación porcentual del cociente entre consumo futuro y presente ante una variación porcentual de su precio relativo, que depende del tipo de interés, y en la utilidad esperada estándar con aversión relativa constante es la inversa del coeficiente de aversión relativa al riesgo.

Elasticidad del coste total Medida de cuánto aumentan los costes cuando la empresa produce más. Técnicamente, es la variación porcentual del coste total ante una variación porcentual de la producción, y sirve para medir las economías de escala.

Elasticidad precio cruzada Medida de cómo reacciona la demanda de un bien cuando cambia el precio de otro. Técnicamente, es la variación porcentual de la demanda de un bien ante una variación porcentual del precio de otro. Es positiva para sustitutivos brutos, nula para bienes independientes y negativa para complementarios.

Elasticidad precio de la demanda Medida de la sensibilidad de la cantidad demandada de un bien a su propio precio. Técnicamente, es la variación porcentual de la cantidad ante una variación porcentual del precio. La demanda es elástica si su valor absoluto supera la unidad, inelástica si es inferior y de elasticidad unitaria si es igual a uno, caso en que el gasto en el bien no cambia al variar el precio.

Elasticidad producto (de un factor) Sensibilidad de la producción a la cantidad empleada de un factor. Técnicamente, es la variación porcentual del producto ante una variación porcentual de ese factor, con los demás constantes, y la suma de las de todos los factores es la elasticidad de escala.

Elasticidad renta Medida de cómo reacciona la demanda de un bien cuando cambia la renta del consumidor. Técnicamente, es la variación porcentual de la cantidad demandada de un bien ante una variación porcentual de la renta.

Eliminación gratuita Supuesto de que a la empresa no le cuesta nada deshacerse de lo que le sobra. Técnicamente, es el axioma según el cual, si un plan de producción es factible, también lo es cualquier otro con iguales o mayores inputs e iguales o menores outputs, de modo que la empresa puede deshacerse de las mercancías sobrantes.

Elusión fiscal Reducción legal de la carga tributaria. Técnicamente, consiste en aprovechar las disposiciones de la propia normativa para pagar menos impuestos, por lo que, a diferencia de la evasión, es lícita.

Empresas superestrella Grandes empresas muy productivas que concentran una parte creciente de las ventas de su sector. Técnicamente, Autor,

 © 2026 Manuel Cabado García

Dorn, Katz, Patterson y Van Reenen (2020) mostraron que tienen márgenes altos y una participación del trabajo en el valor añadido muy baja, y que su expansión explica buena parte de la caída de la participación del trabajo en la renta del conjunto de la economía, aunque no caiga dentro de cada empresa.

Enfermedad de Baumol Tendencia de los servicios públicos a encarecerse con el tiempo en comparación con el resto de la economía. Técnicamente, es el aumento progresivo del precio relativo de los servicios públicos, intensivos en trabajo, porque su productividad crece menos que en el sector privado mientras los salarios crecen a un ritmo similar.

Enfoque cardinal Forma de estudiar la demanda que supone que la satisfacción del consumidor puede medirse con números, como un peso o una longitud. Técnicamente, es el enfoque, asociado a Marshall, que supone una utilidad medible, aditiva, separable y con utilidad marginal del dinero constante, de modo que la demanda de cada bien depende solo de su propio precio.

Enfoque de direcciones Forma de modelizar productos diferenciados en la que cada variedad ocupa un lugar en un espacio de características y cada consumidor prefiere la más próxima a su ideal. Técnicamente, las preferencias se definen sobre las características, de modo que cada variedad compite sobre todo con sus vecinas, como en los modelos de Hotelling, Salop y Lancaster.

Enfoque de no direcciones Forma de modelizar productos diferenciados en la que todas las variedades compiten por igual entre sí y el consumidor reparte su gasto entre muchas. Técnicamente, las preferencias se definen sobre los propios bienes, sustitutivos imperfectos, como en los modelos de Chamberlin y de Dixit-Stiglitz.

Enfoque ordinal Forma de estudiar la demanda que solo pide al consumidor saber qué prefiere, no cuánto. Técnicamente, es el enfoque que solo exige que el consumidor sea capaz de ordenar sus cestas según sus preferencias, sin medir la utilidad ni compararla entre individuos.

Enfoque TOTEX Forma de retribuir las redes que trata igual un euro de inversión que un euro de gasto corriente. Técnicamente, compara la inversión y los gastos de operación de forma conjunta con unos costes de referencia y reparte con la empresa los márgenes obtenidos, lo que reduce el sesgo hacia el capital que, según el modelo de Averch y Johnson, provoca la regulación de la tasa de rendimiento. En España lo aplica a la distribución eléctrica la Circular 8/2025 de la CNMC.

Equidad horizontal Principio de tratar igual a los iguales. Técnicamente, quienes tienen la misma capacidad de pago, medida normalmente por la renta, deben pagar los mismos impuestos.

Equidad vertical Principio de tratar de forma distinta a quienes están en distinta situación económica. Técnicamente, quienes tienen mayor capacidad de pago deben contribuir más, y cuánto más, de forma proporcional o progresiva, depende de la concepción de justicia distributiva que se adopte.

Equilibrio agrupador Situación en la que todos actúan igual, de modo que quien carece de la información no puede saber quién es quién. Técnicamente, es el equilibrio, también llamado pooling equilibrium, en el que agentes de distinto tipo toman la misma decisión, de modo que la parte no informada no puede distinguirlos.

Equilibrio bayesiano de Nash Equilibrio de Nash de un juego en el que cada jugador conoce sus propias características pero no las de sus rivales. Técnicamente, cada jugador elige una acción para cada uno de sus tipos posibles de forma que maximiza su pago esperado, dadas las reglas de los demás y la probabilidad con que se presenta cada tipo.

Equilibrio con controles de precios Resultado de un mercado cuando la ley fija un precio máximo o mínimo. Técnicamente, es la combinación de precio y cantidad en la que el precio respeta el límite legal y minimiza la diferencia entre oferta y demanda dentro del rango permitido, y la cantidad es la menor entre la ofrecida y la demandada.

Equilibrio de competencia perfecta Situación en la que, al precio vigente, todo lo que se quiere vender se compra y nadie desea cambiar su decisión. Técnicamente, es el precio y el nivel de

producción en los que el mercado se vacía, los consumidores maximizan su utilidad y los productores su beneficio.

Equilibrio de Nash Situación en la que nadie se arrepiente de su decisión al ver lo que han hecho los demás. Técnicamente, es la combinación de estrategias en la que ningún jugador tiene incentivos a cambiar unilateralmente su decisión, dada la estrategia de los demás.

Equilibrio de Nash perfecto en subjuegos Equilibrio en el que nadie amenaza con algo que luego no le convendría cumplir. Técnicamente, es el perfil de estrategias que constituye un equilibrio de Nash en todo subjuego y satisface, por tanto, la racionalidad secuencial.

Equilibrio del consumidor Elección que deja al consumidor lo más satisfecho posible con la renta de que dispone. Técnicamente, es la cesta que maximiza la utilidad sujeta a la restricción presupuestaria y, si la solución es interior, en ella la relación marginal de sustitución se iguala al cociente de precios y la curva de indiferencia es tangente a la recta de balance.

Equilibrio general competitivo Situación en la que todos los mercados de una economía funcionan a la vez sin excedentes ni escaseces. Técnicamente, es un vector de precios y de asignaciones en el que los consumidores maximizan su utilidad, las empresas maximizan beneficios y todos los mercados se vacían simultáneamente.

Equilibrio parcial Análisis de un único mercado como si estuviera aislado del resto de la economía, enfoque asociado a Marshall. Técnicamente, mantiene constantes los demás precios y la renta, lo que se justifica cuando el mercado es pequeño y las preferencias son cuasilineales respecto de un bien numerario que agrupa el gasto en todos los demás bienes.

Equilibrio separador Situación en la que cada tipo de agente actúa de forma distinta, de modo que su comportamiento delata lo que es. Técnicamente, es el equilibrio en el que agentes de distinto tipo toman decisiones diferentes respecto a la señal, lo que revela su tipo a la parte no informada.

Equity premium puzzle Dificultad para explicar por qué las acciones rinden a largo plazo mucho más que la deuda pública. Técnicamente, el diferencial observado exigiría una aversión al riesgo poco realista en la utilidad esperada estándar, y se ha explicado con la aversión miope a las pérdidas o con preferencias de Epstein-Zin y riesgo a largo plazo.

Equivalencia ricardiana Idea de que da igual pagar el gasto público con impuestos hoy o con deuda, porque la deuda de hoy son los impuestos de mañana. Técnicamente, es la proposición, recuperada por Barro, según la cual el efecto de un aumento del gasto público no depende de que se financie con impuestos o con deuda, porque los agentes anticipan los impuestos futuros y no consideran la deuda riqueza neta.

Equivalencias impositivas Combinaciones distintas de impuestos que producen exactamente los mismos efectos. Técnicamente, son resultados del modelo de Harberger según los cuales distintas combinaciones de impuestos sobre bienes y factores tienen los mismos efectos. Por ejemplo, un impuesto general sobre el consumo equivale a impuestos iguales sobre el consumo de todos los bienes y a impuestos iguales sobre todos los factores, y un impuesto selectivo sobre un bien equivale a gravar al mismo tipo todos los factores empleados en producirlo.

Equivalente cierto Cantidad de dinero segura que a una persona le da igual recibir que jugarse una apuesta. Técnicamente, es la cantidad monetaria segura que proporciona la misma utilidad que la utilidad esperada de una lotería.

Escala mínima eficiente (o tamaño mínimo eficiente) Tamaño a partir del cual a la empresa ya no le sale más barato producir más. Técnicamente, es el nivel de producción a partir del cual se agotan las economías de escala técnicas y el coste medio deja de disminuir.

Escuela de Public Choice Corriente que estudia la política con las mismas herramientas con que la economía estudia los mercados. Técnicamente, es la escuela fundada por Buchanan y Tullock que aplica las herramientas del análisis económico a la toma de decisiones políticas, suponiendo que

 © 2026 Manuel Cabado García

votantes, políticos y burócratas actúan guiados por su interés propio.

Esquema de pagos óptimo Regla que fija cuánto cobra el agente según el resultado obtenido. Técnicamente, es la que maximiza el beneficio esperado del principal sujeta a la restricción de participación y, si hay riesgo moral, también a la de incentivos. Con esfuerzo observable, principal neutral y agente averso al riesgo es un salario fijo, y con riesgo moral es un salario que crece con el resultado, de modo que el agente asume parte del riesgo.

Estabilidad del equilibrio Capacidad de un mercado para volver por sí solo al equilibrio cuando algo lo aparta de él. Técnicamente, es la propiedad por la que, tras una perturbación, el mercado vuelve al equilibrio mediante un proceso de ajuste. Es local si solo ocurre para perturbaciones cercanas y global si ocurre con independencia de la posición de partida.

Estabilidad marshalliana Forma de estudiar si un mercado vuelve al equilibrio ajustando las cantidades producidas. Técnicamente, el desequilibrio se mide por la diferencia entre el precio que pagarían los compradores y el que exigen los vendedores, y el equilibrio es estable si esa diferencia es positiva a la izquierda del equilibrio y negativa a su derecha.

Estabilidad walrasiana Forma de estudiar si un mercado vuelve al equilibrio ajustando el precio, que sube cuando falta producto y baja cuando sobra. Técnicamente, el equilibrio es estable si el exceso de demanda es positivo por debajo del precio de equilibrio y negativo por encima.

Estado del bienestar Conjunto de políticas con las que el sector público protege a los ciudadanos frente a la pobreza y a riesgos como la vejez, la enfermedad o el paro. Técnicamente, agrupa los instrumentos redistributivos, sobre todo las prestaciones económicas y la provisión de bienes preferentes, con los que el sector público busca un cierto nivel de equidad y una cobertura mínima de necesidades en la sociedad.

Estado estacionario Situación en la que las magnitudes básicas de una economía dejan de cambiar con el tiempo. Técnicamente, en los clásicos es aquella a la que tiende la economía cuando los beneficios se anulan, el salario permanece en la subsistencia y el resto del ingreso va a la renta de la tierra, con lo que se detiene la acumulación. En los modelos de crecimiento neoclásicos, es la senda en la que el capital y la producción por trabajador permanecen constantes.

Estados de la naturaleza Situaciones del mundo que pueden darse y que el agente no controla, como que llueva o no. Técnicamente, cada estado, combinado con la acción elegida, determina el resultado o pago que obtiene el agente.

Estanflación Coincidencia de estancamiento económico, con paro elevado, e inflación alta. Técnicamente, contradice la disyuntiva estable de la curva de Phillips original. Friedman la explicó con la curva aumentada con expectativas: al subir la inflación esperada, la curva se desplaza hacia arriba y una inflación alta puede convivir con un paro igual o superior a su tasa natural, como ocurrió tras el primer choque del petróleo.

Estrategia Plan que indica qué hará un jugador en cualquier situación que pueda presentarse. Técnicamente, es un plan de acción completo y contingente que especifica la decisión de un jugador en cada posible situación del juego.

Estrategia dominante Estrategia que a un jugador le conviene seguir haga lo que haga el resto. Técnicamente, es la que le proporciona los mayores pagos frente a cualquier combinación de estrategias de sus rivales.

Estrategia trigger o gatillo Estrategia de cooperar mientras el otro coopere y castigarle en cuanto deje de hacerlo. Técnicamente, cada jugador coopera mientras todos cooperen y deja de hacerlo en cuanto alguno se desvía.

Estructura económica En la teoría de Marx, la base material de una sociedad, es decir, su forma de producir. Técnicamente, la forman las fuerzas productivas (el trabajo, las técnicas y los medios de producción) y las relaciones de producción (quién posee esos medios y cómo se reparte lo producido). Es la que determina la superestructura.

Evaluación relativa del desempeño Forma de pagar a alguien según cómo lo hace en comparación con otros que trabajan en condiciones pareci-

das. Técnicamente, al condicionar la retribución a los resultados de agentes comparables se descuenta la suerte común que les afecta a todos y el pago depende más del esfuerzo propio.

Evasión fiscal Reducción ilegal de la carga tributaria. Técnicamente, es el impago voluntario de impuestos debidos, por ejemplo ocultando renta, y es casi nula en la renta que comunican terceros, como los salarios, y mucho mayor en la renta autodeclarada.

Excedente del consumidor Diferencia entre lo que el consumidor estaría dispuesto a pagar por un bien y lo que paga por él. Técnicamente, se mide con el área bajo la demanda marshalliana y por encima del precio, y solo es una medida exacta del cambio de bienestar si el efecto renta es nulo, porque en otro caso queda entre la variación compensatoria y la equivalente.

Excedente del productor Ganancia que obtienen los vendedores por vender a un precio superior al mínimo que aceptarían. Técnicamente, es la diferencia entre el precio de venta y ese precio mínimo, que en competencia coincide con el coste marginal. Para el conjunto del mercado es el área situada por debajo del precio y por encima de la curva de oferta.

Excedente social Ganancia total que obtienen compradores y vendedores por participar en un mercado. Técnicamente, es la suma del excedente del consumidor y del excedente del productor, que alcanza su máximo en el equilibrio de competencia perfecta.

Excepción ibérica Mecanismo con el que España y Portugal abarataron la electricidad en 2022 y 2023 limitando el precio del gas usado para producirla. Técnicamente, fue un ajuste temporal autorizado como ayuda de Estado (Real Decreto-ley 10/2022) que fijaba un precio de referencia del gas para las centrales y les compensaba la diferencia, con lo que bajaba el precio marginal del mercado mayorista.

Exceso de capacidad socioeconómica Parte de la capacidad que se desaprovecha porque la empresa elige una planta más pequeña que la socialmente óptima. Técnicamente, es la diferencia entre la producción que minimiza el coste medio a largo plazo y la que minimiza el coste medio a corto plazo de la planta elegida, porque al competidor monopolístico no le interesa construir una planta mayor.

Exceso de capacidad técnica Capacidad ociosa de una planta que produce menos de lo que le permitiría alcanzar su coste medio mínimo. Técnicamente, es la diferencia entre la producción de equilibrio y la que minimiza el coste medio a corto plazo.

Exceso de gravamen Pérdida de bienestar que causa un impuesto por encima de lo que recauda. Técnicamente, es la pérdida causada por el efecto sustitución de un impuesto distorsionante que no se traduce en recaudación. Equivale a la carga adicional frente a un impuesto de suma fija de igual recaudación.

Exceso de gravamen marginal Pérdida adicional de bienestar que causa recaudar un euro más. Técnicamente, es el aumento del coste de eficiencia del sistema fiscal por cada unidad monetaria adicional recaudada, a diferencia del exceso de gravamen medio, que mide el coste total del sistema, por ejemplo en porcentaje del PIB.

Expectativas adaptativas Forma de prever el futuro mirando al pasado reciente y corrigiendo poco a poco los errores. Técnicamente, cada previsión se ajusta en una fracción del último error cometido, por lo que ante cambios persistentes los agentes se equivocan de forma sistemática.

Expectativas ingenuas Forma de prever en la que se espera que el precio futuro sea igual al actual. Técnicamente, el precio esperado para un periodo coincide con el observado en el periodo anterior, supuesto en el que descansa el modelo de la telaraña.

Expectativas racionales Idea de que la gente hace sus previsiones aprovechando bien toda la información de que dispone, sin equivocarse de forma sistemática. Técnicamente, es la hipótesis según la cual los agentes usan eficientemente toda la información disponible, de modo que sus previsiones coinciden con la esperanza condicionada de la variable y solo cometen errores aleatorios.

Explotación de monopolio Parte del valor de lo que produce el trabajador que no cobra porque la

 © 2026 Manuel Cabado García

empresa tiene poder en el mercado del bien. Técnicamente, en el sentido de Joan Robinson, es la diferencia entre el valor de la productividad marginal del trabajo y el ingreso marginal del factor.

Explotación de monopsonio Parte de su aportación que el trabajador no cobra porque la empresa es la única que contrata. Técnicamente, en el sentido de Joan Robinson, es la diferencia entre el ingreso marginal del factor, igual en el equilibrio al gasto marginal, y el salario pagado, que se lee sobre la curva de oferta.

Exportación impositiva Traslado de parte de la carga de los impuestos de un territorio a personas que viven fuera de él. Técnicamente, su grado depende de la sustituibilidad en la demanda y en la producción, de las intensidades factoriales, de la movilidad de los factores y del tamaño relativo de las economías.

Externalidad Efecto, bueno o malo, que la actividad de alguien tiene sobre otros sin que nadie pague ni cobre por él. Técnicamente, es el efecto, positivo o negativo, de la producción o el consumo de un agente sobre la utilidad o el beneficio de otro, sin reflejo en los precios ni en los costes.

Externalidad pecuniaria Efecto que una empresa causa a las demás a través de los precios sin tenerlo en cuenta al decidir. Técnicamente, en el mercado de trabajo, al bajar el salario todas las empresas producen más y hacen caer el precio del bien, por lo que la demanda agregada de trabajo es más inelástica que la suma horizontal de las individuales. A diferencia de una externalidad en sentido estricto, se transmite a través de los precios y en mercados competitivos no genera ineficiencia.

Externalidades de congestión Efecto por el que cada nuevo participante en la búsqueda dificulta la de los de su mismo lado y facilita la del otro, sin tenerlo en cuenta al decidir. Técnicamente, en el modelo DMP cada vacante adicional eleva la tasa a la que los parados encuentran empleo pero reduce la tasa a la que se cubren las demás vacantes, y por eso el equilibrio solo es eficiente si se cumple la condición de Hosios.

F

Factor cuasifijo Factor que, una vez contratado, no conviene despedir a la primera, porque incorporarlo costó dinero. Técnicamente, es el trabajo cuando contratar y formar exige costes fijos, de modo que la empresa retiene al trabajador mientras su productividad marginal cubra el salario (Oi, 1962).

Factor de descuento Medida de la paciencia, es decir, de cuánto se valora hoy un beneficio que llega más tarde. Técnicamente, es el valor actual de una unidad recibida un periodo después, y en los juegos repetidos la colusión solo es sostenible si supera un umbral, que es más alto cuantas más empresas formen el cártel.

Factor inferior Factor del que la empresa emplea menos cuando dispone de un presupuesto mayor. Técnicamente, es aquel cuya demanda ordinaria disminuye cuando aumenta el coste total que la empresa decide asumir, igual que un bien inferior en la teoría del consumidor, mientras que la de un factor normal aumenta.

Factores cooperantes Factores que se ayudan entre sí, de modo que emplear más de uno hace más productivo al otro. Técnicamente, la productividad marginal de cada factor aumenta con la cantidad empleada del otro, como en una función Cobb-Douglas, y entonces una bajada del salario eleva a largo plazo la demanda de trabajo y también la de capital.

Falacia del celofán Error en que se puede caer al delimitar un mercado cuando la empresa analizada ya cobra precios altos. Técnicamente, es una limitación del test SSNIP por la que, si el precio de partida ya incorpora poder de mercado, el mercado relevante se define de forma artificialmente amplia. Debe su nombre al caso antitrust estadounidense del celofán.

Falacia del coste hundido Error de seguir adelante con algo que no merece la pena solo porque ya se ha invertido mucho en ello. Técnicamente, es un sesgo cognitivo que lleva a seguir invirtiendo en un proyecto inviable por lo ya gastado en él, aunque los costes hundidos deberían ser irrelevantes para la decisión.

Fallo de coordinación Resultado colectivo malo que nadie elige pero al que lleva lo que conviene a cada uno. Técnicamente, en la nueva economía keynesiana, la decisión individual de no ajustar precios, racional para cada empresa, genera una externalidad de demanda agregada que amplifica el efecto de las perturbaciones sobre la producción y el empleo.

Fallo de mercado Situación en la que el mercado, por sí solo, no aprovecha bien los recursos. Técnicamente, se produce cuando falla algún supuesto del Primer Teorema del Bienestar, de modo que el equilibrio competitivo no es un óptimo de Pareto.

Fallos del sector público Problemas que crea la propia intervención del Estado cuando intenta corregir los del mercado. Técnicamente, son los costes e ineficiencias de la intervención pública, que pueden compensar total o parcialmente las ganancias de corregir los fallos de mercado.

Fases de la producción de Cassels División de la curva de producción en tres tramos según lo bien que se aprovecha el factor variable. Técnicamente, la fase I tiene productividad marginal superior a la media, la fase II, productividad marginal inferior a la media pero positiva, y es la única eficiente, y la fase III, productividad marginal negativa.

Federalismo fiscal de segunda generación Revisión de la teoría del federalismo fiscal que deja de suponer un gobierno benevolente. Técnicamente, incorpora la teoría de la elección colectiva, los incentivos de políticos y burócratas y los problemas de información, y contrasta empíricamente las ventajas atribuidas a la descentralización.

Fondo de Garantía de Servicios Públicos Fundamentales Principal fondo del sistema español de financiación autonómica. Técnicamente, busca que todas las comunidades puedan prestar un nivel similar de sanidad, educación y servicios sociales, pero sus transferencias no están condicionadas al gasto en esos servicios, por lo que funciona como una transferencia incondicional.

Fondo de Liquidez Autonómica Mecanismo con el que el Estado financió a las comunidades autónomas cuando perdieron el acceso a los mercados tras la crisis de 2008. Técnicamente, fue un rescate encubierto que sentó un precedente capaz de acentuar el riesgo moral, y es el ejemplo español más reciente de restricción presupuestaria blanda.

Fondo para días de lluvia (rainy day fund) Reserva que un gobierno territorial acumula en las épocas de bonanza para gastarla en las recesiones. Técnicamente, es una forma de autoseguro frente a perturbaciones propias, alternativa a que el gobierno central asuma la función aseguradora, con el riesgo moral que ello conlleva, y a los fondos comunes de reparto del riesgo entre jurisdicciones.

Forma extensiva Representación de un juego como un árbol de decisiones, que muestra quién mueve en cada momento. Técnicamente, el árbol recoge el orden de las decisiones, las acciones posibles y los pagos finales.

Forma normal o estratégica Representación de un juego como una tabla de doble entrada. Técnicamente, es una matriz que recoge las estrategias de cada jugador y los pagos de cada combinación de estrategias.

Formación específica Formación útil solo en la empresa que la imparte. Técnicamente, como eleva la productividad solo en esa empresa, su coste y sus rendimientos se reparten entre empresa y trabajador, lo que reduce la rotación y los despidos de estos trabajadores.

Formación general Formación útil en cualquier empresa. Técnicamente, como eleva por igual la productividad en todas las empresas, la competencia obliga a pagar al trabajador su nueva productividad, por lo que la financia el propio trabajador, normalmente con un salario más bajo mientras se forma.

Free-rider Persona que se aprovecha de algo sin pagarlo, sabiendo que no se la puede excluir. Técnicamente, es el agente que, amparado en la no exclusión, oculta su valoración y no paga por un bien al que puede acceder gratuitamente.

Frontera de posibilidades de producción Combinaciones de bienes que una economía puede obtener como máximo cuando emplea todos sus recursos sin desperdicio. Técnicamente, es la

 © 2026 Manuel Cabado García

curva que da la máxima producción de un bien para cada nivel de producción del otro, dados los factores y la tecnología, y resulta de llevar al espacio de bienes la curva de contrato de la producción. Sus puntos son eficientes, los interiores ineficientes y los exteriores inalcanzables.

Frontera de posibilidades de utilidad Máximas combinaciones de bienestar de dos personas que pueden lograrse repartiendo de todas las formas eficientes una misma cesta de bienes ya producida. Técnicamente, es la curva de contrato del consumo trasladada al espacio de utilidades, de modo que en ella no puede aumentarse la utilidad de un agente sin disminuir la del otro. Hay una frontera distinta para cada punto de la frontera de posibilidades de producción.

Frontera eficiente del conjunto de producción Parte del borde del conjunto de producción formada por los planes técnicamente eficientes. Técnicamente, en ella no es posible obtener más producto con los mismos factores ni el mismo producto con menos factores, y, cuando hay un único producto, coincide con la función de producción.

Fuga de carbono Traslado de producción contaminante a países con normas ambientales más laxas cuando un país encarece las emisiones. Técnicamente, es el aumento de emisiones fuera de la zona regulada que contrarresta en parte la reducción lograda dentro, y para limitarla se asignan derechos gratuitos a los sectores más expuestos.

Función de bienestar social Regla que resume en una sola cifra lo bien que está una sociedad a partir del bienestar de cada uno de sus miembros. Técnicamente, planteada por Bergson y Samuelson, asigna un índice de utilidad social a cada combinación de índices de utilidad individuales y permite elegir entre los distintos óptimos de Pareto.

Función de bienestar social CES Función de bienestar social que permite graduar con un solo parámetro cuánto le importa a la sociedad la desigualdad. Técnicamente, pertenece a una familia paramétrica de funciones de bienestar cuyo parámetro mide la aversión social a la desigualdad y que incluye como casos extremos la utilitarista y la maximin, y como caso intermedio la de Nash.

Función de bienestar social de Bernoulli-Nash Función que mide el bienestar social multiplicando las utilidades de los individuos. Técnicamente, es de tipo Cobb-Douglas, producto de las utilidades individuales, con curvas de indiferencia social convexas y una aversión a la desigualdad finita, intermedia entre la nula de la utilitarista y la infinita de la maximin, de modo que acepta pérdidas de los más favorecidos a cambio de ganancias de los más desfavorecidos, pero no a cualquier tasa.

Función de bienestar social maximin Función que juzga el bienestar de una sociedad solo por cómo le va a su miembro más desfavorecido. Técnicamente, propuesta por Rawls, iguala el bienestar social a la utilidad del individuo peor situado, y sus curvas de indiferencia social tienen forma de L.

Función de bienestar social utilitarista Función que mide el bienestar social sumando las utilidades de todos, sin importar cómo se repartan. Técnicamente, desarrollada por Bentham y Harsanyi, es la suma lineal de las utilidades individuales con igual ponderación, sin aversión a la desigualdad y con curvas de indiferencia social rectas de pendiente igual a menos uno.

Función de costes Relación que indica lo mínimo que le cuesta a la empresa producir cada cantidad. Técnicamente, es el coste mínimo de obtener cada volumen de producción dados los precios de los factores. Es homogénea de grado uno y cóncava en precios y contiene esencialmente la misma información que la función de producción.

Función de costes translogarítmica Forma de la función de costes lo bastante flexible para no imponer de antemano cómo se sustituyen los factores. Técnicamente, propuesta por Christensen, Jorgenson y Lau (1973), es una aproximación de segundo orden al logaritmo de la función de costes que no fija a priori ni la elasticidad de sustitución ni una elasticidad de escala constante.

Función de emparejamiento Relación que indica cuántos contratos se firman según cuántos parados y cuántas vacantes buscan a la vez. Técnicamente, es una función creciente, cóncava y con rendimientos constantes a escala del paro y las

vacantes, que resume las fricciones del mercado sin modelizarlas una a una.

Función de exceso de demanda Mide cuánto producto falta, o sobra si es negativa, a cada precio. Técnicamente, es la cantidad demandada menos la ofrecida a cada precio. Si es continua y cambia de signo, existe al menos un precio de equilibrio, aquel en que se anula.

Función de gasto Función que indica lo mínimo que hay que gastar para lograr un nivel de satisfacción a unos precios dados. Técnicamente, recoge el gasto mínimo necesario para alcanzar un nivel de utilidad dado a unos precios, y es homogénea de grado uno y cóncava en los precios, y creciente en la utilidad.

Función de producción Relación que indica cuánto puede producirse como máximo con unas cantidades dadas de factores. Técnicamente, asigna a cada vector de factores el máximo producto obtenible con una tecnología dada, de modo que solo recoge los métodos técnicamente eficientes.

Función de producción CES Función de producción en la que la facilidad para sustituir un factor por otro es la misma en cualquier punto. Técnicamente, su elasticidad de sustitución es constante y depende solo del parámetro de sustitución, e incluye como casos particulares o límite la función lineal (elasticidad infinita), la Cobb-Douglas (elasticidad unitaria) y la de Leontief (elasticidad nula).

Función de producción Cobb-Douglas Función de producción en la que el producto resulta de multiplicar las cantidades de los factores, cada una elevada a un exponente. Técnicamente, tiene relación marginal de sustitución técnica decreciente, factores imperfectamente sustitutivos y elasticidad de sustitución igual a uno, y, con precedente en Wicksell, la desarrollaron Cobb y Douglas en 1928.

Función de producción de Leontief Función de producción en la que los factores deben usarse siempre en la misma proporción, sin posibilidad de cambiar uno por otro. Técnicamente, es una función de coeficientes fijos en la que los factores son complementarios perfectos, con isocuantas en forma de L y elasticidad de sustitución nula.

Función de producción de los hogares de Becker Idea de que los hogares fabrican su propio bienestar combinando lo que compran con su tiempo. Técnicamente, es el enfoque en el que el hogar combina bienes de mercado y tiempo para producir los bienes que realmente generan utilidad, de modo que el precio efectivo de cada bien incluye el coste de oportunidad del tiempo.

Función de producción lineal Función de producción en la que un factor puede reemplazar por completo a otro a una tasa fija. Técnicamente, el producto es una suma ponderada de las cantidades de los factores, con isocuantas rectas, relación marginal de sustitución técnica constante y elasticidad de sustitución infinita.

Función de producción neoclásica de buen comportamiento Función de producción en la que duplicar todos los factores duplica la producción, pero aumentar uno solo rinde cada vez menos. Técnicamente, es continua, dos veces diferenciable, homogénea de grado 1 y con productividades marginales positivas y decrecientes, lo que la hace cóncava pero no estrictamente cóncava.

Función de reacción Regla que indica la mejor respuesta de una empresa a lo que haga su rival. Técnicamente, es la relación que indica el nivel de producción, o de precio, que maximiza el beneficio de una empresa para cada nivel de producción, o de precio, de su rival.

Función de utilidad Forma de poner números a las preferencias, de modo que las cestas preferidas reciban valores más altos. Técnicamente, en el enfoque ordinal solo importa el orden, por lo que cualquier transformación monótona creciente de una función de utilidad representa las mismas preferencias.

Función de utilidad de Bernoulli (o de von Neumann-Morgenstern) Función que indica cuánto valora un agente cada cantidad segura de riqueza. Técnicamente, se define sobre los resultados ciertos, es única salvo transformaciones afines positivas y su curvatura refleja la actitud ante el riesgo, y la utilidad esperada de una lotería es la media de sus valores ponderada por las probabilidades.

 © 2026 Manuel Cabado García

Función de valor (teoría de las perspectivas) Forma en que una persona valora lo que gana o pierde respecto a su situación de partida. Técnicamente, es la valoración de ganancias y pérdidas respecto a un punto de referencia que sustituye a la utilidad de Bernoulli, y es cóncava en ganancias, convexa en pérdidas y más pronunciada en estas.

Función homogénea Función que responde de forma regular cuando todas sus variables se multiplican por un mismo número. Técnicamente, al multiplicar todos sus argumentos por un escalar, queda multiplicada por ese escalar elevado a su grado de homogeneidad.

Función homotética Función cuyas isocuantas son copias ampliadas unas de otras, de modo que la sustitución entre factores depende solo de su proporción y no de la escala. Técnicamente, es una transformación monótona creciente de una función homogénea, y puede tener rendimientos a escala variables.

Función indirecta de producción Cuánto puede producir como máximo la empresa según lo que cuestan los factores y lo que está dispuesta a gastar. Técnicamente, es la producción máxima alcanzable para cada nivel de precios de los factores y de coste, y se obtiene al sustituir las demandas ordinarias en la función objetivo del problema primal.

Función indirecta de utilidad Función que indica la mayor satisfacción que puede alcanzar el consumidor según los precios y su renta. Técnicamente, recoge la utilidad máxima alcanzable en función de los precios y la renta, y es homogénea de grado cero, no creciente en los precios, creciente en la renta y cuasiconvexa.

Función inversa de demanda Relación que indica el precio que los consumidores están dispuestos a pagar por cada cantidad del bien. Técnicamente, se obtiene despejando el precio en la función de demanda agregada y, con preferencias cuasilineales, mide el beneficio marginal social de cada unidad.

Funcionamientos Lo que una persona consigue ser o hacer con los recursos de que dispone, desde estar bien alimentada hasta participar en la vida de su comunidad. Técnicamente, en el enfoque de Sen son los estados y actividades efectivamente alcanzados que sustituyen a la utilidad como espacio para evaluar el bienestar.

Funciones del sector público de Musgrave Las tres grandes tareas que justifican que el Estado intervenga en la economía. Técnicamente, son la asignación, la distribución y la estabilización, las tres razones de la intervención pública según la síntesis de Musgrave entre la economía del bienestar y el pensamiento keynesiano.

G

Gasto marginal del factor Coste adicional de contratar una unidad más de un factor. Técnicamente, con poder de monopsonio supera al salario, porque para contratar a un trabajador más hay que subir el salario a todos, y es tanto mayor cuanto menos elástica es la oferta de trabajo.

Gasto público Dinero que gastan las Administraciones Públicas para prestar servicios y ayudar a quien lo necesita. Técnicamente, es el conjunto de flujos de recursos que las Administraciones Públicas destinan a la producción de bienes y servicios y a la redistribución de la renta, financiados fundamentalmente con ingresos coactivos.

Gastos divisibles Gastos públicos de los que se puede saber quién se beneficia y en qué medida. Técnicamente, son los gastos públicos en bienes privados provistos por el sector público cuyo beneficio puede imputarse directamente a cada consumidor.

Gastos no divisibles Gastos públicos que benefician a todos sin que pueda saberse cuánto a cada uno. Técnicamente, son los gastos en bienes públicos que benefician al conjunto de la sociedad sin un beneficiario identificable.

Gobernanza de los comunes (Ostrom) Gestión de un recurso común por sus propios usuarios, con reglas que ellos mismos fijan y vigilan, sin privatizarlo ni dejarlo en manos del Estado. Técnicamente, Ostrom documentó que funciona cuando se cumplen ciertos principios de diseño, como fronteras claras, reglas adaptadas al lugar, vigilancia, sanciones graduales y mecanismos baratos de resolución de conflictos.

Grado de economías de alcance Medida de cuánto se ahorra al fabricar varios productos juntos. Técnicamente, es el porcentaje de ahorro de costes obtenido al producir conjuntamente dos o más bienes en lugar de hacerlo individualmente, positivo con economías de alcance y negativo con deseconomías.

Gran frontera de posibilidades de utilidad Límite máximo de bienestar que puede alcanzar cada persona, dado el de las demás, cuando la economía elige tanto lo que produce como su reparto. Técnicamente, es la envolvente de los puntos de las fronteras de posibilidades de utilidad que corresponden a óptimos globales, de modo que todos sus puntos son óptimos de Pareto en la producción, en el consumo y globalmente.

Grupo de interés Asociación de personas que se unen para que las decisiones públicas les favorezcan, como un sindicato. Técnicamente, influye mediante el intercambio de favores, la financiación de campañas o la información que hace llegar a legisladores y burócratas, y su actividad es un bien público para todos los que comparten su interés, estén afiliados o no.

Guardián de acceso Gran plataforma digital por la que empresas y usuarios tienen que pasar para llegar unos a otros. Técnicamente, es la empresa que la Comisión designa con arreglo a la Digital Markets Act por ser una puerta de acceso importante y estable entre empresas y usuarios, lo que se presume si supera unos umbrales de volumen de negocios o capitalización y de usuarios, y a la que se imponen obligaciones y prohibiciones ex ante.

H

Hipótesis de Calmfors y Driffill Relación entre el grado de centralización de la negociación colectiva y el paro. Técnicamente, Calmfors y Driffill (1988) encontraron una relación en forma de U invertida: el paro es menor con negociación muy centralizada, que tiene en cuenta sus efectos sobre el conjunto de la economía, o muy descentralizada, disciplinada por la competencia, y mayor con la negociación por sectores.

Histéresis Situación en la que el paro no vuelve a su nivel anterior cuando desaparece la causa que lo elevó. Técnicamente, el paro de equilibrio depende del paro pasado, de modo que una perturbación transitoria tiene efectos permanentes. En el modelo insiders-outsiders surge porque, tras una caída del empleo, los insiders que quedan fijan salarios que solo protegen sus propios puestos.

Hit and run Entrada rápida en un mercado para aprovechar sus precios altos y salir después. Técnicamente, es la estrategia por la que un entrante accede a un mercado impugnable cuyos precios superan el coste medio, vende más barato, obtiene beneficios extraordinarios y sale.

Homo economicus Imagen del ser humano como alguien que solo busca su propio beneficio de forma calculada. Técnicamente, es el individuo racional que persigue su propio interés maximizando su riqueza, idea que Mill articuló explícitamente por primera vez.

I

Identidad de Roy Resultado que permite deducir cuánto comprará el consumidor sin conocer sus gustos, a partir de la satisfacción máxima que puede alcanzar. Técnicamente, obtiene la demanda marshalliana a partir de la función indirecta de utilidad, como cociente cambiado de signo entre su derivada respecto al precio y su derivada respecto a la renta.

Ideología del partido gobernante (ciclo político partidista) Explicación del gasto público por la ideología del partido que gobierna, y no solo por el calendario electoral. Técnicamente, los partidos tienen funciones de pérdida distintas según su ideología, de modo que los gobiernos progresistas tienden a elevar el gasto social y la redistribución y los conservadores a reducir el gasto y los impuestos. Las restricciones institucionales, las preferencias del votante mediano y las coaliciones moderan estas diferencias.

Ignorancia racional del votante Decisión sensata del votante de no informarse demasiado, porque su voto apenas cuenta. Técnicamente, el votante decide racionalmente permanecer poco informado, porque el coste de informarse es despro-

 © 2026 Manuel Cabado García

porcionado frente a la influencia real de un único voto.

Igualdad de oportunidades Principio según el cual lo que consigue cada persona no debe depender de circunstancias que no ha elegido, como su origen familiar. Técnicamente, es, junto a la equidad horizontal y la vertical, uno de los tres pilares de la función redistributiva del sector público, y se distingue de la igualdad de resultados en que no pretende igualar las rentas finales.

Ilusión fiscal Tendencia de los votantes a creer que pagan menos impuestos de los que realmente pagan. Técnicamente, concepto de Puviani (1903) desarrollado por Buchanan, es la infravaloración por los votantes de la carga fiscal que soportan, favorecida por las retenciones en origen, la imposición indirecta o la opacidad del sistema, que permite aumentar el tamaño del sector público sin que perciban un aumento equivalente de la presión fiscal.

Ilusión monetaria Confusión entre valores nominales y reales. Técnicamente, se produce cuando los agentes reaccionan a cambios en magnitudes nominales, como el salario o los precios, sin tener en cuenta su efecto sobre el poder adquisitivo.

Imposibilidad de producción libre Supuesto de que no se puede producir algo a partir de nada. Técnicamente, es el axioma que impide obtener cantidades positivas de algún output sin emplear ningún input.

Impuesto Pago obligatorio al Estado que no da derecho a recibir nada concreto a cambio. Técnicamente, es una transferencia coactiva del sector privado al sector público, establecida por ley y sin contraprestación directa, lo que la distingue de las tasas y las contribuciones especiales.

Impuesto ad valorem Impuesto que se paga como un porcentaje del precio o del valor de lo gravado. Técnicamente, su cuantía por unidad crece con el precio, por lo que gráficamente hace pivotar la curva gravada en lugar de desplazarla en paralelo. En competencia perfecta tiene los mismos efectos que un impuesto específico de igual recaudación, pero en monopolio recauda más que este para una misma cantidad de equilibrio.

Impuesto de capitación Impuesto en el que todo el mundo paga lo mismo. Técnicamente, es un impuesto de tanto alzado en el que todos los contribuyentes pagan la misma cantidad.

Impuesto de Clarke Pago que se exige a quien, con su participación, cambia la decisión colectiva, por el perjuicio que causa a los demás. Técnicamente, recae sobre los agentes bisagra, aquellos cuya valoración decide la provisión del bien público, y es igual a la pérdida neta que su participación causa al resto.

Impuesto de fuente (sobre el capital) Impuesto sobre el capital situado en un país, sea quien sea su propietario. Técnicamente, en una economía pequeña con capital perfectamente móvil provoca una salida de capitales hasta que el rendimiento neto vuelve a igualar el tipo de interés mundial, de modo que lo acaba soportando el trabajo, con un exceso de gravamen añadido.

Impuesto de residencia (sobre el capital) Impuesto sobre el capital de los residentes de un país, dondequiera que esté invertido. Técnicamente, en una economía pequeña con capital perfectamente móvil no provoca salidas de capital, porque invertir fuera no lo evita, y equivale a un impuesto de suma fija sobre la riqueza que soportan íntegramente los propietarios nacionales del capital.

Impuesto de suma fija Impuesto cuya factura no puede reducirse cambiando de conducta. Técnicamente, es aquel cuya cuota no depende de ninguna decisión del contribuyente, ni de consumo ni de oferta de factores, por lo que no altera los precios relativos, genera solo efecto renta y su exceso de gravamen es nulo. Todo impuesto neutral equivale a uno de suma fija, que se usa poco porque no atiende a la capacidad de pago.

Impuesto distorsionante Impuesto cuya factura se puede reducir cambiando de conducta, por ejemplo trabajando o consumiendo menos de lo gravado. Técnicamente, es aquel cuya cuota depende de decisiones del contribuyente, por lo que altera los precios relativos, provoca un efecto sustitución y genera exceso de gravamen.

Impuesto específico (o unitario) Impuesto de una cantidad fija por unidad vendida o comprada,

sea cual sea su precio. Técnicamente, desplaza en paralelo la curva gravada, a diferencia del impuesto ad valorem, que la hace pivotar. En monopolio reduce el ingreso marginal en toda la cuantía del impuesto, por lo que contrae la producción más que un impuesto ad valorem de igual recaudación.

Impuesto negativo sobre la renta Sistema que integra las ayudas contra la pobreza en el impuesto sobre la renta: por debajo de cierta renta el Estado paga al contribuyente y por encima el contribuyente paga al Estado. Técnicamente, propuesto por Friedman en 1962, combina una transferencia igual para todos, el dividendo social, con un impuesto proporcional sobre la renta, de modo que el impuesto neto cambia de signo en la renta umbral, igual al cociente entre el dividendo social y el tipo impositivo.

Impuesto pigouviano Impuesto que hace pagar a quien perjudica a otros por el daño que causa. Técnicamente, grava una actividad generadora de una externalidad negativa y se fija igual al daño marginal en el nivel de producción eficiente, para que el precio refleje el coste social.

Impuesto progresivo Impuesto en el que quien gana más paga, además, una proporción mayor de lo que gana. Técnicamente, es aquel cuyo tipo medio crece al aumentar la base imponible.

Impuesto progresivo sobre el consumo Impuesto que grava lo que cada persona gasta en todo el año, con tipos más altos para quien más gasta. Técnicamente, es un impuesto directo que grava de forma progresiva el consumo total anual de un individuo, en lugar de cada compra en el momento del gasto.

Impuesto proporcional Impuesto que hace pagar a todos el mismo porcentaje de su base. Técnicamente, es aquel cuyo tipo medio es constante al variar la base imponible y coincide con el tipo marginal. Ya hace pagar más a quien más gana, pero mantiene inalteradas las desigualdades relativas.

Impuesto regresivo Impuesto que se lleva una proporción mayor de la base de quien menos tiene. Técnicamente, es aquel cuyo tipo medio disminuye al aumentar la base imponible, de modo que su tipo marginal es inferior al medio.

Impuesto sobre el valor del suelo Impuesto sobre el valor de los terrenos, sin contar lo que el dueño haya construido o mejorado en ellos. Técnicamente, grava el valor potencial del suelo, excluidas las mejoras del propietario, y no distorsiona porque la oferta de suelo es fija.

Impuestos cedidos Impuestos estatales cuya recaudación se entrega por completo a las regiones, que además los gestionan y pueden modificarlos en parte. Técnicamente, son tributos establecidos y regulados por el Estado cuyo rendimiento corresponde a la comunidad autónoma, que recibe también competencias de gestión y una capacidad normativa limitada.

Impuestos compartidos Impuestos que el Estado regula y gestiona pero en los que las regiones pueden fijar una parte, como la tarifa autonómica del IRPF. Técnicamente, son tributos cedidos parcialmente con capacidad normativa autonómica sobre algunos elementos.

Impuestos participados Impuestos que el Estado regula y cobra por completo y cuya recaudación reparte en parte con las regiones. Técnicamente, son tributos cedidos parcialmente sin capacidad normativa autonómica, como el IVA o los impuestos especiales.

Impuestos propios en sentido estricto Impuestos que una región o un municipio crea y gestiona por su cuenta. Técnicamente, son los impuestos diseñados y gestionados íntegramente por las haciendas territoriales.

Incertidumbre Situación en la que no se sabe con qué probabilidad ocurrirá cada resultado. Técnicamente, el agente no puede asignar probabilidades objetivas a los estados de la naturaleza y, como mucho, puede formarse probabilidades subjetivas, lo que Knight distingue del riesgo.

Incidencia absoluta Efecto de crear o subir un impuesto sobre quién gana y quién pierde, sin tocar nada más. Técnicamente, son los cambios en la distribución de la renta causados por la introducción o el aumento de un tributo, manteniendo constantes el gasto público y los demás tributos. Musgrave la denominó después incidencia específica.

 © 2026 Manuel Cabado García

Incidencia de impacto Primer efecto de un impuesto, el que soporta quien debe ingresarlo antes de que precios y conductas se ajusten. Técnicamente, es el efecto directo de un tributo sobre el contribuyente legal, antes de cualquier proceso de ajuste. Equivale a la incidencia legal.

Incidencia de presupuesto equilibrado Efecto sobre quién gana y quién pierde de subir a la vez los impuestos y el gasto en la misma cuantía. Técnicamente, son los efectos distributivos de un aumento de la tributación acompañado de un aumento equivalente del gasto público.

Incidencia del gasto público Análisis de quién se beneficia realmente del gasto público. Técnicamente, De Wulf distinguió cuatro enfoques: el de impacto, que imputa el gasto a quien lo recibe directamente, el normativo o legal, que atiende a su diseño institucional, el de precios relativos, que estudia su efecto sobre los precios de factores y productos, y el del beneficio, que mide la mejora de bienestar de los beneficiarios.

Incidencia diferencial Efecto sobre quién gana y quién pierde de cambiar un impuesto por otro que recauda lo mismo. Técnicamente, son los cambios en la distribución de la renta al sustituir un tributo por otro de igual rendimiento real, manteniendo constante el gasto público.

Incidencia económica Reparto real de la carga de un impuesto, que puede no coincidir con quién lo ingresa en Hacienda. Técnicamente, es el cambio en la distribución de la renta real que provoca el impuesto a través de los precios de los bienes y de los factores.

Incidencia efectiva Quién acaba pagando de verdad un impuesto cuando todo se ha ajustado. Técnicamente, es la distribución final de la carga de un tributo una vez concluido el proceso de ajuste que este genera.

Incidencia intergeneracional Estudio de si un impuesto lo pagan quienes viven hoy o las generaciones futuras. Técnicamente, es el análisis de qué generación soporta finalmente la carga de un impuesto en un contexto dinámico.

Incidencia legal Respuesta a la pregunta de quién paga un impuesto mirando solo lo que dice la ley, sin tener en cuenta cómo reaccionan los precios. Técnicamente, es la distribución de la carga de un impuesto según quién debe pagarlo de acuerdo con la ley. También se denomina incidencia de impacto.

Incidencia pretendida Reparto de la carga de un impuesto que el legislador quería conseguir. Técnicamente, es el cambio en la distribución de la renta que se produciría si el tributo lo soportaran exactamente las personas a las que la ley pretende gravar, estén o no obligadas a ingresarlo. Musgrave la denominó incidencia de intención legislativa.

Inconsistencia dinámica de la política económica Problema por el que a un gobierno deja de convenirle, llegado el momento, la política que anunció como óptima. Técnicamente, como mostraron Kydland y Prescott (1977), una vez que los agentes han fijado sus expectativas conforme al anuncio, la autoridad tiene incentivos a desviarse, los agentes lo anticipan y el anuncio pierde credibilidad. En política monetaria discrecional genera un sesgo inflacionista y justifica reglas e instituciones independientes. También se llama inconsistencia temporal.

Independencia de las alternativas irrelevantes Exigencia de que la elección entre dos opciones no cambie porque aparezca o desaparezca una tercera. Técnicamente, es el axioma de Arrow según el cual la elección social entre dos alternativas debe depender solo de las preferencias individuales entre ambas y no de su posición frente a terceras.

Índice de Atkinson Medida de desigualdad que indica qué parte de la renta total podría sacrificar la sociedad sin perder bienestar si esa renta se repartiera por igual. Técnicamente, es uno menos el cociente entre la renta equivalente igualmente distribuida y la renta media. Se deriva de una función de bienestar social CES, cuyo parámetro de aversión a la desigualdad hereda.

Índice de autoridad regional Indicador que compara cuánto poder político tienen las regiones de distintos países. Técnicamente, es un indicador comparativo de la autonomía política de las regiones que combina diez dimensiones de autogobierno y de participación en el gobierno central.

Índice de frustración Medida de cuánto se aleja el resultado real de un impuesto de lo que buscaba el legislador. Técnicamente, es la diferencia entre la incidencia efectiva y la incidencia pretendida, y recoge el grado de fracaso de la política fiscal en sus objetivos distributivos. Braña lo denominó incidencia fiscal.

Índice de Gini Número que resume lo desigual que es el reparto de la renta en una sociedad. Técnicamente, es el área entre la curva de Lorenz y la línea de perfecta igualdad dividida por el área total bajo esa línea. Toma valores entre cero, con igualdad perfecta, y uno, con desigualdad extrema.

Índice de Herfindahl-Hirschman Medida de lo concentrado que está un mercado, que aumenta cuanto menos numerosas y más grandes son las empresas. Técnicamente, es la suma de los cuadrados de las cuotas de mercado, en porcentaje, de todas las empresas, y alcanza 10.000 en el monopolio.

Índice de Kakwani Indicador de si un impuesto recae proporcionalmente más sobre quienes más renta tienen. Técnicamente, es el índice de concentración de las cuotas menos el índice de Gini de la renta antes de impuestos, y es positivo si el impuesto es progresivo, negativo si es regresivo y nulo si es proporcional.

Índice de Laspeyres Índice de precios que compara lo que cuesta hoy la misma cesta que se compraba al principio. Técnicamente, mide el cambio relativo del coste de adquirir la cesta de consumo inicial y, ante subidas de precios, sobreestima el aumento del coste de vida al ignorar la sustitución entre bienes.

Índice de Lerner Medida del poder de mercado de una empresa según cuánto supera su precio al coste marginal. Técnicamente, propuesto por Lerner (1934), es la diferencia entre el precio y el coste marginal dividida por el precio, que es nula en competencia perfecta y, en el óptimo del monopolista, igual a la inversa del valor absoluto de la elasticidad-precio de la demanda.

Índice de Malmquist Indicador de cuánto cambia la productividad de una empresa entre dos periodos. Técnicamente, compara la distancia de cada observación a las fronteras eficientes de ambos periodos y se descompone en cambio técnico, por el desplazamiento de la frontera, y cambio en la eficiencia relativa, de modo que un valor superior a uno indica una ganancia de productividad total de los factores.

Índice de Musgrave y Thin Indicador de cuánto reduce un impuesto la desigualdad. Técnicamente, es el cociente entre uno menos el índice de Gini de la renta después de impuestos y uno menos el de antes de impuestos, y es mayor que 1 cuando el impuesto reduce la desigualdad, como ocurre siempre que es progresivo.

Índice de Paasche Índice de precios que compara lo que costaba antes la cesta que se compra ahora. Técnicamente, mide el cambio relativo del coste de adquirir la cesta de consumo final y, ante subidas de precios, subestima el aumento del coste de vida.

Índice de Pechman y Okner Indicador de cuánto reduce un impuesto la desigualdad en términos relativos. Técnicamente, es la reducción porcentual de la desigualdad medida con el índice de Gini, calculada como la diferencia entre el índice de Gini antes y después de impuestos dividida por el índice de Gini antes de impuestos, y es positivo si el impuesto corrige la desigualdad.

Índice de Reynolds-Smolensky Medida de cuánto reduce un impuesto la desigualdad de la renta. Técnicamente, es la diferencia entre el índice de Gini de la renta antes de impuestos y el de la renta después de impuestos, y es positiva si el impuesto redistribuye en sentido igualitario.

Individualismo metodológico Principio de método que explica los fenómenos sociales a partir de las decisiones de los individuos. Técnicamente, es el enfoque que introdujeron los marginalistas frente al análisis clásico por clases sociales y que la escuela austriaca llevó al extremo.

Inducción hacia atrás Forma de resolver un juego empezando por el final y razonando hacia el principio. Técnicamente, es el método de resolución de juegos con información perfecta que determina primero la decisión óptima en los últimos nodos y, anticipándola, retrocede nodo a nodo hasta el inicial.

 © 2026 Manuel Cabado García

Industria de costes constantes Sector cuya expansión no encarece ni abarata los factores que utiliza. Técnicamente, es el sector en el que el precio de los factores no varía con la demanda de factores de la industria, lo que equivale a una oferta de factores perfectamente elástica. Su oferta a largo plazo es horizontal, y no implica costes marginales constantes.

Industria de costes crecientes Sector en el que producir más encarece la producción de todas las empresas, normalmente porque sube el precio de factores escasos. Técnicamente, tiene deseconomías externas y su oferta a largo plazo tiene pendiente positiva.

Industria de costes decrecientes Sector en el que la expansión conjunta abarata la producción de cada empresa, por factores más baratos o por conocimiento compartido. Técnicamente, tiene economías externas y su oferta a largo plazo tiene pendiente negativa.

Industria de red Sector cuyo servicio llega al usuario a través de una infraestructura física en red, como los tendidos eléctricos, las vías férreas o los cables. Técnicamente, combina una red con rasgos de monopolio natural, por su elevada inversión inicial y su bajo coste marginal, con actividades que pueden prestarse en competencia, lo que obliga a regular el acceso a la red o a separarla verticalmente.

Ineficiencia asignativa Pérdida que se produce cuando se fabrica menos de un bien de lo que conviene a la sociedad, porque su precio está por encima de lo que cuesta producir una unidad más. Técnicamente, surge cuando el precio supera al coste marginal, de modo que hay unidades cuya valoración excede su coste de producción y no se producen.

Ineficiencia dinámica (innovación) Pérdida que se produce cuando una empresa con poder de mercado innova e invierte menos de lo que convendría a la sociedad. Técnicamente, es la tercera ineficiencia del poder de mercado, junto a la asignativa y la productiva, aunque la evidencia empírica sobre su importancia no es concluyente. Es lo contrario de la eficiencia dinámica, que juzga un mercado por su capacidad de innovar con el tiempo y no por cómo asigna los recursos en un momento dado.

Ineficiencia productiva Situación en la que un bien se produce más caro de lo técnicamente posible. Técnicamente, ocurre cuando las empresas operan con un coste medio superior al mínimo, como sucede con el exceso de capacidad en competencia monopolística.

Ineficiencia X Despilfarro que se produce dentro de una empresa cuando produce más caro de lo que podría. Técnicamente, es el exceso de coste que soporta respecto al mínimo que le permitiría su tecnología. Según Leibenstein, se debe a que trabajadores y directivos se esfuerzan menos de lo necesario, algo que se agrava cuando la presión competitiva es baja.

Información asimétrica Situación en la que una de las partes de un trato sabe algo relevante que la otra ignora. Técnicamente, las partes disponen de distinta información sobre características (información oculta) o acciones (acciones ocultas), lo que, unido a un conflicto de intereses, origina problemas de selección adversa o de riesgo moral.

Información imperfecta Situación de un juego en la que algún jugador decide sin saber todo lo que han hecho antes los demás. Técnicamente, no conoce todos los movimientos previos de sus rivales, como ocurre cuando los jugadores deciden a la vez.

Información incompleta (contratos) Situación en la que nadie sabe con certeza qué va a ocurrir, aunque todos sepan lo mismo. Técnicamente, las partes no conocen qué estado de la naturaleza se dará, en un contexto de riesgo si pueden asignar probabilidades o de incertidumbre si no pueden. Si además unas saben más que otras, la información es asimétrica.

Información incompleta (juego bayesiano) Situación en la que un jugador desconoce algún rasgo de sus rivales, como sus costes o sus pagos. Técnicamente, cada jugador tiene un tipo que solo él observa, y la transformación de Harsanyi convierte el juego en uno de información imperfecta, llamado juego bayesiano, cuya solución es el equilibrio bayesiano de Nash.

Información oculta Lo que una parte sabe de sí misma, o de lo que vende, y la otra no puede comprobar antes de firmar. Técnicamente, es la asimetría sobre características no observables de una de las partes que existe antes del contrato y que origina la selección adversa.

Información perfecta Situación de un juego en la que quien decide sabe todo lo que ha ocurrido antes. Técnicamente, cada jugador conoce, al mover, todos los movimientos previos de sus rivales, lo que permite resolver el juego por inducción hacia atrás.

Ingreso marginal Ingreso adicional que obtiene la empresa por vender una unidad más de producto. Técnicamente, en monopolio es menor que el precio, porque para vender esa unidad la empresa debe rebajar el precio también en las unidades que ya vendía.

Ingreso marginal del factor Ingreso adicional que obtiene la empresa al contratar una unidad más de un factor, también llamado ingreso del producto marginal. Técnicamente, es el ingreso marginal del bien por la productividad marginal del factor, que con poder de monopolio en el mercado del bien queda por debajo del valor de la productividad marginal.

Innovación de proceso Mejora en la forma de producir un bien que ya existe. Técnicamente, es el progreso técnico que permite obtener más producto con los mismos factores, o el mismo con menos, a diferencia de la innovación de producto, que crea bienes nuevos o mejora su calidad.

Intercambio de regalos Explicación del salario de eficiencia basada en la reciprocidad: quien se siente bien pagado corresponde trabajando más. Técnicamente, Akerlof (1982) la planteó como un intercambio en el que la empresa paga un salario superior al de mercado y los trabajadores responden con un esfuerzo superior al mínimo exigido.

Intercambio de votos Práctica de "hoy por ti, mañana por mí" entre votantes. Técnicamente, es el acuerdo, también llamado logrolling, por el que varios votantes se comprometen a apoyar mutuamente sus proyectos. Permite reflejar la intensidad de las preferencias, pero también puede aprobar proyectos cuyo coste total supera sus beneficios.

Irreversibilidad (de la producción) Supuesto de que la producción no puede deshacerse marcha atrás. Técnicamente, es el axioma según el cual el proceso productivo no puede invertirse transformando los outputs en los inputs que se emplearon para obtenerlos.

Isoclina Línea que une, en distintos niveles de producción, los puntos en los que la facilidad para cambiar un factor por otro es la misma. Técnicamente, une los puntos de distintas isocuantas en los que la relación marginal de sustitución técnica es la misma, y en las funciones homotéticas es una recta que parte del origen.

Isocuanta Curva que reúne las distintas formas de combinar los factores para producir una misma cantidad. Técnicamente, es una curva de nivel de la función de producción que reúne las combinaciones de factores con las que se obtiene un mismo nivel de producto.

J

Juego Descripción ordenada de una situación en la que varias personas deciden teniendo en cuenta a las demás. Técnicamente, es la representación formal de una interacción estratégica entre varios individuos, descrita por sus jugadores, reglas, resultados y pagos.

Juego actuarialmente justo Juego en el que, en promedio, no se gana ni se pierde. Técnicamente, el coste de participar es igual a la ganancia esperada, de modo que su valor neto esperado es nulo.

Juego cooperativo Juego en el que los jugadores pueden firmar acuerdos que les obligan. Técnicamente, la teoría de juegos cooperativa toma como unidad de análisis las coaliciones y estudia qué repartos del excedente conjunto son estables o equitativos, a diferencia del juego no cooperativo, en el que cada jugador decide por su cuenta sin acuerdos vinculantes. No debe confundirse con el oligopolio cooperativo, es decir, con la colusión.

Juego de ofertas alternativas de Rubinstein Negociación en la que las partes se turnan para proponer un reparto hasta que una de ellas acep-

 © 2026 Manuel Cabado García

ta. Técnicamente, Rubinstein (1982) mostró que tiene un único equilibrio perfecto en subjuegos, y Binmore, Rubinstein y Wolinsky mostraron que, cuando el intervalo entre ofertas tiende a cero, converge a la solución de Nash generalizada, en la que el poder de negociación de cada parte es mayor cuanto más paciente es.

Juego repetido Situación en la que los mismos jugadores se enfrentan una y otra vez. Técnicamente, es un juego que se repite varias veces entre los mismos jugadores, lo que genera un historial que permite condicionar la estrategia al comportamiento pasado del rival.

Juego secuencial Juego en el que los jugadores mueven por turnos, de modo que quien decide después puede tener en cuenta lo que han hecho los anteriores. Técnicamente, también llamado dinámico, se representa en forma extensiva y, si hay información perfecta, se resuelve por inducción hacia atrás, frente al juego simultáneo o estático, en el que todos deciden a la vez.

L

Largo plazo Periodo suficiente para que la empresa pueda cambiar todos sus recursos, aunque no la tecnología. Técnicamente, es el horizonte de planificación en el que todos los factores son variables pero la tecnología es constante, sin costes hundidos y con libre entrada y salida de empresas.

Lema de Shephard Resultado de la teoría de la dualidad que permite obtener las demandas compensadas derivando una función de valor. Técnicamente, la derivada de la función de gasto respecto al precio de un bien es la demanda hicksiana de ese bien y, en la empresa, la derivada de la función de costes respecto al precio de un factor es su demanda condicionada.

Lente (caja de Edgeworth) Zona de la caja de Edgeworth en la que las dos partes están mejor, o al menos igual, que en el punto de partida. Técnicamente, es la región delimitada por las dos curvas de indiferencia, o las dos isocuantas, que pasan por la dotación inicial. Su intersección con la curva de contrato es el núcleo.

Leviatán Imagen del sector público como un monstruo que tiende a crecer sin límite a costa del sector privado. Técnicamente, Brennan y Buchanan (1980) tomaron la figura de Hobbes para describir un Estado que, sin restricciones constitucionales, sobreexpande su actividad financiera y fiscal y expulsa a la actividad privada con una carga impositiva excesiva, lo que justifica la constitución fiscal.

Ley de Dalton Regla según la cual un impuesto lo acaba pagando sobre todo la parte del mercado que menos puede esquivarlo. Técnicamente, en un mercado competitivo la carga de un impuesto se reparte entre compradores y vendedores según las elasticidades-precio de la demanda y de la oferta, de modo que recae más sobre el lado relativamente más inelástico, con independencia de la incidencia legal. Solo es válida en equilibrio parcial, con los demás precios constantes.

Ley de hierro de los salarios Idea de que los salarios no pueden mantenerse mucho tiempo por encima de lo justo para sobrevivir. Técnicamente, es la tendencia de los salarios a situarse a largo plazo en el nivel de subsistencia, porque cualquier salario superior eleva la población y, con ella, la oferta de trabajo.

Ley de la demanda Regularidad según la cual la cantidad demandada de un bien disminuye cuando sube su precio. Técnicamente, descansa en el signo negativo del efecto sustitución propio y se cumple siempre en los bienes normales, pero puede fallar en un bien inferior cuyo efecto renta supere al efecto sustitución, como el bien Giffen.

Ley de la tasa decreciente de ganancia Tendencia del capitalismo, según Marx, a que la rentabilidad del capital caiga a medida que se mecaniza la producción, también llamada caída tendencial de la tasa de beneficio. Técnicamente, como solo el trabajo vivo crea plusvalía, el aumento de la composición orgánica del capital reduce la tasa de beneficio salvo que la tasa de explotación crezca lo suficiente.

Ley de Okun Regularidad según la cual, cuando la producción supera su nivel normal, el paro queda por debajo del suyo, y a la inversa. Técnicamente, es la relación empírica negativa entre las

desviaciones del desempleo respecto de su tasa natural y las de la producción real respecto de la potencial.

Ley de rendimientos decrecientes Idea de que, si se añade cada vez más de un factor a una cantidad fija de otro, lo que aporta cada unidad adicional acaba siendo cada vez menor. Técnicamente, la ley, formulada por Turgot y también llamada de las proporciones variables, establece que, dado un factor fijo, llega un punto en que unidades adicionales del factor variable aumentan el producto cada vez menos.

Ley de Say Principio según el cual toda producción genera las rentas necesarias para comprarla, de modo que no puede haber sobreproducción generalizada. Técnicamente, el valor de lo producido se reparte íntegramente en rentas que se consumen o se invierten, y solo caben excesos de oferta en unos mercados compensados por excesos de demanda en otros. La fórmula "la oferta crea su propia demanda" no es de Say, sino de Keynes, que la usó en 1936 para criticar la ley.

Ley de Walras Resultado según el cual el valor de lo que se demanda y lo que se ofrece en el conjunto de la economía siempre cuadra. Técnicamente, la suma del valor monetario de los excesos de demanda de todos los mercados es igual a cero para cualquier vector de precios. En teoría del consumidor, se llama así también a la propiedad de la demanda marshalliana por la que el consumidor gasta toda su renta.

Ley del precio único Principio según el cual un mismo bien debe venderse al mismo precio en todos los mercados. Técnicamente, con producto homogéneo, información perfecta, precios flexibles y sin barreras al comercio, el arbitraje iguala los precios de un mismo bien expresados en la misma moneda.

Leyes de Wagner Idea de que el Estado crece más deprisa que la economía a medida que un país se enriquece. Técnicamente, explican el aumento del peso del gasto público en el PIB por la mayor intervención que exige una sociedad industrializada y por la elasticidad-renta superior a la unidad de la demanda de bienes sociales como la educación o la sanidad.

Línea de certeza Conjunto de situaciones en las que el agente tiene la misma riqueza ocurra lo que ocurra. Técnicamente, en el plano de riquezas contingentes es la recta de 45 grados, y un agente averso al riesgo se sitúa sobre ella, con seguro completo, cuando la prima es actuarialmente justa.

Líneas de contorno Límites de la zona del plano de factores en la que tiene sentido producir. Técnicamente, son las isoclinas que unen los puntos de las isocuantas en los que la productividad marginal de uno de los factores se anula, y entre ellas queda la región económica de producción.

Lotería Situación en la que se sabe qué puede pasar y con qué probabilidad, pero no qué pasará. Técnicamente, es una distribución de probabilidad sobre un conjunto de resultados, con la que se representan las alternativas con riesgo.

M

Maldición del ganador Tendencia de quien gana una subasta a haber sido el más optimista de los participantes, y por eso a ganar menos de lo que esperaba. Técnicamente, en las subastas de valor común el ganador suele ser el licitador que más sobrestimó el valor del objeto o, si se adjudica al precio más bajo, el que más infravaloró sus costes, lo que lleva a los licitadores racionales a pujar con prudencia.

Manipulación de la agenda Poder de quien decide el orden en que se vota para inclinar el resultado a su favor. Técnicamente, es la capacidad de quien fija el orden de votación de las alternativas para obtener el resultado final que desea.

Mano invisible Idea de que, al buscar cada uno su propio beneficio, se acaba favoreciendo a toda la sociedad sin que nadie lo planee. Técnicamente, es el mecanismo por el que el individuo que persigue libremente su propio interés contribuye, sin pretenderlo, al bienestar colectivo.

Margen extensivo Decisión de trabajar o no trabajar. Técnicamente, es la respuesta de la oferta de trabajo al salario a través de la entrada en el mercado de quienes ven superado su salario de reserva, más sensible que la de las horas, sobre todo en las mujeres casadas. En teoría de la pro-

 © 2026 Manuel Cabado García

ducción, la expresión designa también el óptimo técnico.

Margen intensivo Horas que trabajan quienes ya tienen empleo. Técnicamente, es la respuesta de la oferta de trabajo al salario a través de las horas de los ocupados, distinta de la decisión de participar en el mercado. En teoría de la producción, la expresión designa también el máximo técnico.

Máximo técnico Punto a partir del cual añadir más factor variable ya no aumenta la producción, sino que la reduce. Técnicamente, es el punto que separa las fases II y III de la producción, también llamado margen intensivo, en el que la producción total es máxima y la productividad marginal del factor variable se anula.

Mayoría óptima Proporción de votos que conviene exigir para aprobar una decisión, equilibrando el riesgo de imponerse a otros con el esfuerzo de ponerse de acuerdo. Técnicamente, según Buchanan y Tullock, es el porcentaje de votos exigido que minimiza la suma de los costes externos y los costes de decisión de una elección colectiva.

Mecanismo de ajuste automático Regla que recorta automáticamente las pensiones cuando el sistema deja de estar equilibrado, sin necesidad de una nueva ley. Técnicamente, en un sistema de cuentas nocionales reduce el tipo de interés nocional cuando el balance actuarial, cociente entre los activos y los pasivos del sistema, cae por debajo de uno, hasta restaurar el equilibrio sin reforma legislativa. Suecia lo activó por primera vez en 2010.

Mecanismo de Equidad Intergeneracional Cotización social adicional que se aplica en España desde 2023 para reforzar los ingresos del sistema de pensiones. Técnicamente, es un recargo que sube del 0,6 % al 1,2 % del salario hasta 2029 y que, junto con la cuota de solidaridad sobre los salarios que superan la base máxima de cotización, aborda la sostenibilidad del sistema por el lado de los ingresos y no por el del gasto.

Mecanismo de flujo de especie Ajuste automático del saldo comercial bajo el patrón oro, también llamado ley de los metales preciosos. Técnicamente, un superávit atrae oro, aumenta la cantidad de dinero y los precios del país, resta competitividad a sus exportaciones y reduce el superávit, y a la inversa con un déficit.

Mecanismo de Loeb y Magat Regulación que premia al monopolio con lo que ganan sus clientes para que deje de restringir la producción. Técnicamente, el monopolista fija libremente el precio y recibe un subsidio igual al excedente del consumidor generado, lo que le incentiva a fijar el precio igual al coste marginal.

Mecanismo de Vickrey-Clarke-Groves Procedimiento para decidir cuánto bien público producir en el que a nadie le conviene mentir sobre lo que lo valora. Técnicamente, es el mecanismo de revelación de preferencias que elige el nivel de bien público que maximiza la suma de valoraciones declaradas y cobra a cada agente decisivo el daño que causa a los demás, de modo que decir la verdad es estrategia dominante.

Mediana generalizada Regla que elige la opción central entre las declaradas, completadas con unas posiciones ficticias fijadas de antemano. Técnicamente, Moulin mostró que, con preferencias unimodales sobre una sola dimensión, las únicas reglas anónimas, unánimes y no manipulables eligen la mediana de los picos declarados y de un número de votos fantasma igual al de individuos menos uno.

Menú de contratos Lista de contratos entre los que cada parte elige el que más le conviene, revelando así lo que sabe. Técnicamente, es el conjunto de contratos que el principal ofrece para que cada tipo de agente elija el suyo y revele así su información. En regulación, Laffont y Tirole lo aplican combinando un pago fijo con una cobertura parcial de las desviaciones de coste.

Mercado completo Situación ideal en la que cualquier riesgo futuro puede cubrirse porque existe un mercado para cada bien y servicio en cada estado futuro de la naturaleza. Técnicamente, es el ideal teórico en el que cada bien se trata como contingente al estado de la naturaleza y tiene su propio mercado. Con mercados completos, el primer teorema del bienestar se extiende a economías con tiempo e incertidumbre.

Mercado contestable, disputable o impugnable Mercado en el que la simple ame-

naza de entrada obliga a las empresas instaladas a comportarse como en competencia, aunque sean pocas. Técnicamente, según Baumol (1982), es aquel en el que un entrante no sufre desventajas frente a los instalados y puede salir sin costes hundidos, de modo que el precio se sitúa en el coste medio y no hay beneficios extraordinarios.

Mercado de dos caras Mercado en el que una plataforma reúne a dos grupos de usuarios que se necesitan entre sí, como compradores y vendedores o anunciantes y lectores. Técnicamente, es aquel en que el valor que obtiene cada grupo depende del número de usuarios del otro (efectos de red cruzados), por lo que el precio de cada lado no refleja solo su coste y puede ser óptimo cobrar un precio nulo a uno de ellos, como mostraron Rochet y Tirole.

Mercado de los limones Ejemplo clásico de cómo la falta de información sobre la calidad puede hundir un mercado: si el comprador no distingue un coche usado bueno de uno malo, solo paga por la calidad media y los dueños de coches buenos dejan de venderlos. Técnicamente, es el modelo de selección adversa de Akerlof (1970), en el que el precio medio puede caer por debajo de la valoración de los vendedores de calidad alta y expulsarlos, hasta el extremo de que solo se negocien coches malos.

Mercado incompleto Situación habitual en la práctica, en la que hay riesgos o bienes futuros que no se pueden comprar ni asegurar. Técnicamente, falta algún mercado para algún bien en algún estado de la naturaleza, por costes de transacción, información asimétrica o riesgos no asegurables, de modo que el primer teorema del bienestar deja de estar garantizado. Hart (1975) mostró que entonces el equilibrio competitivo puede no ser eficiente ni siquiera respecto a los mercados existentes, es decir, en sentido paretiano restringido.

Mercado relevante Conjunto de productos que compiten de verdad entre sí y que las autoridades de competencia toman como referencia. Técnicamente, es el conjunto de bienes que compiten entre sí y sobre el que se evalúa el poder de mercado. Su delimitación se complica cuando los productos están diferenciados.

Mercados contingentes Mercados en los que se negocian hoy promesas de entrega que solo se cumplen si ocurre un determinado acontecimiento, como un seguro. Técnicamente, con un mercado para cada estado de la naturaleza la asignación con información simétrica sería eficiente en el sentido de Pareto, aunque crearlos suele ser demasiado costoso y no resuelven los problemas de información asimétrica.

Método de producción Forma concreta de combinar los factores para fabricar un bien. Técnicamente, es la combinación de cantidades de factores necesaria para producir una unidad de producto, y la función de producción reúne todos los métodos técnicamente eficientes.

Migración de reemplazo Inmigración necesaria para compensar el descenso o el envejecimiento de la población de un país. Técnicamente, la calculó Naciones Unidas en 2000 como el flujo neto que mantendría constante la población total, la población en edad de trabajar o la relación entre esta y los mayores de 65 años, y concluyó que esta última exigiría volúmenes inalcanzables.

Mínimo de explotación Precio por debajo del cual a la empresa le conviene cerrar a corto plazo, porque no llega a cubrir sus costes variables. Técnicamente, coincide con el mínimo del coste variable medio, donde la empresa es indiferente entre producir y cerrar.

Mochila austriaca Fondo que la empresa va llenando para cada trabajador y que este se lleva consigo al cambiar de empleo. Técnicamente, sustituye la indemnización por despido por una cuenta individual de capitalización financiada con aportaciones periódicas del empleador, disponible en caso de despido o, si no se usa, al jubilarse.

Modelo Beveridge Sistema de protección social que cubre a todos los residentes y se paga con impuestos. Técnicamente, es el modelo universal de protección social, heredero del Informe Beveridge, que cubre a toda la población residente y se financia con impuestos generales.

Modelo Bismarck Sistema de protección social que cubre sobre todo a quienes trabajan y se paga con lo que cotizan. Técnicamente, es el modelo

 © 2026 Manuel Cabado García

profesional de protección social cuya cobertura se limita fundamentalmente a los trabajadores y sus familias y que se financia con cotizaciones sociales proporcionales a la renta.

Modelo de Allingham y Sandmo Modelo que trata la decisión de evadir impuestos como una apuesta. Técnicamente, en el planteamiento de Allingham y Sandmo (1972) el contribuyente elige cuánta renta declarar maximizando su utilidad esperada, dados el tipo impositivo, la probabilidad de inspección y la sanción. Más probabilidad o más sanción reducen la evasión, mientras que el efecto del tipo impositivo es ambiguo porque sus efectos sustitución y renta van en sentidos opuestos.

Modelo de Barro Modelo que explica por qué un sector público ni demasiado pequeño ni demasiado grande maximiza el crecimiento. Técnicamente, Barro (1990) introdujo los servicios públicos como factor de la producción privada en un modelo de crecimiento endógeno, de modo que un mayor peso del gasto eleva la productividad del capital pero exige más impuestos distorsionantes, y la relación entre tamaño del sector público y crecimiento tiene forma de U invertida. Con tecnología Cobb-Douglas, el tamaño que maximiza el crecimiento iguala la participación del gasto en el PIB a la elasticidad de la producción respecto a los servicios públicos.

Modelo de Bertrand Modelo de dos empresas que compiten fijando precios. Técnicamente, es un modelo de duopolio simultáneo en el que las empresas compiten en precios con producto homogéneo y costes marginales constantes e iguales, de modo que el precio de equilibrio se iguala al coste marginal.

Modelo de búsqueda (empleo) Modelo que explica qué empleo acepta un parado cuando las ofertas llegan poco a poco y con salarios distintos. Técnicamente, el parado acepta la primera oferta que supera su salario de reserva, de modo que la duración del paro depende de ese umbral y de la frecuencia de las ofertas (McCall y Mortensen, 1970).

Modelo de Chamberlin Modelo clásico de un mercado con muchas empresas que venden productos parecidos pero no idénticos. Técnicamente, es el modelo de competencia monopolística de grupo numeroso que Chamberlin (1933) planteó con productos sustitutivos cercanos, libre entrada y empresas miopes, y que a largo plazo conduce a beneficios nulos con exceso de capacidad.

Modelo de ciudad circular de Salop Modelo en el que las empresas y los consumidores se reparten alrededor de un círculo, lo que permite estudiar cuántas empresas caben en un mercado. Técnicamente, es un modelo de diferenciación espacial con costes fijos de entrada en el que las empresas se sitúan equidistantes en el círculo y el número de empresas es endógeno.

Modelo de contrato eficiente Negociación en la que se pactan a la vez el salario y el número de empleos, de modo que no queda ningún acuerdo mejor para ambas partes. Técnicamente, el acuerdo está sobre la curva de contrato, a la derecha de la demanda de trabajo, y con trabajadores aversos al riesgo genera más empleo que la competencia (McDonald y Solow, 1981).

Modelo de Cournot Modelo de dos empresas que deciden a la vez cuánto producir. Técnicamente, es un modelo de duopolio en el que las empresas eligen simultáneamente cantidades con variación conjetural nula, y cuyo equilibrio se sitúa en la intersección de sus funciones de reacción.

Modelo de derecho a gestionar Negociación en la que sindicato y empresa pactan el salario y la empresa decide después cuánta gente contrata. Técnicamente, el acuerdo está sobre la demanda de trabajo y es ineficiente en el sentido de Pareto, porque la curva de indiferencia del sindicato y la isobeneficio de la empresa se cortan en vez de ser tangentes (Nickell y Andrews, 1983).

Modelo de Dixit de inversión en capacidad Modelo en el que la empresa ya instalada invierte de antemano para convencer a los posibles rivales de que no les conviene entrar. Técnicamente, es un juego en tres etapas en el que la empresa instalada incurre en un coste hundido, como ampliar capacidad, para hacer creíble la amenaza de luchar y disuadir así la entrada.

Modelo de Dixit-Stiglitz Modelo que explica por qué en los mercados más grandes se ofrecen

más variedades de un mismo tipo de producto. Técnicamente, es un modelo de competencia monopolística con consumidor representativo, utilidad de elasticidad de sustitución constante con preferencia por la variedad, economías de escala internas y libre entrada, en el que el número de variedades crece con el tamaño del mercado.

Modelo de Downs-Hotelling Explicación de por qué, cuando compiten dos grandes partidos, ambos tienden a parecerse y a disputarse el centro. Técnicamente, es el modelo de competencia electoral bipartidista en el que, con votantes de preferencias unimodales sobre una ideología unidimensional, los partidos convergen hacia la posición del votante mediano.

Modelo de generaciones solapadas Modelo de una economía en la que conviven en cada momento personas de distintas edades. Técnicamente, los individuos viven dos períodos, juventud y vejez, y en cada período coinciden ambas generaciones. Samuelson (1958) lo planteó con intercambio puro y Diamond (1965) le añadió capital y producción, lo que permite comparar el reparto con la capitalización y muestra que el equilibrio competitivo puede ser dinámicamente ineficiente.

Modelo de Green y Porter Modelo que explica por qué los cárteles atraviesan guerras de precios aunque nadie haya hecho trampa. Técnicamente, es un modelo de colusión con información imperfecta en el que las empresas, al observar solo el precio, castigan con guerras de precios temporales cualquier caída por debajo de un umbral, aunque nadie haya incumplido.

Modelo de grilletes dorados Contrato que paga poco al principio y mucho más después, de modo que solo resulta atractivo para quien piensa quedarse muchos años en la empresa. Técnicamente, es el mecanismo de autoselección de Salop y Salop (1976), en el que un salario creciente con la antigüedad filtra a los trabajadores con baja propensión a abandonar la empresa y reduce la rotación.

Modelo de Harberger Modelo que estudia cómo un impuesto en un mercado acaba afectando al resto de la economía. Técnicamente, es un modelo bisectorial de equilibrio general con dos bienes, dos factores de oferta fija y perfectamente móviles y competencia perfecta, usado para analizar la incidencia de los impuestos entre mercados.

Modelo de Hotelling Modelo que estudia cómo compiten unas empresas que eligen dónde situarse a lo largo de una calle, como forma de representar cuánto se parecen sus productos. Técnicamente, es un modelo de competencia espacial en una ciudad lineal en el que la localización de las empresas representa el grado de una característica del producto y los costes de transporte representan la heterogeneidad de gustos.

Modelo de huelgas de Hicks Modelo que explica el salario pactado por lo que cada parte está dispuesta a aguantar una huelga. Técnicamente, el salario se fija en el cruce entre la curva de concesión de la empresa, creciente con la duración de la huelga, y la curva de resistencia del sindicato, decreciente con ella (Hicks, 1932).

Modelo de insiders-outsiders Teoría según la cual los trabajadores con empleo negocian salarios que protegen sus puestos aunque dejen fuera a los parados. Técnicamente, los costes de rotación dan poder de negociación a los insiders, que logran salarios superiores a los que aceptarían los outsiders sin que a la empresa le compense sustituirlos (Lindbeck y Snower, 1988).

Modelo de Kreps y Scheinkman Modelo en el que las empresas deciden primero cuánto pueden producir y luego compiten en precios. Técnicamente, es un juego en dos etapas, primero de elección de capacidad y después de competencia en precios, cuyo equilibrio reproduce las cantidades y el precio de Cournot.

Modelo de la telaraña Explica los ciclos de precios de mercados en los que la producción tarda en ajustarse, como los agrícolas. Técnicamente, los productores deciden cuánto producir según el precio del periodo anterior, y el precio oscila de forma convergente, divergente o constante según la oferta sea menos, más o igual de elástica que la demanda.

Modelo de Lancaster (demanda de características) Forma de entender el consumo en la que lo que se compra en realidad es un conjunto de cualidades, y cada producto es

 © 2026 Manuel Cabado García

solo un medio de obtenerlas. Técnicamente, es la teoría de la demanda según la cual los consumidores no demandan bienes en sí mismos sino las características que contienen, presentes en cada bien en distintas proporciones.

Modelo de Lucas y Rapping Modelo que explica por qué el empleo fluctúa mucho a lo largo del ciclo aunque el salario real varíe poco. Técnicamente, Lucas y Rapping (1969) fueron los primeros en tratar la oferta de trabajo como una decisión intertemporal: un individuo que vive dos periodos y puede prestar o pedir prestado al tipo de interés real reparte sus horas según el salario relativo entre periodos. Su razonamiento lo extendió después la teoría del ciclo real.

Modelo de Rothschild y Stiglitz Modelo de un mercado de seguros en el que la aseguradora no sabe quién tiene más riesgo y ofrece varias pólizas para que cada cliente elija la suya. Técnicamente, con aseguradoras en competencia, el equilibrio, si existe, es separador: los individuos de alto riesgo obtienen cobertura completa y los de bajo riesgo solo una cobertura parcial, que es el coste de distinguirse.

Modelo de señalización de Spence Explicación de por qué las empresas pagan más a los titulados aunque los estudios no hagan más productivo a nadie: el título cuesta menos a quien ya es más capaz, y por eso le sirve para demostrarlo. Técnicamente, la educación, más barata para los trabajadores productivos, permite un equilibrio separador en el que cada tipo elige un nivel de educación distinto y cobra un salario acorde con su productividad.

Modelo de Shapiro-Stiglitz Modelo en el que la empresa paga más de lo necesario para que perder el empleo cueste y nadie se escaquee. Técnicamente, al no observar el esfuerzo, la empresa fija el salario sobre la curva de no escaqueo, y en equilibrio hay paro involuntario, que actúa como mecanismo de disciplina (Shapiro y Stiglitz, 1984).

Modelo de Stackelberg Modelo de duopolio en el que una empresa mueve antes que la otra y saca ventaja de ello. Técnicamente, es un juego secuencial en cantidades en el que la líder elige su producción incorporando la función de reacción de la seguidora, que actúa con variación conjetural nula.

Modelo del precio límite Estrategia por la que la empresa instalada renuncia a parte de su beneficio para que a nadie le compense entrar. Técnicamente, en el modelo de Bain, Sylos-Labini y Modigliani la instalada produce más que un monopolista y fija el precio más alto con el que la entrada de un competidor a su escala mínima eficiente hundiría el precio hasta el coste medio, bajo el supuesto, poco creíble, de que mantendrá su producción si el rival entra.

Modelo DMP Modelo que explica por qué hay paro aunque el salario sea flexible: encontrar empleo, o a quien contratar, lleva tiempo. Técnicamente, combina una función de emparejamiento, la libre entrada de vacantes y un salario negociado a la Nash para determinar a la vez la tensión, el salario y el paro de equilibrio (Diamond, Mortensen y Pissarides).

Modelo IS-LM Modelo que explica a la vez la producción y el tipo de interés de una economía a corto plazo. Técnicamente, es el modelo de economía cerrada con precios rígidos, formulado por Hicks, que determina la renta y el tipo de interés en la intersección de la curva IS, que recoge el equilibrio del mercado de bienes, y la curva LM, que recoge el equilibrio del mercado de dinero.

Modelo multicapa Relación de encargo con muchos eslabones, como la que va de los accionistas al consejo de administración, de este a la alta dirección y de ella a los mandos intermedios. Técnicamente, es la extensión de la teoría de la agencia a relaciones con varios niveles de delegación o con múltiples agentes y principales interconectados.

Modelos de equilibrio general dinámico estocástico (DSGE) Modelos macroeconómicos que construyen la economía a partir de las decisiones de hogares y empresas que miran al futuro. Técnicamente, son modelos de equilibrio general con agentes que optimizan de forma intertemporal, expectativas racionales y perturbaciones aleatorias, y son el marco común de la teoría de los ciclos reales y de la nueva síntesis neokeynesiana.

Monetarismo Escuela que da al dinero un papel central en la explicación de la inflación. Técnicamente, es la escuela encabezada por Friedman que estudia los efectos de la oferta monetaria sobre precios, empleo y producción y sostiene que la inflación es un fenómeno monetario y que el dinero solo tiene efectos reales a corto plazo.

Monopolio Mercado con una sola empresa vendedora. Técnicamente, es la estructura de mercado con un único oferente de un producto sin sustitutivos cercanos, protegido por barreras de entrada, que fija el precio o la cantidad para maximizar su beneficio.

Monopolio bilateral Mercado en el que solo hay un vendedor y un comprador. Técnicamente, su resultado es indeterminado y depende del poder negociador de las partes.

Monopolio multiplanta Monopolio que fabrica el mismo producto en varias fábricas. Técnicamente, como las plantas tienen distintos costes, reparte la producción igualando los costes marginales de las plantas entre sí y con el ingreso marginal.

Monopolio multiproducto Monopolio que vende varios bienes distintos. Técnicamente, su precio óptimo para cada bien depende de si las demandas son sustitutivas o complementarias y de si los costes son separables, de modo que la sustituibilidad entre sus bienes eleva su poder de mercado y la complementariedad lo reduce.

Monopolio natural Mercado en el que lo más barato es que produzca una sola empresa, como ocurre con una red de distribución de agua. Técnicamente, existe cuando la función de costes es subaditiva en el nivel de producción demandado, para lo que las economías de escala bastan, aunque no son imprescindibles.

Monopsonio Mercado con un único comprador. Técnicamente, como se enfrenta a toda la curva de oferta, su gasto marginal supera al precio, por lo que compra menos y paga menos que en competencia, y en el mercado de trabajo paga el salario que indica la oferta para el empleo en el que el ingreso marginal del factor iguala al gasto marginal.

Monotonía En general, propiedad de lo que varía siempre en el mismo sentido. Técnicamente, en teoría del consumidor es el axioma según el cual el consumidor prefiere más a menos de cualquier bien (“cuanto más, mejor”), lo que hace creciente la función de utilidad y da pendiente negativa a las curvas de indiferencia.

Motivo especulación Deseo de guardar dinero a la espera de mejores oportunidades de inversión. Técnicamente, es la demanda de dinero como depósito de valor, inversamente relacionada con el tipo de interés, porque con tipos bajos se esperan subidas que abaratarían los bonos y se prefiere atesorar liquidez.

Motivo precaución Demanda de dinero para cubrir gastos imprevistos. Técnicamente, es el segundo motivo keynesiano para mantener dinero, crece con la renta y Whalen lo fundamentó con los costes de quedarse sin liquidez cuando ingresos y pagos son inciertos.

Motivo transacción Demanda de dinero para hacer frente a los pagos ordinarios. Técnicamente, es uno de los tres motivos de Keynes para mantener dinero, depende positivamente de la renta, y Baumol y Tobin lo fundamentaron después con modelos de gestión de inventarios.

Multiplicador keynesiano Efecto por el que un aumento del gasto acaba elevando la renta en una cantidad mayor que el propio gasto, porque el gasto de un agente es el ingreso de otro. Técnicamente, es el efecto más que proporcional sobre la renta de un aumento exógeno de un componente de la demanda agregada. Es mayor que uno al estar la propensión marginal a consumir entre cero y uno.

N

NAIRU Tasa de paro por debajo de la cual la inflación empieza a acelerarse. Técnicamente, en la nueva economía keynesiana resulta del conflicto entre los salarios reales que reclaman los trabajadores y los márgenes que fijan las empresas en mercados imperfectos, y depende de factores como la negociación colectiva o la regulación.

Necesidad de gasto Lo que le cuesta a un territorio prestar a sus residentes unos servicios pú-

 © 2026 Manuel Cabado García

blicos de nivel estándar. Técnicamente, depende de su población y de características que encarecen la prestación, como la edad o la dispersión de sus habitantes, y en España se mide mediante la población ajustada.

Neutralidad ante el riesgo Actitud de quien solo se fija en lo que gana en promedio, sin importarle el riesgo. Técnicamente, el agente es indiferente entre jugar una lotería y recibir con certeza su valor esperado, lo que corresponde a una función de utilidad lineal.

Nivelación Reparto que lleva recursos de los territorios con más capacidad para recaudar a los que tienen menos, para que todos puedan prestar servicios parecidos. Técnicamente, son transferencias que cubren la diferencia entre la capacidad fiscal de cada jurisdicción y su necesidad de gasto, de modo que la financiación por habitante ajustado tiende a igualarse.

No apropiación del excedente social Situación en la que quien lanza un producto nuevo no cobra todo el beneficio que este da a los consumidores, lo que le resta incentivos para lanzarlo. Técnicamente, es una externalidad positiva por la que una empresa no puede apropiarse de todo el excedente del consumidor que genera una nueva variedad, lo que lleva a introducir menos variedades que el óptimo social.

No exclusión Propiedad de un bien del que nadie puede quedar fuera, pague o no. Técnicamente, es la imposibilidad, o el coste prohibitivo, de impedir su consumo a quien no paga su precio, lo que da lugar al problema del free-rider.

No rivalidad Propiedad de un bien que pueden disfrutar más personas sin que las demás disfruten menos. Técnicamente, el consumo de un agente no reduce los beneficios que el bien reporta al resto, de modo que el coste marginal de un usuario adicional es nulo.

No saciedad local Supuesto de que el consumidor nunca está del todo satisfecho, porque siempre hay algo cercano que preferiría. Técnicamente, es el supuesto, más débil que la monotonía, según el cual en torno a cualquier cesta existe siempre otra cercana estrictamente preferida.

Novísima Economía del Bienestar Etapa de la economía del bienestar posterior a la Nueva Economía del Bienestar, que estudia qué ocurre cuando no pueden cumplirse los supuestos del óptimo de Pareto y cómo agregar las preferencias individuales en decisiones colectivas. Técnicamente, agrupa las aportaciones de los años cincuenta y sesenta, sobre todo la teoría de la elección colectiva de Arrow y la teoría del second best de Lipsey y Lancaster.

Núcleo (core) Conjunto de repartos que ningún agente, ni ningún grupo de agentes, tiene interés en rechazar para negociar por su cuenta. Técnicamente, son las asignaciones factibles que no puede bloquear ninguna coalición, concepto procedente de la teoría de juegos cooperativos. Con dos agentes coincide con el tramo de la curva de contrato situado dentro de la lente.

Nudge Pequeño cambio en la forma de presentar una decisión que orienta lo que se elige sin obligar a nada, literalmente "empujón". Técnicamente, según Thaler y Sunstein, es cualquier aspecto de la arquitectura de la elección que altera el comportamiento de forma predecible sin prohibir opciones ni cambiar significativamente los incentivos económicos, y que puede evitarse de forma fácil y barata.

Nueva Economía del Bienestar Etapa de la economía del bienestar que intentó elegir entre situaciones eficientes que el criterio de Pareto no permite ordenar. Técnicamente, es la corriente de los años treinta y cuarenta que trató de superar la indeterminación paretiana mediante los criterios de compensación y la función de bienestar social.

Nueva economía keynesiana Corriente que explica por qué el dinero y la demanda influyen en la producción a partir de las dificultades para ajustar precios y salarios. Técnicamente, es la corriente surgida en los años ochenta que, con el mismo rigor microeconómico que la nueva macroeconomía clásica, introduce competencia imperfecta y rigideces de precios y salarios que hacen que el dinero no sea neutral.

Nueva macroeconomía clásica Escuela que, desde los años setenta, rehízo la macroeconomía a partir del comportamiento optimizador de los

individuos. Técnicamente, encabezada por Lucas, Sargent, Barro, Kydland y Prescott, combina expectativas racionales, agentes que optimizan de forma intertemporal y mercados que se vacían continuamente, por lo que solo las perturbaciones no anticipadas tienen efectos reales.

Numerario Bien que sirve de unidad de cuenta para expresar los precios de los demás. Técnicamente, su precio se normaliza a uno y, en el análisis de equilibrio parcial, representa el gasto agregado en todos los bienes distintos del estudiado.

Número de Ramsey Medida de cuánto deben alejarse los precios regulados del coste marginal para cubrir los costes. Técnicamente, es el cociente entre el multiplicador de la restricción de beneficio de la regla de Ramsey-Boiteux y ese mismo multiplicador más uno, y está comprendido entre cero y uno. Vale cero cuando la restricción no obliga y los precios igualan el coste marginal, y se acerca a uno cuando los precios se aproximan a los de monopolio.

O

Oferta de trabajo Horas que una persona está dispuesta a trabajar a cada salario. Técnicamente, se deriva de maximizar una utilidad que depende del consumo y del ocio, sujeta a una restricción presupuestaria y a otra temporal que reparte el tiempo disponible entre ocio y trabajo.

Oligopolio Mercado dominado por unas pocas empresas, cada una pendiente de lo que hacen las demás. Técnicamente, es una estructura con pocos oferentes y muchos demandantes, barreras de entrada e interdependencia estratégica reconocida por las empresas, que les confiere poder de mercado.

Oligopolio natural Mercado en el que solo caben unas pocas empresas aunque la entrada sea libre. Técnicamente, en los modelos de diferenciación vertical de Shaked y Sutton, si los consumidores son homogéneos en su disposición a pagar por la calidad, una empresa adicional con otra calidad no encontraría demanda.

Opción por defecto Lo que ocurre si la persona no decide nada. Técnicamente, es la alternativa que se aplica en ausencia de una elección expresa y, por la inercia, tiende a ser la más elegida, como muestran la afiliación automática a planes de pensiones o el consentimiento presunto en la donación de órganos.

Óptimo de explotación Precio a partir del cual la empresa empieza a ganar dinero en sentido económico. Técnicamente, coincide con el mínimo del coste total medio, donde el beneficio es nulo, y también se llama umbral de beneficios.

Óptimo de Pareto (optimalidad paretiana) Situación en la que ya no se puede mejorar a nadie sin empeorar a otro, también llamada optimalidad paretiana o eficiencia en el sentido de Pareto. Técnicamente, es una asignación factible para la que no existe ninguna otra Pareto superior, lo que exige que sea a la vez óptima en la producción, en el consumo y globalmente. Hay infinitos óptimos, no comparables entre sí, y ninguno garantiza por sí solo el máximo bienestar social ni un reparto equitativo.

Óptimo de primer orden Regulación que obliga al monopolio a vender al precio que fijaría la competencia. Técnicamente, busca restablecer la eficiencia asignativa imponiendo que el precio iguale al coste marginal y cubriendo las pérdidas con subvenciones o eliminando los beneficios con impuestos de suma fija.

Óptimo de segundo orden Regulación que acepta no alcanzar la eficiencia plena para que la empresa regulada no tenga pérdidas. Técnicamente, maximiza el bienestar sujeto a una restricción de beneficio no negativo, como en el precio igual al coste medio o los precios de Ramsey.

Óptimo de tercer orden Regulación que se conforma con poner límite a las ganancias del monopolio. Técnicamente, ante información insuficiente para restablecer la eficiencia, se limita a acotar los beneficios extraordinarios del monopolista.

Óptimo en el consumo Reparto de los bienes ya producidos que no se puede cambiar para contentar más a alguien sin disgustar a otro. Técnicamente, es la distribución de los bienes entre consumidores en la que no es posible aumentar la utilidad de uno sin disminuir la de otro, lo que exige igualar sus relaciones marginales de sustitución.

 © 2026 Manuel Cabado García

Óptimo en la producción Reparto de los recursos productivos que no desperdicia nada, porque ya no se puede fabricar más de algo sin fabricar menos de otra cosa. Técnicamente, es la asignación de los factores entre industrias en la que no es posible producir más de un bien sin reducir la producción de otro, lo que exige igualar la relación marginal de sustitución técnica entre industrias.

Óptimo global Situación en la que se produce sin desperdicio, se reparte sin desperdicio y, además, se fabrica la combinación de bienes que los consumidores valoran. Técnicamente, es una asignación eficiente en producción y en consumo en la que las relaciones marginales de sustitución de todos los consumidores coinciden con la relación marginal de transformación.

Óptimo técnico Punto en el que cada unidad del factor variable rinde, en promedio, lo máximo posible. Técnicamente, es el punto que separa las fases I y II de la producción, también llamado margen extensivo, en el que la productividad media del factor variable es máxima e igual a la marginal.

Ordinalidad Regla de que el reparto solidario no debe dar la vuelta a la clasificación de los territorios: quien más capacidad fiscal tiene antes debe seguir teniendo más recursos después. Técnicamente, exige que las transferencias de nivelación no alteren el orden de las regiones por capacidad fiscal por habitante, aunque reduzcan las diferencias entre ellas. Esta es la versión débil. En su versión fuerte, quien más aporta debe recibir más.

P

Pacto de Estabilidad y Crecimiento Conjunto de reglas fiscales de la Unión Europea que limitan el déficit y la deuda de los Estados miembros. Técnicamente, tiene un brazo preventivo, que desde la reforma de 2024 se basa en planes nacionales a medio plazo con una senda de gasto neto que fija el Consejo a recomendación de la Comisión, y un brazo correctivo, el procedimiento por déficit excesivo. Este puede abrirse si el déficit supera el tres por ciento del PIB, o si la deuda supera el sesenta por ciento y el país se desvía de su senda de gasto neto, y por la vía del déficit exige un ajuste estructural mínimo de medio punto del PIB al año.

Paradoja de Allais Experimento que muestra que la gente no siempre elige entre apuestas como predice la teoría de la utilidad esperada. Técnicamente, es la evidencia experimental de que los individuos violan el axioma de independencia, al preferir un resultado seguro a una lotería de mayor valor esperado pero invertir su elección cuando ambas alternativas son inciertas.

Paradoja de Bertrand Resultado según el cual bastan dos empresas para que un mercado funcione como si fuera competitivo. Técnicamente, si compiten en precios con producto homogéneo, costes marginales constantes e iguales y sin límites de capacidad, el precio de equilibrio se iguala al coste marginal y el beneficio es nulo.

Paradoja de Condorcet Situación en la que una votación por mayoría se queda sin ganador porque cada opción pierde frente a alguna otra. Técnicamente, aunque las preferencias de cada votante sean transitivas, la preferencia social obtenida por mayoría en votaciones por pares es intransitiva y genera mayorías cíclicas.

Paradoja de Diamond Resultado según el cual, si las empresas fijan los salarios y buscar cuesta, todas acaban pagando lo mínimo aceptable y no queda nada que buscar. Técnicamente, la distribución de salarios se reduce a un único valor, el salario de reserva, lo que contradice la dispersión que justificaba la búsqueda (Diamond, 1971).

Paradoja de Ellsberg Experimento que muestra que la gente prefiere apostar cuando conoce las probabilidades que cuando no las conoce. Técnicamente, es la evidencia de aversión a la ambigüedad, según la cual los individuos prefieren loterías con probabilidades conocidas a otras con probabilidades desconocidas, lo que viola el axioma del elemento seguro de Savage.

Paradoja de la cadena de tiendas Resultado según el cual un monopolista no puede frenar la entrada de competidores amenazando con guerras de precios cuando el número de mercados es finito y conocido. Técnicamente, la inducción hacia

atrás muestra que acomodarse es óptimo en el último mercado y, por extensión, en todos, aunque la intuición sugiera que luchar al principio crearía reputación.

Paradoja de San Petersburgo Contradicción entre el valor esperado infinito de un juego de doble o nada y la pequeña cantidad que cualquiera pagaría por jugarlo. Técnicamente, Bernoulli la resolvió suponiendo una utilidad marginal de la riqueza decreciente, aunque la paradoja reaparece con otros premios si la utilidad no está acotada.

Paradoja de Scitovsky Contradicción por la que un mismo criterio puede aconsejar un cambio y, una vez hecho, aconsejar deshacerlo. Técnicamente, es la inconsistencia de los criterios de compensación por la que, cuando las fronteras de posibilidades de utilidad se cortan, un mismo criterio recomienda pasar de un estado a otro y luego volver al inicial.

Paradoja del agua y los diamantes Aparente contradicción de que algo tan útil como el agua sea barato y algo tan prescindible como los diamantes sea caro. Técnicamente, el agua tiene un valor de uso muy superior al de los diamantes y, sin embargo, un valor de cambio mucho menor. Los marginalistas la resolvieron distinguiendo entre utilidad total y utilidad marginal.

Paradoja del ahorro Idea de que lo que es prudente para un hogar puede ser perjudicial para el conjunto. Técnicamente, si todos intentan ahorrar más a la vez, la caída de la demanda agregada reduce la renta y el ahorro total no aumenta, o incluso disminuye.

Paradoja liberal de Sen Conflicto entre respetar la libertad de cada persona en sus asuntos privados y aceptar siempre lo que todos prefieren. Técnicamente, Sen (1970) demostró que ninguna regla de elección social con dominio universal puede cumplir a la vez el criterio de Pareto débil y la libertad mínima, que hace a al menos dos individuos decisivos sobre un par de alternativas de su esfera personal.

Pareto superior Se dice de un cambio, o de la situación a la que conduce, que beneficia a alguien sin perjudicar a nadie. Técnicamente, una asignación es Pareto superior a otra si, al pasar de la segunda a la primera, al menos un individuo mejora y ninguno empeora, lo que no la convierte por sí solo en un óptimo de Pareto.

Paternalismo Intervención pública que corrige las decisiones de los individuos por su propio bien. Técnicamente, parte de que el Estado conoce mejor que los individuos los beneficios de ciertos bienes y seguros, porque estos tienden a subestimar los riesgos futuros. Con ese argumento se justifican la afiliación obligatoria a los seguros sociales y la provisión pública de bienes preferentes como la sanidad.

Paternalismo libertario Postura que defiende orientar las decisiones de las personas por su propio bien sin quitarles la libertad de elegir otra cosa. Técnicamente, sostiene que algún grado de paternalismo es inevitable, porque alguien debe fijar la opción por defecto, y que debe ejercerse sin coacción.

Penalización por hijos Caída de los ingresos que sufren las madres, y no los padres, tras tener un hijo. Técnicamente, se mide con estudios de eventos en torno al nacimiento del primer hijo y recoge la menor participación, el menor número de horas y el menor salario por hora, sea cual sea su causa.

Pensión de jubilación Renta periódica que se cobra al dejar de trabajar por edad. Técnicamente, es la prestación económica que compensa la pérdida de la capacidad del trabajador para generar rentas laborales por vejez, y se financia por reparto, con las cotizaciones de quienes trabajan en ese momento, o por capitalización, con lo acumulado por cada trabajador.

Percusión Momento en que el impuesto recae sobre quien la ley obliga a pagarlo. Técnicamente, es la recaída del tributo sobre el sujeto pasivo contribuyente, obligado por ley a liquidarlo.

Pérdida crítica Máximo porcentaje de ventas que un monopolista hipotético puede perder sin que una subida de precio deje de compensarle. Técnicamente, en el test SSNIP se compara con la pérdida realmente esperada para decidir si el mercado propuesto está bien delimitado.

 © 2026 Manuel Cabado García

Pérdida irrecuperable de eficiencia Pérdida de bienestar que no se convierte en ganancia para nadie. Técnicamente, es la parte del excedente total que desaparece cuando una distorsión, como el poder de monopolio o un impuesto, reduce la cantidad intercambiada por debajo de la eficiente, porque dejan de realizarse intercambios cuya valoración supera su coste.

Periodo de mercado Plazo tan corto que no da tiempo a producir más ni a que entren nuevas empresas. Técnicamente, es el muy corto plazo en el que la producción y el número de empresas están dados, de modo que la oferta es totalmente inelástica y el precio solo raciona la demanda.

Peso significativo en el mercado Posición de una operadora lo bastante fuerte como para actuar sin preocuparse de sus rivales. Técnicamente, es la posición de fuerza económica que permite a una empresa comportarse con independencia apreciable de competidores, clientes y consumidores, y que permite imponerle obligaciones ex ante según el Código Europeo de las Comunicaciones Electrónicas.

Pesos decisorios Importancia que las personas dan de hecho a cada probabilidad, que no coincide con la probabilidad real. Técnicamente, son las ponderaciones que sustituyen a las probabilidades objetivas en la teoría de las perspectivas y que sobreponderan las probabilidades bajas e infraponderan las altas.

Planificador central (dictador benevolente) Figura teórica que decide por toda la sociedad buscando el máximo bienestar, sin usar precios de mercado. Técnicamente, maximiza una función de bienestar social sujeta solo a las restricciones de recursos y de tecnología. Sin distorsiones, su solución coincide con el equilibrio competitivo, y esa coincidencia es la idea de fondo de los teoremas del bienestar.

Plusvalía En la teoría de Marx, parte del valor que crean los trabajadores y que no se les paga. Técnicamente, es la diferencia entre el valor total generado por el trabajo y lo pagado al trabajador como salario, o capital variable, que el capitalista se apropia y que para Marx es la esencia de la explotación.

Población ajustada (o habitante ajustado) Población corregida para reflejar cuánto cuesta de verdad atender a los residentes de un territorio. Técnicamente, pondera a los habitantes según los factores que elevan la necesidad de gasto, como la edad o la dispersión de la población, y es la unidad con la que se compara la financiación por habitante entre comunidades autónomas.

Poder compensador Contrapeso que un grupo organizado opone al poder de mercado de la otra parte. Técnicamente, en el mercado de trabajo es el papel del sindicato frente a una empresa monopsonista, porque un salario negociado entre el de monopsonio y el competitivo puede elevar a la vez el salario y el empleo. La expresión procede de Galbraith, que la aplicó en general a los contrapesos del poder de las grandes empresas.

Poder de mercado Capacidad de una empresa para subir su precio sin perder todas sus ventas. Técnicamente, es la posibilidad de fijar un precio superior al coste marginal, que existe cuando la demanda a la que se enfrenta la empresa no es perfectamente elástica.

Políticas activas de empleo Medidas públicas que ayudan a los parados a encontrar empleo, en lugar de solo sostener su renta. Técnicamente, incluyen la intermediación de los servicios públicos de empleo, la formación de los parados y el fomento de la movilidad geográfica, y buscan mejorar la eficiencia del emparejamiento, es decir, acercar la curva de Beveridge al origen.

Posibilidad de inacción Supuesto de que la empresa siempre puede optar por no producir nada. Técnicamente, es el axioma que permite a la empresa cesar su actividad, de modo que el origen pertenece al conjunto de producción, y deja de cumplirse cuando existen costes hundidos.

Praxeología Método que pretende deducir la teoría económica por pura lógica, sin contrastarla con los datos. Técnicamente, parte del axioma de que el ser humano actúa con un propósito y es la base metodológica de la escuela austriaca a partir de Ludwig von Mises.

Precio-aceptante Quien tiene que conformarse con el precio que marca el mercado. Técnicamen-

te, es el agente que no puede influir en el precio de mercado, de modo que la empresa percibe una demanda horizontal en la que puede vender cuanto desee al precio vigente y nada a un precio superior.

Precio natural Precio en torno al cual oscilan los precios de mercado cuando se dejan a un lado las perturbaciones pasajeras. Técnicamente, es el precio de equilibrio a largo plazo en competencia perfecta, determinado por los costes de producción, en torno al cual gravita el precio de mercado.

Precio relativo Precio de un bien medido en unidades de otro bien, y no en dinero. Técnicamente, es el cociente entre los precios nominales de dos bienes. En el equilibrio competitivo se iguala a la relación marginal de sustitución de todos los consumidores y a la relación marginal de transformación, y así coordina las decisiones de consumo y de producción.

Precio sombra Precio que refleja el verdadero valor social de un bien o de un recurso cuando el de mercado no lo hace. Técnicamente, en el análisis coste-beneficio es el precio de cuenta con el que el sector público valora costes y beneficios para corregir las distorsiones de los precios de mercado. En un problema de optimización coincide con el multiplicador de una restricción, es decir, con lo que mejora el objetivo si esa restricción se relaja en una unidad.

Precios de Lindahl Sistema en el que cada persona paga por un bien público un precio distinto, según lo que lo valore. Técnicamente, son precios personalizados por los que cada individuo paga el bien público en función de su valoración marginal, lo que permite cumplir la condición de Samuelson si se revelan las preferencias.

Precios de Ramsey (regla de Ramsey-Boiteux) Precios que permiten a un monopolio multiproducto cubrir sus costes con la menor pérdida de bienestar posible. Técnicamente, el margen relativo de cada bien sobre su coste marginal es inversamente proporcional a su elasticidad-precio y crece con lo exigente que sea la restricción de beneficio (regla de Ramsey-Boiteux o de la elasticidad inversa).

Precios frente a cantidades (Weitzman) Criterio para elegir entre poner precio a la contaminación o fijar un tope de emisiones cuando no se conocen bien los costes de reducirla. Técnicamente, Weitzman (1974) mostró que, con incertidumbre sobre esos costes, conviene un impuesto si el coste marginal de reducir emisiones es más inclinado que su beneficio marginal y un tope de cantidad en el caso contrario, como cuando el daño se dispara a partir de un umbral.

Precios hedónicos Forma de averiguar cuánto se paga por cada una de las características de un producto. Técnicamente, es el método econométrico que descompone el precio de un bien heterogéneo en los precios implícitos de sus características.

Preferencia por la variedad Gusto del consumidor por repartir su gasto entre muchas variedades en lugar de concentrarlo en una. Técnicamente, la utilidad aumenta, ceteris paribus, con el número de variedades consumidas, de modo que consumir una variedad más siempre reporta más utilidad, como ocurre con la función de utilidad de elasticidad de sustitución constante del modelo de Dixit-Stiglitz.

Preferencia temporal Tendencia a valorar más un bien disponible hoy que el mismo bien en el futuro. Técnicamente, Böhm-Bawerk la explicó por la expectativa de una mayor riqueza futura, por la infravaloración de las necesidades futuras y por la superioridad técnica de los bienes presentes, y en ella basó el interés como precio por disfrutar de los bienes presentes en lugar de esperar.

Preferencias cuasilineales Preferencias en las que la cantidad demandada de un bien no depende de la renta. Técnicamente, la utilidad es la suma de una función del bien estudiado y de la cantidad de numerario, que entra de forma lineal, de modo que no hay efecto renta sobre ese bien y el excedente del consumidor mide exactamente el cambio de bienestar.

Preferencias de Epstein-Zin Preferencias que permiten que la actitud ante el riesgo y la disposición a trasladar consumo en el tiempo sean independientes. Técnicamente, son una utilidad recursiva que separa en dos parámetros indepen-

 © 2026 Manuel Cabado García

dientes la aversión relativa al riesgo y la elasticidad de sustitución intertemporal, que en la utilidad esperada estándar con aversión relativa constante están ligadas, porque una es la inversa de la otra.

Preferencias homotéticas Preferencias en las que el consumidor gasta su renta en las mismas proporciones sea cual sea su nivel. Técnicamente, son aquellas en las que la relación marginal de sustitución depende solo de la proporción en que se consumen los bienes. Todos los bienes tienen elasticidad renta unitaria y la curva de renta-consumo es una recta desde el origen.

Preferencias lexicográficas Preferencias en las que un bien importa siempre más que otro, igual que en un diccionario manda la primera letra. Técnicamente, priorizan absolutamente la cantidad de un bien y solo atienden a otro cuando la del primero coincide. Cumplen el preorden completo pero no la continuidad, por lo que no admiten función de utilidad.

Preferencias unimodales Preferencias en las que cada persona tiene una opción favorita y le gustan menos las demás cuanto más se alejan de ella. Técnicamente, al ordenar las alternativas sobre una sola dimensión, como el nivel de gasto, la utilidad de cada individuo tiene un único máximo y decrece a ambos lados. Es el supuesto de Black que evita los ciclos de la votación por mayoría.

Preorden completo Ordenación de todas las cestas en la que el consumidor siempre sabe decir cuál prefiere o si le son indiferentes, sin incoherencias. Técnicamente, es la relación de preferencia que cumple completitud, reflexividad y transitividad, condición necesaria pero no suficiente para que exista una función de utilidad que la represente.

Presupuesto base cero Forma de presupuestar que obliga a justificar todo el gasto cada año, como si se empezara de nuevo. Técnicamente, es la técnica presupuestaria que exige justificar cada partida desde cero en cada ejercicio y ordenar los programas de gasto por prioridad, sin depender de la inercia del pasado.

Presupuesto incrementalista Forma de presupuestar que parte de lo que se gastó el año anterior y se limita a retocarlo. Técnicamente, es la regla de decisión, estudiada por Wildavsky, que fija cada partida tomando como referencia la dotación del año anterior, de modo que el crecimiento del gasto se explica por decisiones pasadas.

Presupuesto plurianual Presupuesto que planifica el gasto para varios años en lugar de para uno solo. Técnicamente, es la técnica que sustituye la confección anual del presupuesto por una programación a varios años, lo que visibiliza los costes futuros de los programas de gasto.

Price cap Regulación que pone un techo a la subida de los precios de la empresa. Técnicamente, también llamada IPC menos X o techo de precios, permite a los precios crecer a la tasa de inflación menos un factor de productividad fijado de antemano, lo que incentiva a la empresa a reducir costes.

Prima de riesgo (aversión al riesgo) Lo que un agente averso al riesgo está dispuesto a sacrificar, en promedio, con tal de no correr un riesgo. Técnicamente, es la diferencia entre el valor esperado de una lotería y su equivalente cierto, y equivale a lo que pagaría por un seguro por encima de la pérdida esperada, es decir, la prima máxima menos la prima justa.

Prima justa Precio de un seguro que cubre exactamente lo que, en promedio, se espera perder. Técnicamente, es la pérdida esperada, igual a la diferencia entre la riqueza sin siniestro y el valor esperado de la lotería, de modo que el seguro es actuarialmente justo y su valor neto esperado es nulo.

Prima máxima Lo más que alguien pagaría por librarse por completo de un riesgo, por ejemplo con un seguro. Técnicamente, es la cantidad máxima que el agente pagaría por renunciar a la lotería y eliminar el riesgo, igual a la diferencia entre la riqueza sin siniestro y el equivalente cierto.

Primer teorema del bienestar Resultado según el cual una economía de mercados competitivos, sin fallos de mercado, no deja sin aprovechar ninguna mejora posible. Técnicamente, si los mercados son completos, los agentes son precio-aceptantes y no hay externalidades, bienes públicos ni información asimétrica, todo equilibrio

competitivo es un óptimo de Pareto.

Primera ley de Gossen Idea de que cada unidad adicional de un bien satisface menos que la anterior. Técnicamente, es el principio según el cual cada acto sucesivo de consumo produce un placer cada vez menor hasta llegar a la saciedad, lo que equivale a una utilidad marginal decreciente.

Principio de capacidad de pago Idea de que cada uno debe pagar impuestos según sus posibilidades económicas. Técnicamente, es el principio de justicia tributaria según el cual la contribución se realiza en función de la renta, con independencia de los beneficios obtenidos de la intervención del Estado.

Principio de diferencia (Rawls) Principio de justicia de Rawls según el cual las desigualdades solo son aceptables si benefician a quienes están peor. Técnicamente, exige que las desigualdades económicas y sociales redunden en el mayor beneficio posible de los más desfavorecidos. Forma parte del segundo principio de justicia junto con la igualdad equitativa de oportunidades, que tiene prioridad sobre él, y ambos quedan subordinados al principio de libertad.

Principio de diferenciación mínima Tendencia de los competidores a parecerse entre sí situándose en el centro. Técnicamente, es la tendencia de las empresas a localizarse en el centro del espacio de características cuando no compiten en precios y desaparece el efecto competición. En competencia electoral, es la tendencia análoga de dos partidos a converger hacia la posición del votante mediano.

Principio de flexibilidad Idea de que los impuestos deben ajustarse por sí solos a la marcha de la economía. Técnicamente, es el principio impositivo según el cual los impuestos deben adaptarse al contexto económico para actuar como estabilizadores automáticos.

Principio de informatividad Regla según la cual la retribución de un agente debe depender de todo dato que ayude a saber cuánto se ha esforzado. Técnicamente, Holmström (1979) muestra que una señal debe incluirse en el contrato óptimo si aporta información sobre el esfuerzo que no esté ya contenida en el resultado, lo que justifica la evaluación relativa del desempeño.

Principio de la demanda efectiva Idea de que lo que se produce y el empleo dependen de lo que las empresas esperan vender. Técnicamente, es el principio keynesiano según el cual el nivel de producción y empleo no lo fija el mercado de trabajo, sino el punto en que la demanda agregada esperada por las empresas iguala los ingresos mínimos que exigen para contratar ese volumen de empleo.

Principio de la población de Malthus Tesis de que la población tiende a crecer más deprisa que los alimentos disponibles. Técnicamente, Malthus sostuvo que la población crece en progresión geométrica y los alimentos en progresión aritmética, lo que impide elevar de forma permanente el nivel de vida de los trabajadores por encima de la subsistencia.

Principio de Pigou-Dalton Idea de que pasar renta de un rico a un pobre nunca empeora a la sociedad, siempre que no se inviertan sus posiciones. Técnicamente, es el principio según el cual una transferencia de un individuo más rico a otro más pobre que no invierta su posición relativa nunca puede reducir el bienestar social.

Principio de quien contamina paga Principio según el cual el coste de la contaminación debe recaer en quien la provoca. Técnicamente, justifica hacer pagar al emisor el daño que causa para que lo tenga en cuenta en sus decisiones, y es uno de los argumentos para preferir el impuesto corrector a la regulación directa.

Principio de revelación Resultado según el cual, para diseñar contratos cuando hay información privada, basta con buscar entre aquellos en los que a cada parte le conviene decir la verdad. Técnicamente, todo resultado alcanzable mediante un mecanismo cualquiera puede obtenerse con un mecanismo directo e incentivo-compatible, como un menú de contratos en el que cada tipo elige el diseñado para él.

Principio de sencillez administrativa Principio impositivo según el cual el sistema fiscal debe ser fácil y barato de aplicar y de cumplir. Técnicamente, para lograrlo se aconsejan pocos impuestos con bases muy amplias, reglas sencillas de valoración y pocos casos particulares.

 © 2026 Manuel Cabado García

Principio de suficiencia Idea de que los impuestos deben recaudar lo bastante para pagar lo que gasta el Estado. Técnicamente, es el principio impositivo según el cual la recaudación debe bastar para cubrir los gastos previstos del sector público.

Principio del beneficio Idea de que cada uno debe pagar impuestos según lo que recibe del Estado. Técnicamente, es el principio de justicia tributaria según el cual cada individuo paga impuestos en función del beneficio que disfruta de la actuación del Estado.

Problema de agregación de preferencias Dificultad de convertir los gustos de muchas personas en una sola decisión colectiva. Técnicamente, es la dificultad del gobierno para agregar las preferencias individuales en una valoración social del bien público aunque los individuos las revelen sinceramente.

Problema de la transformación Dificultad para pasar de los valores medidos en trabajo a los precios que se observan en el mercado. Técnicamente, es el problema de la teoría marxista para conciliar los valores-trabajo con los precios de producción cuando la competencia iguala la tasa de beneficio entre sectores con distinta composición orgánica del capital.

Problema de revelación de preferencias Tendencia a decir que un bien público se valora menos de lo que de verdad se valora, para pagar menos por él. Técnicamente, es la tendencia de los individuos a declarar una valoración del bien público inferior a la real para pagar menos impuestos, que conduce a su infraprovisión.

Problema dual Planteamiento del consumidor que trata de gastar lo menos posible para alcanzar un nivel de satisfacción fijado. Técnicamente, es el problema de minimizar el gasto necesario para alcanzar un nivel de utilidad dado, y su solución son las demandas hicksianas.

Problema dual del productor Pregunta de cómo puede obtener la empresa una producción dada gastando lo menos posible. Técnicamente, consiste en elegir la combinación de factores que minimiza el coste de obtener una producción dada, y de su solución se obtienen las demandas condicionadas de factores y la función de costes.

Problema primal Planteamiento del consumidor que trata de obtener la mayor satisfacción posible con el dinero de que dispone. Técnicamente, es el problema de maximizar la utilidad sujeto a una restricción presupuestaria fija, y su solución son las demandas marshallianas.

Problema primal del productor Pregunta de cuánto puede producir como máximo la empresa con un presupuesto dado. Técnicamente, consiste en elegir la combinación de factores que maximiza la producción sujeta a un coste dado, y de su solución se obtienen las demandas ordinarias de factores.

Proceso acumulativo Espiral de subida, o de bajada, de los precios que no se corrige por sí sola. Técnicamente, según Wicksell, si el tipo de interés monetario queda por debajo del natural, la inversión supera al ahorro, suben los precios y mejoran las expectativas de beneficio, lo que alimenta nuevas inversiones hasta que una política crediticia restrictiva lo detiene. Un tipo monetario superior al natural genera el proceso inverso.

Producción conjunta Producción en la que una misma empresa obtiene varios bienes con los factores de que dispone. Técnicamente, la tecnología tiene varios productos, y la eficiencia técnica exige repartir los factores entre ellos de modo que no pueda aumentarse la producción de uno sin reducir la de otro, lo que en el caso de dos bienes ocurre cuando se igualan sus relaciones marginales de sustitución técnica.

Productividad marginal Producción extra que se obtiene al emplear una unidad más de un factor, manteniendo los demás constantes. Técnicamente, es la derivada parcial de la función de producción respecto a ese factor, y gráficamente, la pendiente de la curva de producto total.

Productividad media Lo que se produce, en promedio, por cada unidad de un factor. Técnicamente, es la cantidad media de producto por unidad de factor, que gráficamente es la pendiente del rayo vector que une el origen con cada punto de la función de producción.

Programa de clemencia Perdón de la multa a la empresa que delata al cártel del que forma parte. Técnicamente, es un mecanismo de defensa de la

competencia que exime de sanción a la empresa de un cártel que lo denuncie y aporte pruebas, explotando así la inestabilidad interna de la colusión.

Programas de reducción de la pobreza Ayudas que garantizan un ingreso mínimo a quien carece de recursos, haya cotizado o no. Técnicamente, son prestaciones no contributivas, sujetas a comprobación de recursos y financiadas con impuestos generales, como las pensiones no contributivas o las rentas mínimas.

Programas de sustitución de rentas Prestaciones que reponen el salario perdido por jubilación, desempleo o incapacidad. Técnicamente, son seguros sociales obligatorios, financiados con cotizaciones previas, cuya prestación crece con la renta que el trabajador deja de percibir.

Progresión de la carga Medida de en qué porcentaje suben los impuestos pagados cuando la renta sube un 1 %. Técnicamente, es la elasticidad de la cuota respecto a la renta antes de impuestos, igual al cociente entre tipo marginal y tipo medio, y el impuesto es progresivo si es mayor que 1.

Progresión del tipo marginal Indicador de progresividad que atiende a cuánto sube el tipo marginal al aumentar la renta. Técnicamente, mide el crecimiento del tipo marginal en cada punto de la función impositiva, y se considera un mal indicador porque califica de proporcional un impuesto lineal con mínimo exento, que en realidad es progresivo.

Progresión del tipo medio Medida de cuánto sube el tipo medio cuando aumenta la renta. Técnicamente, es la derivada del tipo medio respecto a la renta, igual a la diferencia entre el tipo marginal y el tipo medio dividida por la renta, y el impuesto es progresivo si es positiva en todo punto.

Progresión residual Medida de en qué porcentaje sube la renta que queda tras pagar impuestos cuando la renta antes de impuestos sube un 1 %. Técnicamente, es la elasticidad de la renta después de impuestos respecto a la renta antes de impuestos, igual al cociente entre uno menos el tipo marginal y uno menos el tipo medio, y el impuesto es progresivo si es inferior a uno. Es la única medida local coherente con la idea de redistribución como reducción de la desigualdad tras impuestos.

Progresividad en frío Subida de impuestos encubierta que provoca la inflación cuando no se actualiza el impuesto sobre la renta. Técnicamente, es el aumento de la carga real del impuesto progresivo sobre la renta que causa la inflación cuando no se ajustan las cantidades exentas, las deducciones y la tarifa nominal. También se denomina rémora fiscal.

Progreso técnico Mejora de los conocimientos técnicos que permite producir más con los mismos recursos. Técnicamente, desplaza hacia arriba la función de producción y hacia el origen las isocuantas.

Progreso técnico incorporado Progreso técnico que solo mejora las máquinas nuevas y no las ya instaladas. Técnicamente, la mejora tecnológica llega a la producción a través de la inversión en capital nuevo, a diferencia del progreso no incorporado, que eleva la productividad de todo el capital existente.

Progreso técnico neutral en el sentido de Harrod Progreso técnico que hace a cada trabajador más eficiente. Técnicamente, no altera el reparto de la renta entre capital y trabajo mientras la relación capital-producto se mantenga constante, lo que equivale a que la tecnología actúe como un aumento de la eficiencia de cada trabajador.

Progreso técnico neutral en el sentido de Hicks Progreso técnico que mejora por igual el rendimiento del trabajo y del capital. Técnicamente, eleva en la misma proporción la productividad marginal de ambos factores, de modo que la relación marginal de sustitución técnica no cambia para una relación capital-trabajo dada.

Progreso técnico neutral en el sentido de Solow Progreso técnico que hace a cada máquina más eficiente. Técnicamente, no altera el reparto de la renta entre capital y trabajo mientras la relación trabajo-producto se mantenga constante, lo que equivale a que la tecnología

 © 2026 Manuel Cabado García

actúe como un aumento de la eficiencia de cada unidad de capital.

Progreso técnico sesgado Progreso técnico que mejora más el rendimiento de un factor que el del otro. Técnicamente, aumenta la productividad marginal de un factor proporcionalmente más que la del otro y altera así la relación marginal de sustitución técnica. Es ahorrador de trabajo si favorece al capital y ahorrador de capital si favorece al trabajo.

Proliferación de marcas Estrategia de una empresa que lanza muchas variedades de su producto para no dejar hueco a posibles competidores. Técnicamente, es una estrategia de disuasión de la entrada por la que una empresa instalada lanza nuevas variedades para ocupar los nichos más rentables.

Propensión marginal a consumir Cuánto aumenta el consumo cuando la renta disponible crece en una unidad. Técnicamente, en la función keynesiana es la pendiente de la función de consumo, está comprendida entre cero y uno y es inferior a la propensión media debido al consumo autónomo.

Propiedad de la ratio de verosimilitud monótona Propiedad por la que un buen resultado es indicio de que el agente se ha esforzado. Técnicamente, la ratio de verosimilitud disminuye a medida que aumenta el resultado, porque los resultados altos son relativamente más probables con el esfuerzo óptimo, lo que justifica un salario creciente con el resultado.

Proposición de inefectividad de la política Idea de que una política monetaria previsible no logra cambiar la producción ni el empleo. Técnicamente, es el resultado de Sargent y Wallace según el cual, con expectativas racionales y vaciado continuo de los mercados, cualquier regla monetaria sistemática es neutral incluso a corto plazo y solo las sorpresas monetarias tienen efectos reales.

Provisión pública Financiación con dinero público de un bien o servicio, lo produzca quien lo produzca. Técnicamente, el sector público decide de la cantidad y la paga con ingresos coactivos, mientras que la producción puede encargarse a empresas privadas, distinción clave para corregir la infraprovisión de los bienes públicos.

Puesta en común del riesgo (risk pooling) Reducción del riesgo que soporta cada asegurado cuando muchas personas agrupan riesgos independientes, porque el coste medio del grupo se vuelve predecible. Técnicamente, por la ley de los grandes números, la varianza de la pérdida media de un grupo de riesgos independientes es la varianza individual dividida por el número de asegurados, y tiende a cero cuando este aumenta.

Punto de felicidad Situación que la sociedad considera la mejor de todas las eficientes, también llamada óptimo optimorum o punto de máxima felicidad. Técnicamente, es el punto de máximo bienestar social, dado por la tangencia entre la gran frontera de posibilidades de utilidad y la curva de indiferencia social más alejada del origen. Es un punto puramente tecnocrático, porque no depende de los precios ni del sistema económico con el que se alcance.

Puzle de la volatilidad del paro Constatación de que el paro y las vacantes fluctúan mucho más de lo que predice el modelo de búsqueda y emparejamiento. Técnicamente, Shimer (2005) mostró que la volatilidad de la tensión del mercado es casi veinte veces la de la productividad, mientras que el modelo DMP con salarios negociados predice volatilidades parecidas.

Q

Q de Tobin Indicador de si a una empresa le compensa invertir, según cuánto valora el mercado su capital frente a lo que costaría reponerlo. Técnicamente, es el cociente entre el valor de mercado de una empresa y el coste de reposición de su capital. Si es mayor que uno, a la empresa le conviene invertir.

R

Racionalidad (axioma de Arrow) Exigencia de que las decisiones de la sociedad sean coherentes entre sí, de modo que, si se prefiere una opción a otra y esta a una tercera, se prefiera también la primera a la tercera. Técnicamente, es el axioma de Arrow que exige que la preferencia social

resultante sea transitiva, evitando las votaciones cíclicas.

Racionalidad secuencial Exigencia de que cada jugador actúe lo mejor posible en cualquier momento del juego, incluso en situaciones a las que no se llega. Técnicamente, el comportamiento prescrito por una estrategia de equilibrio debe ser óptimo en cada punto del juego, se alcance o no, lo que descarta las amenazas no creíbles.

Racionamiento del crédito Situación en la que hay quien quiere pedir prestado al tipo de interés vigente y no lo consigue, aunque esté dispuesto a pagar más. Técnicamente, en el modelo de Stiglitz y Weiss (1981) es un equilibrio con exceso de demanda de crédito, porque subir el tipo por encima del que maximiza el rendimiento esperado del banco atraería a prestatarios más arriesgados (selección adversa) y fomentaría proyectos más arriesgados (riesgo moral).

Ratio de verosimilitud Medida de hasta qué punto un resultado es más propio de un agente que se ha esforzado poco que de uno que se ha esforzado lo debido. Técnicamente, es el cociente entre la probabilidad de obtener un resultado con un esfuerzo inferior al óptimo y la probabilidad de obtenerlo con el esfuerzo óptimo.

Recta de balance Línea que muestra todas las combinaciones de bienes que el consumidor puede comprar gastando toda su renta. Técnicamente, es la frontera del conjunto presupuestario, cuya pendiente es el cociente de precios y representa el coste de oportunidad de un bien en términos del otro.

Recta isocoste Conjunto de combinaciones de factores que le cuestan a la empresa lo mismo. Técnicamente, es el lugar de las combinaciones de factores que, dados sus precios, tienen el mismo coste total, y su pendiente es el cociente de los precios de los factores cambiado de signo.

Recurso común o comunal Recurso que todos pueden usar, pero que se agota con el uso, como un caladero. Técnicamente, es un bien rival y no excluible, y por ello tiende a sobreexplotarse.

Redistribución como bien público local Idea de que la redistribución puede descentralizarse porque los ciudadanos de renta alta valoran sobre todo la mejora de los pobres de su propio territorio. Técnicamente, Pauly planteó en 1973 la redistribución como un bien público local cuya demanda varía entre jurisdicciones, de modo que su provisión territorial puede ser eficiente si la movilidad es imperfecta.

Reflexividad Exigencia mínima de coherencia según la cual toda cesta se compara consigo misma. Técnicamente, es el axioma por el que cualquier cesta es al menos tan buena como ella misma, y junto con la completitud y la transitividad forma los axiomas de racionalidad.

Región económica de producción Zona de combinaciones de factores en la que tiene sentido producir, porque usar más de cualquier factor no reduce el producto. Técnicamente, es la zona del plano de factores comprendida entre las líneas de contorno, en las que las productividades marginales se anulan, dentro de la cual se cumple la condición de eficiencia técnica.

Regla de Borda Sistema de votación en el que se puntúan las opciones en lugar de elegir solo una. Técnicamente, cada votante asigna puntos a las alternativas según su orden de preferencia y resulta elegida la que suma más puntos.

Regla de Corlett-Hague Regla que aconseja gravar más lo que se consume junto con el tiempo libre. Técnicamente, como el ocio no puede gravarse, la imposición óptima exige tipos más altos sobre los bienes complementarios del ocio y más bajos sobre los sustitutivos, lo que reduce la distorsión sobre la oferta de trabajo.

Regla de Dorfman-Steiner Regla que indica cuánto debe gastar una empresa en publicidad en proporción a lo que ingresa. Técnicamente, establece que el gasto óptimo en publicidad sobre los ingresos es igual al cociente entre la elasticidad de la demanda respecto a la publicidad y la elasticidad-precio. Por eso conviene gastar más cuanto más eficaz es la publicidad y cuanto menos sensible al precio es el consumidor.

Regla de elección colectiva Procedimiento para pasar de lo que prefiere cada persona a lo que decide la sociedad. Técnicamente, es el mecanismo que agrega preferencias individuales para obtener una preferencia social, como las reglas de votación

 © 2026 Manuel Cabado García

o las funciones de bienestar social.

Regla de elección excluyente Regla que, entre dos impuestos posibles, aconseja el que menos altera la conducta de los consumidores. Técnicamente, es la regla de imposición eficiente según la cual, entre dos impuestos alternativos, es preferible el que grava el bien con la demanda compensada más inelástica.

Regla de elección no manipulable Regla de votación en la que a nadie le compensa mentir sobre sus preferencias. Técnicamente, también llamada a prueba de estrategia, es aquella en la que ningún individuo puede lograr un resultado mejor para él declarando preferencias falsas, de modo que declarar las verdaderas es una estrategia dominante.

Regla de gasto Norma que impide que el gasto público crezca más deprisa que la economía. Técnicamente, en España la ley de estabilidad presupuestaria de 2012 limita el crecimiento del gasto computable de las Administraciones Públicas a la tasa de referencia de crecimiento del PIB a medio plazo, y en las reglas fiscales europeas de 2024 una senda de gasto neto pactada con cada país es la variable central del brazo preventivo.

Regla de igualación de distorsiones marginales Regla que aconseja repartir la recaudación entre todos los impuestos disponibles en lugar de concentrarla en uno solo. Técnicamente, los tipos óptimos son los que igualan el exceso de gravamen marginal por euro adicional recaudado en todos los impuestos.

Regla de la elasticidad inversa Regla que aconseja gravar más los bienes cuya demanda reacciona poco al precio. Técnicamente, si los bienes son independientes, los tipos que minimizan el exceso de gravamen son inversamente proporcionales a las elasticidades-precio de sus demandas compensadas. La planteó Pigou y la resolvió Ramsey en 1927.

Regla de la mayoría simple Regla de votación en la que gana la opción que reúne más apoyos. Técnicamente, entre dos alternativas resulta elegida la que obtiene la mitad más uno de los votos. Cumple el criterio de Pareto de Arrow, pero no garantiza un nivel eficiente de bien público porque ignora la intensidad de las preferencias.

Regla de la unanimidad Regla según la cual una propuesta solo sale adelante si nadie se opone. Técnicamente, es la regla de votación, propuesta por Wicksell para financiar eficientemente los bienes públicos, que exige el acuerdo de todos los individuos para aprobar una decisión.

Regla de oro de la acumulación de capital Nivel de capital que permite el mayor consumo posible de forma permanente. Técnicamente, es el capital por trabajador que maximiza el consumo por trabajador en el estado estacionario del modelo de Solow, donde la productividad marginal del capital iguala la suma de las tasas de crecimiento de la población y de la tecnología más la depreciación.

Regla de oro de la Hacienda Pública clásica Excepción que permite al Estado endeudarse solo para pagar inversiones que se financian a sí mismas. Técnicamente, es la excepción al equilibrio presupuestario que admite el déficit para financiar inversiones autoliquidables, cuya rentabilidad permite pagar los intereses y reembolsar los préstamos.

Regla de Ramsey-Diamond-Mirrlees Regla según la cual los impuestos deben reducir el consumo de todos los bienes en la misma proporción, también llamada regla de Ramsey o de la reducción equiproporcional. Técnicamente, el sistema que minimiza el exceso de gravamen reduce todas las demandas compensadas en la misma proporción aunque haya complementariedad y sustituibilidad entre bienes, y la regla de la elasticidad inversa es su caso particular con bienes independientes.

Regla de Taylor Pauta sencilla que describe cómo fija los tipos un banco central: los sube cuando la inflación supera su objetivo o la producción supera su potencial, y los baja en el caso contrario. Técnicamente, para estabilizar la economía, el tipo nominal debe subir más de un punto por cada punto adicional de inflación (principio de Taylor).

Regla del precio eficiente de los componentes Regla para fijar lo que un competidor paga por usar la red de un incumbente. Técnicamente, iguala el precio de acceso al coste de dar acce-

so más el margen que el incumbente deja de ganar en el mercado final, de modo que solo entran rivales al menos tan eficientes como él.

Reglas de Marshall-Hicks Condiciones de las que depende cuánto empleo se pierde cuando se encarece el trabajo. Técnicamente, la demanda de un factor es más elástica cuanto más fácil es sustituirlo por otros, cuanto más elástica es la demanda del bien que produce, cuanto mayor es su peso en los costes, siempre que la demanda del bien sea más elástica que la sustitución entre factores, y cuanto más elástica es la oferta de los demás factores.

Regulación de la tasa de rendimiento Regulación que limita lo que una empresa puede ganar por cada euro invertido. Técnicamente, también llamada regulación por coste del servicio, fija precios que cubren los costes de operación y una rentabilidad máxima sobre el capital superior a su coste de oportunidad, lo que incentiva la sobrecapitalización.

Regulación económica Intervención pública que fija de antemano precios, entrada o calidad en actividades en las que la competencia no es deseable. Técnicamente, actúa ex ante y se reserva para el poder de mercado que procede de la propia estructura de costes, como en el monopolio natural, mientras que el que nace de conductas estratégicas o de barreras legales se combate con la defensa de la competencia.

Regulación óptima Conjunto de reglas de precios, entrada y calidad que el sector público impone a una empresa con poder de mercado para lograr el mejor resultado posible para la sociedad. Técnicamente, es la que maximiza la suma del excedente de los consumidores y del beneficio de la empresa, sujeta a que esta sea viable y a la información de que dispone el regulador.

Regulación por incentivos (performance-based regulation) Regulación que premia a la empresa por reducir costes o mejorar el servicio. Técnicamente, desvincula, total o parcialmente, el precio o los ingresos permitidos de los costes declarados por la empresa, como hacen el price cap, el reparto de beneficios o la competencia por comparación.

Relación de preferencia Criterio con el que el consumidor compara dos cestas cualesquiera. Técnicamente, es la relación binaria “es al menos tan bueno como”, definida sobre el conjunto de elección, de la que se derivan la preferencia estricta y la indiferencia.

Relación marginal de sustitución Cantidad de un bien a la que el consumidor está dispuesto a renunciar a cambio de una unidad más de otro sin que varíe su utilidad. Técnicamente, es igual al cociente de utilidades marginales y a la pendiente de la curva de indiferencia en valor absoluto.

Relación marginal de sustitución entre ocio y consumo Consumo al que una persona está dispuesta a renunciar a cambio de una hora más de ocio. Técnicamente, es el cociente entre la utilidad marginal del ocio y la del consumo, que en el óptimo del trabajador se iguala al salario real.

Relación marginal de sustitución estocástica Riqueza que una persona está dispuesta a perder en una situación posible a cambio de tener más en otra. Técnicamente, es la tasa a la que el agente está dispuesto a sacrificar riqueza en un estado de la naturaleza a cambio de riqueza en otro, igual a la relación marginal de sustitución ponderada por el cociente de probabilidades de ambos estados.

Relación marginal de sustitución técnica Cantidad de un factor de la que puede prescindirse al emplear una unidad más de otro sin que cambie la producción. Técnicamente, es el cociente de las productividades marginales y el valor absoluto de la pendiente de la isocuanta.

Relación marginal de transformación Coste de oportunidad de producir una unidad más de un bien, medido en unidades del otro bien a las que hay que renunciar. Técnicamente, es el valor absoluto de la pendiente de la frontera de posibilidades de producción.

Remoción Reacción de quien, ante un impuesto, trabaja o produce más para compensar lo que paga. Técnicamente, es el aumento del esfuerzo productivo del contribuyente para paliar el impacto de un tributo.

Rendimientos a escala Forma en que responde la producción cuando la empresa aumenta de tama-

 © 2026 Manuel Cabado García

ño ampliando todos sus factores por igual. Técnicamente, es la respuesta de la producción cuando todos los factores varían en la misma proporción, y son constantes, crecientes o decrecientes según la elasticidad de escala sea igual, mayor o menor que uno.

Rendimientos globales a escala Rendimientos a escala que no dependen del tamaño de la empresa. Técnicamente, son los que la tecnología presenta sea cual sea el tamaño de partida y el factor por el que se amplíe. Las funciones homogéneas los tienen siempre, con elasticidad de escala igual a su grado, pero no son las únicas.

Rendimientos locales a escala Rendimientos a escala que pueden cambiar según el tamaño de la empresa. Técnicamente, se evalúan mediante la elasticidad de escala en un punto concreto, de modo que una misma tecnología puede presentarlos crecientes a niveles bajos de producción y decrecientes a niveles elevados.

Renta básica universal Pago periódico que recibiría cada ciudadano por el mero hecho de serlo, trabaje o no. Técnicamente, es una transferencia universal e incondicional que no se retira al aumentar la renta, por lo que no genera trampa de pobreza.

Renta completa Renta que tendría un hogar si dedicara a trabajar todo su tiempo. Técnicamente, en el modelo de Becker es la renta no laboral más el valor del tiempo total disponible al salario vigente, y resume en una sola las restricciones de dinero y de tiempo.

Renta diferencial Ingreso extra que obtienen los dueños de las tierras más fértiles. Técnicamente, según Ricardo, es la renta de la tierra que obtienen las tierras más fértiles porque el precio del cereal lo fija el coste de la tierra marginal, la menos fértil de las cultivadas. No es un coste de producción, sino un excedente.

Renta informativa Beneficio extra que hay que dejar a quien sabe más para que diga la verdad. Técnicamente, es el beneficio que el regulador debe conceder a la empresa eficiente para que no le compense declararse ineficiente. En el modelo de dos tipos es igual a la diferencia de costes multiplicada por la cantidad asignada al tipo ineficiente.

Renta permanente Ingreso que una persona espera tener de forma estable, frente a los ingresos pasajeros. Técnicamente, es el flujo de renta que puede mantenerse en el tiempo y consumirse íntegramente sin mermar la riqueza de la que deriva. Para Friedman, el consumo es proporcional a ella.

Reparto de beneficios o escala móvil Regulación en la que la empresa comparte con los consumidores parte de lo que gana por encima de lo normal. Técnicamente, la empresa retiene solo una fracción de los beneficios que superan una banda de rentabilidad, en un punto intermedio entre la regulación por coste del servicio y el price cap.

Reparto óptimo del riesgo Forma de repartir el riesgo de un resultado incierto que no puede mejorarse para una parte sin empeorar a la otra. Técnicamente, con información simétrica iguala entre estados el cociente de utilidades marginales del principal y del agente, de modo que, si una parte es neutral, asume todo el riesgo y la otra queda plenamente asegurada.

Reserva de capacidad Capacidad de producción que la empresa deja sin usar a propósito para poder reaccionar ante cambios. Técnicamente, es la capacidad productiva que el empresario planifica deliberadamente para ganar flexibilidad ante fluctuaciones o una expansión futura. No debe confundirse con el exceso de capacidad no planificado de la teoría tradicional.

Responsabilidad conjunta Fórmula de préstamo en la que los miembros de un grupo responden unos por las deudas de otros, de modo que se vigilan entre ellos. Técnicamente, es el mecanismo de crédito que condiciona la deuda de un grupo al pago de cada miembro, trasladando la supervisión de la entidad financiera a los propios prestatarios.

Restricción de incentivos Condición de que al agente le convenga hacer lo que el principal desea, sin necesidad de vigilarlo. Técnicamente, con riesgo moral exige que la utilidad esperada del esfuerzo deseado no sea inferior a la de cualquier otro, y con selección adversa, también llamada restricción de autoselección, que cada tipo prefie-

ra el contrato diseñado para él.

Restricción de participación Condición de que el contrato deje al agente al menos tan bien como su mejor alternativa, porque de lo contrario no lo firmaría. Técnicamente, la utilidad esperada del contrato, descontada la desutilidad del esfuerzo, debe ser mayor o igual que la utilidad de reserva.

Restricción presupuestaria blanda Situación en la que un gobierno cuenta con que, si gasta o se endeuda de más, otro nivel de gobierno acabará haciéndose cargo. Técnicamente, es un problema de riesgo moral: el gobierno subcentral espera un rescate ex post del gobierno central y por ello no soporta todo el coste de sus decisiones, lo que incentiva el gasto y el endeudamiento excesivos.

Restricción presupuestaria (conjunto presupuestario) Límite que la renta impone a lo que el consumidor puede comprar. Técnicamente, reúne las cestas cuyo gasto a los precios vigentes no supera la renta, y forma un conjunto presupuestario cerrado, acotado, convexo y no vacío cuya frontera es la recta de balance.

Retorno de técnicas (reswitching) Posibilidad de que una misma técnica de producción sea la más rentable con tipos de interés altos y con tipos bajos, y deje de serlo en los intermedios. Técnicamente, contradice la relación decreciente entre tipo de interés e intensidad de capital en que se basa la función de producción agregada neoclásica, y fue el argumento central de la escuela de Cambridge británica en la controversia del capital.

Retraso regulatorio (o paréntesis regulador) Tiempo que pasa entre dos revisiones de las tarifas reguladas. Técnicamente, es el intervalo durante el cual el precio no se ajusta a los costes observados, lo que deja a la empresa quedarse con sus ahorros y refuerza el incentivo a reducir costes, a costa de la eficiencia asignativa.

Revolución marginalista Giro de la teoría económica que pasó a explicar el valor de las cosas por la utilidad de su última unidad. Técnicamente, es el cambio iniciado hacia 1871 por Jevons, Menger y Walras de forma simultánea e independiente, que fundamentó el valor en la utilidad marginal e introdujo el individualismo metodológico y el cálculo diferencial.

Riesgo Situación en la que no se sabe qué va a ocurrir, pero sí con qué probabilidad puede ocurrir cada cosa. Técnicamente, el agente desconoce qué estado de la naturaleza se materializará, pero puede asignar una probabilidad a cada uno.

Riesgo moral Tendencia a ser menos cuidadoso cuando otro paga las consecuencias, como quien conduce peor por estar asegurado. Técnicamente, es un problema de información asimétrica posterior al contrato, debido a acciones ocultas de una de las partes, que cambia su conducta porque ya no soporta todo el riesgo.

Riesgo social Riesgo que, cuando se materializa, afecta a la vez a gran parte de la población, como el desempleo en una recesión. Técnicamente, al no ser independiente entre asegurados no puede diversificarse, y podría llevar a la quiebra a una aseguradora privada que tuviera que pagar a todos a la vez. El sector público sí puede cubrirlo porque dispone de ingresos distintos de las primas.

Rigidez nominal Lentitud con que se ajustan los precios y salarios fijados en dinero. Técnicamente, por contratos de varios periodos, revisiones escalonadas o costes de cambiar precios, los precios nominales no se adaptan de inmediato a las perturbaciones, por lo que estas afectan a la producción y al empleo y el dinero deja de ser neutral a corto plazo.

Rigidez real Falta de ajuste de los precios relativos, sobre todo del salario real, ante los cambios de la oferta o la demanda. Técnicamente, aparece cuando a trabajadores o empresas les conviene mantener el salario real por encima del que vaciaría el mercado, como en los modelos de insiders y outsiders, de salarios de eficiencia o de contratos implícitos, y es compatible con el paro involuntario aun con expectativas racionales.

Rigidez salarial Resistencia de los salarios a ajustarse cuando cambian la oferta o la demanda de trabajo. Técnicamente, la rigidez del salario real la explican los modelos de negociación, los salarios de eficiencia y los contratos implícitos, en los que el salario deja de ser un precio que se ajusta libremente, de modo que las perturbaciones re-

 © 2026 Manuel Cabado García

caen sobre todo en el empleo.

RIIO Modelo británico de regulación de las redes energéticas que sustituyó al RPI–X a partir de 2013. Técnicamente, fija los ingresos permitidos para periodos de varios años en función de incentivos, innovación y resultados, trata conjuntamente la inversión y los gastos de operación y reparte con los consumidores las desviaciones de gasto.

Robo de negocio Situación en la que una empresa que lanza un producto nuevo gana clientes quitándoselos a las que ya estaban en el mercado. Técnicamente, es una externalidad negativa por la que una nueva variedad reduce la demanda y los beneficios de las empresas existentes, lo que lleva a introducir demasiadas variedades.

S

Salario de eficiencia Salario más alto de lo necesario para contratar, que la empresa paga porque así sus trabajadores rinden más. Técnicamente, es el salario que minimiza el coste por unidad de esfuerzo cuando la productividad depende del salario. Si supera el que vaciaría el mercado, genera paro involuntario.

Salario de reserva Salario más bajo por el que una persona acepta trabajar. Técnicamente, en el modelo de oferta de trabajo es la relación marginal de sustitución entre ocio y consumo en el punto en el que no se trabaja, y en los modelos de búsqueda es la oferta mínima que el parado acepta en lugar de seguir buscando.

Second best (información asimétrica) Lo mejor que se puede conseguir cuando algún obstáculo impide alcanzar la situación ideal. Técnicamente, en economía de la información es la asignación óptima sujeta a las restricciones de participación y de incentivos, inferior a la de información perfecta por el coste de la señalización, el filtro o los incentivos.

Segunda ley de Gossen Regla que indica cómo repartir el gasto entre bienes para sacarle el máximo partido. Técnicamente, es el principio de asignación óptima de la renta según el cual el consumidor reparte su gasto hasta que la utilidad marginal por unidad monetaria es igual en todos los bienes.

Segundo teorema del bienestar Resultado según el cual el mercado puede llevar a cualquier reparto eficiente que se desee si antes se redistribuyen adecuadamente los recursos. Técnicamente, en una economía sin fallos de mercado y con preferencias y conjuntos de producción convexos, cualquier óptimo de Pareto puede alcanzarse como equilibrio competitivo redistribuyendo las dotaciones iniciales mediante transferencias de suma fija, lo que permite separar los objetivos de eficiencia y de equidad.

Selección adversa Problema que surge cuando una parte no puede conocer, antes de contratar, rasgos de la otra que le importan, de modo que acaban contratando sobre todo los peores. Técnicamente, es la información asimétrica anterior al contrato, debida a información oculta, que lleva a tratar a todos según la calidad media y puede expulsar del mercado a los agentes de mejor tipo.

Señalización Forma de demostrar una cualidad que los demás no pueden ver haciendo algo que sí ven y que a quien carece de ella le costaría demasiado imitar. Técnicamente, es la solución a la selección adversa por la que la parte informada elige una acción observable cuyo coste es menor para los tipos buenos, lo que permite un equilibrio separador.

Senda de expansión Camino que sigue la combinación de factores de la empresa cuando aumenta su producción al menor coste. Técnicamente, es el lugar geométrico de las combinaciones técnica y económicamente eficientes de factores para cada nivel de producción, formado por las tangencias entre isocuantas e isocostes.

Separación vertical División entre la empresa que posee o gestiona una red y las que prestan servicios sobre ella. Técnicamente, es la solución estructural para las industrias de red, que va desde encomendar la gestión de la red a un operador independiente hasta separar su propiedad, y que evita que el dueño de la red discrimine a sus rivales en las actividades competitivas a costa de perder economías de integración.

Sesgo de habilidad Error por el que se atribuye a la educación lo que se debe a la capacidad de

quien estudia. Técnicamente, si los más capaces estudian más y ganarían más con cualquier nivel educativo, la estimación por mínimos cuadrados del rendimiento de la educación en la ecuación de Mincer está sesgada al alza.

Sesgo de publicación Tendencia a que se publiquen sobre todo los estudios con resultados favorables. Técnicamente, sesga al alza el efecto medio que recoge la literatura, como mostraron los metaanálisis sobre la eficacia de los nudges.

Sesgo deficitario Tendencia de los gobiernos a endeudarse más de lo conveniente. Técnicamente, en el modelo de Alesina y Tabellini surge porque el gobierno que prevé ser sustituido por otro de distinta ideología usa la deuda para condicionar las decisiones de su sucesor.

Sindicato monopolista Sindicato que actúa como único vendedor de trabajo frente a las empresas. Técnicamente, el resultado depende de su objetivo. Si maximiza la renta neta de sus afiliados, fija el empleo donde su ingreso marginal se iguala al coste marginal del trabajo, que es la oferta, y el salario sobre la curva de demanda, con más salario, menos empleo y paro involuntario que en competencia, mientras que si maximiza el empleo el equilibrio es el competitivo.

Síntesis neoclásica Corriente que combinó las ideas de Keynes con la economía neoclásica anterior. Técnicamente, fue la corriente dominante entre los años cuarenta y setenta que combinaba el modelo keynesiano, válido a corto plazo con precios y salarios rígidos, y el modelo neoclásico, válido a largo plazo con precios flexibles.

Sistema casi ideal de demanda (AIDS) Modelo para estimar con datos cómo reparten los hogares su gasto entre bienes de forma coherente con la teoría. Técnicamente, es el sistema de demanda flexible de Deaton y Muellbauer, derivado de una función de gasto que agrega exactamente a consumidores distintos, que expresa las cuotas presupuestarias en función de los precios y de la renta real y permite imponer o contrastar las restricciones teóricas de la demanda, como la homogeneidad y la simetría.

Sistema de capitalización Sistema de pensiones en el que cada trabajador ahorra para su propia jubilación. Técnicamente, las cotizaciones de cada trabajador se acumulan e invierten, y la pensión depende de lo aportado y de su rendimiento. Supone ahorro forzoso pero no redistribución.

Sistema de contribución definida Sistema de pensiones en el que lo que está fijado es lo que se aporta y no lo que se cobrará. Técnicamente, la pensión resulta de convertir en una renta periódica, según la esperanza de vida, lo acumulado en la cuenta del trabajador, que es real en la capitalización y ficticia en las cuentas nocionales. Se opone a los sistemas de prestación definida, en los que la pensión se calcula con una fórmula sobre el salario y los años cotizados.

Sistema de derechos de emisión Sistema que limita la contaminación total y permite a las empresas comprar y vender permisos para contaminar. Técnicamente, es un mecanismo de mercado, según el principio de limitar y comerciar, que fija un tope total de emisiones, reparte derechos negociables entre las empresas y deja que su precio se ajuste.

Sistema de recargos Forma de dar a los territorios cierta capacidad fiscal dejando que suban o bajen un impuesto del Estado. Técnicamente, es la fórmula de descentralización parcial en la que el gobierno central fija la base y la liquidación del tributo y los gobiernos territoriales establecen recargos o bonificaciones sobre los tipos.

Sistema de reparto Sistema de pensiones en el que quienes trabajan hoy pagan las pensiones de los jubilados de hoy. Técnicamente, se basa en un contrato implícito intergeneracional por el que las prestaciones de cada período se pagan con las cotizaciones de ese mismo período.

Sistema impositivo óptimo Combinación de impuestos que obtiene lo que necesita el Estado de la mejor forma posible. Técnicamente, si la eficiencia es el único criterio, es la que logra la recaudación requerida con el mínimo exceso de gravamen total. Si también importa la equidad, es la que, con esa recaudación, maximiza una función de bienestar social, aunque acepte más coste de eficiencia a cambio de una distribución más igualitaria.

 © 2026 Manuel Cabado García

Sobreacumulación de capital Situación en la que una economía ahorra e invierte tanto que acaba consumiendo menos de lo que podría. Técnicamente, es una situación de ineficiencia dinámica en la que el capital por trabajador supera al de la regla de oro, de modo que todas las generaciones podrían consumir más reduciendo el ahorro agregado.

Sobretraslación Situación en la que el precio que paga el consumidor sube más que el impuesto. Técnicamente, el grado de traslación supera la unidad, como ocurre en monopolio con demanda de elasticidad constante y coste marginal constante, donde el precio aumenta en el impuesto multiplicado por el cociente entre el valor absoluto de la elasticidad y ese mismo valor menos uno. Con una elasticidad de dos, por ejemplo, el precio sube el doble del impuesto.

Socialismo de mercado Sistema en el que los medios de producción son públicos pero se siguen usando precios para asignar los recursos. Técnicamente, en la propuesta de Lange un organismo central fija los precios y los corrige por tanteo, al alza si hay exceso de demanda y a la baja si hay exceso de oferta, hasta vaciar cada mercado.

Solución de negociación de Nash Forma de predecir el resultado de una negociación a partir de lo que cada parte gana con el acuerdo y de su poder para imponerse. Técnicamente, es la única solución que cumple los axiomas de Nash (1950) y maximiza el producto de las ganancias de cada parte respecto a su situación sin acuerdo, ponderadas en la versión generalizada por su poder de negociación.

Sostenibilidad del monopolio natural Capacidad de una empresa que abastece sola un mercado con menor coste que varias para mantenerse sin que la entrada de competidores acabe con ella. Técnicamente, la sostenibilidad del monopolio natural depende de la posición de la demanda respecto a las curvas de coste medio y, con varios bienes, de la ausencia de subsidios cruzados.

Spending review Revisión a fondo del gasto público para ver qué funciona y qué sobra. Técnicamente, es el examen sistemático de la eficacia y la eficiencia de las políticas de gasto para identificar partidas prescindibles y mejorar la asignación de recursos.

Suavización fiscal Recomendación de mantener los impuestos estables a lo largo del ciclo y usar la deuda para absorber los vaivenes. Técnicamente, es la teoría de Barro según la cual conviene mantener constante el tipo impositivo y dejar que la deuda absorba los shocks, con déficit en las recesiones y superávit en las expansiones, para minimizar las pérdidas de eficiencia. También llamada suavización impositiva, la expuso Barro (1979).

Subaditividad de costes Situación en la que una sola empresa puede abastecer un mercado más barato que varias. Técnicamente, es la propiedad de una tecnología por la que resulta menos costoso producir una cantidad dada en una sola empresa que repartirla entre varias. Las economías de escala son condición suficiente pero no necesaria para que exista.

Subasta de Vickrey Subasta en la que gana quien más puja pero paga lo que ofreció el segundo. Técnicamente, es una subasta de segundo precio a sobre cerrado en la que declarar la verdadera valoración es una estrategia dominante, porque lo que paga el ganador no depende de su propia puja.

Subastador walrasiano Personaje imaginario que fija a la vez los precios de todos los mercados. Técnicamente, es un agente ficticio que recoge todas las ofertas y demandas, busca por tanteo los precios que vacían los mercados y solo permite intercambiar a esos precios de equilibrio.

Subsidios cruzados Financiación de un producto deficitario con los beneficios de otro dentro de la misma empresa. Técnicamente, existen cuando los ingresos de un bien no cubren su coste incremental y otro bien genera ingresos superiores a su coste aislado, lo que invita a la entrada selectiva y hace insostenible el monopolio multiproducto.

Subvención pigouviana Ayuda pública a una actividad que beneficia a terceros para que se realice en la cantidad que conviene a la sociedad. Técnicamente, es un pago por unidad igual al beneficio marginal externo en el nivel eficiente, que acerca el beneficio privado al social e internaliza una externalidad positiva.

Superelasticidad Elasticidad de la demanda de un bien que tiene en cuenta su efecto sobre la demanda de los demás bienes de la empresa. Técnicamente, mide cómo varía la demanda conjunta de los bienes de un monopolio multiproducto ante el precio de uno de ellos y sustituye a la elasticidad propia en la regla de Ramsey cuando las demandas no son independientes.

Superestructura En la teoría de Marx, el conjunto de instituciones e ideas que se levanta sobre la estructura económica. Técnicamente, comprende el Estado, el derecho, la política y las formas de conciencia social, como la religión o la ideología. Está condicionada por la estructura y contribuye a mantenerla.

Supuesto heroico Simplificación de Chamberlin que trata a todas las empresas del grupo como si fueran iguales. Técnicamente, supone que sus curvas de demanda y de costes son idénticas, lo que permite hablar de un precio único para la industria pese a que los productos están diferenciados.

Sustitución intertemporal del trabajo Tendencia a trabajar más en los periodos en que trabajar compensa más. Técnicamente, es la reasignación de horas entre periodos en respuesta al salario relativo, corregido por el tipo de interés y la impaciencia, que en los modelos del ciclo real transmite los shocks transitorios al empleo.

Sustitutivos brutos Bienes que se reemplazan entre sí en el consumo, de modo que si uno se encarece se compra más del otro. Técnicamente, lo son cuando la elasticidad precio cruzada de la demanda marshalliana es positiva, por lo que el criterio recoge a la vez el efecto sustitución y el efecto renta, a diferencia de los sustitutivos netos.

Sustitutivos netos Bienes que el consumidor usa uno en lugar del otro, de modo que, si uno se encarece, compra más del otro aunque se le compense la pérdida de renta. Técnicamente, son bienes cuyo efecto sustitución cruzado, medido con la demanda hicksiana, es positivo.

Sustitutivos perfectos Bienes que al consumidor le dan igual, de modo que siempre está dispuesto a cambiar uno por otro en la misma proporción. Técnicamente, son bienes con relación marginal de sustitución constante y curvas de indiferencia lineales, cuya elasticidad de sustitución es infinita.

Sustitutos estratégicos Variables de decisión en las que a una empresa le conviene moverse en dirección contraria a su rival. Técnicamente, son aquellas cuyas funciones de reacción tienen pendiente negativa, como las cantidades en el modelo de Cournot.

T

Tarifa en dos partes Precio formado por una cuota fija por el derecho a consumir y un pago por cada unidad consumida. Técnicamente, la usa el monopolista como discriminación de segundo grado y, en la regulación del monopolio natural, la cuota reparte los costes fijos entre los consumidores y el precio por unidad iguala el coste marginal, lo que logra la producción eficiente con beneficio nulo. La analizó por primera vez Oi (1971).

Tarifa pico-valle Tarifa más cara en las horas de más demanda y más barata en las de menos. Técnicamente, se aplica a servicios con demanda fluctuante y producción no almacenable, de modo que los consumidores de periodos punta pagan además los costes de capacidad y los de periodos valle solo los costes de uso.

Tarificación dinámica Cambio automático y continuo de los precios según las condiciones del momento. Técnicamente, combina el ajuste a la demanda y a la capacidad disponible con la segmentación mediante datos de los clientes, y en su versión más personalizada se aproxima a la discriminación de primer grado.

Tasa de dependencia Proporción de personas que, por su edad, dependen de quienes están en edad de trabajar. Técnicamente, es el cociente entre la población mayor de 65 años, a veces sumada a la menor de 16, y la población en edad de trabajar. Su aumento por el envejecimiento reduce el número de cotizantes por pensionista en un sistema de reparto.

Tasa de explotación En la teoría de Marx, medida de cuánto trabajo no pagado obtiene el capitalista por cada unidad pagada en salarios. Téc-

 © 2026 Manuel Cabado García

nicamente, es el cociente entre la plusvalía y el capital variable y, junto con la composición orgánica del capital, determina la tasa de beneficio.

Tasa de reposición Parte del sueldo perdido que se sigue cobrando durante el paro. Técnicamente, es el cociente entre la prestación por desempleo y el salario anterior a la pérdida del empleo, expresado en porcentaje.

Tasa de retribución financiera Rentabilidad que el regulador reconoce a las redes eléctricas y gasistas por el capital invertido. Técnicamente, es el coste medio ponderado del capital, expresado antes de impuestos, que la CNMC aplica al valor neto de los activos regulados.

Tasa natural de paro Paro que persiste incluso cuando la economía funciona con normalidad, por los cambios de empleo y los desajustes propios del mercado de trabajo. Técnicamente, es el desempleo de equilibrio, fijado por factores estructurales, e intentar reducirlo con políticas de demanda solo acelera la inflación.

Tâtonnement Búsqueda por prueba y error de los precios que equilibran los mercados. Técnicamente, es el proceso de tanteo por el que el subastador walrasiano eleva los precios de los bienes con exceso de demanda y reduce los de los bienes con exceso de oferta hasta alcanzar el equilibrio general.

Techo a la capacidad de gravamen Límite al poder del Estado para recaudar, que sustituye al control directo del gasto. Técnicamente, Brennan y Buchanan propusieron restringir el diseño de los impuestos, con bases reducidas y elásticas, tipos uniformes, recaudación ligada a servicios públicos complementarios y límites a la imposición del capital, para que un sector público maximizador de ingresos, el Leviatán, no pueda apropiarse del excedente de los contribuyentes. Estas condiciones chocan con los principios de la imposición óptima.

Techo de déficit público Límite legal que impide que los gastos públicos superen a los ingresos. Técnicamente, es la modalidad más antigua de constitución fiscal, la regla de equilibrio presupuestario defendida por los clásicos y por Buchanan. Para no impedir la estabilización a lo largo del ciclo puede formularse en términos de déficit estructural, como hace el artículo 135 de la Constitución Española.

Techo de gasto público Límite máximo a lo que puede gastar el sector público. Técnicamente, es la modalidad de constitución fiscal que limita el gasto en términos absolutos o su crecimiento en relación con algún índice.

Tecnologías de agregación Reglas que indican cómo se combinan los esfuerzos de cada país para lograr un bien público común. Técnicamente, son las formas en que las contribuciones de cada país determinan el nivel de un bien público, que puede depender de su suma, del que menos contribuye (eslabón más débil) o del que más consigue (mejor tiro).

Tensión del mercado de trabajo Número de vacantes que hay por cada parado. Técnicamente, es el cociente entre la tasa de vacantes y la tasa de paro, y de él dependen la rapidez con que los parados encuentran empleo y la rapidez con que se cubren las vacantes.

Teorema básico de la dualidad Resultado que garantiza que los dos modos de plantear la elección, maximizar la satisfacción o minimizar el gasto, llevan a la misma decisión. Técnicamente, los problemas primal y dual, del consumidor o del productor, tienen la misma solución cuando la restricción de cada uno coincide con el valor óptimo del otro, de modo que, en el caso del consumidor, en ese punto coinciden las demandas marshallianas y las hicksianas.

Teorema de Arrow-Debreu con incertidumbre Resultado según el cual la competencia sigue asignando eficientemente los recursos aunque el futuro sea incierto, siempre que todo riesgo pueda negociarse. Técnicamente, si los agentes maximizan su utilidad esperada con probabilidades subjetivas independientes de sus actos y existe un mercado para cada bien en cada estado de la naturaleza, el equilibrio competitivo es óptimo de Pareto, resultado de Arrow (1953) y Debreu (1959) que recoge el manual de Arrow y Hahn (1971).

Teorema de Coase Idea de que, si los afectados por una externalidad pueden negociar sin coste,

llegarán por sí solos a la solución eficiente. Técnicamente, con derechos de propiedad bien definidos, costes de transacción nulos e información perfecta, la negociación privada alcanza un óptimo de Pareto y, sin efecto renta, la asignación resultante no depende de a quién se otorguen los derechos, que solo afecta al reparto.

Teorema de Debreu Resultado que asegura cuándo las preferencias de un consumidor pueden expresarse con números. Técnicamente, establece que unas preferencias completas, transitivas, reflexivas y continuas pueden representarse mediante una función de utilidad continua.

Teorema de descentralización de Oates Resultado que justifica dejar que cada territorio elija cuánto bien público local quiere en lugar de imponer el mismo nivel a todos. Técnicamente, si no hay ahorros de coste por la provisión centralizada ni externalidades interjurisdiccionales, el bienestar alcanzado cuando cada jurisdicción elige su nivel de provisión de forma independiente es siempre igual o superior al de un nivel uniforme.

Teorema de Euler Resultado que relaciona el producto con lo que aporta cada factor en el margen. Técnicamente, en una función homogénea la suma de las productividades marginales multiplicadas por las cantidades de cada factor es igual al producto multiplicado por el grado de homogeneidad, de modo que, con grado uno, pagar a cada factor su productividad marginal agota exactamente el producto.

Teorema de Gibbard-Satterthwaite Resultado según el cual casi cualquier sistema de votación con tres o más opciones da incentivos a mentir sobre las propias preferencias. Técnicamente, con al menos tres alternativas y dominio universal, no existe ninguna regla de elección no manipulable que cumpla el criterio de Pareto, la racionalidad y la ausencia de dictador.

Teorema de Henry George Resultado que justifica financiar los bienes públicos locales con un impuesto sobre el suelo. Técnicamente, en los modelos de bienes públicos locales, cuando la población de la ciudad es la óptima, la suma de las rentas del suelo iguala el gasto en el bien público, de modo que un impuesto que las grave íntegramente podría financiarlo sin distorsiones.

Teorema de imposibilidad de Arrow Resultado según el cual no existe ningún procedimiento de decisión colectiva que cumpla a la vez unas pocas exigencias que parecen razonables. Técnicamente, con al menos tres alternativas y un número finito de individuos, ninguna regla de elección colectiva cumple a la vez dominio universal, racionalidad, criterio de Pareto, ausencia de dictador e independencia de las alternativas irrelevantes.

Teorema de la integrabilidad Resultado que indica cuándo unas demandas observadas pueden atribuirse a un consumidor racional con unos gustos determinados. Técnicamente, un sistema de demandas marshallianas procede de una función de utilidad creciente y cuasicóncava si y solo si cumple la ley de Walras y su matriz de efectos sustitución es simétrica y semidefinida negativa.

Teorema de la separación Resultado según el cual se puede decidir cuánto invertir sin pensar en cuánto consumir. Técnicamente, formulado por Fisher, establece que con mercados de capitales perfectos la inversión óptima, también en capital humano si el ocio no da utilidad, es la que maximiza el valor actual de los ingresos, con independencia de las preferencias de consumo.

Teorema de Le Chatelier-Samuelson Resultado según el cual la empresa reacciona menos a un cambio de precios a corto plazo que a largo plazo. Técnicamente, cuantos más factores permanecen fijos, más rígidas son tanto la oferta del bien como la demanda de cada factor variable, por ejemplo la de trabajo.

Teorema de no sustitución Resultado según el cual, con cierta estructura tecnológica, los precios del resto de bienes no se ven afectados por lo que ocurra en un mercado concreto. Técnicamente, obtenido de forma independiente por Samuelson y Georgescu-Roegen (1951), establece que, si el numerario es el único factor primario, los demás bienes se producen con rendimientos constantes a escala y no hay producción conjunta, sus precios no dependen del mercado analizado, lo que justifica el equilibrio parcial con independencia del tamaño de ese mercado.

 © 2026 Manuel Cabado García

Teorema del exceso de capacidad Idea de que, en un mercado con muchas empresas de productos diferenciados, cada una produce menos de lo que le permitiría abaratar al máximo sus costes. Técnicamente, es el resultado según el cual, en el equilibrio a largo plazo de competencia monopolística, cada empresa produce a un coste medio superior al mínimo y deja economías de escala sin explotar.

Teorema del límite del núcleo Resultado según el cual, cuando el número de agentes crece, los únicos repartos que ningún grupo puede mejorar por su cuenta son los de equilibrio competitivo. Técnicamente, Debreu y Scarf (1963) demostraron, generalizando una intuición de Edgeworth, que al replicar la economía el núcleo se reduce hasta coincidir en el límite con el conjunto de asignaciones de equilibrio walrasiano.

Teorema del votante mediano Resultado según el cual, en ciertas votaciones por mayoría, gana lo que prefiere el votante situado justo en el centro. Técnicamente, según Black, con preferencias unimodales y alternativas unidimensionales, el resultado de la votación por mayoría coincide con la alternativa preferida por el votante mediano.

Teorema folk Resultado según el cual repetir indefinidamente un juego permite sostener acuerdos que serían inestables en una sola jugada. Técnicamente, con horizonte infinito y factor de descuento suficientemente alto, cualquier pago medio factible superior al del equilibrio de Nash estático puede sostenerse como equilibrio del juego repetido.

Teoría cuantitativa del dinero Teoría según la cual el nivel de precios depende de la cantidad de dinero en circulación. Técnicamente, parte de la ecuación cuantitativa, según la cual la cantidad de dinero multiplicada por su velocidad de circulación iguala el valor nominal de la producción. En la versión de Fisher, si la velocidad y la producción real no cambian, una variación de la cantidad de dinero se traduce en una variación proporcional del nivel general de precios, de modo que el dinero es neutral.

Teoría de juegos Estudio de las situaciones en las que lo que gana cada uno depende también de lo que hagan los demás. Técnicamente, es la rama de las matemáticas que estudia el comportamiento de agentes, llamados jugadores, cuyas decisiones se afectan mutuamente, lo que les obliga a considerar las posibles acciones del resto al elegir sus estrategias.

Teoría de la acción colectiva de Olson Explicación de por qué los grupos pequeños defienden sus intereses mejor que los grandes. Técnicamente, la actividad de un grupo de interés tiene carácter de bien público y genera incentivos al free-rider, por lo que los grupos pequeños se organizan con más facilidad que los grandes.

Teoría de la agencia Estudio de las relaciones en las que alguien encarga algo a otra persona que sabe más que él, como los accionistas al directivo de su empresa. Técnicamente, analiza el diseño de contratos entre un principal y un agente con información asimétrica e intereses enfrentados, tanto antes de la firma (selección adversa) como después (riesgo moral).

Teoría de la burocracia de Niskanen Explicación de por qué las oficinas públicas tienden a gastar más de lo necesario. Técnicamente, es el modelo de Niskanen en el que el burócrata maximiza el tamaño de su presupuesto y aprovecha la información asimétrica frente al político para imponer un nivel de provisión superior al eficiente.

Teoría de la dualidad Forma de estudiar las decisiones del consumidor planteando la misma elección desde dos ángulos que llevan al mismo resultado. Técnicamente, analiza dos problemas equivalentes, la maximización de la utilidad (problema primal) y la minimización del gasto (problema dual), y permite relacionar las funciones de utilidad con las de demanda.

Teoría de la elección colectiva Estudio de cómo una sociedad puede tomar decisiones conjuntas a partir de lo que quiere cada uno de sus miembros. Técnicamente, analiza cómo se agregan las preferencias individuales en una decisión colectiva coherente y las propiedades de los mecanismos que lo hacen posible.

Teoría de la preferencia revelada Forma de averiguar los gustos del consumidor observando lo que compra. Técnicamente, es el enfoque in-

ductivo de Samuelson que deduce las preferencias del comportamiento observado del consumidor y obtiene las mismas predicciones que la teoría neoclásica con supuestos menos restrictivos.

Teoría de la productividad marginal Idea de que cada factor de producción cobra según lo que aporta su última unidad. Técnicamente, es la teoría de la segunda generación marginalista según la cual cada factor productivo se remunera según su productividad marginal, del mismo modo que el valor de los bienes depende de su utilidad marginal.

Teoría de la renta relativa Idea de que cada familia decide cuánto consumir mirando a su entorno y a su propio pasado. Técnicamente, según Duesenberry, el consumo depende de la renta propia en relación con la del grupo de referencia y con la renta máxima alcanzada antes, lo que lo hace rígido a la baja (efecto trinquete).

Teoría de la utilidad esperada Teoría que explica cómo elige una persona racional entre opciones arriesgadas, valorando cada una por la satisfacción media que espera obtener. Técnicamente, según von Neumann y Morgenstern, si las preferencias sobre loterías cumplen completitud, transitividad, continuidad e independencia, el agente elige la lotería que maximiza la esperanza de la utilidad de los resultados.

Teoría de las perspectivas Teoría que describe cómo deciden de verdad las personas ante el riesgo, valorando ganancias y pérdidas respecto a su situación de partida. Técnicamente, es la teoría de Kahneman y Tversky que sustituye la utilidad esperada por una función de valor en forma de S, definida sobre ganancias y pérdidas respecto a un punto de referencia, y por pesos decisorios que distorsionan las probabilidades.

Teoría de los ciclos reales Explicación de las expansiones y las recesiones como la reacción sensata de la gente a cambios en la tecnología y en otros factores reales. Técnicamente, es la teoría de Kydland y Prescott que explica los ciclos como respuesta óptima de los agentes a perturbaciones reales, sobre todo de productividad, por lo que son un fenómeno de equilibrio eficiente en el sentido de Pareto que no requiere estabilización.

Teoría de los clubes Analogía que trata un territorio como un club cuyos socios comparten un servicio, para decidir cuántos deben ser. Técnicamente, en el planteamiento de Buchanan (1965) el tamaño óptimo de la jurisdicción es aquel en el que el ahorro de impuestos por un socio adicional iguala al coste marginal de atenderlo, lo que ocurre en el mínimo del coste medio de provisión.

Teoría de los costes de Stigler Idea de que las empresas construyen sus plantas para producir a un coste parecido en un amplio margen de cantidades. Técnicamente, es una crítica a la curva de costes en U según la cual las empresas diseñan plantas flexibles, de modo que el coste medio a corto plazo es plano en un amplio tramo de producción. La teoría moderna de costes completa esta idea con una curva de coste medio a largo plazo en forma de L.

Teoría del acelerador Idea de que las empresas invierten cuando esperan vender más. Técnicamente, en su versión simple, la inversión neta es un múltiplo fijo de la variación de la producción, y ese múltiplo, el coeficiente acelerador, es la relación entre el capital y la producción.

Teoría del capital humano Teoría que trata la educación como una inversión que eleva la productividad y el salario. Técnicamente, formulada por Schultz y Becker, supone que cada individuo se forma mientras el valor actual de los mayores salarios futuros supere los costes directos y el salario al que renuncia mientras estudia.

Teoría del ciclo vital Idea de que las personas reparten su consumo a lo largo de la vida para mantenerlo estable. Técnicamente, es la teoría de Modigliani según la cual el consumo depende de la renta vitalicia esperada, de modo que los individuos ahorran en las etapas de ingresos altos y desahorran en las de ingresos bajos para mantener un consumo estable.

Teoría del federalismo fiscal Estudio de cómo repartir tareas e ingresos entre el Estado, las regiones y los municipios. Técnicamente, es la parte de la Economía del Sector Público que analiza cómo se adoptan las decisiones descentralizadas: justificación, reparto de competencias, niveles de gobierno y financiación.

 © 2026 Manuel Cabado García

Teoría del second best Idea de que, cuando una economía tiene una distorsión que no puede eliminarse, acercarse a las demás condiciones de eficiencia no garantiza más bienestar. Técnicamente, Lipsey y Lancaster (1956) mostraron que, si no se cumple una condición de óptimo de Pareto, las restantes dejan en general de ser deseables y el objetivo pasa a ser un óptimo restringido o de segundo orden.

Teoría del valor-trabajo Teoría según la cual el valor de una mercancía depende del trabajo necesario para producirla. Técnicamente, Ricardo lo identificó con el trabajo incorporado, directo e indirecto a través del capital, y Marx con el tiempo de trabajo socialmente necesario, a partir del cual explicó la plusvalía.

Teoría positiva de la regulación Explicación de por qué se regula como se regula, sin suponer que el regulador busca el interés general. Técnicamente, es el enfoque de la Escuela de Chicago, de Stigler, Peltzman y Becker, que trata la regulación como un bien intercambiado en un mercado político en el que los grupos de presión compiten por las rentas.

Teoría subjetiva del valor Idea de que las cosas valen lo que las personas creen que valen para ellas. Técnicamente, es la teoría según la cual el valor de los bienes no depende de su naturaleza objetiva ni de su coste de producción, sino de la estimación subjetiva de su utilidad por los individuos.

Tercera ley de Gossen Idea de que solo tiene valor económico lo que es escaso. Técnicamente, es el tercero de los enunciados atribuidos a Gossen, también llamado ley de la escasez, según el cual un bien solo tiene valor cuando su demanda supera a la cantidad disponible.

Test de replicabilidad económica Prueba de que un rival eficiente puede igualar las ofertas del operador dominante pagando los precios mayoristas regulados. Técnicamente, comprueba que el margen entre el precio mayorista de acceso y el precio minorista del incumbente permite competir a un operador alternativo eficiente, de modo que no se le expulse estrechando ese margen. La CNMC lo aplicaba a los servicios mayoristas de fibra de Telefónica.

Test SSNIP Prueba que usan las autoridades de competencia para decidir qué productos forman un mismo mercado. Técnicamente, es un criterio de delimitación del mercado relevante que pregunta si un monopolista hipotético podría subir de forma rentable y sostenida el precio un margen pequeño pero significativo, habitualmente un 5 por ciento.

Tipo de interés monetario Tipo de interés que se cobra en los mercados de crédito, también llamado efectivo. Técnicamente, en el análisis de Wicksell lo fijan los bancos con cierta autonomía respecto al tipo natural, y la diferencia entre ambos pone en marcha el proceso acumulativo.

Tipo de interés natural Rentabilidad que el capital obtiene en la producción, con independencia de lo que cobren los bancos. Técnicamente, para Wicksell viene determinado por la productividad marginal del capital y es el tipo con el que el ahorro iguala a la inversión y los precios permanecen estables.

Tipo marginal Parte que se paga en impuestos de un euro adicional de base. Técnicamente, es la variación de la cuota ante un aumento muy pequeño de la base imponible, y el impuesto es progresivo en un punto cuando el tipo marginal supera al tipo medio.

Tipo medio Proporción de la base imponible que se paga en impuestos. Técnicamente, es el cociente entre la cuota y la base imponible, y cómo evoluciona al aumentar la base define si un impuesto es progresivo, proporcional o regresivo.

Tiranía de la mayoría Peligro de que los más numerosos impongan su criterio a los demás. Técnicamente, es el riesgo de que la mayoría imponga su voluntad a la minoría y la presione para aceptar propuestas que no la benefician.

Tiranía de la minoría Peligro de que unos pocos bloqueen una decisión para sacar ventaja. Técnicamente, es el riesgo, señalado por Baumol, de que una minoría con poder de veto bloquee las decisiones de la mayoría hasta obtener concesiones desproporcionadas.

Tragedia de los anticomunes Situación en la que un recurso se aprovecha menos de lo debi-

do porque son demasiados los que pueden impedir su uso. Técnicamente, es la infrautilización de un recurso causada por un exceso de derechos de propiedad o de exclusión en manos de muchas partes.

Tragedia de los comunes Agotamiento de un recurso que todos pueden usar, como un pasto o un caladero, porque a cada usuario le conviene aprovecharlo al máximo. Técnicamente, descrita por Hardin (1968), es la sobreexplotación de un recurso común, rival y no excluible, porque cada usuario no tiene en cuenta el coste que su uso impone a los demás, de modo que la acción individualmente racional resulta perjudicial para la comunidad a largo plazo.

Trampa de liquidez Situación en la que inyectar más dinero en la economía no consigue bajar los tipos de interés, porque el público lo retiene en lugar de gastarlo o invertirlo. Técnicamente, la demanda de dinero se vuelve infinitamente elástica en un tipo de interés mínimo y la política monetaria pierde eficacia.

Trampa de pobreza Situación en la que ganar más apenas mejora la renta de quien recibe una ayuda, porque la ayuda se reduce casi en la misma cuantía. Técnicamente, la retirada de prestaciones condicionadas a la renta actúa como un tipo marginal implícito cercano o superior al 100 %, que desincentiva la oferta de trabajo de las rentas bajas.

Transferencia condicional compensatoria Subvención que crece con lo que el propio territorio gasta en el fin subvencionado, de modo que le abarata ese gasto. Técnicamente, el donante aporta una cantidad fija por cada unidad monetaria que aporta el beneficiario, lo que reduce el coste unitario del bien público para este y añade al efecto renta un efecto sustitución, por lo que, a igual coste para el donante, es la transferencia que más estimula el gasto. También se denomina cofinanciación.

Transferencia condicional no compensatoria Ayuda de una cuantía fija que debe dedicarse a un fin concreto. Técnicamente, es una transferencia condicionada a un uso concreto cuyo importe no depende del gasto del beneficiario en ese fin. Como el dinero es intercambiable, salvo que el territorio gastara en ese fin menos de lo transferido, produce el mismo efecto renta que una transferencia incondicional.

Transferencia incondicional Dinero que un gobierno da a otro para que lo gaste como quiera. Técnicamente, es la transferencia concedida sin condiciones sobre su destino, que genera un efecto renta puro y es la opción aconsejable para maximizar el bienestar territorial.

Transferencia intergubernamental Dinero que un gobierno da a otro para que pague sus gastos. Técnicamente, es la cantidad de dinero que un gobierno donante entrega a otro beneficiario para que financie sus gastos.

Transformación de Harsanyi Procedimiento para analizar juegos en los que un jugador desconoce algún rasgo de sus rivales. Técnicamente, convierte un juego de información incompleta en uno de información imperfecta al introducir un movimiento de la naturaleza que asigna a cada jugador un tipo que solo él observa.

Transformaciones de Musgrave Cambios que trae consigo el desarrollo y que empujan al alza el gasto del Estado. Técnicamente, son los cambios demográficos, tecnológicos y urbanísticos asociados al desarrollo económico que, según Musgrave, tienden a elevar el gasto público.

Transitividad Coherencia en las preferencias: si se prefiere una cosa a otra y esta a una tercera, se prefiere la primera a la tercera. Técnicamente, es el axioma según el cual, si una cesta es al menos tan buena como otra y esta como una tercera, la primera es al menos tan buena como la tercera.

Traslación impositiva Paso de la carga de un impuesto desde quien debe pagarlo según la ley hacia otros agentes. Técnicamente, es la transferencia, total o parcial, de la carga desde el contribuyente legal, que hace que la incidencia legal y la económica no coincidan. Puede ser hacia delante, cuando el contribuyente legal eleva el precio de lo que vende, o hacia atrás, cuando rebaja lo que paga por lo que compra, como los salarios o los insumos.

 © 2026 Manuel Cabado García

Triángulo de Harberger Área que representa gráficamente la pérdida irrecuperable de eficiencia de una distorsión. Técnicamente, es el triángulo comprendido entre la demanda y la oferta, o el coste marginal, entre la cantidad eficiente y la que resulta del monopolio o de un impuesto, y Harberger (1954) lo estimó en torno al 0,1 % de la renta nacional para el monopolio en la industria manufacturera estadounidense.

Triángulo de hierro Alianza estable entre quienes hacen las leyes, quienes las aplican y quienes se benefician de ellas. Técnicamente, es la relación entre políticos, burócratas y grupos de interés que, según Adams, determina las decisiones políticas y actúa en favor de los buscadores de rentas.

U

Utilidad Satisfacción que un bien proporciona a quien lo consume. Técnicamente, es la capacidad de los bienes para satisfacer necesidades, que el enfoque cardinal trata como una magnitud medible y el enfoque ordinal solo como un índice que ordena las preferencias.

Utilidad de reserva Lo mínimo que alguien tiene que obtener de un trato para aceptarlo, porque es lo que conseguiría en otra parte. Técnicamente, es la utilidad mínima por la que el agente estaría dispuesto a aceptar el contrato, igual a la que obtendría con su mejor alternativa.

Utilidad esperada subjetiva Extensión de la utilidad esperada a las situaciones sin probabilidades objetivas. Técnicamente, en el planteamiento de Savage el agente se comporta como si asignara a los estados de la naturaleza unas probabilidades subjetivas, determinadas de forma única, y maximizara con ellas la utilidad esperada, de modo que el riesgo es el caso particular en que coinciden con las objetivas.

Utilidad marginal Satisfacción adicional que proporciona la última unidad consumida de un bien. Técnicamente, es la variación de la utilidad ante un aumento muy pequeño de la cantidad consumida de ese bien, es decir, la derivada de la función de utilidad respecto a ella, y es decreciente según la primera ley de Gossen.

Utilidad marginal de la renta Satisfacción adicional que obtendría el consumidor si dispusiera de un euro más. Técnicamente, es el aumento de la utilidad máxima alcanzable ante una unidad adicional de renta, recogido por el multiplicador de Lagrange del problema de maximización de la utilidad, que actúa como precio sombra de la restricción presupuestaria.

V

Valor de Arrow-Debreu Título teórico que solo paga si ocurre un acontecimiento determinado. Técnicamente, es un instrumento financiero hipotético que paga una unidad de riqueza si ocurre un evento futuro concreto y cero en cualquier otro caso.

Valor de cambio Cantidad de otras cosas que se pueden obtener a cambio de un bien. Técnicamente, es el poder que tiene una mercancía para comprar otros bienes, equivalente a su precio.

Valor de la productividad marginal del trabajo Lo que ingresa la empresa por el producto de un trabajador adicional. Técnicamente, es el precio del bien multiplicado por la productividad marginal del trabajo, y en competencia perfecta es la curva de demanda de trabajo a corto plazo.

Valor de Shapley Forma de repartir lo que ganan juntos varios jugadores según lo que aporta cada uno. Técnicamente, es el reparto del excedente conjunto que asigna a cada jugador el promedio de su contribución marginal a todas las coaliciones posibles, cumpliendo axiomas de simetría, eficiencia y aditividad.

Valor de uso Utilidad que tiene un bien para quien lo consume. Técnicamente, es la capacidad de un bien para satisfacer una necesidad, que Smith distinguió del valor de cambio.

Valor estadístico de la vida Cantidad que una población está dispuesta a pagar, en conjunto, para evitar una muerte. Técnicamente, se deduce de la prima salarial que los trabajadores exigen por aceptar un pequeño aumento del riesgo de accidente mortal y se usa para evaluar la regulación de la seguridad laboral.

Variación compensatoria Cantidad de dinero que compensa exactamente un cambio de precios una vez producido. Técnicamente, es la variación de renta que, a los precios finales, permitiría al consumidor mantener su utilidad inicial, y se calcula integrando la demanda hicksiana asociada a la utilidad inicial.

Variación conjetural Lo que una empresa espera que hagan sus rivales cuando ella cambia su conducta. Técnicamente, es la creencia que una empresa tiene sobre cómo reaccionarán sus competidores, en cantidad o en precio, si ella modifica su propia producción o precio.

Variación equivalente Cantidad de dinero que equivale a un cambio de precios sin que este llegue a producirse. Técnicamente, es la variación de renta que, a los precios iniciales, proporcionaría al consumidor la utilidad que alcanza tras el cambio, y se calcula integrando la demanda hicksiana asociada a la utilidad final.

Velo de la ignorancia Experimento mental en el que se eligen las reglas de la sociedad sin saber qué lugar se va a ocupar en ella. Técnicamente, es la situación hipotética de Rawls en la que los individuos desconocen de antemano su lugar en la sociedad, lo que les lleva a adoptar la regla maximin.

Ventaja absoluta Capacidad de un país para producir un bien más barato que los demás. Técnicamente, es la situación de un país que produce un bien con un coste menor que cualquier otro país. Para Smith, cada país debía especializarse en los bienes en que la posee.

Ventaja comparativa Capacidad de un país para producir un bien renunciando a menos de otros bienes que su socio comercial. Técnicamente, es la situación de un país que produce un bien con menor coste relativo, o coste de oportunidad, que otro. Según Ricardo, es la base de la especialización y de las ganancias del comercio.

Ventana de Overton Conjunto de ideas políticas que la opinión pública está dispuesta a aceptar en cada momento. Técnicamente, es el rango de políticas que una sociedad considera aceptables en un momento dado, desde lo impensable hasta la política vigente, y que se desplaza con el tiempo.

Voto con los pies Forma de mostrar qué servicios públicos se prefieren mudándose al municipio que los ofrece. Técnicamente, es el mecanismo descrito por Tiebout por el que los individuos revelan sus preferencias por los bienes públicos locales desplazándose a las jurisdicciones cuyas políticas más se aproximan a ellas.

Voto útil Voto que se da a quien tiene opciones de ganar en lugar de al candidato que más gusta. Técnicamente, es el voto a un candidato distinto del preferido que tiene más probabilidades de ganar, lo que puede alejar el resultado del votante mediano.

W

Welfarismo Enfoque que juzga lo bien o lo mal que está una sociedad solo por la utilidad de sus miembros. Técnicamente, es el marco informativo de la economía del bienestar tradicional, en el que la evaluación social depende únicamente de los niveles de utilidad individuales, y que Sen criticó proponiendo valorar las capacidades.

Z

Zona de Malinvaud Conjunto de repartos de la producción con los que las dos empresas ganarían más si se pusieran de acuerdo. Técnicamente, es la región en forma de lente delimitada por las curvas isobeneficio que pasan por un equilibrio no cooperativo, en la que ambas partes obtendrían mayores beneficios cooperando.

© 2026 Manuel Cabado García