# Lecciones de economía general e internacional

**Lecciones de economía, vol. 1 · Macroeconomía**

Arturo Lorenzo

Edición web 1.0 · noviembre de 2026

# Tabla de contenidos

**Nota** **9**

**A29. Modelos de horizonte infinito y generaciones solapadas** **10**
Introducción . . . 10
1. Fundamentos de la elección intertemporal . . . 12
1.1. El conjunto de elección y la ordenación de las preferencias intertemporales . . . 12
1.2. La función de utilidad intertemporal y sus curvas de indiferencia . . . 13
1.3. La restricción intertemporal de recursos . . . 15
1.4. La decisión de consumo intertemporal (Fisher, 1930) . . . 18
1.5. Conclusiones . . . 26
2. Decisiones intertemporales de la empresa . . . 27
2.1. El conjunto de elección y el objetivo de la empresa . . . 27
2.2. La función de valor y el precio sombra del capital . . . 28
2.3. La restricción tecnológica intertemporal . . . 29
2.4. La decisión de inversión . . . 30
2.5. Conclusiones . . . 37
3. Modelización dinámica del comportamiento colectivo: horizonte infinito y generaciones solapadas . . . 38
3.1. Modelos de horizonte infinito (MHI) . . . 38
3.2. Modelos de generaciones solapadas (MGS) . . . 41
3.3. Comparación entre ambos marcos . . . 46
3.4. Conclusiones . . . 46
Conclusiones . . . 48
Anexos . . . 49
A.1. La regla de Hotelling . . . 49
A.2. Ejemplos resueltos . . . 50

**A30. Magnitudes macroeconómicas y contabilidad nacional** **52**
Introducción . . . 52
1. Fundamentos de la contabilidad nacional . . . 54
1.1. Fundamento teórico: el flujo circular de la renta (Quesnay, 1758; Knight, 1933) 54
1.2. Concepto y base normativa . . . 56
1.3. Contenido: unidades, flujos y *stocks* . . . 56
1.4. Principios del marco contable . . . 59
1.5. Conclusiones . . . 60
2. Las cuentas nacionales . . . 62
2.1. El PIB y el triple enfoque . . . 62
2.2. Cuentas no financieras . . . 63
2.3. Cuentas financieras . . . 68
2.4. Lectura económica de los saldos: identidades ahorro-inversión . . . 69
2.5. Conclusiones . . . 70
3. Otras cuentas, indicadores y reflexiones . . . 71
3.1. Otras cuentas . . . 71
3.2. Indicadores de la economía en su conjunto: crecimiento, renta y ciclo . . . 72

3.3. Indicadores de los mercados de trabajo y de bienes: empleo y precios . . . . 73
3.4. Indicadores sectoriales: sector público, sector financiero y sector exterior . . . 75
3.5. Reflexiones: el PIB como medida del bienestar . . . . . 78
Conclusiones . . . . . 79
Anexo. Ejemplos resueltos . . . . . 80

**A32. Precios y output en economía abierta** **82**
Introducción . . . . . 82
1. La demanda agregada en economía abierta: el modelo IS-LM-BP (Fleming, 1962; Mundell, 1963) . . . . . 84
1.1. Supuestos . . . . . 84
1.2. Modelo . . . . . 84
1.3. Políticas monetaria y fiscal con precios fijos . . . . . 91
1.4. Conclusiones . . . . . 96
2. El modelo OA-DA en economía abierta: determinación de la renta y los precios . . . 97
2.1. Supuestos . . . . . 97
2.2. Modelo . . . . . 97
2.3. Políticas monetaria y fiscal . . . . . 103
2.4. *Shocks* y políticas de oferta . . . . . 107
2.5. Conclusiones . . . . . 110
3. Inflación y brecha de producción: el modelo IS-PC-MR en economía abierta (Galí y Monacelli, 2005) . . . . . 111
3.1. Supuestos . . . . . 111
3.2. Modelo en economía cerrada . . . . . 111
3.3. Modelo en economía abierta . . . . . 114
3.4. Conclusiones . . . . . 117
Conclusiones . . . . . 119

**A33. Teorías de demanda de consumo** **120**
Introducción . . . . . 120
1. Antecedentes a las teorías modernas sobre el consumo . . . . . 122
1.1. Aportación neoclásica . . . . . 122
1.2. Aportación keynesiana . . . . . 123
1.3. Conclusiones . . . . . 129
2. El consumo como función intertemporal en horizontes amplios . . . . . 130
2.1. Fundamentos del consumo intertemporal (Fisher, 1930) . . . . . 130
2.2. Hipótesis del ciclo vital (Modigliani y Brumberg, 1954; Ando y Modigliani, 1963) . . . . . 133
2.3. Hipótesis de la renta permanente (Friedman, 1957) . . . . . 136
2.4. Paseo aleatorio de Hall (1978) . . . . . 140
2.5. Conclusiones . . . . . 142
3. Desarrollos recientes y demanda de bienes de consumo duradero . . . . . 143
3.1. Exceso de sensibilidad y exceso de suavidad . . . . . 143
3.2. Cambios en los supuestos sobre el entorno y los agentes . . . . . 144
3.3. Cambios en los supuestos sobre la función de utilidad . . . . . 147
3.4. La demanda de bienes de consumo duradero (Mankiw, 1982; Bernanke, 1985) 149
3.5. Evidencia empírica e implicaciones de política económica . . . . . 150
3.6. Conclusiones . . . . . 150
Conclusiones . . . . . 152
Anexo. Análisis gráfico de la hipótesis de la renta permanente . . . . . 153

**A34. Teorías de demanda de inversión** **155**
Introducción . . . 155
1. Fundamentos de la teoría moderna sobre la demanda de inversión . . . 157
1.1. Aportación neoclásica: la tasa de retorno sobre el coste (Fisher, 1930) . . . 157
1.2. Aportación keynesiana: la eficiencia marginal del capital (Keynes, 1936) . . . 160
1.3. Revitalización del principio del acelerador . . . 163
1.4. Perspectiva austriaca: inversión como ajuste óptimo (Hayek, 1941) . . . 167
1.5. Conclusiones . . . 167
2. Teorías modernas de la inversión . . . 169
2.1. Modelo de demanda de inversión sin costes de ajuste (Jorgenson, 1963; 1967) 169
2.2. Fundamentos de los modelos con costes de ajuste . . . 172
2.3. Modelos con costes de ajuste: modelos de inversión basados en la ratio $q$ (Summers, 1981; Abel, 1982; Hayashi, 1982) . . . 174
2.4. Conclusiones . . . 182
3. Incertidumbre, irreversibilidad y financiación de la inversión . . . 183
3.1. Incertidumbre e irreversibilidad: la inversión como opción real (Bernanke, 1983; Dixit y Pindyck, 1994) . . . 183
3.2. Importancia de las decisiones de financiación . . . 185
3.3. Implicaciones de política económica . . . 188
Conclusiones . . . 189

**A35. Teorías de la demanda de dinero** **190**
Introducción . . . 190
1. Fundamentos de la teoría sobre la demanda de dinero . . . 192
1.1. Aportación clásica y neoclásica: desarrollos basados en la teoría cuantitativa del dinero . . . 192
1.2. Teoría keynesiana: la preferencia por la liquidez (Keynes, 1936) . . . 195
1.3. Síntesis neoclásica . . . 198
1.4. Monetarismo . . . 207
1.5. Conclusiones . . . 212
2. Desarrollos posteriores . . . 213
2.1. Modelos de generaciones solapadas (Allais, 1947; Samuelson, 1958; Diamond, 1965) . . . 213
2.2. Dinero en la utilidad y sus microfundamentos . . . 215
2.3. Otras aportaciones . . . 220
2.4. Conclusiones . . . 221
3. Evidencia empírica e implicaciones de política económica . . . 222
3.1. Hechos estilizados . . . 222
3.2. Implicaciones de política económica . . . 222
3.3. Debate: economía sin efectivo y dinero digital . . . 222
3.4. Conclusiones . . . 224
Conclusiones . . . 226

**A36. Política monetaria I: diseño e instrumentos** **227**
Introducción . . . 227
1. Fundamentos del diseño de la política monetaria . . . 229
1.1. Concepto de dinero: definición teórica y operativa . . . 229
1.2. Oferta de dinero: el papel del sistema monetario . . . 230
1.3. Demanda de dinero . . . 231
1.4. Equilibrio en el mercado monetario: el proceso de creación de dinero . . . 232
1.5. Conclusiones . . . 236

2. Diseño de la política monetaria . . . 237
2.1. Definición y elementos del diseño . . . 237
2.2. Debate sobre los objetivos y la variable operativa . . . 237
2.3. Debate sobre la discrecionalidad y la credibilidad . . . 242
2.4. Política óptima: el modelo IS-PC-MR . . . 243
2.5. Conclusiones . . . 245
3. Instrumentación de la política monetaria . . . 246
3.1. Instrumentos cuantitativos convencionales . . . 246
3.2. Instrumentos cuantitativos no convencionales . . . 249
3.3. Instrumentos cualitativos y medidas transversales . . . 253
3.4. Debates recientes sobre la instrumentación . . . 256
3.5. Conclusiones . . . 257
Conclusiones . . . 258

**A37. Política monetaria II: transmisión y efectos** **259**
Introducción . . . 259
1. Mecanismos de transmisión de la política monetaria convencional: los canales . . . 261
1.1. Canal de los tipos de interés . . . 261
1.2. Canal del crédito (Bernanke y Gertler, 1995) . . . 262
1.3. Canales de los precios de los activos: riqueza, tipo de cambio y $q$ de Tobin . . . 266
1.4. Canal de la señalización: el papel de las expectativas . . . 267
1.5. Canal de la redistribución . . . 268
1.6. Conclusiones . . . 268
2. Efectos de la política monetaria: de la neutralidad a las rigideces y fricciones financieras 269
2.1. Teoría clásica: la neutralidad del dinero . . . 269
2.2. El debate entre el keynesianismo, la síntesis y el monetarismo . . . 270
2.3. Nueva macroeconomía clásica: vuelta a la no intervención monetaria . . . 273
2.4. Nueva economía keynesiana: vuelta a la defensa del intervencionismo . . . 277
2.5. Conclusiones . . . 281
3. Evidencia empírica y transmisión de la política monetaria no convencional . . . 282
3.1. Metodología . . . 282
3.2. Limitaciones de carácter general . . . 282
3.3. Limitaciones de carácter específico . . . 283
3.4. Hechos observados vinculados al horizonte temporal . . . 284
3.5. Transmisión de la política monetaria no convencional . . . 285
Conclusiones . . . 289

**A38. Política fiscal, crecimiento y ahorro** **290**
Introducción . . . 290
1. Efectos vía gasto público no productivo y gravamen . . . 292
1.1. Antecedentes: el modelo IS-LM . . . 292
1.2. Perspectiva NMC: el modelo del ciclo real con sector público . . . 296
1.3. Perspectiva NEK: el modelo IS-PC-MR con gasto público . . . 300
1.4. Evidencia empírica . . . 305
1.5. Conclusiones . . . 305
2. Efectos vía diseño del sistema de pensiones . . . 306
2.1. Sistemas de pensiones . . . 306
2.2. Modelo base: generaciones solapadas (Allais, 1947; Samuelson, 1958; Diamond, 1965) . . . 308
2.3. Evidencia empírica . . . 310
2.4. Conclusiones . . . 311

3. Efectos vía gasto público productivo . . . 312
3.1. Gasto público productivo . . . 312
3.2. Crecimiento exógeno: el modelo de Ramsey-Cass-Koopmans con sector público 312
3.3. Crecimiento endógeno: el modelo de Barro (1990) . . . 317
3.4. Evidencia empírica . . . 320
3.5. Conclusiones . . . 321
Conclusiones . . . 322

**A39. Déficit público, su financiación y sostenibilidad** **323**
Introducción . . . 323
1. Fundamentos del análisis del déficit público y de la deuda . . . 325
1.1. Conceptos esenciales . . . 325
1.2. Debates sobre la efectividad de la política fiscal . . . 326
1.3. Conclusiones . . . 329
2. Financiación del déficit y su sostenibilidad . . . 330
2.1. Financiación vía deuda: la dinámica de la deuda pública y su sostenibilidad . 330
2.2. Financiación vía monetización en una economía cerrada . . . 337
2.3. Financiación en una economía abierta . . . 341
2.4. Conclusiones . . . 342
3. Dominancia monetaria y dominancia fiscal . . . 344
3.1. El modelo de Leeper (1991) . . . 344
3.2. Dominancia fiscal . . . 348
3.3. Dominancia monetaria . . . 350
3.4. Otros desarrollos y evidencia empírica . . . 351
3.5. Conclusiones . . . 353
Conclusiones . . . 354

**A40. Interrelación de políticas macroeconómicas** **355**
Introducción . . . 355
1. Análisis de la efectividad de las políticas monetaria y fiscal . . . 357
1.1. Multiplicador monetario . . . 357
1.2. Multiplicador fiscal . . . 359
1.3. Conclusiones . . . 362
2. Análisis de la interrelación de las políticas monetaria y fiscal . . . 363
2.1. Interrelación a corto plazo: el modelo IS-PC-MR y el límite inferior de los tipos de interés . . . 363
2.2. Interrelación intertemporal: el modelo de Leeper (1991) . . . 365
2.3. Teorías de dominancia monetaria y fiscal . . . 366
2.4. Conclusiones . . . 367
3. Evidencia empírica desde la Gran Recesión: el valor de los multiplicadores y la interrelación de las políticas . . . 368
3.1. Técnicas de medición . . . 368
3.2. Evidencia empírica sobre el multiplicador fiscal . . . 369
3.3. Evidencia empírica sobre el multiplicador monetario . . . 371
3.4. Evidencia empírica sobre la interrelación entre ambas políticas . . . 373
Conclusiones . . . 376

**A41. Inflación** **377**
Introducción . . . 377
1. La inflación: concepto, costes y efectos sobre la eficiencia y el bienestar . . . 379
1.1. Concepto y medición . . . 379
1.2. Los costes de la inflación . . . 380

1.3. Los beneficios de la inflación y la inflación óptima . . . 382
1.4. Conclusiones . . . 384
2. Las causas de la inflación: las escuelas y sus curvas de Phillips . . . 385
2.1. Los fundamentos: teoría clásica y teoría keynesiana . . . 385
2.2. Síntesis neoclásica y monetarismo: la curva de Phillips y sus expectativas . . . 388
2.3. Nueva macroeconomía clásica: información imperfecta e inconsistencia dinámica 391
2.4. Nueva economía keynesiana: rigideces reales y nominales . . . 395
2.5. Conclusiones . . . 399
3. Hiperinflación y deflación . . . 401
3.1. La hiperinflación . . . 401
3.2. La deflación . . . 405
3.3. Conclusiones . . . 407
Conclusiones . . . 408

**A42. Ciclos económicos** **409**
Introducción . . . 409
1. Concepto y antecedentes . . . 411
1.1. Concepto de ciclo económico . . . 411
1.2. Primeras aportaciones . . . 412
1.3. Modelos del acelerador y ciclos endógenos . . . 413
1.4. Conclusiones . . . 418
2. Teorías modernas del ciclo . . . 419
2.1. Ciclos nominales: el modelo de información imperfecta (Lucas, 1972; 1973; 1975) . . . 419
2.2. Ciclos reales: la teoría del ciclo real (Kydland y Prescott, 1982; Long y Plosser, 1983) . . . 423
2.3. Ciclos de desequilibrio y fricciones financieras . . . 428
2.4. La nueva síntesis neoclásica: el modelo IS-PC-MR . . . 433
2.5. Conclusiones . . . 437
3. Análisis empírico del ciclo . . . 438
3.1. Medición de los ciclos . . . 438
3.2. Hechos estilizados . . . 439
3.3. Las teorías frente a los hechos . . . 440
3.4. Conclusiones . . . 440
Conclusiones . . . 442
Anexo. Ecuación característica, raíces y ciclo . . . 443
A.1. La ecuación en diferencias y su solución . . . 443
A.2. La ecuación característica . . . 443
A.3. Raíces reales: basta el valor absoluto . . . 444
A.4. Raíces complejas: hace falta el módulo . . . 444
A.5. Resumen . . . 445
A.6. Representación gráfica del modelo de Kaldor (1940) . . . 445

**A43. Crecimiento económico exógeno** **447**
Introducción . . . 447
1. Modelo de Harrod-Domar (1939, 1946) . . . 449
1.1. Supuestos . . . 449
1.2. Modelo . . . 449
1.3. Conclusiones . . . 451
2. Modelo de Solow-Swan (1956) . . . 452
2.1. Supuestos . . . 452
2.2. Modelo sin progreso técnico . . . 453

2.3. Modelo con progreso técnico . . . 460
2.4. Dinámica de transición y convergencia . . . 461
2.5. Extensiones . . . 462
2.6. Conclusiones . . . 465
3. Modelo de Ramsey-Cass-Koopmans (1928, 1965) . . . 466
3.1. Supuestos . . . 466
3.2. Modelo . . . 466
3.3. Conclusiones . . . 472
Conclusiones . . . 473
Anexo. De los niveles a las tasas de crecimiento . . . 474

**A44. Crecimiento económico endógeno** **475**
Introducción . . . 475
1. Función de producción no neoclásica . . . 477
1.1. Modelo AK (Rebelo, 1991) . . . 477
1.2. Modelo Sobelow (Jones y Manuelli, 1990) . . . 479
1.3. Externalidades del capital (Romer, 1986; Lucas, 1988) . . . 481
1.4. Gasto público productivo (Barro, 1990) . . . 485
1.5. Conclusiones . . . 488
2. Capital humano con tasa de ahorro endógena: el modelo de Uzawa-Lucas (1965; 1988) 489
2.1. Supuestos . . . 489
2.2. Modelo . . . 490
2.3. Dinámica . . . 493
2.4. Conclusiones . . . 493
3. Innovación tecnológica: modelos de I+D con competencia imperfecta . . . 494
3.1. Aumento del número de variedades (Romer, 1990) . . . 494
3.2. Enfoque schumpeteriano (Aghion y Howitt, 1992) . . . 497
3.3. La crítica del efecto escala y el crecimiento semiendógeno (Jones, 1995) . . . 501
Conclusiones . . . 502
Anexo. La función de producción CES y la condición de Inada . . . 503

**A45. Evidencia empírica sobre el crecimiento** **504**
Introducción . . . 504
1. Predicciones teóricas y contabilidad del crecimiento . . . 506
1.1. Las predicciones del modelo de Solow-Swan . . . 506
1.2. Contabilidad del crecimiento . . . 507
1.3. Conclusiones . . . 510
2. Herramientas para el análisis de la convergencia y la distribución . . . 511
2.1. Análisis de convergencia . . . 511
2.2. Análisis de la distribución personal de la renta . . . 514
2.3. Conclusiones . . . 516
3. Evidencia empírica . . . 517
3.1. Evidencia empírica sobre el crecimiento . . . 517
3.2. Evidencia empírica sobre convergencia . . . 521
3.3. Evidencia empírica sobre distribución de la renta: desigualdad y crecimiento 522
3.4. Evidencia empírica sobre la distribución factorial: la participación del trabajo 523
3.5. Evidencia empírica sobre la transformación estructural . . . 524
Conclusiones . . . 526

# Nota

Este PDF es un extracto de *Lecciones de economía general e internacional* (Lecciones de economía, vol. 1), de Arturo Lorenzo, con el bloque «Macroeconomía» (temas A29-A45). Se distribuye gratis bajo licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 (CC BY-NC-SA 4.0).

Versión de noviembre de 2026, generada a partir de la edición web. La edición completa y actualizada está en https://librovol1.arturolorenzo.com.

# A29. Modelos de horizonte infinito y generaciones solapadas

Modelización dinámica de las tomas de decisiones. Modelos de horizonte infinito y modelos de generaciones solapadas.

## Introducción

> **Introducción común.** Los temas A14 y A29 comparten una introducción, que se desarrolla al comienzo del tema **A14**. Presenta las dos herramientas analíticas que atraviesan la teoría económica moderna, la teoría de juegos, que modeliza la interdependencia estratégica, y el análisis intertemporal, que modeliza el tiempo.

El **análisis dinámico o intertemporal** estudia las relaciones entre variables económicas cuyos valores están referidos a diferentes instantes del tiempo[1]. Cada individuo divide su renta en dos usos. El consumo es la finalidad última de toda actividad económica, mientras que el ahorro traslada recursos en el tiempo y financia la formación de capital físico o humano. Los motivos de esa reasignación son la suavización del consumo, la precaución, el deseo de dejar herencias y las restricciones de crédito.

La dimensión temporal está presente en casi todas las decisiones económicas. El consumo y el ahorro de los hogares (tema A33), la inversión de las empresas (tema A34), la sostenibilidad de la deuda pública y la equivalencia ricardiana (tema A39), las pensiones (tema A38) y el crecimiento (temas A43 y A44) se analizan con las herramientas de este tema.

Fisher (1930) planteó la elección entre consumo presente y futuro en dos periodos, y la macroeconomía la extendió después en dos direcciones. Los **modelos de horizonte infinito**, que parten de Ramsey (1928), suponen un agente representativo que vive para siempre o que se preocupa por sus descendientes. Los **modelos de generaciones solapadas** de Samuelson (1958) y Diamond (1965) suponen, en cambio, individuos de vida finita que conviven con otras generaciones. La elección entre ambos marcos tiene consecuencias, porque solo con generaciones solapadas puede la economía acumular demasiado capital, la **ineficiencia dinámica**, y solo entonces la deuda pública o las pensiones de reparto pueden mejorar el bienestar de todas las generaciones. Los modelos dinámicos de equilibrio general que utilizan los bancos centrales parten, no obstante, del horizonte infinito.

La empresa sigue la misma lógica, porque invertir es renunciar a beneficios presentes a cambio de beneficios futuros, y el precio sombra del capital, la $q$ marginal, mide el valor de esa renuncia. El envejecimiento de la población y el aumento de la deuda pública han devuelto estas herramientas

[1] El análisis dinámico no es estática comparativa, aunque el principio de correspondencia (Samuelson, 1947) establece una relación estrecha. En términos formales, los modelos dinámicos incluyen alguna variable endógena retardada.

al centro del debate, porque de ellas dependen la sostenibilidad de los sistemas de pensiones y el reparto de las cargas entre generaciones.

El tema se organiza en tres partes:

- *Fundamentos*: el tema presenta los fundamentos de la elección intertemporal del consumidor.

- *Decisiones de la empresa*: la misma lógica se aplica a las decisiones intertemporales de la empresa.

- *Horizonte infinito y generaciones solapadas*: se exponen los modelos de horizonte infinito (MHI) y de generaciones solapadas (MGS), se comparan ambos marcos y se valoran sus resultados.

# 1. Fundamentos de la elección intertemporal

La teoría del consumidor de los temas A08 y A09 es estática, porque el agente elige entre bienes disponibles en un mismo momento. La elección intertemporal, en cambio, introduce el tiempo en la decisión, y con él tres novedades: los bienes son consumos fechados, la dotación es una corriente de rentas y el precio relativo entre el consumo de hoy y el de mañana es $1 + r$, con $r$ el tipo de interés.

## 1.1. El conjunto de elección y la ordenación de las preferencias intertemporales

Al fechar los bienes, el tiempo convierte cada bien físico en tantas mercancías como periodos. El consumo de hoy y el de mañana son así bienes distintos y sustituibles entre sí.

> Este apartado se presenta para facilitar la comprensión conceptual. En la exposición oral del tema puede abreviarse para centrar la sección en los apartados 1.2, 1.3 y 1.4.

### 1.1.1. El conjunto de elección

**Definición**: en la teoría estática del consumidor los objetos de elección son cestas de bienes. En la intertemporal, sin embargo, son *sendas de consumo*, $\mathbf{c} = (c_0, c_1, \dots, c_T) \in \mathbb{R}_+^{T+1}$, donde $c_t$ es el consumo del periodo $t$.

**Propiedades**: el conjunto tiene tres rasgos que la teoría estática no tiene:

- *Dimensión*: la fija el horizonte, que es finito, $T + 1$ periodos, en los modelos de ciclo vital, e infinito en los de dinastías, lo que exige condiciones adicionales para que el problema esté bien definido (apartado 3.1).

- *Dotación*: no es una cesta sino una corriente de rentas, $\mathbf{y} = (y_0, \dots, y_T)$, que el consumidor recibe periodo a periodo y que puede transformar en consumo de otros periodos solo a través del ahorro y el endeudamiento.

- *Orden*: la senda $(1, 2)$ no es la $(2, 1)$, porque el consumidor vive el tiempo en una dirección.

### 1.1.2. La ordenación de las preferencias

**Definición**: sobre el conjunto de sendas se define una relación de preferencias $\succsim$ que cumple los axiomas de orden y regularidad del tema A08, reinterpretados en el tiempo:

- *Completitud y transitividad*: el consumidor puede ordenar cualquier par de sendas y lo hace de forma coherente, lo que garantiza que la ordenación sea racional.

- *Continuidad*: pequeños cambios en una senda producen pequeños cambios en su valoración, lo que permite representar las preferencias por una función de utilidad.

- *Monotonía*: más consumo en cualquier periodo, sin menos en ningún otro, es preferido. En términos intertemporales, el consumidor no se sacia en ningún periodo.

- *Convexidad*: las sendas intermedias son preferidas a las extremas, lo que en el tiempo significa preferencia por la *suavización*, y de este axioma deriva el ahorro por motivo de suavización.

**Propiedades**: a esos axiomas se añaden tres propiedades intertemporales, que Koopmans (1960) axiomatizó y que llevan a la función de Samuelson del apartado siguiente:

▪ *Separabilidad*: la valoración del consumo de un periodo no depende del consumo de los demás, de modo que la utilidad de la senda es la suma de las utilidades de cada periodo. Excluye los hábitos (que el consumo de ayer condicione el valor del de hoy) y los bienes duraderos.

▪ *Estacionariedad*: la ordenación entre dos sendas no cambia si ambas se desplazan un periodo en el tiempo, lo que implica que la función de utilidad de cada periodo es la misma y que el descuento entre dos periodos consecutivos es constante.

▪ *Impaciencia*: si dos sendas solo difieren en que intercambian el consumo de dos periodos, el consumidor prefiere la que sitúa antes el mayor, lo que implica que el futuro se descuenta a una tasa positiva. La impaciencia es un juicio sobre las preferencias, no una consecuencia de la incertidumbre ni del tipo de interés, y su magnitud, frente al tipo de interés, decide el perfil de la senda de consumo.

## 1.2. La función de utilidad intertemporal y sus curvas de indiferencia

Las preferencias intertemporales se representan mediante una función de utilidad, cuyas curvas de nivel son las curvas de indiferencia intertemporales.

### 1.2.1. La función de utilidad intertemporal

**Definición**: los axiomas anteriores garantizan que las preferencias intertemporales pueden representarse mediante la *función de utilidad descontada de Samuelson* (1937):

$$U_d = \sum_{t=0}^{T} \beta^t u(c_t), \quad \beta = \frac{1}{1+\rho} \in (0,1)$$

donde $\beta$ es el *factor de descuento*, $\rho > 0$ la *tasa de descuento subjetiva*, que refleja la preferencia por el presente (mortalidad, impaciencia, sesgo de visibilidad)[2], y $u(\cdot)$ es la *función de utilidad instantánea* o función de "felicidad".

**Propiedades**:

1. *Continuidad y diferenciabilidad*: es continua y dos veces diferenciable en el consumo de cada periodo, $c_t$, lo que permite caracterizar el óptimo por condiciones de primer orden.

2. *Transformaciones*: las preferencias sobre sendas de consumo son ordinales, pero la representación aditiva exige que la utilidad instantánea sea *cardinal*, es decir, $u(\cdot)$ es única salvo transformaciones afines positivas, $a + b\,u$ con $b > 0$, como la utilidad de Bernoulli (tema A10). Una transformación monótona creciente cualquiera de $U_d$ representa las mismas preferencias intertemporales, pero pierde la forma de suma descontada.

3. *Inclinación*: es estrictamente creciente en el consumo de cada periodo, pues $u' > 0$.

4. *Forma*: es estrictamente cóncava en el consumo de cada periodo, pues $u'' < 0$, de modo que las curvas de indiferencia intertemporales son convexas y el consumidor prefiere suavizar.

[2] El descuento exponencial garantiza la consistencia dinámica de las preferencias. No explica, sin embargo, la mayor impaciencia en comparaciones a corto plazo (hoy-mañana) que a largo plazo, fenómeno captado por el descuento hiperbólico.

5. *Otras propiedades*, heredadas de los axiomas intertemporales del apartado anterior y que la forma de Samuelson hace explícitas:

- *Agotamiento de la riqueza*: por la monotonía, el consumidor no se sacia en ningún periodo y agota su riqueza a lo largo de la senda, de modo que la restricción en valor presente se cumple con igualdad y la condición terminal es $a_{T+1} = 0$.

- *Impaciencia*: $0 < \beta < 1$, el futuro pesa menos que el presente.

- *Estacionariedad*: $u(\cdot)$ y $\beta$ son invariables en el tiempo.

- *Separabilidad aditiva*: la utilidad de la senda es la suma de las utilidades de cada periodo, sin hábitos ni durabilidad[3].

- *Recursividad*: la utilidad de la senda es la de hoy más la utilidad descontada de la senda que empieza mañana, $U_d(c_0, \dots, c_T) = u(c_0) + \beta\, U_d(c_1, \dots, c_T)$, lo que permite resolver el problema hacia atrás y garantiza la consistencia dinámica.

### 1.2.2. Las curvas de indiferencia intertemporales

**Definición**: las curvas de nivel de la función de Samuelson constituyen las curvas de indiferencia intertemporales, que son decrecientes y estrictamente convexas en el plano $(c_t, c_{t+1})$.

**Propiedades**:

- *Pendiente*: su valor absoluto es la *relación marginal de sustitución intertemporal* (RMSI), el precio subjetivo del consumo presente, expresado en unidades de consumo futuro, para un nivel de utilidad dado:

$$\text{RMSI}_{t+1,t} = \frac{1}{\beta} \cdot \frac{du/dc_t}{du/dc_{t+1}}$$

Equivale al cociente de utilidades marginales corregido por la impaciencia. Es mayor cuanto más escaso es el consumo presente respecto al futuro (por la concavidad de $u$) y cuanto mayor es la impaciencia (menor $\beta$). En particular, a lo largo de una senda constante, $c_t = c_{t+1}$, vale exactamente $1/\beta = 1 + \rho$.

- *Convexidad*: su grado viene dado por la *elasticidad de sustitución intertemporal* (ESI), que mide la respuesta del consumo relativo entre dos periodos ante variaciones de su precio relativo, es decir, de la RMSI o, en el óptimo, del tipo de interés:

$$\sigma = \frac{\partial \ln(c_{t+1}/c_t)}{\partial \ln(1+r)}$$

- $\sigma$ *alta*: las curvas de indiferencia son poco curvadas y una pequeña subida del tipo de interés induce un gran desplazamiento del consumo hacia el futuro, de modo que el ahorro responde mucho al tipo de interés.

- $\sigma$ *baja*: las curvas son muy curvadas y el perfil de la senda apenas cambia con el tipo de interés, porque el consumidor prefiere suavizar a aprovechar la recompensa.

[3] Que la separabilidad resulte verosímil depende de cuánto dure cada periodo. Si los periodos son muy cortos, incluso un bien perecedero sigue prestando servicios después de consumirse y tiene, por tanto, algo de duradero.

- *Casos límite*: con $\sigma \to 0$ las curvas son de ángulo recto y el perfil del consumo no responde al tipo de interés. Con $\sigma \to \infty$, en cambio, son rectas y el consumidor concentra todo el consumo en el periodo relativamente más barato.

- *Función isoelástica*: con $u(c) = c^{1-\theta}/(1-\theta)$, la ESI es constante e igual a $1/\theta$, el inverso del coeficiente de aversión relativa al riesgo. Así, la misma curvatura que gobierna la actitud ante el riesgo (tema A10) gobierna la disposición a sustituir consumo entre periodos.

## 1.3. La restricción intertemporal de recursos

La restricción intertemporal recoge los recursos de que dispone el consumidor a lo largo de toda la senda, con mercados de capitales perfectos o con restricciones de liquidez.

### 1.3.1. Restricción intertemporal

**Definición**: la restricción de recursos puede plantearse de dos formas equivalentes, como *ecuación de estado*, periodo a periodo, o en valor presente, sobre la senda completa:

▪ *Restricción periodo a periodo o ecuación de estado*: en cada periodo, el gasto en consumo, $c_t$, más la acumulación de activos financieros, $a_{t+1} - a_t$, no puede superar la renta exógena, $y_t$, más los intereses de los activos:

$$c_t + (a_{t+1} - a_t) = y_t + r\, a_t, \qquad \forall\, t = 0, 1, \dots, T$$

donde $a_t \in \mathbb{R}$ son los activos financieros al inicio del periodo $t$ (positivos si el consumidor presta y negativos si se endeuda) y $r$ es el tipo de interés real, que se supone constante.

▪ *Condiciones de frontera*: la restricción exige dos para poder resolver problemas de optimización intertemporal.

- *Inicial*: fija la riqueza de partida, $a_0$, que puede ser nula o no según el problema.

- *Terminal*: con horizonte finito, exige $a_{T+1} = 0$, es decir, el consumidor no deja activos sin consumir ni deudas sin pagar, pues lo primero sería subóptimo y lo segundo, imposible. Con horizonte infinito, en cambio, la condición se desdobla. La de no-Ponzi excluye las deudas sin pagar y la de transversalidad, los activos sin consumir (apartado 3.1).

▪ *Restricción en valor presente*: encadenando las restricciones de todos los periodos y usando la condición terminal se obtiene una única restricción sobre la senda completa:

$$\sum_{t=0}^{T} \frac{c_t}{(1+r)^t} \leq (1+r)\, a_0 + \sum_{t=0}^{T} \frac{y_t}{(1+r)^t} \equiv W_0$$

El valor presente del consumo no puede superar la *riqueza* $W_0$, la suma de los activos iniciales y del valor presente de las rentas futuras. Esta restricción es la contrapartida de la restricción presupuestaria estática, con los precios sustituidos por factores de descuento. Así, el precio del consumo de $t$ en términos del de hoy es $1/(1+r)^t$.

**Ejemplo: de la ecuación de estado a la restricción en valor presente con dos periodos**

Sean $T = 1$ y $a_0 = 0$. La ecuación de estado da una restricción para cada periodo:

$$c_0 + (a_1 - a_0) = y_0 + r\,a_0$$
$$c_1 + (a_2 - a_1) = y_1 + r\,a_1$$

**De la ecuación de estado al valor presente**: como $a_0 = 0$, la primera restricción fija los activos al final del primer periodo en la renta no consumida, $a_1 = y_0 - c_0$, positivos si el consumidor presta y negativos si se endeuda. La condición terminal, $a_2 = 0$, reduce la segunda a $c_1 = y_1 + (1+r)\,a_1$. Al sustituir $a_1$ y dividir por $1+r$, las dos restricciones se funden en una:

$$c_0 + \frac{c_1}{1+r} = y_0 + \frac{y_1}{1+r}$$

La ecuación resultante es la restricción en valor presente con $T = 1$ y $a_0 = 0$, y se cumple con igualdad porque la condición terminal impide dejar activos sin consumir.
**Del valor presente a la ecuación de estado**: el camino inverso define los activos al final del primer periodo como la renta no consumida, $a_1 \equiv y_0 - c_0$, que es la primera restricción con $a_0 = 0$. Al sustituir $c_0 = y_0 - a_1$ en la restricción en valor presente y multiplicar por $1 + r$ se obtiene $c_1 = y_1 + (1 + r)\,a_1$, que es la segunda restricción con $a_2 = 0$.
**Comprobación numérica**: con $r = 0{,}25$, $y_0 = 80$, $y_1 = 125$ y $c_0 = 100$, el consumidor se endeuda en el primer periodo, $a_1 = -20$, y paga en el segundo unos intereses de 5 ($r\,a_1 = -5$). Su consumo del segundo periodo es entonces $c_1 = 125 + 1{,}25 \cdot (-20) = 100$, y la senda cumple también la restricción en valor presente, $100 + 100/1{,}25 = 180 = 80 + 125/1{,}25$.

Gráficamente, en el plano $(c_t, c_{t+1})$ (figura A29.1), con dos periodos y sin activos iniciales ($a_0 = 0$), la restricción es la *recta de riqueza intertemporal* (RP):

$$c_{t+1} = y_{t+1} + (1 + r)\,(y_t - c_t)$$

donde $y_t$ e $y_{t+1}$ son las rentas de cada periodo y $(1+r)\,(y_t - c_t)$ es el ahorro del primer periodo capitalizado, negativo si el consumidor se endeuda. La recta tiene tres propiedades, y el óptimo se sitúa sobre ella:

1. *Dotación*: pasa por la dotación $D = (y_t, y_{t+1})$, que es siempre asequible sin acudir al mercado de capitales. Con activos iniciales, la dotación se desplaza hacia la derecha en $(1 + r)\,a_0$.

2. *Pendiente*: es $-(1+r)$, el coste de oportunidad del consumo presente en términos de consumo futuro.

3. *Puntos de corte*: corta a los ejes en la riqueza $W_0$ (todo el consumo hoy) y en $(1+r)\,W_0$ (todo el consumo mañana).

4. *Óptimo*: el punto 0 es la tangencia con la curva de indiferencia más alta, $u_0$. En la figura queda a la derecha de la dotación, de modo que el consumidor se endeuda en $c_t^0 - y_t$. La dotación divide la recta en dos tramos:

- *Izquierda de la dotación*: el consumidor consume menos que su renta corriente y ahorra (es prestamista).

- *Derecha de la dotación*: consume más que su renta corriente y se endeuda (es prestatario).

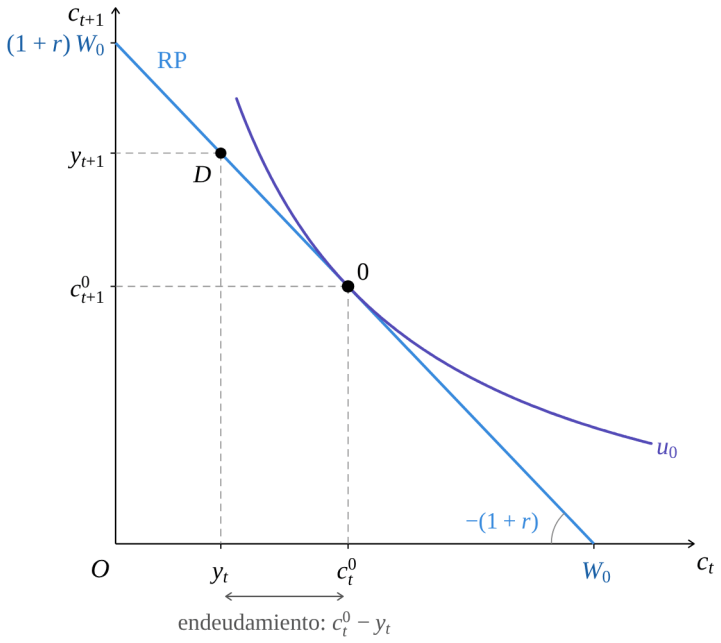



Figura A29.1: Restricción presupuestaria intertemporal en dos periodos. La recta, de pendiente $-(1+r)$, pasa por la dotación $D = (y_t, y_{t+1})$ y corta a los ejes en $W_0$ y $(1+r)\,W_0$. En el óptimo 0, a la derecha de la dotación, el consumidor se endeuda en $c^0_t - y_t$.

### 1.3.2. Restricciones de liquidez

**Definición**: la restricción anterior supone mercados de capitales perfectos, donde el consumidor puede prestar o endeudarse sin límite al mismo tipo $r$. Con mercados incompletos, en cambio, el crédito puede faltar del todo o tener un tope. Esos límites, llamados *restricciones de liquidez*, dejan fuera del conjunto asequible el tramo de la recta situado a la derecha de la dotación (o del tope).

La restricción solo es *efectiva* cuando el consumidor querría consumir más hoy de lo que su renta corriente permite, y en ese caso altera la CPO.

Gráficamente, en el plano $(c_t, c_{t+1})$ (figura A29.2), la restricción de liquidez quiebra la recta de riqueza en la dotación:

1. *Frontera del conjunto asequible*: a la izquierda de $D$ sigue siendo la recta RP, de pendiente $-(1+r)$, porque el consumidor puede ahorrar al tipo $r$. En $D$, en cambio, se hace vertical, porque no puede consumir hoy más que su renta corriente, $c_t \leq y_t$. Con un límite de endeudamiento $\bar{a} > 0$, el quiebro se desplaza hacia la derecha de la dotación, hasta $c_t = y_t + \bar{a}$.

2. *Curvas de indiferencia*: unen las sendas de igual utilidad. El óptimo sin restricción 0, la tangencia con $u_0$, está en el tramo de la recta situado a la derecha de la dotación, que deja de ser alcanzable.

3. *Óptimo*: el consumidor que querría endeudarse se queda en la dotación (punto 1), sobre una curva de indiferencia más baja, $u_1$, más inclinada que la recta en ese punto (RMSI $> 1 + r$).

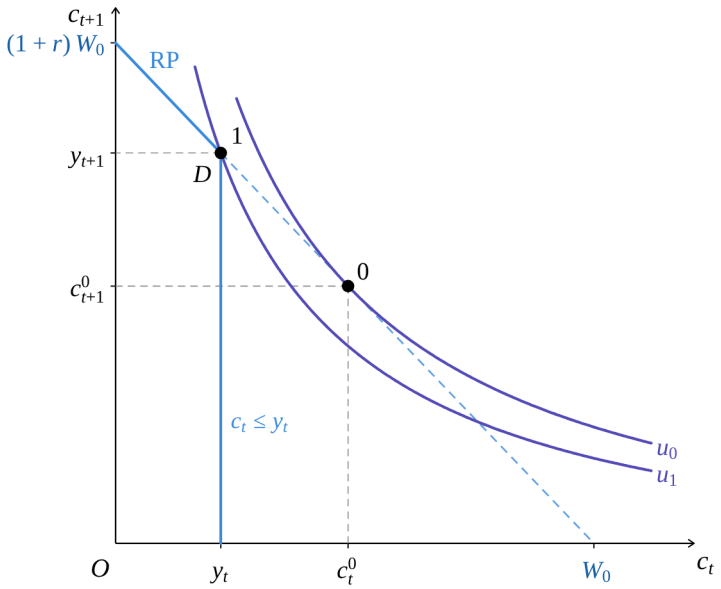



Figura A29.2: Restricción de liquidez sin posibilidad de endeudamiento. La recta de riqueza se quiebra en la dotación $D$ y el tramo a su derecha deja de ser alcanzable. El óptimo sin restricción 0 queda fuera del conjunto asequible y el consumidor se sitúa en la dotación (punto 1), con RMSI $> 1 + r$.

## 1.4. La decisión de consumo intertemporal (Fisher, 1930)

**Origen**: Irving Fisher (1930) trasladó al tiempo el aparato de la teoría del consumidor, con las tres novedades señaladas al inicio de la sección.

**Idea**: el consumidor elige la senda de consumo que maximiza su utilidad intertemporal sujeto a que el valor presente del consumo no supere al de la riqueza. El consumo de cada periodo depende así de la riqueza total, no de la renta corriente, y de la relación entre el tipo de interés y la impaciencia.

### 1.4.1. Modelo

**Planteamiento**: con la función de utilidad y la restricción intertemporal de recursos anteriores, el problema es

$$\max_{\{c_t\}} \sum_{t=0}^{T} \beta^t u(c_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \sum_{t=0}^{T} \frac{c_t}{(1+r)^t} = (1+r)\,a_0 + \sum_{t=0}^{T} \frac{y_t}{(1+r)^t} \equiv W_0$$

donde $c_t$ es la variable de decisión y las demás variables son las ya definidas.

**CPO**: con solución interior y la restricción agotada, derivando el lagrangiano respecto al consumo de cada periodo y eliminando el multiplicador entre dos periodos consecutivos, resulta

$$\text{RMSI}_{t+1,t} = \frac{1}{\beta}\frac{du/dc_t}{du/dc_{t+1}} = 1 + r \tag{1}$$

$$\frac{du}{dc_t} = \beta\,(1+r)\,\frac{du}{dc_{t+1}} = \frac{1+r}{1+\rho}\,\frac{du}{dc_{t+1}} \tag{2}$$

La condición (1) establece que la tasa subjetiva de intercambio intertemporal coincide con el coste de oportunidad del consumo presente o rentabilidad bruta del ahorro, dada por $1 + r$. La condición (2), que expresa la misma CPO de otro modo, es la *ecuación de Euler*. En el óptimo, renunciar hoy a una unidad de consumo cuesta en utilidad exactamente lo que rinde ahorrarla al tipo de interés y gastarla en el periodo siguiente, una vez descontada la impaciencia.

La misma ecuación se obtiene por *programación dinámica* (Bellman, 1957), que aprovecha la recursividad de la utilidad. Sea $V_t(a_t)$ la *función de valor*, la utilidad máxima que puede alcanzarse desde el periodo $t$ con los activos $a_t$. Por el *principio de optimalidad*, el problema de toda la senda se reduce a una sucesión de problemas de un solo periodo, que se resuelven hacia atrás desde $T$. En cada uno, el consumidor elige solo el consumo de hoy, y el futuro queda resumido en $V_{t+1}$. Esa recursión es la *ecuación de Bellman*:

$$V_t(a_t) = \max_{c_t}\{u(c_t) + \beta\,V_{t+1}(a_{t+1})\} \qquad \text{s.a.: } a_{t+1} = (1+r)\,a_t + y_t - c_t$$

Su CPO, $du/dc_t = \beta\,dV_{t+1}/da_{t+1}$, iguala la utilidad marginal de consumir hoy al valor marginal descontado de ahorrar, y la condición de la envolvente, $dV_t/da_t = (1+r)\,du/dc_t$, mide el valor marginal de la riqueza por la utilidad marginal del consumo. Adelantando esta última un periodo y sustituyéndola en la CPO se obtiene de nuevo la ecuación de Euler (2). La ventaja del método aparece con incertidumbre, pues basta con sustituir $V_{t+1}$ por su valor esperado para que la ecuación de Euler se cumpla en esperanza, como en el modelo de Hall.

**CSO**: la concavidad estricta de $u$ hace estrictamente cóncava la utilidad intertemporal y la restricción es lineal, de modo que la ecuación de Euler, junto con la restricción presupuestaria, es necesaria y suficiente.

**Solución**: de la ecuación de Euler se siguen dos propiedades de la senda óptima:

▪ *Perfil de la senda*: la ecuación de Euler ordena el consumo en el tiempo según la relación entre el tipo de interés y la tasa de preferencia temporal[4]:

- $r = \rho$: el consumo es constante en el tiempo.
- $r > \rho$: la recompensa por esperar supera a la impaciencia, la senda es creciente y, con una renta constante en el tiempo, el individuo ahorra, ofreciendo capital.
- $r < \rho$: la senda es decreciente y, con renta constante, el individuo, impaciente, demanda capital.

▪ *Suavización del consumo* (*consumption smoothing*): el individuo desea sendas estables, sean crecientes o decrecientes, lo que requiere $u'' < 0$ y mercados financieros que permitan prestar y endeudarse.

[4] Con función isoelástica, la tasa de crecimiento del consumo depende solo del tipo de interés y de las preferencias, no de la renta: $g_c \approx (r-\rho)/\theta$.

La ecuación de Euler fija el perfil de la senda, pero no su nivel, que lo fija la restricción de riqueza. Combinando ambas se obtiene la función de consumo. Con $r = \rho$ la senda es constante y, sustituyendo $c_t = c$ en la restricción, resulta

$$c_t = \frac{r}{1 + r - (1+r)^{-T}}\, W_0 \xrightarrow{T\to\infty} \frac{r}{1+r}\, W_0$$

donde la fracción que multiplica a $W_0$ es la propensión a consumir la riqueza. El consumo de cada periodo es así una fracción constante de la riqueza, la *anualidad* que esa riqueza puede financiar durante el horizonte, y no de la renta corriente.

La expresión anterior es la forma general de la que las teorías del consumo del tema A33 son casos particulares:

- *Renta permanente* (Friedman, 1957): llama renta permanente, $y^p$, a esa anualidad.

- *Ciclo vital* (Modigliani y Brumberg, 1954): distingue la riqueza en activos de la humana y sustituye el horizonte infinito por la vida restante, por lo que la propensión a consumir la riqueza crece con la edad.

- *Paseo aleatorio* (Hall, 1978): añade incertidumbre y expectativas racionales.

**Ejemplo: elección de Fisher con utilidad logarítmica**

Sean $T = 1$, $a_0 = 0$ y $u(c) = \ln c$, de modo que $u'(c) = 1/c$. La ecuación de Euler (2) fija el perfil, $c_1 = \beta(1+r)\, c_0$, y la restricción, $c_0 + c_1/(1+r) = W_0 = y_0 + y_1/(1+r)$, el nivel:

$$c_0 = \frac{W_0}{1+\beta}, \qquad c_1 = \frac{\beta\,(1+r)}{1+\beta}\, W_0$$

Con $\beta = 0{,}8$ ($\rho = 0{,}25$), $r = 0{,}25$, $y_0 = 80$ e $y_1 = 125$, la riqueza es $W_0 = 80 + 125/1{,}25 = 180$ y

$$c_0 = \frac{180}{1{,}8} = 100 = c_1$$

El consumidor se endeuda en 20 hoy y devuelve 25 mañana, porque con $r = \rho$ la senda es plana y depende de $W_0$, no del reparto temporal de la renta. Con utilidad logarítmica ($\sigma = 1$), además, la propensión a consumir la riqueza, $1/(1+\beta)$, no depende de $r$, ya que los efectos sustitución y renta ordinario se compensan y $r$ solo actúa a través de $W_0$ (efecto renta dotación).

**Desarrollo gráfico**: la decisión de consumo intertemporal se representa en el plano $(c_t, c_{t+1})$ (figura A29.3):

1. *Recta presupuestaria* (RP): recoge las sendas cuyo valor presente iguala la riqueza. Tiene pendiente $-(1+r)$ y corta a los ejes en $W_0$ y $(1+r)\, W_0$.

2. *Curva de indiferencia intertemporal* $u_0$: une las sendas de igual utilidad. Es convexa, por la concavidad de $u$, y su pendiente en valor absoluto es la RMSI.

3. *Óptimo*: el punto 0, con consumos $c_t^0$ y $c_{t+1}^0$, es la tangencia entre la curva de indiferencia más alta y la recta, donde RMSI $= 1 + r$.

- *Prestatario*: si la renta presente queda a la izquierda del óptimo ($y_t < c_t^0$), el consumidor se endeuda.

- *Prestamista*: si queda a su derecha ($y_t > c_t^0$), ahorra.

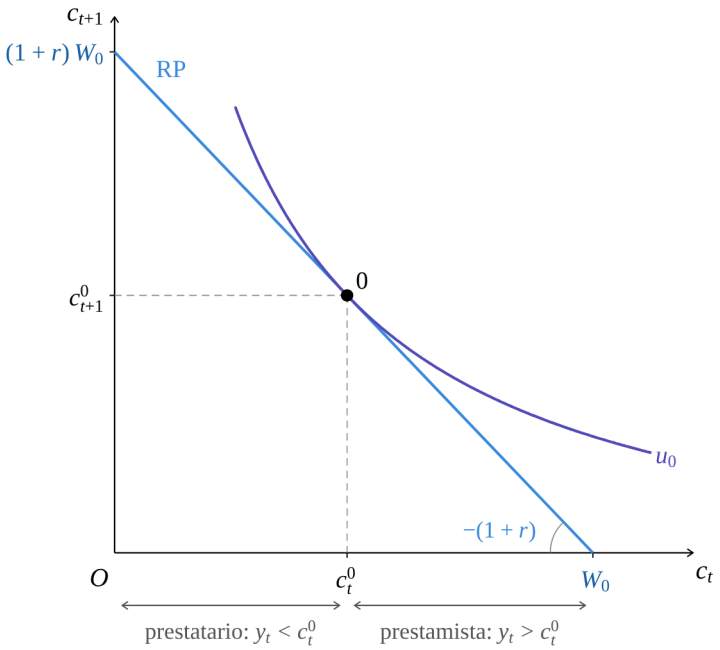



Figura A29.3: Elección intertemporal (Fisher, 1930). El óptimo se sitúa en la tangencia entre la curva de indiferencia intertemporal y la restricción presupuestaria de pendiente $-(1+r)$.

**Estática comparativa**: dadas las preferencias (la impaciencia y la ESI), se estudian los cambios en la dotación y en el tipo de interés.

*Cambios en la dotación*: un aumento de la renta de cualquier periodo, o de la riqueza inicial, desplaza la recta presupuestaria paralelamente hacia fuera, y si el consumo es normal en ambos periodos, eleva el consumo presente y el futuro. La magnitud depende de si el cambio es permanente o transitorio:

- *Aumentos permanentes*: con $r = \rho$, elevan la anualidad y, con ella, el consumo de cada periodo en la misma cuantía.

- *Aumentos transitorios*: se reparten entre todos los periodos y elevan el consumo corriente en una fracción pequeña, que es la propensión marginal a consumir de la renta transitoria.

*Cambios en el tipo de interés*: el efecto sobre el consumo presente se descompone con la ecuación de Slutsky con dotaciones:

$$\frac{\partial c_t}{\partial r} = \underbrace{\left.\frac{\partial c_t}{\partial r}\right|_{u_0}}_{\text{ES}<0} + \underbrace{(y_t - c_t)\frac{\partial c_t}{\partial y}}_{\text{ERT: signo ambiguo}}$$

- *Efecto sustitución* (ES): es inequívoco. Un aumento de $r$ encarece el consumo presente respecto al futuro, de modo que $c_t$ cae y el ahorro aumenta.

- *Efecto renta total* (ERT): combina dos efectos de signo opuesto:

  - *Efecto renta ordinario* (ERO): $-c_t\,\partial c_t/\partial y$, negativo, porque el encarecimiento del consumo presente, con la dotación dada, reduce el poder adquisitivo del consumidor.

  - *Efecto renta dotación* (ERD): $y_t\,\partial c_t/\partial y$, positivo, porque la renta presente de la dotación se revaloriza en términos de consumo futuro.

El signo del efecto total depende, por tanto, de la posición del individuo:

- *Prestamista* ($y_t > c_t$): el efecto renta es positivo y el total ambiguo.

- *Prestatario* ($y_t < c_t$): el efecto renta es negativo y el total inequívocamente negativo.

La evidencia empírica no es concluyente. La ESI estimada es baja y el ahorro agregado responde poco al tipo de interés, de modo que ninguno de los dos efectos se impone con claridad[5].

Gráficamente, en el plano $(c_t, c_{t+1})$ (figura A29.4), un cambio en el tipo de interés hace pivotar la recta sobre la dotación. Si $r$ sube ($\Delta r > 0$), la recta pasa de $\mathrm{RP}_0$ a $\mathrm{RP}_1$, más inclinada, lo que reduce las posibilidades de quien se endeuda y amplía las de quien presta. En la figura, para un prestamista, el efecto sustitución lleva del punto 0 al $0'$ a lo largo de $u_0$, sobre la recta compensada $\mathrm{RP}_c$, paralela a $\mathrm{RP}_1$, y el efecto renta total, del $0'$ al 1, sobre $u_1$. Como domina el efecto sustitución, el consumo presente cae de $c_t^0$ a $c_t^1$.

[5] En el agregado, el *stock* de riqueza es generalmente positivo, es decir, los hogares son, en promedio, más prestamistas que prestatarios (Romer, 2012, cap. 8).

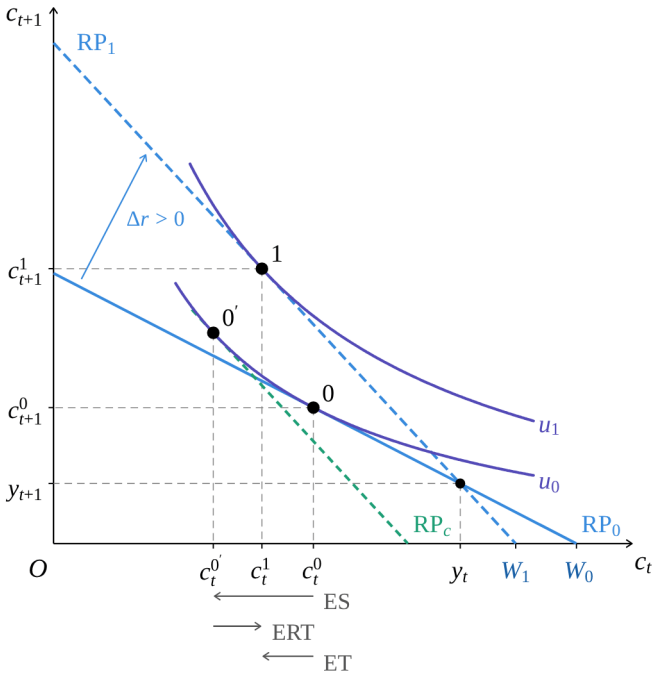



Figura A29.4: Subida del tipo de interés para un prestamista. El efecto sustitución reduce el consumo presente y el efecto renta total lo aumenta (para un prestatario, ambos lo reducen).

### 1.4.2. Extensión: consumo con incertidumbre, el paseo aleatorio (Hall, 1978)

> El paseo aleatorio de Hall, sus supuestos y su contrastación se desarrollan en el tema A33, apartado 2.4. Aquí se presenta como extensión del problema de Fisher.

**Origen**: Robert Hall (1978) resolvió el problema de Fisher bajo incertidumbre y expectativas racionales, y obtuvo una predicción contrastable.

**Idea**: si el consumidor maximiza la utilidad esperada y usa toda la información disponible, el consumo de hoy ya incorpora todo lo que se sabe sobre la renta futura, de modo que solo la información nueva puede alterarlo. El consumo sigue, por ello, un paseo aleatorio.

**Planteamiento**: con expectativas racionales, utilidad cuadrática[6] y $r = \rho = 0$, el consumidor maximiza la utilidad esperada de su senda de consumo sujeto a su riqueza esperada:

$$\max_{c_{t+j}} \mathbb{E}_t \left[ \sum_{j=0}^{T-t} \left( c_{t+j} - \frac{\eta}{2} c_{t+j}^2 \right) \right], \ \eta > 0$$

$$\text{s.a.:} \begin{cases} \sum_{j=0}^{T-t} \mathbb{E}_t[c_{t+j}] \leq a_t + \sum_{j=0}^{T-t} \mathbb{E}_t[y_{t+j}] \\ a_{T+1} = 0 \end{cases}$$

[6] La utilidad cuadrática tiene dos inconvenientes propios. La utilidad marginal puede ser negativa a partir de cierto nivel de consumo, y la aversión absoluta al riesgo crece con la renta.

donde $\eta$ es el parámetro de la utilidad cuadrática, $a_t$ la riqueza no humana al inicio del periodo, $y$ la renta, $T-t+1$ el número de periodos de vida que restan y $\mathbb{E}_t$ la expectativa condicionada a la información disponible en $t$.

**CPO**: la utilidad marginal es lineal, $u'(c) = 1 - \eta c$, de modo que la ecuación de Euler esperada, $du/dc_t = \mathbb{E}_t[du/dc_{t+1}]$, se convierte en

$$\mathbb{E}_t[c_{t+1}] = c_t \quad \Rightarrow \quad c_{t+1} = c_t + \varepsilon_{t+1}, \qquad \mathbb{E}_t[\varepsilon_{t+1}] = 0$$

donde $\varepsilon_{t+1}$ es la innovación, es decir, la nueva información sobre la riqueza esperada que llega en $t+1$. El consumo presente es, por tanto, el mejor predictor del consumo futuro.

**CSO**: la utilidad cuadrática es estrictamente cóncava y la restricción es lineal, de modo que la CPO es necesaria y suficiente en el tramo en que la utilidad marginal es positiva ($c_t < 1/\eta$).

**Solución**: llevada la CPO a la restricción presupuestaria, el individuo consume en cada periodo la anualidad de su riqueza esperada:

$$c_t = \frac{1}{T-t+1}\left(a_t + \sum_{j=0}^{T-t} \mathbb{E}_t[y_{t+j}]\right)$$

La implicación es doble:

- *Imprevisibilidad*: la variación del consumo no debería ser predecible con ninguna información pasada.
- *Política económica*: solo afecta al consumo si genera sorpresas o altera la riqueza esperada.

**Conclusiones**:

- *Aspectos positivos*: lleva las expectativas racionales a la teoría del consumo y convierte la hipótesis de la renta permanente en una restricción contrastable, la ecuación de Euler, que no exige estimar la función de consumo. Por ello, ha sido el punto de partida de la literatura empírica moderna sobre el consumo (tema A33, apartado 2.4).
- *Limitaciones*: la evidencia rechaza parcialmente esa imprevisibilidad, tanto por exceso de sensibilidad a la renta corriente, atribuido a restricciones de liquidez y a consumidores que siguen reglas simples, como por exceso de suavidad (tema A33, apartado 3.1).

### 1.4.3. Extensión: restricciones de liquidez (Deaton, 1991)

> Las restricciones de liquidez y su contrastación empírica se desarrollan en el tema A33, apartado 3.2.2. Aquí se presentan como modificación del problema de Fisher.

**Origen**: Angus Deaton (1991) incorporó al problema intertemporal la imposibilidad de endeudarse, para dar cuenta del exceso de sensibilidad del consumo a la renta que el modelo de Hall no puede explicar.

**Idea**: si el consumidor no puede endeudarse, o solo hasta un límite, la restricción de liquidez puede ser efectiva, y entonces su consumo deja de depender de la riqueza y pasa a seguir a la renta corriente. Además, la mera posibilidad de que la restricción sea efectiva en el futuro altera el comportamiento presente.

**Razonamiento**: el problema añade a las restricciones del modelo un límite de endeudamiento en cada periodo,

$$a_{t+1} \geq -\bar{a}$$

donde $\bar{a} \geq 0$ es la deuda máxima que el consumidor puede contraer ($\bar{a} = 0$ si no puede endeudarse). Si el límite llega a operar, la igualdad de Euler se rompe y se convierte en

$$\frac{du}{dc_t} > \beta\,(1+r)\,\frac{du}{dc_{t+1}} = \frac{1+r}{1+\rho}\,\frac{du}{dc_{t+1}}$$

lo que significa que el consumidor querría trasladar consumo del futuro al presente y no puede. Consume, por tanto, todos los recursos de que dispone en el periodo, y su consumo sigue a la renta corriente.

Cuando la restricción no es efectiva pero puede serlo en el futuro, el consumidor prudente acumula un *colchón* de activos (el ahorro *buffer-stock* de Deaton) para no verse forzado a reducir el consumo ante una caída de la renta. Además, su propensión marginal a consumir la renta transitoria es mayor que la del modelo de Fisher, porque, cerca del límite, cada unidad adicional de recursos reduce mucho la necesidad de ahorro precautorio y se consume en mayor proporción.

**Conclusiones**:

▪ *Aspectos positivos*: las restricciones de liquidez reconcilian el modelo intertemporal con la evidencia, porque explican tres hechos:

- *Consumo*: por qué el consumo agregado sigue a la renta corriente más de lo que la renta permanente predice.
- *Política fiscal*: por qué la política fiscal transitoria tiene efectos reales sobre el consumo de los hogares restringidos.
- *Distribución*: por qué la distribución de la riqueza importa para el multiplicador.

▪ *Limitaciones*: el coste de incorporarlas es la pérdida de la solución cerrada, pues con restricciones ocasionalmente efectivas el problema solo se resuelve numéricamente.

### 1.4.4. Extensión: bienes duraderos (Mankiw, 1982)

> El consumo de bienes duraderos y sus reglas de ajuste se desarrollan en el tema A33, apartado 3.4.

**Origen**: Gregory Mankiw (1982) extendió el modelo de Hall al caso en que el bien consumido no se agota en el periodo, y mostró que las predicciones cambian.

**Idea**: un bien duradero no proporciona utilidad cuando se compra sino mientras dura. Por eso, lo que entra en la función de utilidad es el *flujo de servicios* del *stock* de duraderos, no el gasto en ellos, que es la inversión que mantiene y amplía ese *stock*.

**Razonamiento**: sea un *stock* de duraderos $d_t$ que se deprecia a la tasa $\delta$ y se repone con el gasto en duraderos, $x_t$:

$$x_t = d_t - (1-\delta)\, d_{t-1}$$

donde $d_{t-1}$ es el *stock* heredado del periodo anterior. La utilidad del periodo, $u[c_t, s(d_t)]$, depende del consumo no duradero, $c_t$, y de los servicios del *stock*, $s(d_t)$, de modo que el problema es el del modelo, ahora con dos bienes. La condición de primer orden para los duraderos iguala la utilidad marginal de sus servicios, en unidades de consumo no duradero, al *coste de uso* del duradero, $p_d\,(r + \delta - \pi_d)$, con $p_d$ su precio relativo y $\pi_d$ su revalorización esperada. Ese coste tiene la misma forma que el coste de uso del capital de la empresa (apartado 2.4.3).

La suavización se aplica al *stock*, no al gasto. El gasto, que es la variación del *stock*, es por ello mucho más volátil que el consumo de no duraderos y responde con fuerza al tipo de interés y al crédito. Con utilidad cuadrática y expectativas racionales, el *stock* sigue un paseo aleatorio, como el consumo en Hall, y la primera diferencia del gasto es un proceso de medias móviles de orden uno,

$$\Delta x_t = \varepsilon_t - (1-\delta)\, \varepsilon_{t-1}$$

donde $\varepsilon_t$ es la innovación del *stock*. Con costes fijos de ajuste (comprar un coche no admite fracciones), la compra sigue además una regla $(S, s)$ (Grossman y Laroque, 1990), en la que el consumidor deja que el *stock* se deprecie hasta un umbral y entonces lo repone de una vez.

**Conclusiones**:

- *Aspectos positivos*: los duraderos explican la parte más volátil del consumo agregado y su sensibilidad a las condiciones financieras. También unifican la teoría del consumo con la de la inversión, porque el hogar que compra una vivienda o un automóvil resuelve el mismo problema que la empresa que compra una máquina.

- *Limitaciones*: empíricamente, el gasto en duraderos no muestra la fuerte reversión a la media que el modelo predice, sino que se comporta como si la depreciación fuera casi total ($\delta \approx 1$), igual que el de un bien no duradero (tema A33).

## 1.5. Conclusiones

**Aspectos positivos**: el modelo de Fisher traslada el aparato del consumidor al tiempo, con la ecuación de Euler como condición de óptimo. De él derivan la separación entre decisiones de consumo y de inversión y las teorías modernas del consumo (renta permanente, ciclo vital, paseo aleatorio).

**Limitaciones**: supone mercados de capitales perfectos y preferencias aditivamente separables, y con incertidumbre el resultado de Hall exige además utilidad cuadrática. Las extensiones con restricciones de liquidez y bienes duraderos relajan lo primero y lo segundo, y son el puente hacia la teoría de la demanda de consumo del tema A33.

# 2. Decisiones intertemporales de la empresa

La decisión de inversión de la empresa tiene la misma estructura que la decisión de consumo intertemporal del apartado 1 y que la decisión estática de contratación de capital del tema A11.

La diferencia con respecto a la decisión de consumo intertemporal está en el objetivo y en la restricción. Por un lado, la empresa no maximiza una utilidad descontada a una tasa subjetiva, sino su valor, descontado al tipo de interés de mercado. Por otro lado, su restricción no es una corriente de rentas dada, sino una tecnología que convierte la inversión de hoy en capital productivo mañana. El consumidor elige cuándo consumir una riqueza dada, mientras que la empresa elige cuánta riqueza crear.

La diferencia con respecto a la decisión estática de contratación de capital está en el tiempo. En el tema A11 el capital es un factor que se alquila cada periodo al precio de mercado. Aquí, en cambio, es un *stock* que se acumula, se deprecia y cuesta ajustar, así que la decisión de hoy condiciona las de mañana y la empresa ha de anticipar toda la senda futura de precios y productividades. Esa dependencia convierte la contratación de capital en una decisión de inversión.

## 2.1. El conjunto de elección y el objetivo de la empresa

La empresa elige sendas de variables de control y de estado para maximizar su valor.

### 2.1.1. El conjunto de elección

**Definición**: en la teoría estática de la empresa los objetos de elección son combinaciones de factores, que la empresa contrata en los mercados de factores productivos. En la intertemporal, en cambio, son *sendas de inversión y de empleo*, $\{I_t, L_t\}_{t=0}^{T}$, es decir, las cantidades de bienes de capital y de trabajo que la empresa contrata en esos mercados en cada periodo. La inversión $I_t$ y el empleo $L_t$ son las variables de *control*, los flujos que la empresa decide cada periodo. El capital $K_t$ es la variable de *estado*, el *stock* que resulta de esas decisiones, que la empresa no elige directamente sino a través de la inversión acumulada.

**Propiedades**: el conjunto tiene tres rasgos:

- *Dimensión*: la fija el horizonte de planificación de la empresa, que es finito, $T+1$ periodos, si tiene fecha de cierre, e infinito en caso contrario.

- *Mercados*: los mercados de factores son perfectos, de modo que la empresa puede contratar cualquier cantidad de trabajo y de bienes de capital a los precios de cada periodo, $w_t$ y $P_{K,t}$, que toma como dados.

- *Orden*: las decisiones de los distintos periodos no son independientes, porque el capital de cada periodo resulta de la inversión de los anteriores.

### 2.1.2. El objetivo de la empresa

**Definición**: la empresa maximiza su valor, el valor presente de la corriente de beneficios futuros, y no el beneficio de cada periodo.

**Propiedades**: con mercados de capitales perfectos, el objetivo tiene dos propiedades:

- *Separación*: por el teorema de separación de Fisher, todos los propietarios, sean cuales sean sus preferencias, quieren que la empresa maximice su valor, porque es lo que maximiza la riqueza de los propietarios.

- *Descuento*: la tasa de descuento de la empresa es, por ello, el tipo de interés de mercado, y no una tasa subjetiva como la del consumidor.

## 2.2. La función de valor y el precio sombra del capital

El objetivo de la empresa se representa mediante la función de valor, cuya derivada respecto al capital es su precio sombra.

### 2.2.1. La función de valor

**Definición**: el objetivo, el valor presente de los beneficios netos del gasto en inversión, $\Pi_t = P_t f(K_t, L_t) - w_t L_t - P_{K,t} I_t$, se escribe como la *función de valor*:

$$V_0 = \sum_{t=0}^{T} \frac{\Pi_t}{(1+r)^t} \qquad \text{o, en tiempo continuo,} \qquad V_0 = \int_0^T e^{-rt}\, \Pi(t)\, dt$$

Equivale a la función de utilidad descontada de Samuelson, con el beneficio en lugar de la utilidad instantánea y $1/(1+r)$ en lugar de $\beta$. Evaluada en la senda óptima de inversión, pasa a depender únicamente del capital heredado y de los precios, $V_t = V_t(K_t)$, y es entonces una función de valor en sentido estricto, como las de los problemas estáticos.

**Propiedades**:

1. *Continuidad y diferenciabilidad*: es continua y diferenciable en el beneficio de cada periodo. Evaluada en la senda óptima, es diferenciable en el capital cuando el beneficio es cóncavo y los costes de ajuste convexos, lo que permite definir el precio sombra.

2. *Transformaciones*: es una función cardinal, medida en unidades monetarias, de modo que, a diferencia de la utilidad de Samuelson (apartado 1.2.1), no admite transformaciones. Una subida equiproporcional de todos los precios eleva el valor en la misma proporción sin alterar la senda óptima, como en la función de máximo beneficio (tema A16, apartado 1.3.4).

3. *Inclinación*: es creciente en el beneficio de cada periodo y, si los beneficios futuros son positivos, decreciente en el tipo de interés.

4. *Forma*: es lineal en los beneficios de cada periodo. Evaluada en la senda óptima, es cóncava en el capital cuando el beneficio es cóncavo y los costes de ajuste convexos (CSO del modelo de inversión), de modo que el precio sombra no aumenta con el capital.

5. *Otras propiedades*:

▪ *Separabilidad aditiva*: es la suma de los beneficios descontados de cada periodo, sin interacciones entre ellos, como la función de Samuelson.

▪ *Envolvente*: en la senda óptima, su derivada respecto al capital es el precio sombra, que coincide con el multiplicador $q$ del hamiltoniano del modelo de inversión.

▪ *Observabilidad*: con mercados de capitales perfectos coincide con el valor de mercado de la empresa, que es observable, a diferencia de su derivada, la $q$ marginal.

### 2.2.2. El precio sombra del capital

**Definición**: el *precio sombra* o valor marginal del capital, $q_t = \partial V_t / \partial K_t$, es el equivalente del valor marginal de la riqueza del consumidor, $dV_t/da_t$ (apartado 1.4.1), y muestra cuánto aumenta el valor de la empresa una unidad adicional de capital, medido en unidades de beneficio descontado.

**Propiedades**: resume toda la información sobre el futuro que necesita la decisión de invertir y unifica los modelos de inversión. Así, en Jorgenson coincide con el precio del bien de capital, en Tobin es la $q$ marginal y en Hotelling (Anexo A.1) es la renta de escasez del recurso.

## 2.3. La restricción tecnológica intertemporal

**Definición**: a la función de producción del periodo, $Y_t = f(K_t, L_t)$, se añade la *ley de acumulación del capital*, que es la ecuación de estado de la empresa:

$$K_{t+1} = (1-\delta)\,K_t + I_t \qquad \text{o, en tiempo continuo,} \qquad \dot{K} = I - \delta K$$

donde $\delta$ es la tasa de depreciación. La ley de acumulación hace el papel que en el consumidor hacía la restricción periodo a periodo, pues liga la decisión de hoy con el estado de mañana. Como en el consumidor, el problema exige además dos condiciones de frontera, una inicial, el capital heredado $K_0$, y una terminal:

▪ *Horizonte finito*: el capital final se liquida o carece de valor.

▪ *Horizonte infinito*: la condición de transversalidad, $\lim_{t\to\infty} e^{-rt} q(t) K(t) = 0$[7], excluye acumular capital cuyo valor descontado no se recupera nunca, y hace el papel de la condición de transversalidad del consumidor (apartado 3.1).

**Propiedades**: los distintos modelos difieren en dos supuestos sobre la ley de acumulación:

▪ *Reversibilidad*: si el capital puede venderse al mismo precio al que se compró, lo que separa los modelos que siguen de los de inversión irreversible.

▪ *Costes de ajuste*: si su ajuste tiene costes además del precio de los bienes de capital, lo que distingue el modelo de Jorgenson del de la $q$ de Tobin.

[7] La condición de transversalidad es la holgura complementaria de la restricción de no negatividad del capital llevada al límite (tema A43, apartado 3.2).

## 2.4. La decisión de inversión

La decisión de inversión aplica a la empresa la teoría del control óptimo, con la inversión como senda de control.

> Las teorías de la demanda de inversión se desarrollan en el tema A34. Aquí se presentan como aplicaciones de la elección intertemporal.

### 2.4.1. Modelo

**Planteamiento**: el problema de la empresa se resuelve con el *principio del máximo de Pontryagin*:

$$\max_{\{I(t),\,L(t)\}} V_0 = \int_0^{\infty} e^{-rt}\,\Pi(K,I,L)\,dt \qquad \text{s.a.:} \begin{cases} \dot{K} = I - \delta K \\ K_0 \text{ dado} \end{cases}$$

donde $I$ y $L$ son las variables de control, $K$ la variable de estado, $K_0$ el capital heredado y $\Pi(K,I,L)$ el beneficio instantáneo neto del gasto en inversión. El hamiltoniano en valor corriente, con $q$ el precio sombra del capital, es

$$H = \Pi(K,I,L) + q\,(I - \delta K)$$

**CPO**: sus condiciones de primer orden son tres:

- *Control*: la de las variables de control, $\partial H/\partial I = 0$, que iguala el coste marginal de invertir al precio sombra $q$.
- *Estado*: la de la variable de estado, $\partial H/\partial K = rq - \dot{q}$, que es la ecuación de Euler de la empresa y dice cómo evoluciona el precio sombra.
- *Transversalidad*: la condición ya enunciada, que fija el nivel de la senda.

**CSO**: las condiciones del principio del máximo son necesarias, y son además suficientes con beneficios cóncavos y costes de ajuste convexos[8].

Los modelos siguientes son casos particulares de este problema. La regla de Hotelling (1931), que lo aplica a la extracción de un recurso agotable, se desarrolla en el Anexo A.1.

[8] Son suficientes, junto con la condición de transversalidad, si el hamiltoniano es conjuntamente cóncavo en el estado y el control (Mangasarian, 1966) o, más en general, si el hamiltoniano maximizado es cóncavo en la variable de estado (Arrow, 1968).

### 2.4.2. Extensión: el consumidor-productor (Fisher, 1930; Hirshleifer, 1958)

**Origen**: Fisher (1930) extendió su análisis del consumidor al caso en que la renta no es una dotación sino el resultado de una decisión de producción, y Jack Hirshleifer (1958) lo formalizó como teoría de la inversión.

**Idea**: un agente que es a la vez consumidor y productor decide cuánto consumir, cuánto ahorrar y cuánto invertir. Si tiene acceso a un mercado de capitales, la decisión de inversión se separa de la de consumo y depende solo del tipo de interés.

**Supuestos**:

▪ *Entorno*: tiempo discreto, horizonte finito y mundo neoclásico (información perfecta, no atesoramiento, precio-aceptación). Los precios están normalizados, de modo que no se tiene en cuenta la inflación.

▪ *Agentes*: un Robinson Crusoe o consumidor-productor, dueño de los factores y de la tecnología y perfectamente racional, que toma la triple decisión de consumo, ahorro e inversión.

▪ *Tecnología*: se produce un solo bien de consumo con un único factor, el capital, mediante una función de producción con productividad marginal positiva ($f' > 0$) y decreciente ($f'' < 0$), es decir, con una frontera de posibilidades de producción intertemporal cóncava. La acumulación de capital es reversible.

▪ *Restricciones*: existe un capital preexistente $k_0$ que permite obtener la renta del primer periodo, y en el último periodo no quedan activos financieros ni productivos.

**Planteamiento**: el consumidor-productor maximiza su utilidad intertemporal eligiendo cuánto consumir y cuánto invertir en cada periodo, sujeto a la tecnología y, si existe, al mercado de capitales:

$$\max_{\{c_t,\, i_t\}} \sum_{t=0}^{T} \beta^t u(c_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} c_t + i_t + (a_{t+1} - a_t) = f(k_t) + r\, a_t & \text{(restricción de recursos)} \\ k_{t+1} = (1-\delta)\, k_t + i_t & \text{(ley de acumulación)} \\ k_0 \text{ dado}, \quad k_{T+1} = a_{T+1} = 0 & \text{(condiciones de frontera)} \end{cases}$$

donde $f(k_t)$ es la renta del periodo, que ahora es endógena, e $i_t$ la inversión en capital productivo. Sin acceso a mercados financieros, $a_t = 0$ en todos los periodos y la única forma de trasladar consumo al futuro es invertir.

**CPO**: en ese caso, con solución interior y sustituyendo las restricciones en la función objetivo y derivando respecto a $k_{t+1}$, la condición de primer orden iguala la relación marginal de sustitución intertemporal a la *relación marginal de transformación intertemporal*, el rendimiento bruto de invertir una unidad de consumo:

$$\text{RMSI}_{t+1,t} = 1 + \frac{df}{dk_{t+1}} - \delta \equiv \text{RMTI}$$

Las decisiones de consumo y de inversión están ligadas, porque el consumidor invierte hasta que el rendimiento marginal del capital iguala su impaciencia corregida por la escasez, y para consumir más mañana ha de renunciar a consumo hoy.

Gráficamente, la elección del consumidor-productor sin mercado de capitales se representa en el plano $(c_t, c_{t+1})$ (figura A29.5):

1. *Frontera de posibilidades de producción intertemporal* (FPP): es cóncava, porque la productividad marginal del capital es decreciente, y su pendiente en valor absoluto es la relación marginal de transformación intertemporal (RMTI). Sin mercado de capitales es a la vez la frontera de posibilidades de consumo.

2. *Curva de indiferencia* $u_0$: es convexa, y su pendiente en valor absoluto es la RMSI.

3. *Óptimo*: el punto 0, con consumos $c_t^0$ y $c_{t+1}^0$, es la tangencia entre la curva de indiferencia más alta y la frontera, donde RMSI = RMTI.

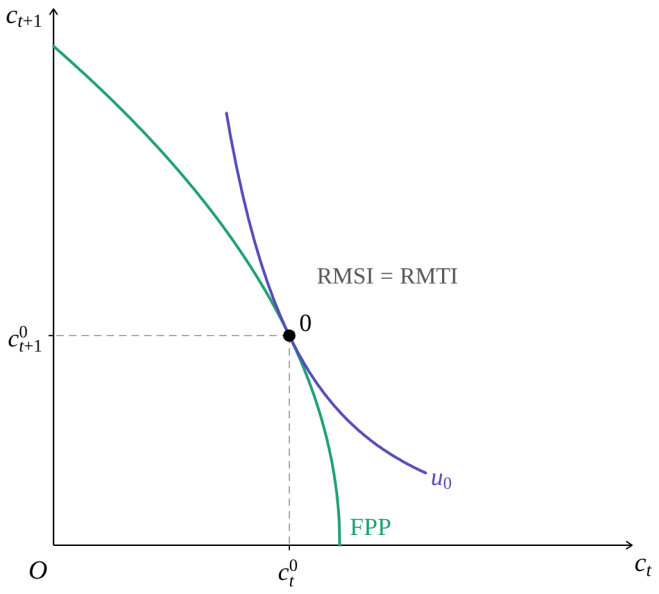



Figura A29.5: Consumidor-productor sin mercado de capitales. El óptimo es la tangencia entre la frontera de producción intertemporal y la curva de indiferencia.

*Con acceso a mercados financieros*: si el consumidor-productor puede prestar y endeudarse al tipo $r$, las condiciones de primer orden se separan en dos:

$$\underbrace{\frac{df}{dk_{t+1}}}_{\text{Pmg}_k \text{ neta}} - \delta = r \qquad \text{y} \qquad \text{RMSI}_{t+1,t} = 1 + r$$

La primera fija la inversión. Se invierte hasta que el rendimiento neto del capital iguala al tipo de interés, con independencia de las preferencias. La segunda fija el consumo, pues, dado el valor presente de la renta que esa inversión genera, el consumidor elige su senda como en el apartado 1.4.

El *teorema de separación de Fisher* (1930) recoge este resultado. Con mercados de capitales perfectos, la decisión de inversión, que maximiza la riqueza, es independiente de la de consumo, que la reparte en el tiempo. La segunda parte del teorema establece que la inversión es independiente de cómo se financie, con ahorro propio o con endeudamiento.

**CSO**: concavidad de $u$ y de $f$ ($f'' < 0$), por la que la frontera intertemporal es cóncava y las curvas de indiferencia convexas, de modo que las tangencias descritas son máximos.

**Desarrollo gráfico**: el teorema de separación de Fisher se representa en el plano $(c_t, c_{t+1})$ (figura A29.6):

1. *Frontera de posibilidades de producción intertemporal* (FPP): es la misma de la figura A29.5.

2. *Recta de riqueza* $\mathrm{RP}_1$: tiene pendiente $-(1+r)$ y es tangente a la frontera en el óptimo de producción P, con rentas $y_t$ e $y_{t+1}$, donde $\mathrm{RMTI} = 1+r$ y el valor presente de la producción, $W_1$, es máximo.

3. *Óptimo de consumo*: el punto 1, con consumos $c_t^1$ y $c_{t+1}^1$, es la tangencia entre esa misma recta y la curva de indiferencia $u_1$, donde $\mathrm{RMSI} = 1+r$. La distancia entre P y 1 es el préstamo $(c_t^1 - y_t)$ o el ahorro con que el mercado de capitales financia la diferencia.

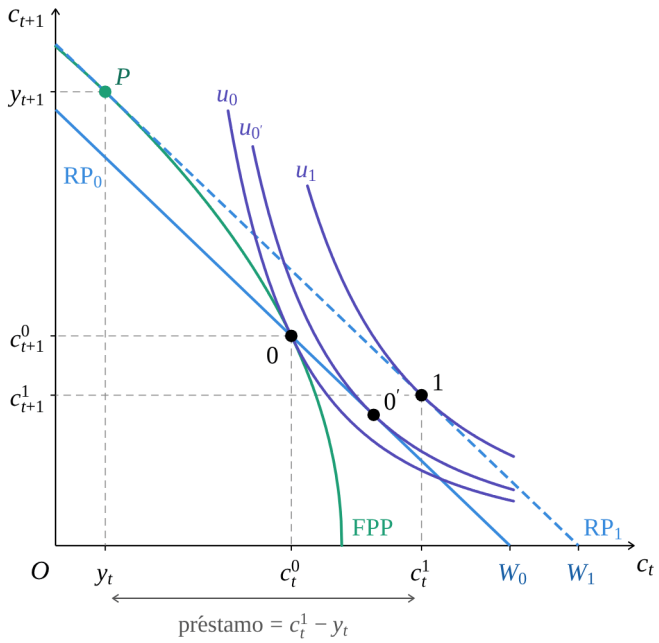



Figura A29.6: Teorema de separación de Fisher. Con mercado de capitales, la decisión de inversión (maximizar el valor presente) se separa de la de consumo.

**Estática comparativa**: el acceso al mercado permite alcanzar una curva de indiferencia más alta que sin él, porque la recta de riqueza, tangente a la frontera en el óptimo de producción, queda por encima de ella en todos los demás puntos. El mercado amplía así el conjunto de consumo alcanzable, y la mejora es estricta salvo que ambos óptimos coincidan. La ganancia puede descomponerse en dos etapas:

1. *Ganancia de intercambio*: manteniendo la producción en el óptimo autárquico 0, la recta de mercado $\mathrm{RP}_0$ que pasa por él permite alcanzar el punto $0'$ sobre $u_{0'}$.

2. *Ganancia de producción*: desplazando además la producción hasta P, donde la frontera es tangente a $\mathrm{RP}_1$, se alcanza el punto 1 sobre $u_1$.

**Conclusiones**:

- *Aspectos positivos*: el teorema de separación fundamenta la teoría financiera de la empresa. Permite valorar los proyectos de inversión por su valor actual neto al tipo de interés de mercado, con independencia de las preferencias de los propietarios, y justifica que la empresa maximice su valor de mercado (tema B02).

- *Limitaciones*: la separación exige mercados de capitales perfectos, con un único tipo de interés para prestar y endeudarse y sin restricciones de crédito. Cuando fallan, las decisiones de inversión vuelven a depender de la liquidez y de las preferencias de los propietarios, punto de partida de la teoría financiera con imperfecciones (tema B03).

### 2.4.3. Extensión: demanda de inversión sin costes de ajuste (Jorgenson, 1963)

**Origen**: Dale Jorgenson (1963) dedujo la función de demanda de inversión del problema con el que la empresa elige su senda óptima de capital.

**Idea**: la empresa quiere tener el capital cuya productividad marginal coincide con su coste de uso, y la inversión es el ajuste hacia ese *stock*. Ambos, capital deseado e inversión, varían por tanto con los precios relativos de los factores.

**Supuestos**: los de la restricción tecnológica intertemporal, con tres precisiones:

- *Entorno*: horizonte infinito, tiempo continuo y tipo de interés real constante y exógeno.

- *Tecnología*: función de producción neoclásica de buen comportamiento, depreciación constante $\delta$ y ausencia de costes de ajuste.

- *Mercado de factores*: oferta de bienes de capital perfectamente elástica.

**Planteamiento**: la empresa maximiza su valor con la inversión bruta $I$ y el empleo $L$ como variables de control:

$$\max_{\{I,L\}} V = \int_0^\infty e^{-rt}[P \cdot f(K,L) - wL - P_K I]\,dt \quad \text{s.a.:} \begin{cases} \dot{K} = I - \delta K \\ K_0 \text{ dado} \end{cases}$$

donde $K$ es el *stock* de capital (la variable de estado)[9], $P$ el precio del bien, $w$ el salario nominal y $P_K$ el precio del bien de capital.

**CPO**: como no hay costes de ajuste, el precio sombra del capital coincide con su precio de mercado, $q = P_K$. Las condiciones de primer orden del hamiltoniano se reducen así a dos condiciones estáticas:

[9] En tiempo continuo se omite el argumento temporal: $K \equiv K(t)$, $I \equiv I(t)$, $L \equiv L(t)$ y $q \equiv q(t)$; se escribe solo en las condiciones inicial y terminal.

$$\underbrace{P\frac{\partial f}{\partial L}}_{\text{VPmg}_L} = w$$

$$\underbrace{P\frac{\partial f}{\partial K}}_{\text{VPmg}_K} = P_K\,(r+\delta) \equiv r_K$$

La primera iguala el valor de la productividad marginal del trabajo al salario nominal. La segunda iguala el del capital a su *coste de uso*, $r_K$, que desglosa el precio de adquisición ($P_K$), el coste de oportunidad de los fondos ($r$) y la depreciación ($\delta$).

La inversión neta sale de comparar el capital deseado, $K^*$, con el capital instalado. Cuando ambos coinciden, la empresa solo repone lo que se deprecia. Cuando $K^*$ queda por debajo, desinvierte, y cuando lo supera, la inversión neta es positiva.

**CSO**: sin costes de ajuste el problema es estático en cada instante y la CSO es la de la empresa competitiva, con $f$ estrictamente cóncava en $(K, L)$.

**Extensiones**: con inflación del capital, el coste de uso incluye $\pi_K$ y pasa a ser $r_K = P_K(r+\delta-\pi_K)$, con $\pi_K$ medida en términos reales (en exceso de la inflación general), pues $r$ es real. Hall y Jorgenson (1967) muestran que, con imposición (impuesto sobre sociedades y desgravaciones fiscales), $r_K = \chi_t P_K(r+\delta)$, donde $\chi_t$ depende de parámetros como el tipo de gravamen sobre beneficios. Del coste de uso así ampliado se derivan las implicaciones de política económica del modelo:

- *Política monetaria*: es efectiva en la medida en que afecte al tipo de interés real (a corto plazo, con rigideces nominales) y, con él, al coste del capital.

- *Política fiscal*: tiene efectos diferenciados según el instrumento. Si se basa en el gasto público, el efecto es adverso en la medida en que ese gasto eleve el tipo de interés real, un canal de equilibrio general, pues el modelo toma $r$ como dado. En cambio, si se basa en deducciones impositivas, el efecto es favorable, porque las deducciones reducen el coste de uso.

**Conclusiones**:

- *Aspectos positivos*: es el modelo canónico para explicar la inversión.

- *Limitaciones*:

  - *Costes de ajuste*: al no haberlos, el futuro de las variables relevantes no condiciona la decisión de la empresa.

  - *CPO estática*: no admite sustitución intertemporal del capital.

  - *Discontinuidad*: ante un cambio discreto de $r$ la inversión puede dar saltos, algo que la producción agregada, al ser limitada, no permite (no cabe un "salto" infinito).

### 2.4.4. Extensión: costes de ajuste y la $q$ de Tobin (Tobin, 1969; Hayashi, 1982)

**Origen**: James Tobin (1969) propuso como determinante de la inversión la ratio entre lo que vale el capital en el mercado y su coste de reposición, y Fumio Hayashi (1982) le dio fundamento con los costes de ajuste.

**Idea**: las incorporaciones netas de capital conllevan costes de ajuste, de modo que la empresa no salta al *stock* deseado sino que se acerca a él gradualmente (véase la figura del tema A34, apartado 2.3.2). El resultado es un ritmo óptimo de inversión, y no un nivel óptimo de capital, gobernado por el valor que el mercado asigna a una unidad adicional de capital.

**Supuestos**: los de la restricción tecnológica intertemporal, con cuatro precisiones:

- *Entorno*: precio del bien de capital constante y unitario.
- *Tecnología*:

- *Depreciación y factores*: por simplicidad, la depreciación es nula ($\dot{k} = I$) y el único factor productivo es el capital.
- *Costes de ajuste*: el supuesto central es que existen costes internos y convexos, $\varphi(I)$, que afectan también a la desinversión:

$$\varphi(0) = 0, \qquad \varphi'(0) = 0, \qquad \varphi''(I) > 0$$

  de modo que $\varphi'(I)$ tiene el signo de $I$ (p. ej. la forma cuadrática $\varphi = (b/2)\, I^2$).
- *Beneficios*: en la función de beneficios instantánea, $\Pi(K)k - I - \varphi(I)$, el beneficio aumenta con el capital de la empresa y con el ingreso marginal del capital, y disminuye con los costes de ajuste. Ese ingreso marginal se reduce al crecer el capital de la industria, ya que más capital agregado significa más oferta y precios más bajos.

**Planteamiento**: la empresa maximiza su valor con la inversión $I$ como variable de control:

$$\max_{\{I\}} V = \int_0^{\infty} e^{-rt}[\Pi(K)k - I - \varphi(I)]\, dt \quad \text{s.a.:} \begin{cases} \dot{k} = I \\ k_0 \text{ dado} \end{cases}$$

donde $k$ es el capital de la empresa (la variable de estado), con $k_0$ dado, $K$ el capital de la industria, $\Pi(K)$ el ingreso marginal del capital y $\varphi(I)$ la función de costes de ajuste.

**CPO**: el precio sombra del capital, $q$, es ahora la *$q$ marginal de Tobin*, el valor que una unidad adicional de capital añade al valor de la empresa, y las condiciones de primer orden del hamiltoniano son tres:

$$1 + \frac{d\varphi}{dI} = q \tag{3}$$

$$\Pi(K) = rq - \dot{q} \tag{4}$$

$$\lim_{t\to\infty} e^{-rt}\, q(t)\, k(t) = 0 \tag{5}$$

Según la condición (3), la empresa invierte hasta que una unidad adicional de capital vale, medida por la $q$ marginal de Tobin, lo mismo que cuesta adquirirla e instalarla, $1 + d\varphi/dI$, es decir, su precio más el coste de ajuste marginal. Los costes de ajuste suavizan así la inversión y generan una relación continua y decreciente entre inversión y tipo de interés, a diferencia de la relación discontinua de Jorgenson.

La condición (4), la ecuación de Euler de la empresa, establece que se invierte hasta que el ingreso marginal del capital iguala su coste de uso, es decir, el rendimiento que se obtendría invirtiendo $q$ al tipo de interés menos la ganancia de capital esperada. La condición (5) es la de transversalidad.

**CSO**: la convexidad de $\varphi(I)$ y la linealidad de $\Pi(K)k$ en $k$ hacen cóncavo el hamiltoniano, de modo que las condiciones (3) a (5) son suficientes (Mangasarian, 1966).

**Conclusiones**:

- *Aspectos positivos*: el modelo dota a la inversión de una dinámica propia y de un vínculo con los mercados financieros, ya que la $q$ resume en un solo precio las expectativas sobre la rentabilidad futura del capital. Por ello, la política monetaria y las noticias afectan a la inversión a través de la valoración de las empresas.

- *Limitaciones*: la $q$ marginal no es observable y solo coincide con la $q$ media, el cociente entre el valor bursátil y el coste de reposición, bajo competencia perfecta y rendimientos constantes a escala tanto en la producción como en los costes de ajuste (Hayashi, 1982). Además, la inversión responde al flujo de caja una vez controlada la $q$, sobre todo en las empresas con más difícil acceso a la financiación externa (Fazzari, Hubbard y Petersen, 1988), lo que apunta a restricciones financieras. Finalmente, los datos de planta se reproducen mejor con costes de ajuste convexos y no convexos combinados (Cooper y Haltiwanger, 2006).

## 2.5. Conclusiones

Las decisiones intertemporales de la empresa comparten estructura con las del consumidor. El teorema de separación de Fisher da a los modelos de Jorgenson y de Tobin su objetivo común, el valor de la empresa descontado al tipo de interés de mercado, que solo es independiente de las preferencias de los propietarios con mercados financieros perfectos. La regla de Jorgenson iguala la productividad marginal del capital a su coste de uso, pero carece de dinámica propia. Los costes de ajuste añaden esa dinámica, porque convierten la inversión en un proceso gradual gobernado por la $q$ de Tobin. El precio sombra del capital unifica ambos modelos, porque en Jorgenson coincide con el precio del bien de capital y en Tobin se separa de él en el coste de ajuste marginal.

# 3. Modelización dinámica del comportamiento colectivo: horizonte infinito y generaciones solapadas

Los modelos intertemporales son los cimientos de la microfundamentación de la macroeconomía. El interés por microfundamentar, que se afianza con la crítica de Lucas (1976), nace de la necesidad de modelos robustos, dependientes de parámetros estructurales (dotaciones, preferencias y tecnología).

La fundamentación se basa en el concepto de **agente representativo**, que admite tres interpretaciones:

- *Agente cualquiera*: cualquier individuo representa a la colectividad, lo que requiere que todos los agentes sean idénticos. Esta interpretación es teóricamente coherente, pero irreal.

- *Agente medio*: su comportamiento es el promedio de los individuales en una colectividad de agentes heterogéneos.

- *Agente ficticio*: su comportamiento coincide exactamente con el de la colectividad. Esta interpretación es la más habitual, aunque pueden darse problemas de agregación[10].

Frente a los modelos macroeconómicos previos, los modelos intertemporales permiten comparar situaciones y políticas en términos de bienestar. Los dos marcos que siguen difieren en un solo supuesto, la duración de la vida del agente representativo, y esa diferencia decide su propiedad más importante, la eficiencia del equilibrio.

## 3.1. Modelos de horizonte infinito (MHI)

**Origen**: Frank Ramsey (1928) planteó el problema del ahorro óptimo de una economía con horizonte infinito, y David Cass (1965) y Tjalling Koopmans (1965) lo convirtieron en el modelo de crecimiento óptimo que lleva sus nombres. De él parten los modelos del ciclo real y los de equilibrio general dinámico y estocástico (DSGE).

**Idea**: el agente representativo vive indefinidamente. La hipótesis se justifica de dos formas:

- *Dinastías* (Barro, 1974): familias cuyos miembros están conectados por altruismo intergeneracional, de modo que cada generación valora el bienestar de la siguiente y, si las herencias son operativas (positivas), la cadena de decisiones es equivalente a la de un único agente inmortal.

- *Muerte de Poisson* (Yaari, 1965; Blanchard, 1985): una vida finita pero de duración desconocida, que introduce en el descuento una probabilidad constante de morir, de modo que el problema de cada individuo es formalmente de horizonte infinito con una tasa de descuento mayor.

Con muerte de Poisson, a diferencia del caso dinástico, hay recambio de individuos aunque la población sea constante, y el modelo tiende un puente hacia los MGS.

> El modelo de horizonte infinito con acumulación de capital, el de Ramsey-Cass-Koopmans, se desarrolla en el tema A43, apartado 3, y sus aplicaciones al ciclo, en el tema A42, apartado 2.2. Aquí se presenta el problema del consumidor de vida infinita, común a todos

[10] Teorema de Sonnenschein-Mantel-Debreu (1972-1974): los comportamientos optimizadores individuales no se trasladan necesariamente a la demanda agregada. La agregación requiere preferencias de forma polar de Gorman (teorema de Gorman). Del lado de las empresas, sin embargo, es más sencilla (teorema de la empresa representativa: si no hay externalidades de producción y los factores se intercambian competitivamente, existe una empresa representativa).

ellos.

### 3.1.1. Supuestos

**Entorno**: tiempo discreto e infinito y mercados financieros perfectos, de modo que el agente puede prestar y endeudarse al tipo de interés $r$. La información es perfecta o, con incertidumbre, las expectativas son racionales.

**Agentes**: la economía la forma un agente representativo con preferencias aditivamente separables en el tiempo, $\sum_{t=0}^{\infty} \beta^t u(c_t)$, con $\beta = 1/(1+\rho) < 1$ y $u$ creciente y cóncava. El agente recibe una corriente de rentas $y_t$ y mantiene activos $a_t$.

**Restricciones**: la restricción presupuestaria de cada periodo es

$$a_{t+1} = (1+r)\,(a_t + y_t - c_t)$$

donde $a_t$ son los activos al inicio del periodo $t$. A diferencia del apartado 1.3, incluyen ya los intereses devengados en el periodo anterior, de ahí que en valor presente figure $a_0$ y no $(1+r)\,a_0$. Las restricciones de todos los periodos se encadenan en una restricción intertemporal única, que exige además una *condición de no-Ponzi*, $\lim_{T\to\infty} a_T/(1+r)^T \geq 0$. Esta condición impide al agente financiar indefinidamente su consumo emitiendo deuda nueva para pagar la antigua. Sin ella el problema no tiene solución, porque el consumo podría ser infinito.

### 3.1.2. Modelo

**Planteamiento**: el problema del consumidor de vida infinita en tiempo discreto es el de Fisher extendido a infinitos periodos:

$$\max_{\{c_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} \displaystyle\sum_{t=0}^{\infty} \frac{c_t}{(1+r)^t} = a_0 + \sum_{t=0}^{\infty} \frac{y_t}{(1+r)^t} & \text{(restricción intertemporal)} \\ \displaystyle\lim_{T\to\infty} \frac{a_T}{(1+r)^T} \geq 0 & \text{(condición de no-Ponzi)} \\ a_0 \text{ dado} & \text{(condición inicial)} \end{cases}$$

donde $a_0$ son los activos iniciales, $a_T$ los activos al inicio del periodo $T$ y las demás variables son las ya definidas.

**CPO**: con solución interior y la restricción agotada, las condiciones de primer orden son dos. La primera se obtiene derivando el lagrangiano respecto al consumo de cada periodo:

$$\frac{\partial \mathcal{L}}{\partial c_t} = \beta^t \frac{du}{dc_t} - \frac{\lambda}{(1+r)^t} = 0, \qquad \forall\, t$$

donde $\lambda$ es el multiplicador de la restricción intertemporal. Eliminándolo entre dos periodos consecutivos resulta la ecuación de Euler, que rige el perfil de la senda:

$$\frac{du}{dc_t} = \beta\,(1+r)\,\frac{du}{dc_{t+1}}$$

La segunda es la *condición de transversalidad*[11], que fija su nivel:

$$\lim_{T\to\infty} \beta^T \frac{du}{dc_T}\, a_T = 0$$

Esta condición exige que el valor presente de los activos que quedan sin consumir sea nulo, porque dejar riqueza sin usar no es óptimo. Equivale, en horizonte infinito, a la condición terminal de los problemas finitos y selecciona, entre las infinitas sendas que cumplen la ecuación de Euler, la única compatible con la restricción presupuestaria.

**CSO**: concavidad estricta de $u$, restricción lineal y transversalidad, de modo que el único óptimo es la senda que cumple tanto la condición de Euler como la de transversalidad.

**Solución**: con $r = \rho$ la senda es constante, y el consumo es la anualidad de la riqueza:

$$c = \frac{r}{1+r}\left[a_0 + \sum_{t=0}^{\infty} \frac{y_t}{(1+r)^t}\right]$$

donde el corchete es la riqueza del agente, activos iniciales más valor presente de las rentas. Esta solución es la hipótesis de la renta permanente en su forma más pura.

**Equilibrio**: el problema anterior toma $r$ como dado. En una economía cerrada sin sector público, el único activo es el capital físico, de modo que la riqueza de los hogares iguala al capital, $a_t = k_t$, y las empresas contratan el capital hasta que su productividad marginal iguala su coste, $r = df/dk - \delta$. En el estado estacionario el consumo es constante, lo que por la ecuación de Euler exige $r = \rho$, y la productividad marginal del capital cubre la impaciencia y la depreciación:

$$\left.\frac{df}{dk}\right|_{k^*} = \rho + \delta$$

donde $f(k)$ es la producción por trabajador y $\delta$ la tasa de depreciación. Esta condición es la *regla de oro modificada*. El capital estacionario es inferior al de la regla de oro, $df/dk|_{k_{\text{oro}}} = n + \delta$, con $n$ la tasa de crecimiento de la población, porque se supone $\rho > n$, condición para que la utilidad de la dinastía no diverja, de modo que la productividad marginal exigida es mayor y, por rendimientos decrecientes, el capital menor (tema A43, apartado 3.2).

**Estática comparativa**: para el consumidor, que toma $r$ como dado, los efectos dependen del tipo de perturbación:

- *Aumento permanente de la renta*: eleva el consumo en toda su cuantía.

- *Aumento transitorio de la renta*: lo eleva solo en su anualidad, $r/(1+r)$ veces el aumento, y el resto se ahorra.

- *Aumento del tipo de interés*: tiene los efectos sustitución y renta del apartado 1.4. En horizonte infinito, donde la mayor parte de la riqueza es renta futura, cobra especial peso, dentro del efecto renta, un efecto riqueza negativo, porque la subida de $r$ reduce el valor presente de las rentas futuras (la riqueza humana).

[11] La condición de transversalidad es la holgura complementaria de la no-Ponzi llevada al límite (tema A43, apartado 3.2).

**Bienestar**: como el agente representativo internaliza el bienestar de todas las generaciones, el equilibrio competitivo es Pareto-eficiente y coincide con el del planificador, de modo que el papel del Gobierno se reduce a corregir fallos de mercado. Además, como el mercado lleva a la regla de oro modificada y no a la regla de oro, no puede darse la ineficiencia dinámica del exceso de ahorro.

### 3.1.3. Conclusiones

**Aspectos positivos**: los MHI se acercan a ser modelos genuinamente estructurales, dependientes solo de preferencias, tecnología y dotaciones, y por ello en gran medida inmunes a la crítica de Lucas. Además, su formalismo dinámico admite reglas simples de comportamiento (*rules of thumb*)[12] y métodos de resolución bien establecidos (inspección de las condiciones de primer orden, estado estacionario, calibración y funciones de respuesta al impulso).

Su propiedad central es la eficiencia, que sirve de referencia, y sus aplicaciones cubren, además, el conjunto de la macroeconomía:

- *Oferta de trabajo*: sustitución intertemporal de Lucas y Rapping (tema A25).
- *Consumo*: renta permanente y paseo aleatorio de Hall (tema A33).
- *Inversión*: Jorgenson y $q$ de Tobin (apartado 2 y tema A34).
- *Demanda de dinero*: modelos de dinero en la función de utilidad (Sidrauski) y de restricción de efectivo (Lucas) (tema A35).
- *Crecimiento*: Ramsey-Cass-Koopmans (tema A43).
- *Ciclo*: ciclo real y modelos DSGE (tema A42).

**Limitaciones**: la hipótesis de vida infinita, o su justificación dinástica, es fuerte, y con ella desaparecen la heterogeneidad por edad y los conflictos entre generaciones. Además, la resolución analítica es posible solo en casos simples y suele exigir métodos numéricos.

## 3.2. Modelos de generaciones solapadas (MGS)

**Origen**: Maurice Allais (1947) y Paul Samuelson (1958) introdujeron la estructura de generaciones solapadas, y Peter Diamond (1965) la incorporó a un modelo con capital, que es la versión de referencia.

**Idea**: el agente representativo vive un número finito de periodos, y cada nueva generación llega antes de que se extinga la anterior (hipótesis de recambio poblacional). Los hogares son así un flujo, no un conjunto fijo, y en cada periodo conviven personas de distinta edad. Esa heterogeneidad por edad es la razón de ser del marco. Los jóvenes ahorran y los ancianos desahorran, y su encuentro en el mercado de capital produce efectos que los MHI no pueden recoger.

> Las aplicaciones de los modelos de generaciones solapadas a las pensiones se desarrollan en el tema A38, apartado 2, y la deuda pública y los límites de la equivalencia ricardiana, en el tema A39.

[12] Las *rules of thumb* permiten incorporar conductas observadas que no se derivan del comportamiento optimizador.

### 3.2.1. Supuestos

**Entorno**: el tiempo es discreto e infinito para la economía, pero finito para cada agente, y los mercados de bienes y factores son competitivos. La población crece a la tasa $n$, de modo que en cada periodo conviven $L_t$ jóvenes y $L_{t-1} = L_t/(1+n)$ ancianos.

**Agentes**: cada individuo vive dos periodos. En el primero (joven) ofrece una unidad de trabajo, recibe el salario $w_t$, consume $c_{1t}$ y ahorra $s_t$. En el segundo (anciano) no trabaja y consume su ahorro capitalizado, $c_{2t+1}$. No hay herencias ni altruismo, es decir, cada generación se ocupa solo de sí misma.

### 3.2.2. Modelo

**Empresas**: producen con una función neoclásica de rendimientos constantes, en términos por trabajador (por joven) $y = f(k)$.

▪ *Planteamiento*: en cada periodo maximizan su beneficio, $F(K_t, L_t) - w_t L_t - r_t K_t$, tomando como dados el salario, $w_t$, y el tipo de interés real, $r_t$, con $f(k_t) = F(k_t, 1)$ y $k_t = K_t/L_t$.

▪ *CPO*: cada factor cobra su productividad marginal:

$$w_t = \underbrace{f(k_t) - k_t \frac{df}{dk_t}}_{\text{Pmg}_l}, \qquad r_{t+1} = \underbrace{\frac{df}{dk_{t+1}}}_{\text{Pmg}_k}$$

donde el tipo que cobran en $t+1$ los ahorradores de $t$ iguala la productividad marginal del capital de ese periodo, pues se supone depreciación nula.

▪ *CSO*: con $f'' < 0$ y rendimientos constantes, el beneficio es cóncavo y las CPO caracterizan el máximo, con beneficio nulo.

**Hogares**: ahorran de jóvenes para la vejez.

▪ *Planteamiento*: el joven de la generación $t$ resuelve

$$\max_{c_{1t},\, c_{2t+1}} \; u(c_{1t}) + \frac{1}{1+\rho}\, u(c_{2t+1})$$

$$\text{s.a.:} \begin{cases} c_{1t} + s_t = w_t & \text{(primer periodo)} \\ c_{2t+1} = (1 + r_{t+1})\, s_t & \text{(segundo periodo)} \end{cases}$$

donde $c_{1t}$ y $c_{2t+1}$ (el consumo de joven y de anciano) son las variables de decisión y $\rho$ la tasa de descuento subjetiva.

▪ *CPO*: con solución interior, sustituyendo las restricciones en la función objetivo, la única variable de decisión es el ahorro, $s_t$. Derivando el objetivo, $u(w_t - s_t) + u((1 + r_{t+1})\, s_t)/(1+\rho)$, respecto a ella e igualando a cero:

$$-\frac{du}{dc_{1t}} + \frac{1 + r_{t+1}}{1+\rho}\, \frac{du}{dc_{2t+1}} = 0$$

que es la ecuación de Euler de dos periodos. De ella resulta una función de ahorro $s_t = s(w_t, r_{t+1})$, creciente en el salario y de signo ambiguo respecto del tipo de interés, por los efectos sustitución y renta contrapuestos del apartado 1.4.

▪ *CSO*: $u'' < 0$ y restricciones lineales, por lo que la ecuación de Euler de dos periodos es suficiente y la función de ahorro está bien definida.

**Equilibrio**: como los ancianos consumen todo su capital, el *stock* de capital del periodo siguiente es exclusivamente el ahorro de los jóvenes, $K_{t+1} = L_t\, s_t$. En términos por trabajador,

$$(1+n)\, k_{t+1} = s(w_t, r_{t+1})$$

que, con las condiciones de los mercados de factores, cierra el modelo en una ecuación en diferencias para $k_t$.

En particular, con utilidad logarítmica y tecnología Cobb-Douglas, $y = k^\alpha$, el joven ahorra la fracción $1/(2+\rho)$ de su salario y la solución es explícita:

$$k_{t+1} = \frac{1-\alpha}{(1+n)(2+\rho)}\, k_t^\alpha$$

Esta ecuación converge monótonamente a un estado estacionario $k^*$ (figura A29.7, panel (a)), como en el modelo de Solow, pero con una tasa de ahorro determinada por las preferencias y no exógena.

**Estática comparativa**: el capital del estado estacionario es

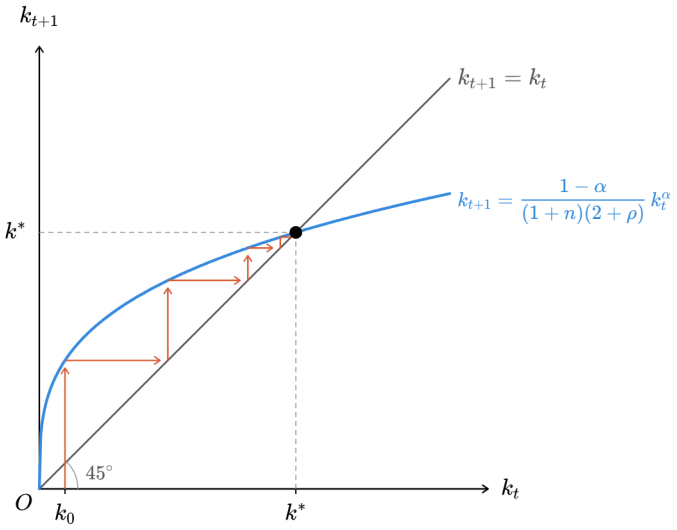
$$k^* = \left[\frac{1-\alpha}{(1+n)(2+\rho)}\right]^{1/(1-\alpha)}$$

y de él se derivan los efectos de tres perturbaciones:

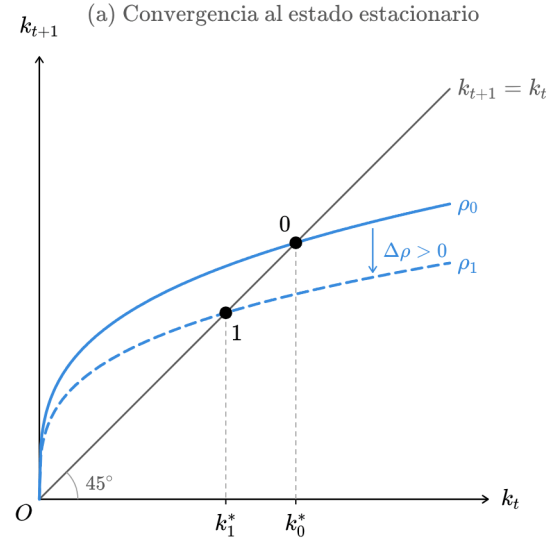
▪ *Impaciencia*: un aumento de $\rho$ reduce la fracción del salario que ahorran los jóvenes, $1/(2+\rho)$, y con ella $k^*$ (figura A29.7, panel (b)), de modo que el tipo de interés de equilibrio, $r^* = df/dk|_{k^*}$, sube.

▪ *Crecimiento de la población*: un aumento de $n$ reparte el ahorro de cada generación entre más trabajadores de la generación siguiente, reduce $k^*$ y eleva también $r^*$.

▪ *Transferencias intergeneracionales*: una transferencia de los jóvenes a los ancianos, mediante deuda pública o pensiones de reparto, reduce el ahorro y $k^*$, y solo mejora el bienestar de todas las generaciones si la economía es dinámicamente ineficiente.




(a) Convergencia al estado estacionario




(b) Aumento de la tasa de descuento

Figura A29.7: Dinámica del capital en el modelo de Diamond. En el panel (a), la curva $k_{t+1} = \frac{1-\alpha}{(1+n)(2+\rho)} k_t^{\alpha}$, creciente y cóncava, corta la bisectriz en el estado estacionario, $k^*$, al que la economía converge monótonamente desde el capital inicial, $k_0$. En el panel (b), un aumento de la tasa de descuento ($\Delta\rho > 0$), como uno de $n$, desplaza la curva hacia abajo y reduce el capital del estado estacionario de $k_0^*$ (punto 0) a $k_1^*$ (punto 1).

**Bienestar**: la diferencia esencial con el horizonte infinito está en la eficiencia. Nada garantiza que $k^*$ sea inferior al capital de la regla de oro, porque el ahorro de cada generación responde a su propio ciclo vital y no a una valoración del bienestar de las generaciones futuras. La condición de *ineficiencia dinámica* (*condición de Malinvaud-Cass*) es

$$r^* < n$$

donde $r^* = df/dk|_{k^*}$ es el tipo de interés real del estado estacionario (con depreciación, la condición es $df/dk|_{k^*} - \delta < n$). Cuando se cumple esa condición, la economía ha *sobreacumulado* capital, por lo que una reducción permanente del ahorro elevaría el consumo de todas las generaciones, presentes y futuras, y el equilibrio competitivo no es Pareto-eficiente.

La posible ineficiencia se debe a la combinación de dos rasgos[13]:

- *Externalidades pecuniarias intergeneracionales*: cada generación se enfrenta a una rentabilidad del capital determinada por un *stock* que no eligió, el heredado de la anterior.

- *Infinitud de agentes y bienes*: con un número finito de generaciones se cumpliría el primer teorema del bienestar, pese a esas externalidades, mientras que con infinitas puede dejar de cumplirse.

En la economía dinámicamente ineficiente aparece un papel para la política que en los MHI no existe, porque la deuda pública o un sistema de pensiones de reparto acercan el capital a la regla de oro.

### 3.2.3. Conclusiones

**Aspectos positivos**: los MGS son el instrumento para toda cuestión en la que la edad importa:

- *Consumo*: la hipótesis del ciclo vital (tema A33).

- *Pensiones*: la comparación entre sistemas de capitalización y de reparto (tema A38).

- *Deuda pública*: como transferencia entre generaciones (tema A39).

- *Economía abierta*: la demografía y la cuenta corriente (tema B12, apartado 3.2.3).

Su resultado distintivo, la posible ineficiencia dinámica, muestra que la eficiencia del equilibrio competitivo no es una propiedad general de las economías dinámicas sino del supuesto de vida infinita.

**Limitaciones**: las principales son tres:

- *Equilibrio*: puede no existir o estar indeterminado (equilibrios múltiples, incluidos equilibrios con burbujas, en los que un activo sin valor intrínseco, como el dinero fiduciario, se sostiene porque cada generación espera venderlo a la siguiente).

- *Simplificación*: el modelo básico reduce el ciclo vital a solo dos periodos.

- *Evidencia*: indica, en general, que las economías avanzadas no son dinámicamente ineficientes (Abel, Mankiw, Summers y Zeckhauser, 1989), lo que acota el alcance práctico del resultado.

[13] Con infinitos periodos, el primer teorema del bienestar exige que el valor presente de la dotación agregada sea finito ($\sum_t p_t^* Y_t < \infty$, con $p_t^*$ el precio de equilibrio en valor presente del bien del periodo $t$ y $Y_t$ la dotación agregada). En los MHI esa condición se cumple y el equilibrio es eficiente. En los MGS, en cambio, puede fallar, y entonces la eficiencia depende de la condición de Malinvaud-Cass del texto ($r^* \geq n$ en estado estacionario).

## 3.3. Comparación entre ambos marcos

Los dos marcos comparten agentes racionales optimizadores, formalismo dinámico microfundamentado e inmunidad, al menos parcial, a la crítica de Lucas.

**Diferencias normativas**: a la diferencia de eficiencia se suma la de la equivalencia ricardiana. En un MHI, una rebaja de impuestos de suma fija financiada con deuda no altera el consumo, porque el agente sabe que pagará la deuda con impuestos futuros y ahorra la rebaja. En un MGS, en cambio, la equivalencia falla, porque quienes reciben la rebaja no son quienes pagarán los impuestos, y la deuda pública tiene efectos reales, desplazando capital y redistribuyendo entre generaciones (tema A39).

El modelo de *juventud perpetua* de Blanchard (1985), basado en la muerte de Poisson, tiende un puente entre ambos:

- *Probabilidad de muerte nula*: se reduce al MHI y la equivalencia ricardiana se cumple.
- *Probabilidad positiva*: se comporta como un MGS de tratamiento analítico sencillo, y la equivalencia falla tanto más cuanto mayor es esa probabilidad.

La elección del marco depende de la pregunta. El MHI es el instrumento cuando la heterogeneidad por edad es secundaria (ciclo, crecimiento, política monetaria), y su eficiencia sirve de referencia. Por su parte, el MGS es imprescindible en las cuestiones distributivas entre generaciones (pensiones, deuda pública, demografía). Ambos marcos son los cimientos de la macroeconomía moderna microfundamentada.

## 3.4. Conclusiones

La tabla A29.1 resume las diferencias entre los dos marcos.

Tabla A29.1: Comparación entre los modelos de horizonte infinito y de generaciones solapadas.

| | MHI | MGS |
|---|---|---|
| Horizonte | Infinito | Finito (2+ periodos) |
| Hogares | Número fijo (dinastías). Con muerte de Poisson, recambio a población constante | Flujo (recambio) |
| Heterogeneidad | No (agente idéntico) | Sí (edad) |
| Eficiencia | Pareto-eficiente sin fallos de mercado | No necesariamente |
| Ineficiencia dinámica | No | Posible (sobreacumulación) |
| Rol del Gobierno | Limitado | Relevante (pensiones, redistribución) |
| Resolución | CPO, estado estacionario, numérica | Puede presentar equilibrios múltiples |
| Justificación | Dinastías (Barro) o muerte de Poisson | Recambio poblacional (Diamond) |

**Aspectos positivos**: la modelización dinámica sustituye las funciones de comportamiento *ad hoc* de la macroeconomía keynesiana por decisiones intertemporales explícitas, lo que permite evaluar las políticas en términos de bienestar y hace a los modelos robustos a los cambios de régimen.

**Limitaciones**: el agente representativo esconde la heterogeneidad dentro de cada generación, que la literatura reciente de modelos neokeynesianos con agentes heterogéneos (HANK) ha vuelto a introducir. Además, la microfundamentación tiene un coste de tratabilidad que obliga a supuestos funcionales fuertes, como la utilidad cuadrática, la de aversión relativa al riesgo constante (CRRA) o la tecnología Cobb-Douglas, de los que dependen muchos resultados.

## Conclusiones

Los modelos intertemporales muestran que los agentes deciden en función de todo su horizonte y no solo de las variables del periodo. En el consumidor de Fisher, la ecuación de Euler fija el perfil de la senda de consumo y la restricción de riqueza su nivel, de modo que el consumo depende de la riqueza y no de la renta corriente. Con incertidumbre y expectativas racionales el consumo sigue un paseo aleatorio (Hall), del que la evidencia se aparta por las restricciones de liquidez y los bienes duraderos. En la empresa, el teorema de separación de Fisher desliga la inversión de las preferencias de los propietarios, y la $q$ de Tobin resume el valor futuro de invertir. Sobre esos fundamentos, los MHI son eficientes en ausencia de fallos de mercado, mientras que los MGS pueden sobreacumular capital ($r^* < n$) y romper la equivalencia ricardiana.

Estos fundamentos sostienen buena parte de la macroeconomía del temario. La ecuación de Euler es el núcleo de las teorías del consumo (tema A33) y, junto con el coste de uso y la $q$ de Tobin, de las de la inversión (tema A34). El horizonte infinito es la base del modelo de Ramsey-Cass-Koopmans (tema A43) y de los modelos del ciclo real y DSGE (tema A42). Las generaciones solapadas, por su parte, son el marco del análisis de las pensiones (tema A38), de la deuda pública y la equivalencia ricardiana (tema A39) y del dinero como activo que una generación vende a la siguiente (tema A35).

Blanchard (2019) documenta que en las economías avanzadas el tipo de interés seguro ha estado a menudo por debajo de la tasa de crecimiento ($r < g$). Esa diferencia reduce el coste de la deuda pública y reabre, con el lenguaje de los MGS, la cuestión de la eficiencia dinámica. Los modelos de agentes heterogéneos, como el modelo HANK de Kaplan, Moll y Violante (2018), sustituyen al consumidor de horizonte infinito por hogares con restricciones de liquidez. Según estos modelos, la política monetaria influye en el consumo sobre todo a través de la renta, y no de la sustitución intertemporal de la ecuación de Euler.

# Anexos

## A.1. La regla de Hotelling

La regla de Hotelling se obtiene como solución del problema de control óptimo de una empresa que extrae un recurso agotable. Este problema es un caso particular del problema de inversión del apartado 2.4, en el que el *stock* que la empresa gestiona es un yacimiento y no su capital.

**Origen**: Harold Hotelling (1931) fundó la economía de los recursos naturales al plantear la extracción de un recurso no renovable como un problema de elección intertemporal.

**Idea**: un recurso agotable es un activo, y su propietario lo extrae o lo conserva según que la rentabilidad de extraerlo e invertir el producto supere o no a la revalorización de dejarlo en el yacimiento.

**Supuestos**: los del apartado 2.3, con dos salvedades:

- *Entorno*: competencia perfecta en el mercado del recurso, cuyo precio $p(t)$ la empresa toma como dado, y horizonte finito $T$.

- *Tecnología*: la empresa es propietaria de un yacimiento $S_0$ de un recurso no renovable, cuyo *stock* se reduce con la extracción $x(t)$ y se agota al final del horizonte. Extraer tiene un coste marginal constante $c$. La variable de estado es el *stock* restante $S(t)$, que hace el papel del capital, mientras que la extracción es la variable de control.

**Planteamiento**: la empresa maximiza el valor presente de las rentas de extracción sujeta a la ley de agotamiento del recurso:

$$\max_{\{x(t)\}} V = \int_0^T e^{-rt}\,[p(t) - c]\,x(t)\,dt$$

$$\text{s.a.:} \begin{cases} \dot{S} = -x(t) & \text{(ley de agotamiento)} \\ S_0 \text{ dado}, \quad S(T) = 0 & \text{(condiciones de frontera)} \end{cases}$$

donde $V$ es el valor presente de las rentas de extracción, $x(t)$ la extracción (la variable de control), $S(t)$ el *stock* restante del recurso (la variable de estado), $S_0$ el yacimiento inicial, $p(t)$ el precio del recurso, $c$ el coste marginal de extracción, $r$ el tipo de interés y $T$ el horizonte. El hamiltoniano en valor corriente, con $q(t)$ el precio sombra del recurso en el yacimiento, es

$$H = [p(t) - c]\,x(t) - q(t)\,x(t)$$

**CPO**: sus condiciones de primer orden establecen:

1. *Condición estática*: el precio de mercado del recurso neto de su coste de extracción, el *royalty*, iguala al valor marginal del recurso en el yacimiento:

$$p(t) - c = q(t)$$

El *royalty* es una renta de escasez o renta de Hotelling, distinta de la renta ricardiana diferencial (debida a diferencias de calidad o coste entre yacimientos). La renta de escasez es un pago por encima del coste marginal de extracción que existe porque la oferta total es limitada y que refleja el coste de oportunidad de extraer hoy en lugar de mañana (coste de uso).

2. *Condición dinámica*: como el hamiltoniano no depende del *stock*, el precio sombra crece al tipo de interés y, con él, el *royalty*, pues $c$ es constante:

$$\frac{\dot{q}}{q} = r \quad \Rightarrow \quad \frac{\dot{p}}{p-c} = r$$

Esta igualdad es la *regla de Hotelling*, por la que el *royalty* ha de crecer a la tasa de interés de mercado. La regla es una condición de arbitraje entre dos activos. Si el recurso en el yacimiento se revalorizara más deprisa que $r$, convendría conservarlo, lo que reduciría la oferta presente y elevaría el precio hoy hasta restablecer la igualdad. Si se revalorizara más despacio, convendría extraerlo e invertir el producto. En equilibrio no hay beneficios extraordinarios en el sector, y el precio del recurso sigue una senda creciente que raciona su uso en el tiempo.

3. *Condición de transversalidad*: el yacimiento se agota exactamente al final del horizonte, $S(T) = 0$, porque mientras el *royalty* sea positivo dejar recurso sin extraer no sería óptimo.

**CSO**: el hamiltoniano es lineal en la extracción, de modo que la empresa solo elige una senda interior si el *royalty* crece exactamente a la tasa $r$, y en otro caso extraería todo ahora o todo al final. La regla de Hotelling es, por tanto, la condición de equilibrio del mercado más que la de un óptimo interior de la empresa.

**Conclusiones**:

- *Aspectos positivos*: la regla tiene aplicaciones más allá de la economía de los recursos, en cualquier activo que pueda conservarse o liquidarse. La regla es también una de las bases de la economía del cambio climático, donde el carbono en el subsuelo es el recurso y su renta de escasez el *royalty* (tema B31).

- *Limitaciones*: empíricamente, los precios de la mayoría de las materias primas no crecen sistemáticamente al tipo de interés. Esta desviación se atribuye al progreso técnico en la extracción, al descubrimiento de nuevas reservas y al poder de mercado de los productores.

## A.2. Ejemplos resueltos

### Ejemplo: demanda de consumo intertemporal con elasticidad de sustitución constante

Sea $u(c) = c^{1-1/\sigma}/(1-1/\sigma)$, cuya ESI es constante e igual a $\sigma$ (con $\sigma = 1$, $u = \ln c$). Con $T = 1$ y $a_0 = 0$, la ecuación de Euler (2) da $c_1 = [\beta(1+r)]^{\sigma} c_0$ y, llevándola a la restricción, $c_0 + c_1/(1+r) = W_0$, las funciones de demanda de consumo presente y futuro:

$$c_0 = \frac{W_0}{1 + \beta^{\sigma}(1+r)^{\sigma-1}} \qquad c_1 = [\beta(1+r)]^{\sigma}\, c_0$$

Con los datos del ejemplo del apartado 1.4.1 ($\beta = 0{,}8$, $y_0 = 80$, $y_1 = 125$) y $r = 0{,}25$, el consumo es $c_0 = c_1 = 100$ para cualquier $\sigma$, porque $r = \rho$. Si el tipo de interés sube a $r = 0{,}5$, la riqueza cae de 180 a 163,3 y el consumo presente y el ahorro, $s_0 = y_0 - c_0$, pasan a ser:

Tabla A29.2: Consumo presente y ahorro con $r = 0{,}5$.

| | $\sigma = 0{,}5$ | $\sigma = 1$ | $\sigma = 2$ |
|---|---|---|---|
| $c_0$ | 94,4 | 90,7 | 83,3 |

| | $\sigma = 0{,}5$ | $\sigma = 1$ | $\sigma = 2$ |
|---|---|---|---|
| $s_0$ | $-14{,}4$ | $-10{,}7$ | $-3{,}3$ |

Para este prestatario, la subida del tipo de interés reduce siempre el consumo presente, porque lo encarece frente al futuro y además empobrece al prestatario. La caída es mayor cuanto mayor es $\sigma$, pues el consumidor desplaza más consumo hacia el futuro. Para un ahorrador, la subida lo enriquece, y el signo de la respuesta del consumo presente dependería de $\sigma$.

# A30. Magnitudes macroeconómicas y contabilidad nacional

Magnitudes macroeconómicas y contabilidad nacional.

## Introducción

La macroeconomía toma como objeto la economía en su conjunto y la analiza con modelos rigurosos en lo formal y capaces de explicar los datos. El análisis de coyuntura y la política económica se interesan, por un lado, por el **nivel de actividad** (la renta, la producción y su desviación respecto del potencial, o brecha de producción) y, por otro, por los **precios** (su nivel y su tasa de variación, la inflación). A esas dos variables se suman el empleo, el ahorro y la inversión y las relaciones con el exterior, que completan el cuadro con el que se juzga el desempeño de una economía.

Explicar esas variables exige antes medirlas. Siguiendo a Arthur Conan Doyle en *Escándalo en Bohemia*: "es un error garrafal teorizar antes de tener datos; sin darse cuenta, uno comienza a forzar los hechos para que encajen en las teorías". La cuantificación y presentación ordenada de estos agregados corresponde a la **contabilidad nacional**.

Las cuentas nacionales tratan de dar respuesta a cuál es el tamaño y composición de la producción de la economía, cuál es el origen y destino de la renta nacional y cuáles son las fuentes del ahorro e inversión para la producción futura. La cuantificación de estas variables da lugar a distintos indicadores o **macromagnitudes** (p. ej. PIB, IPC o tasa de desempleo) que permiten analizar el comportamiento económico y el desempeño de los países.

Históricamente, la contabilidad nacional comienza a desarrollarse a partir de la década de 1930, dada la necesidad de cuantificar el impacto económico de la Gran Depresión (1929-1933) y, después, de financiar y planificar el esfuerzo bélico de la Segunda Guerra Mundial (1939-1945)[1]. Desde entonces, la exigencia de comparabilidad internacional ha impulsado una metodología común, primero en el ámbito de la ONU (Sistema de Cuentas Nacionales, SCN) y después en el de la UE (Sistema Europeo de Cuentas, SEC). Este último es la base de la contabilidad nacional de los Estados miembros. Las cifras que produce tienen consecuencias directas, porque de ellas dependen el recurso propio de la Unión basado en la renta nacional bruta, el procedimiento de déficit excesivo y el reparto de los fondos estructurales. Con todo, los límites del PIB como medida del bienestar, que señaló la Comisión Stiglitz-Sen-Fitoussi (2009), mantienen abierto el debate sobre qué y cómo medir.

El objetivo del tema es definir las principales variables macroeconómicas y presentarlas dentro de un sistema contable. Para ello, el tema se organiza en tres partes:

[1] Los trabajos pioneros son los de Colin Clark y Simon Kuznets en los años veinte y treinta del s. XX.

- *Fundamentos*: el flujo circular de la renta, del que se derivan las identidades básicas, y la contabilidad nacional según el SEC-2010 (concepto y base normativa, contenido, principios).

- *Cuentas nacionales*: la secuencia de cuentas, no financieras y financieras, y los saldos que la articulan.

- *Indicadores y reflexiones*: otras cuentas, los principales indicadores macroeconómicos y algunas reflexiones sobre lo que el PIB mide y lo que no.

# 1. Fundamentos de la contabilidad nacional

La contabilidad nacional descansa en un fundamento teórico, el flujo circular de la renta, que el SEC-2010 concreta en un sistema normalizado, del que se exponen el concepto y la base normativa, el contenido y los principios del marco contable.

## 1.1. Fundamento teórico: el flujo circular de la renta (Quesnay, 1758; Knight, 1933)

> Este apartado recoge el antecedente teórico de la contabilidad nacional, el flujo circular y la equivalencia entre producción, renta y gasto que se deriva de él. No hace falta exponerlo con detalle en el ejercicio oral, pero conviene entenderlo, porque sostiene las identidades contables del resto del tema.

**Origen**: el fundamento teórico de la contabilidad nacional se encuentra en trabajos de tres siglos:

- *Siglo XVIII*: el *Tableau économique* (1758) del fisiócrata François Quesnay[2], un médico familiarizado con la circulación de la sangre, avanzó la idea del flujo circular del producto neto.

- *Siglo XIX*: Léon Walras (1874) formalizó la interdependencia general entre los mercados.

- *Siglo XX*: Frank Knight (1933) dibujó el primer diagrama del flujo circular, y los trabajos de John Hicks, Wassily Leontief, Simon Kuznets y Richard Stone sentaron las bases de la contabilidad nacional moderna.

**Idea**: el diagrama del flujo circular ilustra los distintos flujos de actividad de una economía (figura A30.1). Intuitivamente, cada unidad física de producto tiene su correspondencia en una unidad monetaria de ingreso en el bolsillo de alguien, y cada unidad monetaria de ingreso termina como unidad monetaria de gasto en un mercado.

**Supuestos**:

- *Entorno*: economía simplificada[3] y cerrada, sin sector público.

- *Agentes*: hogares y empresas, cada uno con una doble función. Las empresas son oferentes de productos y demandantes de factores, mientras que los hogares son demandantes de bienes y oferentes de factores.

**Razonamiento**: entre ambos nacen dos corrientes de sentido opuesto:

1. *Real*: un flujo de bienes, servicios y factores entendidos como entidades físicas (kilos, horas de trabajo o de máquina).

2. *Monetaria*: el ingreso que se obtiene en las transacciones y el gasto, que es su imagen especular.

El modelo desvela dos intuiciones:

[2] Para los fisiócratas, la "tabla económica" de Quesnay constituía su principal aportación teórica, con la que anticiparon la conocida tabla *input-output* de Wassily Leontief.

[3] Al ser cerrada, el producto interior coincide con el nacional. Al tratarse las empresas como un único productor integrado, no hay consumos intermedios, y el valor de la producción coincide con el valor añadido y con la renta generada.

1. *Identidades contables*: las interrelaciones entre los sectores se recogen mediante identidades, que se cumplen siempre por construcción y no implican equilibrio económico.

2. *Tres enfoques*: la producción, la renta y el gasto son tres medidas complementarias de un mismo flujo y los tres puntos del circuito en los que se sitúan las cuentas de la sección 2.

**Desarrollo gráfico**: en la figura A30.1, los hogares (a la izquierda) y las empresas (a la derecha) se unen por las dos corrientes:

1. *Corriente real* (arcos exteriores): el arco superior lleva los factores productivos (trabajo, capital) de los hogares a las empresas, y el inferior, los bienes y servicios que las empresas producen con ellos hacia los hogares.

2. *Corriente monetaria* (flechas interiores): las rentas (salarios, beneficios, intereses) con que las empresas pagan esos factores van hacia los hogares, y el gasto con que estos compran los bienes y servicios vuelve a las empresas.

3. *Puntos de medición*: el PIB se mide en la producción de las empresas (1), en la renta que perciben los hogares (2) y en su gasto (3). Las tres medidas coinciden porque recogen el mismo flujo en tres puntos del circuito.

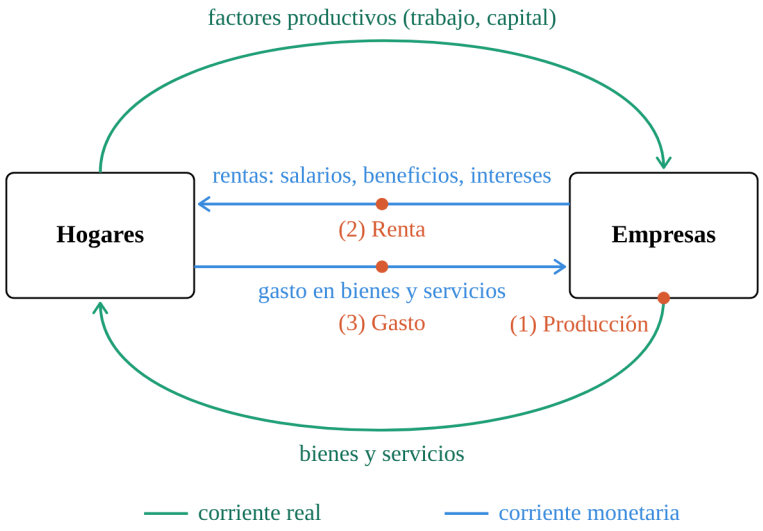



Figura A30.1: Flujo circular de la renta. Entre hogares y empresas circulan una corriente real (bienes y factores) y una corriente monetaria (gasto e ingreso), y el PIB se mide en tres puntos (producción, renta y gasto).

**Conclusiones**:

- *Aspectos positivos*: ofrece una representación sencilla de la interdependencia entre hogares y empresas y fundamenta la equivalencia entre producción, renta y gasto, sobre la que descansa la medición del PIB en la contabilidad nacional.

- *Limitaciones*: en su versión básica omite el ahorro, la inversión, el sector público y el resto del mundo, que hay que añadir para describir una economía real. Además, como sus relaciones

son identidades, describe cómo circula la renta, pero no explica qué determina su nivel ni sus fluctuaciones.

## 1.2. Concepto y base normativa

La **contabilidad nacional** es el registro sistemático y estadístico del conjunto de la actividad macroeconómica de un país durante un periodo, que ofrece una imagen de su situación económica. Es una técnica de representación que permite obtener una descripción cuantitativa y simplificada de la actividad económica y observar cómo se alcanza el equilibrio contable entre las principales magnitudes agregadas.

**Objetivo**: su fin esencial es disponer de un marco estadístico completo, coherente y flexible que permita el análisis económico comparado y la formulación de la política económica.

**Base normativa**: la metodología se ha ido normalizando en dos ámbitos:

1. *Ámbito internacional (SCN)*: en 1953, un grupo de expertos presidido por Richard Stone elaboró para la ONU el primer Sistema de Cuentas Nacionales normalizado (SCN-53), actualizado en 1968, 1993, 2008 y 2025. Las tres últimas revisiones se elaboraron conjuntamente con la Comisión Europea (Eurostat), el FMI, la OCDE y el Banco Mundial. El SCN-2008[4], basado en el SCN-93, se concibió para todos los países, independientemente de su etapa de desarrollo económico. Además, está armonizado con el Manual de Balanza de Pagos (MBP6) del FMI, de modo que las cuentas del resto del mundo y la balanza de pagos comparten conceptos. En marzo de 2025, la Comisión Estadística de la ONU aprobó el SCN-2025 y el FMI publicó la séptima edición del Manual de Balanza de Pagos (MBP7). La UE incorporará el SCN-2025 en una futura revisión del SEC.

2. *Ámbito de la Unión (SEC)*: en línea con el SCN, el Sistema Europeo de Cuentas Nacionales y Regionales (SEC-2010)[5] sustituye al SEC-95 y tiene carácter vinculante para los Estados miembros. Se trata de un reglamento que fija reglas, definiciones, nomenclaturas y normas contables comunes para elaborar cuentas y tablas sobre bases comparables. De esas cuentas y tablas dependen decisiones de la Unión como el recurso propio basado en la renta nacional bruta (RNB), el procedimiento de déficit excesivo o los fondos estructurales.

**Evolución del SEC-95 al SEC-2010**: la estructura básica del sistema apenas se modifica, si bien existen dos tipos de cambios:

1. *Con impacto sobre el PIB*: reconocimiento como inversión, es decir, como formación bruta de capital fijo (FBKF), de los gastos en I+D[6] y de los sistemas de armamento[7].

2. *Sin impacto sobre el PIB*: mayor detalle en la delimitación del sector público y de las instituciones financieras, y armonización con el MBP6 en el registro de flujos con el resto del mundo (comercio triangular, bienes enviados al extranjero para su transformación).

## 1.3. Contenido: unidades, flujos y *stocks*

El sistema describe la economía a partir de las unidades que realizan operaciones (flujos) y poseen activos y pasivos (*stocks*), y de los precios a los que se valora la producción.

[4] Aprobado por la Comisión Estadística de la ONU en 2008 (volumen 1) y 2009 (volumen 2).
[5] Aprobado en 2013 y aplicable desde septiembre de 2014 (Reglamento (UE) n.º 549/2013 del Parlamento Europeo y del Consejo).
[6] En el SEC-95 los gastos en I+D se registraban como consumo intermedio.
[7] El SEC-95 solo consideraba activos fijos las estructuras y el equipamiento militar con equivalente de uso civil.

### 1.3.1. Unidades y sectores institucionales

**Unidades institucionales**: el SEC-2010 las define como entes económicos capaces de ser propietarios de bienes y activos, contraer pasivos y participar en actividades y operaciones económicas con otras unidades en nombre propio. Se agrupan en cinco sectores nacionales mutuamente excluyentes, definidos por el tipo de productor y su función principal, más el resto del mundo:

Tabla A30.1: Sectores institucionales del SEC-2010.

| Sector | Tipo de productor | Función principal | Incluye |
|---|---|---|---|
| Sociedades no financieras (S.11) | De mercado | Producción de bienes y servicios no financieros | Privadas y públicas (p. ej. empresas públicas de correos o ferrocarriles) |
| Instituciones financieras (S.12) | De mercado | Intermediación financiera y actividades auxiliares | Privadas y públicas, incluidos los bancos centrales |
| Administraciones públicas (S.13) | De no mercado | Producción de bienes y servicios no de mercado para consumo individual o colectivo; redistribución de la renta y la riqueza | Administración central, regional, local y fondos de la Seguridad Social |
| Hogares (S.14) | De mercado y para uso final propio | Consumo; producción de mercado (empresarios individuales) y para uso final propio | Personas físicas y empresarios individuales |
| Instituciones sin fines de lucro al servicio de los hogares (ISFLSH, S.15) | De no mercado privado | Producción de servicios no de mercado para los hogares | Organizaciones no gubernamentales, sindicatos, partidos, fundaciones, iglesias |
| Resto del mundo (S.2) | — | Operaciones con unidades no residentes | Unidades no residentes e instituciones de la UE |

La distinción entre productores de mercado y de no mercado separa a las empresas públicas de las Administraciones públicas (AA. PP.). En particular, una unidad pública es productor de mercado, y se clasifica en S.11 o S.12, si, entre otros criterios, sus ventas cubren al menos el 50 % de sus costes de producción a lo largo de varios años.

**Unidades de actividad económica locales (UAE)**: permiten la agrupación por ramas de actividad para el análisis del proceso productivo (tablas de origen y destino e *input-output*). Una actividad comprende los insumos de bienes y servicios, el proceso de producción y los productos obtenidos.

### 1.3.2. Flujos y *stocks*

**Flujos**: para el SEC-2010, expresan cualquier creación, transformación, intercambio, transferencia o extinción del valor económico y modifican lo que valen los activos o pasivos de la unidad institucional afectada. Se dividen en *operaciones*, de cuatro tipos (de bienes y servicios,

distributivas, financieras y otras operaciones), y *otras variaciones de activos* (no derivadas de operaciones: catástrofes, descubrimientos, revalorizaciones).

***Stocks*** o **posiciones**: son los activos y pasivos que se poseen en un momento concreto. Los *activos* son entidades sobre las que se ejercen derechos de propiedad y de las que se derivan beneficios económicos. Pueden ser no financieros (producidos, como el capital fijo y las existencias, o no producidos, como la tierra) o financieros, que tienen como contrapartida un *pasivo* de otra unidad. La diferencia entre activos y pasivos es el *patrimonio neto*. Las posiciones se registran al principio y al final de cada ejercicio en los balances.

### 1.3.3. Precios básicos y de adquisición

La producción se valora en distintos puntos de la cadena que va del productor al comprador. Para una unidad del producto $j$, se distingue entre:

**Precio básico** ($p^b$): lo que retiene el productor por unidad vendida.

$$p_j^b = c_j + t_{p,j}$$

donde $c_j$ es el coste unitario de los consumos intermedios y de los factores (remuneración de asalariados y excedente) y $t_{p,j}$ los otros impuestos sobre la producción netos de subvenciones (p. ej. impuestos sobre los inmuebles de la empresa o licencias de actividad).

**Precio de adquisición** ($p^a$): lo que paga el comprador por unidad.

$$p_j^a = p_j^b + t_{c,j} + \mu_{c,j}$$

donde $t_{c,j}$ son los impuestos sobre los productos netos de subvenciones (IVA no deducible, impuestos especiales, derechos de importación) y $\mu_{c,j}$ los márgenes de distribución y transporte.

**Agregados**: se obtienen multiplicando por las cantidades y sumando. Así, para una rama $i$, la producción a precios básicos es $\mathrm{VP}_i = \sum_j p_j^b\, x_{ij}$, donde $x_{ij}$ son las unidades del producto $j$ producidas por la rama, y el valor añadido bruto es la producción menos los consumos intermedios, valorados estos a precios de adquisición:

$$\mathrm{VAB}_i^{\mathrm{pb}} = \mathrm{VP}_i - \mathrm{CI}_i$$

donde $\mathrm{CI}_i$ son los consumos intermedios de la rama. Deduciendo los otros impuestos netos sobre la producción se obtiene el *valor añadido a coste de los factores*, $\mathrm{VAB}_i^{\mathrm{cf}} = \mathrm{VAB}_i^{\mathrm{pb}} - T_{p,i}$ (con $T_{p,i}$ los otros impuestos netos sobre la producción de la rama), que el SEC-2010 no utiliza formalmente[8][9][10].

Para la economía total, los márgenes de distribución y transporte son a su vez producción de las ramas de comercio y transporte, por lo que se cancelan al agregar. Por tanto, solo queda por añadir la recaudación por impuestos sobre los productos, que no es valor añadido de ninguna rama:

[8] En la teoría económica se distingue la valoración a coste de los factores (lo que retienen los factores) de la valoración a precios de mercado (lo que paga el comprador).

[9] Los impuestos sobre la producción y las importaciones (D.2) se dividen en impuestos sobre los productos (D.21) y otros impuestos sobre la producción (D.29). Las subvenciones (D.3) siguen la misma división.

[10] La denominación "impuestos indirectos" ha caído en desuso y no se utiliza ni en el SCN ni en el SEC.

$$\mathrm{PIB}_{\mathrm{pm}} = \sum_i \mathrm{VAB}_i^{\mathrm{pb}} + T_c$$

donde $T_c = \sum_j t_{c,j}\, x_j$ es la recaudación neta por impuestos sobre los productos, con $x_j$ las unidades vendidas del producto $j$. Esta expresión es la primera de las tres vías de medición del PIB (apartado 2.1) y el saldo con el que se cierra la cuenta de producción (apartado 2.2.1).

## 1.4. Principios del marco contable

El sistema descansa sobre cinco decisiones metodológicas, que fijan qué se contabiliza, cómo se registra, cuándo, con qué criterio de valoración y en qué unidad.

### 1.4.1. Delimitación de la economía nacional

La delimitación se basa en el **criterio de residencia**, y no en el de nacionalidad. El SEC-2010 considera residente en un país a la unidad institucional cuyo centro de interés económico predominante se sitúa en su territorio económico, condición que en la práctica se cumple cuando la unidad realiza allí actividades económicas durante un año como mínimo. Así, la economía nacional agrupa las unidades residentes, y el resto del mundo, las no residentes que operan con ellas.

### 1.4.2. Sistema de registro: partida doble

El criterio básico es la **partida doble**. Cada operación se registra dos veces con el mismo valor para cada unidad que interviene. Una anotación figura como **recurso** (operaciones que aumentan el valor económico de la unidad, como la producción o las rentas recibidas) o como variación de pasivos. La otra figura como **empleo** (operaciones que reducen ese valor, como los consumos intermedios o las rentas pagadas) o como variación de activos[11].

Puesto que cada operación afecta a dos unidades, el sistema es en realidad de **partida cuádruple**, lo que garantiza la coherencia de las cuentas. Las cuentas se presentan en forma de T, con los empleos a la izquierda y los recursos a la derecha, y se cierran con un **saldo contable** que pasa como recurso a la cuenta siguiente.

### 1.4.3. Momento del registro: devengo

Rige el **principio de devengo**: los flujos se anotan cuando se crea, transforma o extingue el valor económico, o cuando nacen, se transforman o se cancelan derechos y obligaciones, al margen de cobros y pagos. En su aplicación cabe distinguir tres casos:

- *Producción*: se registra cuando se obtiene, no cuando se vende (de ahí la variación de existencias).
- *Operaciones con activos*: se registran cuando se transfiere la propiedad económica del activo.
- *Rentas*: los intereses se registran en el ejercicio en que se devengan, aunque se paguen después.

[11] Los términos "empleo" y "recurso" sustituyen a los términos "debe" y "haber". En las cuentas de acumulación se habla de "variaciones de activos" y "variaciones de pasivos y patrimonio neto".

La única desviación práctica son los impuestos y otros flujos en los que intervienen AA. PP., para los que, a falta de datos por devengo, se admite un registro por caja **ajustado en el tiempo** (desplazado al periodo en que se generó la obligación).

### 1.4.4. Criterios de valoración

El SEC-2010 valora los flujos y *stocks* según tres criterios:

1. *Precios de mercado*: es la referencia general, es decir, el valor al que los flujos y *stocks* se cambian o podrían cambiarse por efectivo. Dentro de ella, la producción se valora a precios básicos y los empleos (consumos intermedios, consumo final, formación de capital) a precios de adquisición.

2. *Costes de producción*: cuando no existen precios de mercado propios ni análogos (servicios no de mercado de las AA. PP. y las ISFLSH), la producción se valora por la suma de sus costes. Esos costes son los consumos intermedios, la remuneración de asalariados, el consumo de capital fijo y los otros impuestos netos sobre la producción[12].

3. *Valor franco a bordo* (*free on board*, FOB): las exportaciones e importaciones de bienes se registran por su valor en la frontera del país exportador, sin fletes ni seguros[13]. Por eso hay que ajustar las importaciones aduaneras, que se declaran a valor de coste, seguro y flete (*cost, insurance and freight*, CIF).

Con independencia del criterio, los agregados se valoran a **precios corrientes** del periodo. Por ello, las medidas en términos de **volumen** (el "PIB real") no son una valoración alternativa sino una descomposición posterior de la variación nominal en precios y cantidades, que se obtiene mediante índices encadenados a precios del año anterior (apartado 3.2.1)[14].

### 1.4.5. Unidad de cuenta

La unidad de cuenta es la moneda nacional. Para las comparaciones internacionales se recurre a tipos de cambio o, preferiblemente, a **paridades de poder adquisitivo** (PPA), que corrigen las diferencias de nivel de precios entre países.

## 1.5. Conclusiones

El flujo circular de la renta da a la contabilidad nacional su fundamento teórico. Producción, renta y gasto son tres medidas del mismo flujo, y las relaciones entre los sectores son identidades que se cumplen por construcción, sin implicar equilibrio. Sobre esa base, el SCN-2008 y el SEC-2010 fijan una metodología común, que en la Unión es un reglamento vinculante del que dependen el recurso propio basado en la RNB y el procedimiento de déficit excesivo.

El sistema agrupa las unidades institucionales en cinco sectores y el resto del mundo y distingue los flujos de los *stocks*. La valoración a precios básicos y de adquisición permite pasar del valor añadido de las ramas al PIB a precios de mercado. Los principios de residencia, partida doble, devengo, valoración y unidad de cuenta hacen, por último, que las cuentas sean coherentes y

[12] Como último recurso, si no es posible calcular el coste de producción, se puede recurrir al valor actual de los rendimientos esperados.

[13] Los fletes y seguros entre la frontera del exportador y la del importador no se incluyen en el valor de los bienes, sino que se registran como servicios.

[14] La variación de volumen recoge tanto las cantidades como los cambios de calidad.

comparables. La sección 2 recorre ese flujo en la secuencia de cuentas nacionales, de la cuenta de producción a los balances.

# 2. Las cuentas nacionales

El PIB, que puede medirse por tres vías, es el primero de los saldos de una secuencia de cuentas, no financieras y financieras, cuya lectura económica se expresa en términos de ahorro e inversión.

## 2.1. El PIB y el triple enfoque

El **producto interior bruto** a precios de mercado es el valor de los bienes y servicios finales producidos por las unidades residentes durante un periodo[15]. Es una magnitud flujo, medida en dinero para poder sumar bienes heterogéneos, que solo computa bienes finales y evita así contar más de una vez los intermedios. Por eso, el concepto relevante es el **valor añadido**, la diferencia entre la producción y los consumos intermedios de cada unidad[16]. El PIB puede medirse por tres vías, que coinciden entre sí:

$$\mathrm{PIB}_{\mathrm{pm}} = \underbrace{\sum_i \mathrm{VAB}_i^{\mathrm{pb}} + T_c}_{\text{Oferta}} = \underbrace{\mathrm{RA} + T + \mathrm{EBE} + \mathrm{RM}}_{\text{Rentas}} = \underbrace{C + I + G + (X - M)}_{\text{Demanda}}$$

donde $\mathrm{VAB}_i^{\mathrm{pb}}$ es el valor añadido bruto a precios básicos de la rama $i$, $T_c$ los impuestos sobre los productos netos de subvenciones, RA la remuneración de los asalariados, $T$ los impuestos sobre la producción y las importaciones netos de subvenciones (sobre los productos y otros), EBE el excedente bruto de explotación, RM la renta mixta bruta de los empresarios individuales, $C$ el gasto en consumo final de los hogares e ISFLSH, $I$ la formación bruta de capital (fija, variación de existencias y adquisiciones menos cesiones de objetos valiosos), $G$ el gasto en consumo final de las AA. PP., $X$ las exportaciones y $M$ las importaciones.

Las tres vías son los tres puntos del flujo circular (apartado 1.1). En la secuencia de cuentas, la primera es el saldo de la cuenta de producción, la segunda el reparto que hace la cuenta de explotación y la tercera la suma de los empleos finales. Del PIB, magnitud **interior**, se pasa a las magnitudes **nacionales**, la RNB y la renta nacional bruta disponible (RNBD), a través de las rentas y transferencias con el resto del mundo, en las cuentas de distribución. La diferencia entre interior y nacional no es de nacionalidad sino de residencia, ya que el PIB atiende a la residencia de las unidades que producen y la RNB, a la de las unidades que perciben las rentas[17].

Todas las magnitudes pueden expresarse en términos **brutos** o **netos** según incluyan o no el **consumo de capital fijo** (depreciación, CCF)[18], de modo que $\mathrm{PIN} = \mathrm{PIB} - \mathrm{CCF}$ y $I_N = I - \mathrm{CCF}$. La inversión neta es así la que amplía el *stock* de capital, $I_N = K_{t+1} - K_t$, donde $K_t$ es el *stock* de capital al inicio del periodo.

[15] El SCN, en su versión en español, habla de producto interno; el SEC, de producto interior.

[16] Las materias primas y los productos semiacabados obtenidos en el periodo que todavía no han entrado en el proceso productivo también forman parte del PIB, a través de la variación de existencias.

[17] Ejemplo: en EE. UU. la RNB apenas difiere del PIB. En Irlanda, la RNB es un 20-25 % inferior al PIB por los beneficios, repatriados o reinvertidos, de las multinacionales extranjeras.

[18] La renta relevante para la remuneración de los factores es la neta, pues el consumo de capital fijo no es renta de ningún factor.

## 2.2. Cuentas no financieras

Las cuentas se presentan en forma de T (apartado 1.4.2) y se elaboran para cada sector institucional y para la economía total. Aquí se exponen las de la economía total, en una secuencia que va de la producción a la acumulación y de esta a los balances.

### 2.2.1. Cuentas corrientes

Las **cuentas corrientes** registran la producción, la generación y distribución de la renta y su utilización en consumo y ahorro.

**Cuenta de producción** (C.I): muestra cómo se crea valor en el proceso productivo. Los recursos son la producción a precios básicos y, para la economía total, los impuestos sobre los productos netos de subvenciones, que se incluyen para pasar de la producción a precios básicos al PIB a precios de mercado. Los empleos son los consumos intermedios. El saldo es el *valor añadido bruto* de cada sector y, para la economía total, el producto interior bruto, que se obtiene por agregación de los valores añadidos, no de las producciones.

Tabla A30.2: Cuenta de producción (economía total).

| Empleos | | Recursos | |
|---|---|---|---|
| Consumos intermedios (P.2) | xxx | Producción a precios básicos (P.1) | xxx |
| | | Impuestos sobre los productos netos de subvenciones (D.21 − D.31) | xxx |
| **Saldo: PIB a precios de mercado (B.1*g)** | **xxx** | | |

Las **cuentas de distribución primaria de la renta** (C.II.1) muestran cómo la renta generada en la producción se reparte entre el trabajo, el capital y el sector público, y cómo se ajusta con el exterior. Constan de dos cuentas:

**Cuenta de explotación** (C.II.1.1): parte del PIB como recurso y lo distribuye en los empleos. Los empleos son la remuneración de los asalariados interior (pagada por empleadores residentes, incluidas las cotizaciones sociales a su cargo) y los impuestos sobre la producción y las importaciones netos de subvenciones. El saldo es el *excedente bruto de explotación* (EBE, excedente antes de deducir intereses, rentas de la tierra o de la propiedad) y la *renta mixta bruta* (RM) de los empresarios individuales, que lleva implícita la remuneración del trabajo del propietario y su familia.

Tabla A30.3: Cuenta de explotación (economía total).

| Empleos | | Recursos | |
|---|---|---|---|
| Remuneración de los asalariados (D.1) | xxx | PIB a precios de mercado (B.1*g) | xxx |
| Impuestos sobre la producción y las importaciones netos de subvenciones (D.2 − D.3) | xxx | | |
| **Saldo: EBE + RM (B.2g + B.3g)** | **xxx** | | |

**Cuenta de asignación de la renta primaria** (C.II.1.2): recoge la renta primaria desde la óptica de los residentes que la perciben. Los recursos son el EBE y la RM, la remuneración de asalariados recibida por los hogares residentes (de empleadores residentes y no residentes), los impuestos netos de subvenciones recibidos por las AA. PP. y las rentas de la propiedad recibidas[19]. Los empleos son las rentas de la propiedad pagadas, incluidas las pagadas al resto del mundo. El saldo es la *renta nacional bruta*.

Tabla A30.4: Cuenta de asignación de la renta primaria (economía total).

| Empleos | | Recursos | |
|---|---|---|---|
| Rentas de la propiedad pagadas (D.4) | xxx | EBE + RM (B.2g + B.3g) | xxx |
| | | Remuneración de los asalariados recibida (D.1) | xxx |
| | | Impuestos sobre la producción y las importaciones netos de subvenciones (D.2 − D.3) | xxx |
| | | Rentas de la propiedad recibidas (D.4) | xxx |
| **Saldo: renta nacional bruta (B.5*g)** | **xxx** | | |

Comparando ambas cuentas, la RNB difiere del PIB en las rentas primarias netas recibidas del resto del mundo (remuneración de asalariados, rentas de la propiedad e impuestos netos con instituciones de la UE), que son el saldo de rentas primarias de la balanza de pagos:

$$\text{RNB} = \text{PIB} + \text{SBR}_1$$

donde $\text{SBR}_1$ es el saldo de la balanza de rentas primarias. En los países con posición deudora neta frente al exterior, $\text{SBR}_1$ suele ser negativo (RNB < PIB) por el pago neto de intereses y dividendos al exterior.

**Cuenta de distribución secundaria de la renta** (C.II.2): muestra la redistribución de la renta vía transferencias corrientes. Las transferencias incluyen los impuestos corrientes sobre la renta y el patrimonio, las cotizaciones y prestaciones sociales y otras transferencias corrientes (entre hogares, con AA. PP. y cooperación internacional). Los recursos son la RNB y las transferencias corrientes recibidas, y los empleos, las transferencias corrientes pagadas. El saldo es la *renta nacional bruta disponible*.

Tabla A30.5: Cuenta de distribución secundaria de la renta (economía total).

| Empleos | | Recursos | |
|---|---|---|---|
| Transferencias corrientes pagadas (D.5, D.61, D.62, D.7) | xxx | Renta nacional bruta (B.5*g) | xxx |
| | | Transferencias corrientes recibidas (D.5, D.61, D.62, D.7) | xxx |
| **Saldo: renta nacional bruta disponible (B.6g)** | **xxx** | | |

[19] Las rentas de la propiedad comprenden las rentas de la inversión (intereses, dividendos, beneficios reinvertidos) y las rentas de la tierra y otros recursos naturales. Los derivados financieros no generan rentas de la inversión.

Para la economía total, las transferencias entre residentes se cancelan y solo quedan las cruzadas con el exterior, que son el saldo de rentas secundarias de la balanza de pagos (remesas, cooperación al desarrollo, recursos propios IVA y RNB de la UE):

$$\text{RNBD} = \text{RNB} + \text{SBR}_2$$

donde $\text{SBR}_2$ es el saldo de la balanza de rentas secundarias. Adicionalmente, la *cuenta de redistribución de la renta en especie* (C.II.3) traslada a los hogares las transferencias sociales en especie (educación, sanidad) que financian las AA. PP. y las ISFLSH. El saldo de esta cuenta es la *renta disponible ajustada*. Esta redistribución no altera el total de la economía pero sí su reparto entre sectores[20].

**Cuenta de utilización de la renta disponible** (C.II.4.1): recoge cómo se reparte la renta disponible entre consumo final y ahorro. Solo los hogares, las ISFLSH y las AA. PP. realizan gasto en consumo final[21], que es de dos tipos:

- *Consumo individual*: el de los hogares y las ISFLSH (bienes y servicios rivales, incluidas las transferencias sociales en especie que financian las ISFLSH) y parte del de las AA. PP. (educación, sanidad, bienes preferentes)[22].

- *Consumo colectivo*: el resto del de las AA. PP. (bienes públicos puros como defensa, justicia u orden público, que se prestan simultánea y pasivamente a toda la comunidad).

Los recursos son la RNBD y el ajuste por la variación de los derechos de pensión[23], que reciben los hogares, y los empleos, el gasto en consumo final y ese mismo ajuste, que registran los gestores de los sistemas de pensiones. En la economía total, ambas anotaciones solo difieren por las operaciones con el resto del mundo. El saldo es el *ahorro nacional bruto*.

Tabla A30.6: Cuenta de utilización de la renta disponible (economía total).

| Empleos | | Recursos | |
|---|---|---|---|
| Gasto en consumo final individual (P.31): hogares, ISFLSH y AA. PP. | xxx | Renta nacional bruta disponible (B.6g) | xxx |
| Gasto en consumo final colectivo (P.32): AA. PP. | xxx | Ajuste por variación de los derechos de pensión (D.8) | xxx |
| Ajuste por variación de los derechos de pensión (D.8) | xxx | | |
| **Saldo: ahorro nacional bruto (B.8g)** | **xxx** | | |

El gasto en consumo final de los hogares representa en torno al 55-60 % del PIB en los países de la OCDE y es un componente estable y procíclico. El gasto en consumo final de las AA. PP. se sitúa en torno a una quinta parte del PIB en las economías europeas. Deduciendo el consumo

[20] De ahí la distinción entre gasto en consumo final (según quién lo paga) y consumo final efectivo (según quién lo disfruta, incluidas las transferencias sociales en especie).

[21] Las compras de bienes y servicios de las sociedades son consumos intermedios o formación de capital, nunca consumo final.

[22] El concepto de bienes preferentes (*merit goods*) se debe a Richard Musgrave (1959) y designa bienes privados cuyo consumo el Estado fomenta por encima de lo que elegirían los consumidores, porque considera que las preferencias individuales infravaloran esos bienes (p. ej. la educación obligatoria).

[23] El ajuste por variación de derechos de pensiones recoge la variación de los derechos por pensiones de los hogares debida a las cotizaciones menos las prestaciones, sin incluir revalorizaciones.

de capital fijo del ahorro bruto se obtiene el *ahorro nacional neto*, que es el que efectivamente incrementa el patrimonio de la nación.

### 2.2.2. Cuentas de acumulación

Las **cuentas de acumulación** registran las variaciones de los activos, pasivos y patrimonio neto. La cuenta de capital y la cuenta financiera recogen las debidas a operaciones, mientras que la cuenta de otras variaciones de los activos recoge las restantes.

La **cuenta de capital** (C.III.1) muestra cómo el ahorro y las transferencias de capital financian la adquisición de activos no financieros. Se divide en dos subcuentas:

**Cuenta de variaciones del patrimonio neto debidas al ahorro y a las transferencias de capital** (C.III.1.1): parte del ahorro neto (recurso) y añade las transferencias de capital recibidas menos las pagadas. Las transferencias de capital comprenden los impuestos sobre el capital, las ayudas a la inversión (como los fondos estructurales de la UE recibidos por los Estados miembros) y las condonaciones de deuda acordadas. El saldo es la *variación del patrimonio neto debida al ahorro y a las transferencias de capital*, que puede entenderse como el ahorro nacional neto corregido por las transferencias de capital.

Tabla A30.7: Cuenta de variaciones del patrimonio neto debidas al ahorro y a las transferencias de capital.

| Variaciones de activos | | Variaciones de pasivos y patrimonio neto | |
|---|---|---|---|
| Transferencias de capital pagadas (D.9) | xxx | Ahorro nacional neto (B.8n) | xxx |
| | | Transferencias de capital recibidas (D.9) | xxx |
| **Saldo: variación del patrimonio neto (PN) debida al ahorro y a las transferencias de capital (B.10.1)** | **xxx** | | |

**Cuenta de adquisiciones de activos no financieros** (C.III.1.2): muestra en qué activos no financieros se emplea ese saldo: formación bruta de capital fijo (incluida la vivienda nueva de los hogares[24]), variación de existencias, adquisiciones menos cesiones de objetos valiosos y adquisiciones netas de activos no producidos (tierra, licencias). Se deduce el consumo de capital fijo para trabajar en términos netos. El saldo es la *capacidad (+) o necesidad (−) de financiación.*

Tabla A30.8: Cuenta de adquisiciones de activos no financieros.

| Variaciones de activos | | Variaciones de pasivos y patrimonio neto | |
|---|---|---|---|
| Formación bruta de capital fijo (P.51g) | xxx | Variación del PN debida al ahorro y a las transferencias de capital (B.10.1) | xxx |

[24] En puridad, la vivienda sería un bien de consumo duradero. Sin embargo, la contabilidad nacional la considera un bien de capital. La compraventa de terrenos no se cuenta como FBKF, pero la mejora de estos sí.

| Variaciones de activos | | Variaciones de pasivos y patrimonio neto |
|---|---|---|
| Consumo de capital fijo (P.51c) | (xxx) | |
| Variación de existencias (P.52) | xxx | |
| Adquisiciones menos cesiones de objetos valiosos (P.53) | xxx | |
| Adquisiciones menos cesiones de activos no producidos (NP) | xxx | |
| **Saldo: capacidad (+) / necesidad (−) de financiación (B.9)** | **xxx** | |

La inversión (FBKF) supone en torno al 20 % del PIB y es mucho más volátil y procíclica que el consumo. En conjunto es coincidente o ligeramente retrasada respecto al ciclo, y solo la inversión residencial suele adelantarse a él. Por su parte, la variación de existencias es pequeña (en torno al 0,5-1 % del PIB) pero de gran varianza. La inversión en sentido macroeconómico es la creación de capital productivo, no la inversión financiera[25], y es el componente que vincula el flujo del producto con el *stock* de riqueza.

La capacidad o necesidad de financiación (CIN) es el saldo central del sistema. Para la economía total, CIN $> 0$ indica que la nación ha ahorrado, sumadas las transferencias netas de capital recibidas, más de lo que ha invertido, y que presta al resto del mundo, lo que mejora su posición de inversión internacional salvo efectos de valoración. En cambio, un CIN $< 0$ indica que ha necesitado financiación exterior.

Por sectores, la suma de las capacidades y necesidades de hogares, ISFLSH, sociedades no financieras, instituciones financieras y AA. PP. es igual a la de la nación, lo que permite leer el saldo exterior como el reflejo de los desequilibrios sectoriales internos. En varias economías de la zona del euro, por ejemplo, la elevada necesidad de financiación previa a la Gran Recesión procedía de hogares y sociedades no financieras. El ajuste posterior descansó en el aumento del ahorro privado, mientras crecía la necesidad de financiación de las AA. PP.

**Cuenta de otras variaciones de los activos** (C.III.3): registra las variaciones de activos y pasivos que no se deben a operaciones. Se divide en la cuenta de otras variaciones del volumen (descubrimiento o agotamiento de recursos naturales, catástrofes, cancelaciones unilaterales de deuda, cambios de clasificación) y la cuenta de revalorización (ganancias y pérdidas de posesión por variaciones de precios de los activos reales y financieros existentes).

### 2.2.3. Balances

Los **balances** recogen, tanto al inicio como al cierre del ejercicio, el patrimonio neto de las unidades junto con sus activos y pasivos, así como las variaciones que se producen a lo largo de este. Se componen de balance inicial, variaciones del balance (que son, exactamente, las registradas en las cuentas de acumulación: operaciones, otras variaciones de volumen y revalorizaciones) y balance final:

$$\mathrm{PN}_t = \mathrm{PN}_{t-1} + \underbrace{\mathrm{B.10.1}}_{\text{ahorro y TK}} + \underbrace{\mathrm{OVV}}_{\text{volumen}} + \underbrace{\mathrm{REV}}_{\text{revalorización}}$$

[25] P. ej. comprar un palacete ya existente no es inversión en términos nacionales; construirlo, sí.

donde $\mathrm{PN}_t$ es el patrimonio neto al cierre, B.10.1 la variación debida al ahorro y a las transferencias de capital, OVV las otras variaciones de volumen y REV las revalorizaciones. Para la economía total, el patrimonio neto es la riqueza nacional (apartado 3.2.2).

## 2.3. Cuentas financieras

**Lema**: cada operación de la economía, ya sea real o financiera, tiene una contrapartida financiera, salvo el trueque, las operaciones en especie y las operaciones internas. Esto tiene tres implicaciones:

1. *Doble registro*: una operación real (venta de un bien) tiene una contrapartida financiera (cobro en efectivo o crédito comercial), y una operación financiera (concesión de un préstamo) tiene otra contrapartida financiera (depósito).

2. *Saldo real y financiero*: el saldo de las cuentas reales de cada unidad es igual al saldo de sus cuentas financieras, porque quien gasta más de lo que ingresa financia la diferencia con pasivos o desprendiéndose de activos financieros.

3. *Economía total*: la suma de las capacidades y necesidades de financiación de los sectores es igual a la suma de sus saldos financieros.

**Cuenta financiera** (C.III.2): recoge en qué instrumentos se ha materializado la capacidad o necesidad de financiación. Su saldo, la *adquisición neta de activos financieros menos el incremento neto de pasivos*, es conceptualmente idéntico al saldo de la cuenta de capital, y la diferencia estadística entre ambos es el ajuste entre las cuentas real y financiera:

$$\mathrm{SB}_{\mathrm{CF}} \equiv \Delta\mathrm{AF} - \Delta\mathrm{PF} = \mathrm{CIN}$$

donde $\mathrm{SB}_{\mathrm{CF}}$ es el saldo de la cuenta financiera, $\Delta$AF la adquisición neta de activos financieros y $\Delta$PF el incremento neto de pasivos. Los instrumentos se clasifican en ocho categorías:

Tabla A30.9: Instrumentos financieros del SEC-2010.

| Código | Instrumento | Ejemplos |
|---|---|---|
| F.1 | Oro monetario y derechos especiales de giro (DEG) | Reservas del banco central |
| F.2 | Efectivo y depósitos | Billetes, depósitos a la vista y a plazo |
| F.3 | Valores representativos de deuda | Letras, bonos, obligaciones |
| F.4 | Préstamos | Hipotecas, crédito bancario |
| F.5 | Participaciones en el capital y en fondos de inversión | Acciones cotizadas y no cotizadas, fondos |
| F.6 | Sistemas de seguros, pensiones y garantías estandarizadas | Reservas técnicas, derechos de pensión |
| F.7 | Derivados financieros y opciones de compra de acciones de los asalariados | Futuros, opciones, permutas |
| F.8 | Otras cuentas pendientes de cobro o pago | Créditos comerciales, anticipos |

Para la economía total, los activos y pasivos financieros entre residentes se cancelan, de modo que el saldo de la cuenta financiera coincide conceptualmente con el saldo de la cuenta financiera de la balanza de pagos:

$$\mathrm{SB}_{\mathrm{CF}}^{\mathrm{CN}} \simeq \mathrm{SB}_{\mathrm{CF}}^{\mathrm{BP}}$$

donde $\mathrm{SB}_{\mathrm{CF}}^{\mathrm{CN}}$ es el saldo de la cuenta financiera de la economía total en las cuentas nacionales y $\mathrm{SB}_{\mathrm{CF}}^{\mathrm{BP}}$ el de la balanza de pagos, que registra las mismas operaciones desde la óptica de los residentes frente a los no residentes.

**Cuenta de otras variaciones de activos y pasivos financieros** (COVAPF): con la misma lógica que la cuenta de otras variaciones de los activos, recoge las variaciones de activos y pasivos financieros no debidas a operaciones. Esas variaciones comprenden los efectos de valoración (tipo de cambio, precio de los activos) y las otras variaciones de volumen (cancelaciones, reclasificaciones):

$$\mathrm{COVAPF} = \mathrm{REV} + \mathrm{OVV}$$

donde REV es el efecto valoración (revalorizaciones) y OVV las otras variaciones de volumen. La COVAPF pertenece a las cuentas internacionales del MBP6 y concilia los flujos de la balanza de pagos con la variación de la posición de inversión internacional neta (PIIN), $\Delta\mathrm{PIIN} = \mathrm{SB}_{\mathrm{CF}}^{\mathrm{BP}} + \mathrm{COVAPF}$.

## 2.4. Lectura económica de los saldos: identidades ahorro-inversión

Los saldos de la secuencia de cuentas tienen una lectura económica directa en términos de ahorro e inversión, cuya forma depende de los agentes considerados. Si se define el ahorro como la renta disponible no destinada al consumo, $S = \mathrm{RNBD} - C - G$:

Tabla A30.10: Identidades ahorro-inversión según el tipo de economía.

| Economía | Renta disponible | Identidad | Lectura |
|---|---|---|---|
| Cerrada sin sector público | $\mathrm{RNBD} = \mathrm{PIB} = C + I$ | $S = I$ | El ahorro privado financia íntegramente la inversión privada[26] |
| Cerrada con sector público | $\mathrm{RNBD} = C+I+G$ | $(S_P - I_P) + (S_G - I_G) = 0$ | Lo que un sector invierte por encima de su ahorro lo cubre el ahorro sobrante del otro |
| Abierta con sector público | $\mathrm{RNBD} = C + I + G + \mathrm{SB}_{\mathrm{CC}}$ | $(S_P - I_P) + (S_G - I_G) = \mathrm{SB}_{\mathrm{CC}}$ | Si la inversión supera al ahorro nacional, la diferencia se cubre con financiación exterior |

donde $S_P$ y $S_G$ son el ahorro privado (hogares, ISFLSH y sociedades) y público, $I_P$ e $I_G$ la inversión privada y pública y $\mathrm{SB}_{\mathrm{CC}}$ el saldo de la balanza por cuenta corriente, suma de los saldos de bienes y servicios $(X - M)$, de rentas primarias y de rentas secundarias.

En la economía abierta, si el ahorro nacional es insuficiente para financiar la inversión nacional, se registra un déficit por cuenta corriente. El saldo de la balanza por cuenta de capital ($\mathrm{SB}_{\mathrm{CK}}$) está formado por las transferencias netas de capital recibidas del exterior (ahorro aportado

[26] Sin sector público ni exterior, la renta disponible de los hogares, $\mathrm{RNBD}_H$, es la renta nacional menos el ahorro de las empresas, $S_F$ (beneficios no distribuidos), y se reparte entre el consumo y el ahorro de los hogares, $S_H$: $\mathrm{RNBD}_H = \mathrm{RNBD} - S_F = C + S_H$. Como $\mathrm{RNBD} = C + I$, resulta $I = S_H + S_F = S$.

desde fuera sin contrapartida) y por las cesiones netas al exterior de activos no financieros no producidos. Si se suma a ambos lados de la identidad de la economía abierta, esta pasa a ser:

$$S + \text{SB}_{\text{CK}} - I = \text{SB}_{\text{CC}} + \text{SB}_{\text{CK}} \equiv \text{CIN}$$

donde $S = S_P + S_G$ e $I = I_P + I_G$ son el ahorro y la inversión nacionales y CIN la capacidad ($> 0$) o necesidad ($< 0$) de financiación de la nación. Con este saldo se cierra la cuenta de capital, y la cuenta financiera muestra cómo se materializa. Leído desde el exterior, es también la suma de los saldos corriente y de capital de la balanza de pagos, cuya interpretación se desarrolla en el tema B11, apartado 3.

De estas identidades se derivan dos precisiones:

1. *Saldo público*: $S_G - I_G$, más las transferencias netas de capital recibidas, es la capacidad o necesidad de financiación de las AA. PP., es decir, el saldo presupuestario, que no debe confundirse con el ahorro público[27]. La descomposición del saldo presupuestario en saldo primario, estructural y cíclico se desarrolla en el tema A39, apartado 1.1.

2. *Identidad y equilibrio*: la identidad ahorro-inversión se cumple *ex post* por definición, cualquiera que sea el valor de las variables. El equilibrio económico, en cambio, exige que los planes de los agentes coincidan *ex ante*, y la diferencia se manifiesta en la variación no deseada de existencias.

## 2.5. Conclusiones

**Aspectos positivos**: la secuencia de cuentas ofrece una descripción completa y coherente de la economía, desde la producción hasta la riqueza. En esa secuencia, cada saldo (PIB, RNB, RNBD, ahorro, capacidad o necesidad de financiación, patrimonio neto) enlaza con el siguiente y cada operación tiene su contrapartida. También su presentación en las *cuentas económicas integradas*, que reúnen en una sola tabla las cuentas de todos los sectores y del resto del mundo, permite una lectura directa de los desequilibrios sectoriales y de su reflejo exterior.

**Limitaciones**: la contabilidad nacional registra lo que pasa por el mercado o puede valorarse por sus costes. Por ello deja fuera el trabajo doméstico no remunerado, cuya valoración en las cuentas satélite de la producción doméstica representa una parte considerable del PIB[28], y las externalidades. Solo capta de forma indirecta e imperfecta la economía sumergida (de peso considerable en muchas economías), pues el SEC-2010 exige exhaustividad y los institutos de estadística incluyen estimaciones de la economía no observada (desde 2014, también de ciertas actividades ilegales). La producción de no mercado, a su vez, se valora por sus costes, lo que no refleja las preferencias sociales por los servicios públicos. Además, los agregados de la contabilidad nacional son indicadores de actividad, no de bienestar, cuestión sobre la que se vuelve en el apartado 3.5.

[27] El ahorro público es la renta disponible de las AA. PP. menos su consumo final.

[28] La producción de servicios por los hogares para su autoconsumo se excluye de la contabilidad, salvo los servicios de las viviendas ocupadas por sus propietarios y los del personal doméstico remunerado.

# 3. Otras cuentas, indicadores y reflexiones

Junto al marco central, el sistema comprende otras cuentas, y sobre él se construyen los principales indicadores macroeconómicos. La sección se cierra con una reflexión sobre el PIB como medida del bienestar.

## 3.1. Otras cuentas

Además de la secuencia de cuentas de los sectores institucionales, el SEC-2010 comprende la cuenta del resto del mundo, las cuentas regionales y otras tablas y cuentas complementarias.

### 3.1.1. Cuenta del resto del mundo (S.2)

La **cuenta del resto del mundo** registra las operaciones entre unidades residentes y no residentes, con las no residentes agrupadas en una única unidad que no constituye un sector institucional pero desempeña un papel análogo. Incluye la *cuenta de intercambios exteriores de bienes y servicios*, la *cuenta exterior de rentas primarias y transferencias corrientes*, la *cuenta de capital* y la *cuenta financiera*.

La cuenta del resto del mundo se corresponde con las cuentas internacionales del MBP6 (balanza de pagos, posición de inversión internacional y COVAPF). La diferencia es que las cuentas del resto del mundo se elaboran desde la óptica de los no residentes y la balanza de pagos desde la de los residentes, de modo que los signos se invierten. Por eso, en las cuentas corriente y de capital, un crédito de la balanza de pagos (una exportación o una renta recibida por un residente) es en la cuenta del resto del mundo un empleo, y un débito, un recurso.

> La estructura de la balanza de pagos y su relación con las cuentas nacionales se desarrollan en el tema B11, apartados 2.1 y 2.3.

### 3.1.2. Contabilidad regional (CRE)

El SEC-2010 es también un sistema de cuentas regionales. El Instituto Nacional de Estadística (INE) elabora la Contabilidad Regional de España, que distribuye el PIB, el valor añadido por ramas, el empleo y la renta de los hogares entre comunidades autónomas y provincias, con base en la localización de las UAE.

De esta contabilidad proceden las comparaciones de PIB per cápita regional que determinan la elegibilidad para los fondos estructurales de la UE (regiones menos desarrolladas, por debajo del 75 % de la media de la UE). La Contabilidad Regional alimenta, además, el debate sobre balanzas fiscales y financiación autonómica.

### 3.1.3. Tablas *input-output* y cuentas satélite

**Tablas de origen y destino e *input-output*** (TIO): describen la producción y los empleos por productos y ramas de actividad y las relaciones intersectoriales de consumos intermedios. Son la base del marco de Leontief y permiten calcular los multiplicadores de producción y empleo de cada rama y el contenido importado de la demanda final.

**Cuentas satélite**: responden a necesidades de información específicas aportando mayor detalle sin alterar el marco central. Destacan las cuentas de la agricultura, del medio ambiente, de la salud, de la producción de los hogares, del turismo, de la I+D y de la economía digital.

## 3.2. Indicadores de la economía en su conjunto: crecimiento, renta y ciclo

Un primer grupo de indicadores se refiere a la economía total: cuánto crece, qué nivel de renta y riqueza alcanza, cómo se reparte y en qué fase del ciclo se encuentra.

### 3.2.1. Crecimiento: índices de volumen encadenados

Tradicionalmente, el crecimiento del PIB real se calculaba como un índice de **ponderaciones fijas**, para el que se fijaba un año base y se valoraba la producción de los años siguientes a los precios de ese año. Este método permite separar el efecto de los precios y de las cantidades, pero el año base pierde representatividad con el tiempo (hay que actualizarlo cada cinco o diez años). Además, con la innovación tecnológica y la creciente sustituibilidad entre bienes, los precios relativos cambian demasiado deprisa como para mantener una base fija.

Ante estas limitaciones, desde mediados de los años noventa (Estados Unidos en 1996 y los países de la UE en la década de 2000) se adoptaron los **índices de volumen encadenados**. Estos valoran la producción de cada año a los precios del año anterior y encadenan los resultados. La metodología se apoya en dos conceptos, el **eslabón** y el **índice encadenado**:

$$\mathrm{EV}_t = \frac{\sum_i p_{i,t-1}\, x_{i,t}}{\sum_i p_{i,t-1}\, x_{i,t-1}} \qquad \mathrm{IV}_t = \mathrm{IV}_{t-1} \cdot \mathrm{EV}_t$$

donde $\mathrm{EV}_t$ es el eslabón, cociente entre el PIB del año $t$ a precios del año anterior y el PIB del año anterior, $p_{i,t-1}$ y $x_{i,t}$ el precio del año anterior y la cantidad del año $t$ del producto $i$, e $\mathrm{IV}_t$ el índice encadenado, producto de todos los eslabones desde el año de referencia, con $\mathrm{IV}_{\mathrm{ref}} = 100$.

**Conclusiones**:

- *Aspectos positivos*: la base se actualiza cada año, lo que resuelve el problema de representatividad y reduce las revisiones.

- *Limitaciones*: los índices encadenados no son aditivos, por lo que la suma de los componentes del PIB en volumen no coincide con el PIB en volumen, salvo en el año de referencia y el siguiente. Esto plantea problemas de difusión e interpretación, por lo que se ha extendido el uso de las *contribuciones al crecimiento*, que sí son aditivas en términos anuales.

### 3.2.2. Renta y riqueza: PIB per cápita, riqueza nacional y distribución

**Riqueza nacional**: es el patrimonio neto de la economía total, una variable *stock* que coincide con la suma de los activos no financieros situados en el territorio y la posición de inversión internacional neta:

$$\mathrm{AN} = K + \mathrm{PIIN} = K + (\mathrm{AF}^{\mathrm{ext}} - \mathrm{PF}^{\mathrm{ext}})$$

donde AN es la riqueza nacional, $K$ el valor de los activos no financieros (capital fijo, existencias, vivienda, tierra y recursos naturales) situados en el país, PIIN la posición de inversión internacional neta, $\text{AF}^{\text{ext}}$ los activos financieros de los residentes frente al exterior y $\text{PF}^{\text{ext}}$ sus pasivos.

Los activos financieros entre residentes se cancelan al consolidar, pues son simultáneamente activo de su poseedor y pasivo de su emisor. En un país con posición de inversión internacional neta deudora, esta reduce la riqueza nacional respecto a los activos no financieros situados en él.

**PIB per cápita y paridades de poder adquisitivo**: el PIB per cápita es el indicador habitual del nivel de desarrollo, y para las comparaciones internacionales se expresa en PPA (apartado 1.4.5). Eurostat, por ejemplo, lo publica en estándares de poder adquisitivo (*purchasing power standards*, PPS), una unidad monetaria artificial que elimina las diferencias de nivel de precios entre países y que sirve de referencia a la política de cohesión.

**Distribución de la renta**: el PIB per cápita es una media que no informa del reparto. Los indicadores habituales son el *índice de Gini* (cociente entre el área comprendida entre la curva de Lorenz y la diagonal y el área total bajo la diagonal, entre 0 y 1) y la *ratio S80/S20* (renta del quintil superior sobre la del inferior). A ellos se suma la *tasa de riesgo de pobreza* (población con renta disponible equivalente inferior al 60 % de la mediana). En las economías europeas, la desigualdad de la renta de mercado es muy superior a la de la renta disponible, lo que refleja el papel redistributivo de las transferencias, especialmente de las pensiones.

### 3.2.3. Ciclo: indicadores de confianza y adelantados

**Indicadores de confianza**: engloban encuestas cualitativas a consumidores y empresas que preguntan por la situación actual y las expectativas (situación financiera del hogar, intención de compra de bienes duraderos, cartera de pedidos, existencias, empleo previsto).

Destacan el *indicador de confianza del consumidor* y el *indicador de sentimiento económico* (*Economic Sentiment Indicator*, ESI) de la Comisión Europea, y los *índices de gestores de compras* (PMI), cuyo umbral de 50 separa expansión y contracción. Su ventaja es la rapidez (se publican antes que las cuentas trimestrales) y su limitación, que miden opiniones y no operaciones.

**Indicadores adelantados** (*leading indicators*): índices compuestos que combinan variables que tienden a anticipar los puntos de giro del ciclo (pedidos industriales, permisos de construcción, pendiente de la curva de tipos, cotizaciones bursátiles, confianza). Destacan los *Composite Leading Indicators* de la OCDE y los modelos de predicción en tiempo real (*nowcasting*) de bancos centrales e instituciones fiscales independientes.

## 3.3. Indicadores de los mercados de trabajo y de bienes: empleo y precios

Un segundo grupo se refiere al funcionamiento del mercado de trabajo, medido en personas, y del de bienes y servicios, medido en precios, los dos grandes mercados sobre los que se asienta la oferta agregada.

### 3.3.1. Mercado de trabajo

A diferencia de los agregados anteriores, el desempleo no se mide en unidades monetarias sino en personas o en porcentaje. Existen dos fuentes para su medición:

**Encuesta a los hogares**: clasifica a la población en edad de trabajar en tres grupos:

- *Ocupados*: personas que en la semana de referencia trabajaron al menos una hora por cuenta ajena o propia, o tenían empleo y estaban temporalmente ausentes.

- *Parados*: personas sin empleo, disponibles para trabajar y que han buscado activamente empleo en las cuatro semanas anteriores.

- *Inactivos*: el resto (estudiantes, jubilados, personas dedicadas al hogar y trabajadores desanimados).

En España es la Encuesta de Población Activa (EPA) del INE, y en EE. UU., la *Current Population Survey* (CPS). De la encuesta se obtienen:

$$u = \frac{U}{N_A} \qquad a = \frac{N_A}{N_{16+}} \qquad e = \frac{E}{N_{16+}}$$

donde $u$ es la tasa de paro, $U$ los parados, $N_A$ la población activa (ocupados más parados), $a$ la tasa de actividad, $N_{16+}$ la población de 16 años o más, $e$ la tasa de empleo y $E$ los ocupados. La tasa de empleo es el indicador de referencia de la Estrategia Europea de Empleo, pues no depende de los movimientos entre paro e inactividad.

**Registros administrativos**: recogen el paro registrado en el Servicio Público de Empleo Estatal (SEPE) y la afiliación a la Seguridad Social. Pueden divergir de la encuesta porque los autónomos y los pluriempleados se contabilizan de forma distinta, porque no todo parado se inscribe (ni todo inscrito es parado en sentido EPA) y por errores de estimación.

**Productividad aparente del trabajo**: se mide como cociente entre el valor añadido en volumen y el empleo (personas, puestos equivalentes a tiempo completo u horas trabajadas). Se denomina "aparente" porque atribuye al trabajo todo el valor añadido, con independencia del capital y de la utilización de la capacidad instalada. Junto a la remuneración por asalariado da lugar a los *costes laborales unitarios* (CLU), indicador central de competitividad.

### 3.3.2. Precios e inflación

**Índice de precios al consumo** (IPC): es la medida del nivel de precios de uso más extendido. Compara lo que cuesta una cesta fija de bienes y servicios de consumo con lo que costaba esa misma cesta en el periodo base. Es un índice de *Laspeyres*:

$$P_L = \frac{\sum_{i=1}^{n} p_{it}\, x_{i0}}{\sum_{i=1}^{n} p_{i0}\, x_{i0}}$$

donde $p_{it}$ y $p_{i0}$ son los precios del bien $i$ en el periodo corriente y en el base y $x_{i0}$ las cantidades de la cesta del periodo base. Para las comparaciones en la UE y como referencia del objetivo del BCE se utiliza el *IPC armonizado* (IPCA), que homogeneiza cobertura y metodología. El IPC presenta tres limitaciones:

1. *Sesgo inflacionista*: suele exagerar el aumento del coste de la vida, porque no recoge el efecto sustitución (supone que los consumidores no alteran sus pautas de consumo ante las variaciones de precios) ni recoge plenamente las mejoras de calidad y los nuevos productos. La Comisión Boskin (1996) estimó este sesgo en torno a 1,1 puntos porcentuales anuales para EE. UU. Lebow y Rudd (2003) lo situaron en torno a 0,9 puntos anuales, tras las mejoras metodológicas de la Oficina de Estadísticas Laborales (BLS) de EE. UU.

2. *Índice plutocrático*: el IPC puede reescribirse como una media de los IPC de los hogares ponderada por su gasto, de modo que las pautas de consumo de los hogares ricos pesan más que las de los pobres[29].

3. *Inflación importada*: el IPC incluye bienes importados (energía, alimentos), por lo que recoge perturbaciones de oferta externas que no reflejan presiones inflacionistas internas. De ahí procede el uso de la *inflación subyacente*, que excluye energía y alimentos no elaborados.

**Deflactor implícito del PIB**: el PIB nominal valora la producción a precios corrientes, $Y_t = \sum_i p_{it} x_{it}$, y el PIB real a precios de un periodo de referencia (hoy, en volumen encadenado)[30]. Su cociente es el deflactor del PIB, un índice de *Paasche* (en sentido estricto, cuando el PIB real se valora a precios de un año base fijo, como en la fórmula):

$$P = \frac{\text{PIB}_{\text{nominal}}}{\text{PIB}_{\text{real}}} = \frac{\sum_{i=1}^{n} p_{it}\, x_{it}}{\sum_{i=1}^{n} p_{i0}\, x_{it}}$$

donde $x_{it}$ son las cantidades del periodo corriente, que actúan como ponderaciones. El deflactor se diferencia del IPC en tres aspectos:

1. *Cobertura*: el deflactor mide los precios de todos los bienes y servicios producidos (consumo, inversión, gasto público y exportaciones), mientras que el IPC solo comprende los adquiridos por los hogares.

2. *Importaciones*: el deflactor solo comprende los bienes producidos en el interior, por lo que una subida del precio de las importaciones no afecta directamente al deflactor, pero sí al IPC.

3. *Ponderaciones*: el IPC (Laspeyres, cesta fija) tiende a sobrestimar el aumento del coste de la vida, mientras que el deflactor (Paasche, cesta corriente) tiende a subestimarlo, pues recoge la sustitución pero no su coste en bienestar. El índice de *Fisher*, media geométrica de ambos, es el índice ideal.

### 3.4. Indicadores sectoriales: sector público, sector financiero y sector exterior

Un tercer grupo se refiere a los sectores institucionales cuyos desequilibrios se transmiten al resto de la economía: las AA. PP., las instituciones financieras y el resto del mundo.

[29] Puede construirse un índice "democrático" que pondere por igual a todos los hogares. La diferencia entre ambos índices es el *gap* plutocrático.

[30] El Bureau of Economic Analysis de EE. UU. utiliza desde 1996 índices encadenados de Fisher.

### 3.4.1. Sector público

**Indicadores de gasto**: se utilizan el gasto público total sobre el PIB y el gasto no financiero sobre el PIB, que es el indicador más utilizado del tamaño del sector público (en torno al 45 % en España). Se usa también el gasto no financiero ajustado, que deduce las transferencias para aproximar la absorción directa de recursos.

**Indicadores de ingreso**: la *presión fiscal* es la recaudación efectiva de impuestos y cotizaciones sobre el PIB[31]:

$$\text{PFis} = \frac{T_{\text{tot}} + \text{CS}}{\text{PIB}} \qquad \text{EF} = \frac{\text{PFis}}{\text{PIB}_{\text{pc}}}$$

donde $T_{\text{tot}}$ son los impuestos totales (sobre la producción y las importaciones, sobre la renta y el patrimonio y sobre el capital) y CS las cotizaciones sociales. Por su parte, EF es el *esfuerzo fiscal* o índice de Frank, que relativiza la presión fiscal por el PIB per cápita para aproximar el sacrificio que supone[32]. La presión fiscal mide lo que efectivamente se paga, por lo que una mayor evasión la reduce y hace que los tipos legales sobrestimen la carga efectiva.

**Indicadores de saldo y deuda**: el saldo presupuestario (capacidad o necesidad de financiación de las AA. PP.), sus componentes primario, estructural y cíclico, y la ratio de deuda sobre el PIB miden la salud de las finanzas públicas. En ellos se basan los criterios de Maastricht (déficit inferior al 3 % y deuda inferior al 60 % del PIB)[33] y el procedimiento de déficit excesivo. Los indicadores se elaboran según el SEC-2010, lo que da al sistema contable un papel jurídico directo, porque la delimitación del sector AA. PP. y el registro por devengo determinan qué unidades y qué operaciones computan en el déficit y la deuda.

> La descomposición del saldo presupuestario y la dinámica de la deuda se desarrollan en el tema A39, apartados 1.1 y 2.1, y el marco de gobernanza fiscal de la UE, en el tema B45, apartado 2.1.

### 3.4.2. Sector financiero

**Agregados monetarios y crédito**: los agregados monetarios (M1, M2 y M3 en el Eurosistema) miden la liquidez en manos del público, y su crecimiento se compara con el de la renta nominal. La desviación de la ratio de crédito al sector privado sobre el PIB respecto a su tendencia (*credit-to-GDP gap*) es el indicador de referencia del ciclo financiero y de la activación del colchón de capital anticíclico.

**Tipos de interés y primas de riesgo**: los tipos oficiales, los interbancarios, como el tipo de interés a corto plazo del euro (€STR) y el euríbor, y los rendimientos de la deuda pública describen la curva de tipos, cuya pendiente es un indicador adelantado. La prima de riesgo soberana (diferencial frente al *Bund* a diez años) y los diferenciales corporativos miden el riesgo percibido y la fragmentación financiera.

[31] El sistema fiscal español recauda en torno al 37 % del PIB (37,3 % en 2023), por encima de la media de la OCDE (33,9 %), según la OCDE (*Revenue Statistics 2024*).

[32] Ejemplo: en 2023 la presión fiscal fue similar en Alemania (38,1 %) y en España (37,3 %), pero el PIB per cápita era de unos 49.500 € en Alemania y de unos 31.000 € en España (datos de la OCDE y Eurostat). El esfuerzo fiscal de España es, por tanto, mayor.

[33] Los criterios de Maastricht son cuatro: estabilidad de precios, convergencia de los tipos de interés a largo plazo (rendimiento de la deuda pública), estabilidad de los tipos de cambio como requisito previo a la entrada, y sostenibilidad de las finanzas públicas (déficit $< 3$ % PIB y deuda $< 60$ % PIB).

**Solidez del sistema financiero**: las ratios de capital ordinario de nivel 1 (CET1), de morosidad, de cobertura de liquidez (LCR) y de financiación estable neta (NSFR) aproximan la vulnerabilidad de las entidades de crédito. Los indicadores de endeudamiento de hogares y sociedades sobre la renta disponible o el PIB miden la de sus balances.

> Los agregados monetarios y los instrumentos de política monetaria se desarrollan en el tema A36, apartado 1.1 y sección 3, y la regulación y supervisión financiera, en el tema B27, sección 3.

### 3.4.3. Sector exterior

**Tipo de cambio efectivo nominal** (TCEN): es la media geométrica ponderada de los tipos de cambio nominales bilaterales, expresados en unidades de moneda extranjera por unidad de moneda nacional (al revés que el tipo de cambio $E$ utilizado en el resto del libro). Un aumento indica, por tanto, una apreciación. La ponderación de cada moneda depende del peso del país en el comercio exterior (con doble ponderación para recoger la competencia en terceros mercados)[34].

**Tipo de cambio efectivo real** (TCER): es el TCEN deflactado por la relación entre los precios o costes nacionales y los de los socios comerciales:

$$\text{TCER} = \text{TCEN} \cdot \frac{P}{P^*}$$

donde $P$ es el índice de precios o costes nacional y $P^*$ la media ponderada de los índices de los socios (IPC, deflactor del PIB, CLU o precios de exportación). Un aumento del TCER indica una apreciación real, es decir, una pérdida de competitividad-precio.

**Indicadores *ex post***: miden el resultado de la competitividad en los flujos comerciales observados:

1. *Cuota de mercado*: peso de las exportaciones del país en las importaciones mundiales, $s^j = X^j / \sum_i M^i$.

2. *Tasa de cobertura*: exportaciones sobre importaciones, $X/M$.

3. *Propensión exportadora*: exportaciones sobre PIB, $X/\text{PIB}$.

4. *Grado de apertura*: suma de exportaciones e importaciones sobre PIB, $(X + M)/\text{PIB}$.

Con el auge de las cadenas globales de valor, ganan relevancia las medidas del *contenido importado de las exportaciones* y de las exportaciones en valor añadido (base TiVA de la OCDE-OMC). Una mayor integración eleva la productividad, pero también la exposición a los *shocks* internacionales.

**Indicadores de reservas**: miden la capacidad de las autoridades para hacer frente a salidas de capital:

1. *Cobertura de importaciones*: reservas equivalentes a al menos tres meses de importaciones.

[34] El BCE publica índices de tipos de cambio efectivos del euro frente a tres grupos: TCE-12 (Australia, Canadá, Dinamarca, Hong Kong, Japón, Noruega, Singapur, Corea del Sur, Suecia, Suiza, Reino Unido y EE. UU.), TCE-17 (los seis Estados miembros de la UE no pertenecientes a la zona del euro y once socios comerciales extracomunitarios) y TCE-40 (el TCE-17 más 23 países adicionales).

2. *Regla de Guidotti-Greenspan*: reservas que cubren el 100 % de la deuda externa a corto plazo.

3. *Cobertura del dinero amplio*: reservas equivalentes al 20 % de M2.

> Los criterios de suficiencia de las reservas se desarrollan en el tema B15, apartado 3.1.1.

### 3.5. Reflexiones: el PIB como medida del bienestar

El PIB se ha convertido en la variable flujo más importante de la economía[35], porque indica en qué medida una sociedad logra cubrir sus necesidades materiales con los recursos de que dispone. Sin embargo, es una medida de actividad, no de bienestar, y desde su origen sus propios impulsores lo advirtieron. Kuznets (1934), por ejemplo, señaló que "el bienestar de una nación difícilmente puede inferirse de una medida de la renta nacional".

**Paradoja de Easterlin** (1974): dentro de un país, los individuos más ricos declaran mayor felicidad, pero a lo largo del tiempo el crecimiento del PIB per cápita no se traduce en aumentos apreciables de la felicidad media declarada. Ello apunta a la importancia de la renta relativa y de la adaptación. Estudios posteriores (p. ej. Stevenson y Wolfers, 2008) matizan la paradoja al encontrar una relación positiva entre renta y satisfacción también entre países y en el tiempo, si bien con rendimientos decrecientes.

**Comisión Stiglitz-Sen-Fitoussi** (2009): a instancias del Gobierno francés, recomendó complementar el PIB con medidas de la renta y el consumo de los hogares (más que de la producción), de la distribución, de las actividades no de mercado, de la calidad de vida en sus múltiples dimensiones (salud, educación, medio ambiente, seguridad) y de la sostenibilidad (*stocks* de capital físico, humano y natural). Esta agenda ha reforzado la iniciativa Beyond GDP de la Comisión Europea (lanzada en 2007) y ha inspirado el *Better Life Index* de la OCDE (2011) y los indicadores de la Agenda 2030.

**Índices compuestos**: el índice de desarrollo humano (IDH) del Programa de las Naciones Unidas para el Desarrollo (PNUD) combina renta per cápita, esperanza de vida y educación, y sus variantes ajustan por desigualdad (IDH-D) y por género. Sus limitaciones son la arbitrariedad de las ponderaciones y la alta correlación de sus componentes con el propio PIB per cápita.

En resumen, el PIB no debe abandonarse sino leerse con sus límites:

- *Cobertura*: excluye el trabajo no remunerado y el ocio, y capta de forma imperfecta la economía sumergida.
- *Gasto defensivo*: contabiliza como positiva la actividad defensiva (reparación de daños, gasto sanitario por contaminación).
- *Sostenibilidad*: ignora la depreciación del capital natural.
- *Distribución*: no informa del reparto de la renta.

La respuesta de la contabilidad nacional ha sido ampliar el marco (cuentas satélite, cuentas de distribución de la renta de los hogares, cuentas del capital natural) antes que sustituir el agregado.

[35] Hasta los años noventa se utilizaba sobre todo el producto nacional bruto (PNB), y el PIB se impuso después por su coherencia con otras variables referidas al territorio (p. ej. el desempleo) y por la comparabilidad internacional.

## Conclusiones

La contabilidad nacional proporciona el marco cuantitativo del análisis macroeconómico. El SEC-2010, en línea con el SCN-2008, concreta el flujo circular de la renta en un sistema que registra las operaciones y posiciones de los sectores institucionales con el criterio de residencia, por partida doble, por devengo y a precios de mercado. Su secuencia de cuentas encadena los saldos, del PIB a la RNB y la RNBD y de estas al ahorro, a la capacidad o necesidad de financiación y al patrimonio neto. De ella se derivan las identidades ahorro-inversión, que en una economía abierta vinculan los desequilibrios sectoriales internos con el saldo frente al resto del mundo. Los indicadores construidos sobre las cuentas comparten, sin embargo, la limitación del agregado del que parten, porque miden la actividad y no el bienestar.

Las identidades contables son el punto de partida de la macroeconomía del libro, porque cualquier teoría ha de respetarlas. El enfoque del gasto ordena el estudio del consumo y de la inversión (temas A33 y A34), y la brecha entre ahorro e inversión es la otra cara del saldo por cuenta corriente (tema B11). El saldo de las AA. PP. en términos del SEC-2010 es la medida del déficit que vigilan las reglas fiscales (tema A39). Por último, los índices de precios permiten medir la inflación (tema A41), y el PIB en volumen, descompuesto entre factores y productividad, es la base de la contabilidad del crecimiento (tema A45).

El debate actual se centra en adaptar la medición a la digitalización, la globalización y la sostenibilidad. El SCN-2025 (Naciones Unidas, 2025) reconoce los datos como activo producido y registra el agotamiento de los recursos naturales como un coste de producción. Incorpora, además, desgloses por deciles de la renta, el consumo y la riqueza de los hogares, en la línea de las cuentas nacionales distributivas de Piketty, Saez y Zucman (2018). Brynjolfsson, Collis y Eggers (2019) estiman con experimentos de elección el valor que los usuarios atribuyen a los servicios digitales gratuitos, que el PIB no recoge porque carecen de precio. Por último, como las grandes filiales de multinacionales separan el PIB irlandés de la renta de sus residentes, la oficina estadística de Irlanda publica desde 2017 una renta nacional bruta modificada (*GNI**) que corrige esos efectos.

# Anexo. Ejemplos resueltos

### Ejemplo: una exportación

Una empresa residente produce en el periodo bienes por valor de 100, sin consumos intermedios, y los vende a un no residente, que paga con una transferencia a una cuenta de la empresa en un banco no residente. Para la economía total, y sin otras operaciones:

- *Producción y gasto*: la producción (P.1) y el PIB aumentan en 100. Por el lado del gasto, el aumento son exportaciones (P.6), de modo que $X - M$ sube en 100.

- *Renta y ahorro*: la renta generada (salarios y excedente) eleva en 100 la RNB y la RNBD. Si no se consume en el periodo, el ahorro nacional aumenta en 100.

- *Capital*: sin inversión, la capacidad de financiación (B.9) aumenta en 100.

- *Financiera*: la empresa adquiere un depósito frente al resto del mundo (F.2), un activo financiero de 100, y $\Delta\text{AF} - \Delta\text{PF} = 100 = \text{CIN}$.

En la balanza de pagos, la misma operación se anota dos veces, como exportación de bienes en la cuenta corriente y como adquisición de un activo exterior en la cuenta financiera.

### Ejemplo: inversión bruta e inversión neta

Una empresa residente compra a otra residente una máquina producida en el periodo por 50 y la paga con sus depósitos. El *stock* de capital existente se deprecia en el periodo en 10 (CCF).

- *Producción y gasto*: la producción de la máquina aumenta el PIB en 50, que por el lado del gasto es formación bruta de capital fijo (P.51g).

- *Renta y ahorro*: la renta generada al producirla, si no se consume, eleva en 50 el ahorro nacional bruto.

- *Capital*: la inversión neta es la FBKF menos el CCF, $I_N = 50 - 10 = 40$, que es lo que crece el *stock* de capital, $K_{t+1} - K_t = 40$. El ahorro y la FBKF aumentan en lo mismo, de modo que la capacidad de financiación no cambia, y el CCF, que se resta de ambos en términos netos, tampoco la altera.

- *Financiera*: los depósitos pasan de una empresa residente a otra, y para la economía total se cancelan.

La inversión amplía la capacidad productiva solo en la parte que supera la reposición del capital que se desgasta.

### Ejemplo: la concesión de un préstamo

Un banco residente concede a un hogar residente un préstamo de 30 y lo abona en la cuenta corriente del hogar.

- *Banco*: adquiere un activo, el préstamo (F.4), y contrae un pasivo, el depósito del hogar (F.2), ambos de 30, de modo que $\Delta\text{AF} - \Delta\text{PF} = 0$.

- *Hogar*: adquiere un depósito (F.2) y contrae una deuda (F.4) de 30, y su saldo financiero tampoco cambia.

- *Economía total*: el préstamo no altera la producción, la renta, el ahorro ni la capacidad de financiación, porque es una operación financiera con contrapartida financiera (apartado 2.3), que cambia la composición de los balances, no el patrimonio neto. Si el prestatario fuera un no residente, la economía adquiriría un activo frente al resto del mundo y, a la vez, un pasivo por el depósito del no residente en el banco residente, con el mismo saldo nulo.

Solo cuando el hogar gaste los 30 en consumo aparecerá un efecto real, financiado con deuda en lugar de con renta.

# A32. Precios y output en economía abierta

El modelo de oferta y demanda agregada: determinación de renta e inflación en una economía abierta. Análisis de las políticas monetaria y fiscal, y de los *shocks* y políticas de oferta.

## Introducción

En una economía abierta, la renta y los precios dependen también de lo que ocurre fuera de las fronteras. El grado de **apertura**, comercial y financiera, determina qué parte de la demanda se filtra al exterior a través de las importaciones y cuánto capital entra o sale cuando cambian los tipos de interés. El **tipo de cambio** se convierte así en un precio macroeconómico central, que relaciona los precios interiores con los exteriores y transmite las perturbaciones de un país a otro. Por su parte, el **régimen cambiario** decide qué parte del ajuste recae en el tipo de cambio y qué parte en la renta, el empleo o los precios.

De esa interdependencia se deriva una restricción que condiciona la política macroeconómica de toda economía abierta, el **trilema de Mundell-Fleming** (o trinidad imposible), según el cual un país no puede mantener a la vez un tipo de cambio fijo, la libertad de movimientos de capital y una política monetaria autónoma (tema B15). Por ello, la eficacia de las políticas monetaria y fiscal deja de ser una cuestión interior y pasa a depender del régimen cambiario y del grado de movilidad del capital.

El análisis moderno de estas cuestiones arranca a comienzos de los años sesenta, cuando Fleming (1962) y Mundell (1963) extendieron el modelo IS-LM al sector exterior. Ambos mostraron que, con movilidad perfecta del capital, la política monetaria es eficaz con tipos de cambio flexibles y la fiscal con tipos fijos. Más tarde, Dornbusch (1976) combinó expectativas racionales y precios rígidos para explicar la sobrerreacción del tipo de cambio (tema B14), y las perturbaciones de oferta de los años setenta llevaron a completar el análisis con la oferta agregada, que determina a la vez la renta y el nivel de precios. Desde los años noventa, la nueva macroeconomía de economía abierta dio microfundamentos a ese esquema, y el modelo neokeynesiano de economía pequeña y abierta de Galí y Monacelli (2005) es hoy la referencia para analizar la política monetaria con una regla de tipos de interés.

La exposición sigue el orden en que se construyó la teoría, que es también el de los horizontes temporales:

- *Demanda agregada con precios dados*: a corto plazo, el modelo IS-LM-BP de Mundell y Fleming hace del sector exterior y del régimen cambiario los determinantes de la eficacia de la política económica.

- *Oferta y demanda agregadas*: la ausencia de oferta y de expectativas en ese modelo lleva al modelo de oferta y demanda agregadas (OA-DA) en economía abierta, que determina a la vez la renta y el nivel de precios y permite analizar las perturbaciones y políticas de oferta.

- *Inflación y brecha de producción*: como el modelo OA-DA es estático y determina el nivel de precios y no la inflación, se completa con el modelo IS-PC-MR, dinámico y con una regla de tipos.

# 1. La demanda agregada en economía abierta: el modelo IS-LM-BP (Fleming, 1962; Mundell, 1963)

**Origen**: el modelo IS-LM-BP, debido a Fleming (1962) y Mundell (1963), es el modelo canónico de la economía abierta a corto plazo.

**Idea**: el modelo extiende el IS-LM de la síntesis neoclásica añadiendo un tercer mercado, el sector exterior, y con él dos elementos nuevos, el tipo de cambio y el régimen cambiario.

> El modelo IS-LM en economía cerrada se presenta en el tema A05, apartado 1.2.1.

## 1.1. Supuestos

**Entorno**: la economía es abierta y pequeña, y las variables exteriores (nivel de precios, $P^*$, tipo de interés, $i^*$, y renta, $Y^*$) son exógenas. El horizonte es el corto plazo, con precios fijos ($\bar{P}$) e inflación nula ($\pi = 0$), lo que iguala los tipos de interés nominal y real ($i = r$) y hace proporcionales los tipos de cambio nominal y real.

Con precios dados, la demanda determina la producción (la oferta agregada es perfectamente elástica). El *tipo de cambio*, $E$, se define de forma directa, en unidades de moneda nacional por unidad de moneda extranjera, y un aumento de $E$ es una depreciación.

**Agentes**: son neutrales al riesgo (no exigen prima por mantener activos en distintas monedas) y tienen expectativas estáticas sobre el tipo de cambio.

**Mercados**: el modelo incluye los de bienes (IS), dinero (LM) y sector exterior o mercado de divisas (BP). Los flujos de capital responden al diferencial de tipos de interés con una intensidad $\gamma$, que mide la movilidad del capital. En el límite de movilidad perfecta, cuando los bonos nacionales y extranjeros son sustitutivos perfectos, se cumple la paridad de intereses.

**Sector exterior**: la cuenta corriente se identifica con la balanza comercial ($\text{SB}_{\text{CC}} = \text{BC}$), sin rentas primarias ni transferencias, y se cumple la condición Marshall-Lerner ($\varepsilon_X + |\varepsilon_M| > 1$), de modo que una depreciación mejora la balanza comercial[1]. La cuenta financiera, excluidos los activos de reserva, recoge los flujos autónomos de capital, medidos como entradas netas, con el signo opuesto al del saldo de la cuenta financiera en el Manual de Balanza de Pagos del FMI (MBP6). En régimen de cambio fijo, la variación de reservas ($\Delta R$) cierra la balanza.

> La condición Marshall-Lerner, la curva J y el enfoque absorción se desarrollan en el tema B12, apartado 1; la paridad de intereses, en el tema B13, apartados 2.2.3 y 2.2.4, y en el tema B14, apartado 1.3.

## 1.2. Modelo

**Planteamiento**: el modelo se resume en tres condiciones de equilibrio, una por mercado.

[1] Con precios fijos, las elasticidades de oferta de exportaciones e importaciones son infinitas, de modo que la condición se reduce a que la suma de las elasticidades de demanda, en valor absoluto, supere la unidad, partiendo de una balanza comercial equilibrada.

### 1.2.1. Mercado de bienes: curva IS

**Definición**: la *curva IS* es el lugar geométrico de las combinaciones de renta y tipo de interés ($Y$, $r$)[2] que garantizan el equilibrio en el mercado de bienes, es decir, la igualdad entre producción y demanda agregada, $Y = C + I + G + \text{XN}$. En economía abierta, la demanda agregada incorpora las exportaciones netas o balanza comercial (demanda externa), XN ≡ BC. Sus ecuaciones de comportamiento son:

- *Consumo*: función keynesiana simple, sin efecto riqueza, $C = C_0 + c\,(1-\tau)\,Y$, donde $C_0$ es el consumo autónomo, $c \in (0,1)$ la propensión marginal a consumir sobre la renta disponible y $\tau$ el tipo impositivo sobre la renta.

- *Inversión*: $I = I_0 - v\,r$, con un componente autónomo $I_0$ y una sensibilidad $v > 0$ al tipo de interés.

- *Gasto público*: $G = G_0$, exógeno.

- *Balanza comercial*: $\text{BC} = X - M = x\,Y^* - m\,Y + \phi\,e$. Las exportaciones dependen positivamente de la renta exterior (con propensión marginal a exportar $x > 0$) y del *tipo de cambio real*, $e = EP^*/P$. Las importaciones dependen positivamente de la renta nacional (con propensión marginal a importar $m \in (0,1)$) y negativamente del tipo de cambio real. El parámetro $\phi$ recoge el efecto neto del tipo de cambio real sobre la balanza comercial, positivo bajo la condición Marshall-Lerner. La condición garantiza que los efectos volumen (más exportaciones y menos importaciones) dominen el efecto valoración (las importaciones cuestan más en moneda nacional).

Sustituyendo en la condición de equilibrio y despejando la renta se obtiene la ecuación de la curva IS:

$$Y = \frac{1}{1 - c(1-\tau) + m}\left[C_0 + I_0 + G_0 - v\,r + x\,Y^* + \phi\,e\right]$$

**Determinantes**: la expresión tiene dos factores:

- *Multiplicador*: el primero es el *multiplicador de la economía abierta*, $1/[1 - c(1-\tau) + m]$, menor que el de la economía cerrada, $1/[1-c(1-\tau)]$. Es menor porque a la filtración fiscal ($\tau$) el denominador añade la exterior ($m$, la parte del gasto adicional que se dirige a bienes importados y no genera renta nacional). El multiplicador aumenta con $c$ y disminuye con $\tau$ y con $m$.

- *Gasto autónomo*: el segundo factor suma la demanda autónoma interna ($C_0 + I_0 + G_0$), la externa ($x\,Y^*$ y $\phi\,e$) y el componente que depende del tipo de interés ($-v\,r$), que no es autónomo.

**Desarrollo gráfico**: en el plano $(Y, r)$ (figura A32.1), la IS tiene pendiente negativa, $\partial r/\partial Y = -[1-c(1-\tau)+m]/v < 0$, porque una bajada del tipo de interés estimula la inversión y, a través del multiplicador, eleva la renta de equilibrio. La IS es tanto más inclinada cuanto menores son $v$ y el multiplicador, y el *efecto filtración* de las importaciones ($m > 0$) la hace más inclinada que la de la economía cerrada. Por eso, la misma bajada del tipo de interés, de $r_0$ a $r_1$, lleva la economía del punto 0 al 1 y eleva la renta solo hasta $Y_1$, frente a $Y_1^{\text{c}}$ sobre la IS (cerrada).

[2] Con precios fijos y sin inflación esperada, el tipo de interés real coincide con el nominal, $r = i$, y por eso la LM, la BP y el equilibrio se representan con $i$. La distinción importa en el apartado 2, donde el nivel de precios esperado actúa sobre la demanda a través del tipo de interés real (efecto Fisher).

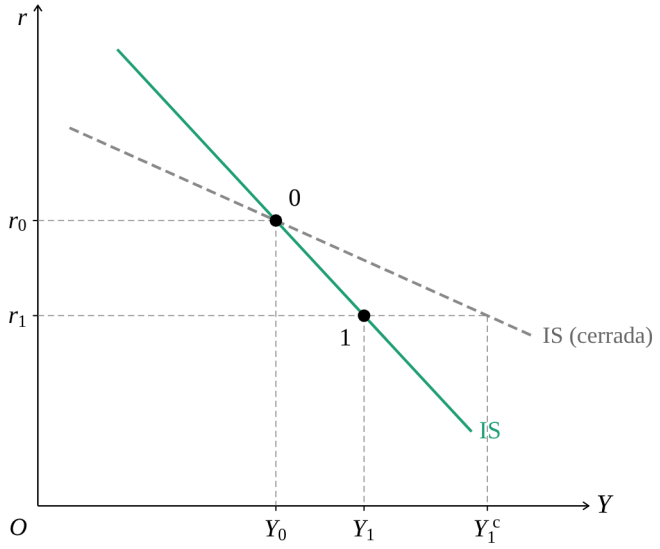



Figura A32.1: Curva IS en economía abierta. La filtración de las importaciones reduce el multiplicador y hace la IS más inclinada que la de la economía cerrada, de modo que la misma bajada del tipo de interés, de $r_0$ a $r_1$, eleva menos la renta ($Y_1 < Y_1^c$).

**Estática comparativa**:

▪ *Cambios de posición* (variables autónomas): desplazan la IS a la derecha los aumentos de $C_0$, $I_0$, $G_0$ e $Y^*$ y una depreciación real ($\uparrow e$), y a la izquierda, los movimientos contrarios.

▪ *Cambios de pendiente* (parámetros): una subida de $\tau$, por ejemplo, reduce el multiplicador y hace rotar la IS hacia dentro.

### 1.2.2. Mercado de dinero: curva LM

**Definición**: la *curva LM* es el lugar geométrico de las combinaciones de renta y tipo de interés ($Y$, $i$) que garantizan el equilibrio en el mercado de dinero, es decir, la igualdad entre la oferta real de dinero y la demanda de saldos reales:

$$\frac{M_0^s}{\bar{P}} = k\,Y - h\,i \quad \Rightarrow \quad i = \frac{1}{h}\left(k\,Y - \frac{M_0^s}{\bar{P}}\right)$$

**Determinantes**:

▪ *Demanda de dinero*: depende positivamente de la renta a través de $k$ (motivos transacción y precaución) y negativamente del tipo de interés a través de $h$ (motivo especulación), como se desarrolla en el tema A35.

▪ *Oferta de dinero*: la oferta monetaria nominal la fija el banco central y, con precios fijos, la oferta real queda determinada. Con tipo fijo, sin embargo, $M^s$ tiene un componente endógeno que depende del saldo de la balanza de pagos. Un superávit expande la base monetaria, porque

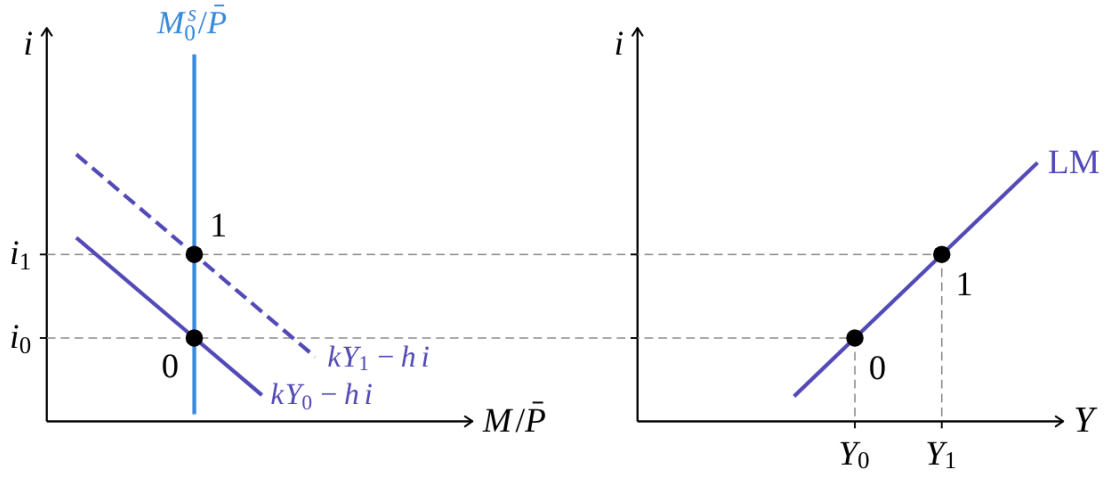
el banco central compra divisas, y un déficit la contrae, salvo que el banco central esterilice[3]. En cambio, con tipo flexible, $M^s$ es exógena.

**Desarrollo gráfico**: la derivación de la curva LM se representa en el plano saldos reales-tipo de interés y en el plano $(Y, i)$, que comparten el eje vertical (figura A32.2, paneles a y b):

1. *Oferta real de dinero* $M_0^s/\bar{P}$: es vertical, porque la fija el banco central y los precios están dados.

2. *Demanda de saldos reales* $kY - h\,i$: es decreciente en el tipo de interés y, al crecer la renta de $Y_0$ a $Y_1$, se desplaza de $kY_0 - h\,i$ a $kY_1 - h\,i$ por el motivo transacción.

3. *Equilibrio y curva LM*: con una oferta real dada, el equilibrio del mercado de dinero pasa del punto 0 al 1 solo si el tipo de interés sube de $i_0$ a $i_1$. La LM une esos pares de renta y tipo de interés en el panel (b), con pendiente positiva, $\partial i/\partial Y = k/h > 0$. La apertura no altera la pendiente, solo el control de la oferta.




(a) Mercado de dinero (b) Curva LM

Figura A32.2: Derivación de la curva LM. Con una oferta real de dinero dada, el aumento de la renta de $Y_0$ a $Y_1$ eleva la demanda de saldos reales y exige un tipo de interés más alto, $i_1$, para mantener el equilibrio del mercado de dinero.

**Estática comparativa**:

- *Cambios de posición*: un aumento de $M^s$ desplaza la LM a la derecha, y una contracción monetaria, a la izquierda.

[3] La esterilización consiste en compensar el efecto de la variación de reservas sobre la base monetaria con operaciones de mercado abierto de signo contrario. Solo es posible de forma transitoria, mientras el banco central disponga de reservas (si el déficit persiste) o de títulos que vender (si el superávit persiste).

▪ *Cambios de pendiente*: una mayor sensibilidad de la demanda de dinero a la renta ($\uparrow k$) hace la LM más rígida, mientras que una mayor sensibilidad al tipo de interés ($\uparrow h$) la hace más plana.

### 1.2.3. Sector exterior: curva BP

**Definición**: la *curva BP* es el lugar geométrico de las combinaciones de renta y tipo de interés ($Y$, $i$) compatibles con el *equilibrio externo*, es decir, con un saldo nulo de la balanza de pagos, suma de la cuenta corriente y la cuenta financiera:

$$\begin{aligned} \text{BP} &= \underbrace{(x\,Y^* - m\,Y + \phi\,e)}_{\text{SB}_{\text{CC}}} + \underbrace{\gamma\,(i - i^*)}_{\text{entradas netas de capital}} = 0 \\ &\Rightarrow\ i = \frac{1}{\gamma}\left[m\,Y - x\,Y^* - \phi\,e\right] + i^* \end{aligned}$$

**Determinantes**:

▪ *Cuenta corriente*: es la balanza comercial, con los mismos determinantes que en la IS.

▪ *Cuenta financiera*: recoge las entradas netas de capital, que crecen con el diferencial de tipos de interés, $i - i^*$, según la movilidad del capital, $\gamma \geq 0$.

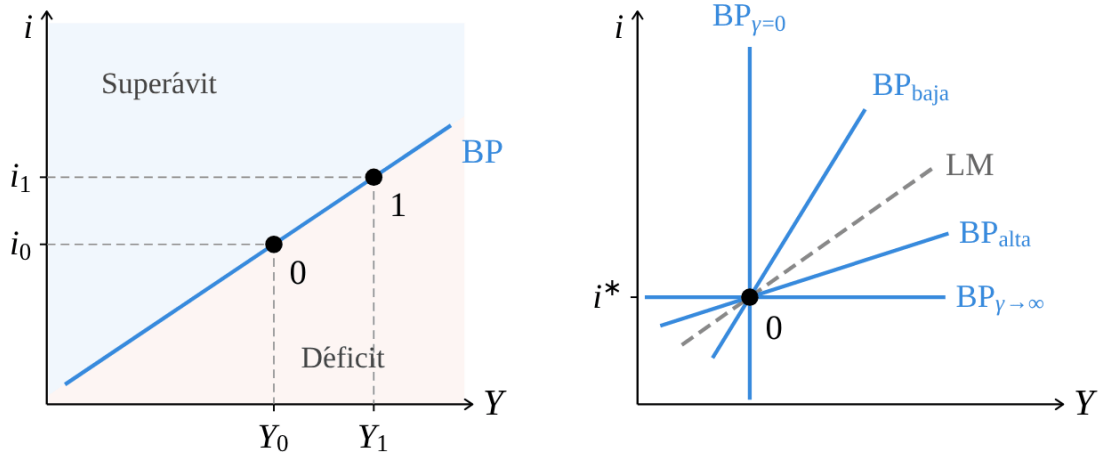
▪ *Desequilibrios*: los puntos por encima de la curva representan superávit de la balanza de pagos (el tipo de interés atrae más capital del preciso para financiar la cuenta corriente), y los puntos por debajo, déficit.

**Desarrollo gráfico**: en el plano $(Y, i)$ (figura A32.3), la BP tiene pendiente $\partial i/\partial Y = m/\gamma \geq 0$. Como las importaciones crecen con la renta, cada aumento de esta empeora la balanza comercial, y solo un tipo de interés más alto, que atraiga capital, restablece el equilibrio externo (paso de 0 a 1 en el panel a). Según la movilidad del capital, se distinguen tres casos (panel b):

1. *Movilidad nula* ($\text{BP}_{\gamma=0}$): la BP es vertical.

2. *Movilidad imperfecta* ($0 < \gamma < \infty$): la BP tiene pendiente positiva. Hay *alta movilidad* ($\text{BP}_{\text{alta}}$) cuando la BP es más plana que la LM ($m/\gamma < k/h$) y *baja movilidad* ($\text{BP}_{\text{baja}}$) cuando es más inclinada ($m/\gamma > k/h$).

3. *Movilidad perfecta* ($\text{BP}_{\gamma\to\infty}$): la BP es horizontal en $i = i^*$.




(a) Saldo de la balanza de pagos

(b) Movilidad del capital

Figura A32.3: Curva BP. En el panel (a), por encima de la curva hay superávit de balanza de pagos y, por debajo, déficit, y un aumento de la renta exige un tipo de interés más alto para mantener el equilibrio externo. En el panel (b), la BP se aplana al aumentar la movilidad del capital, de vertical con movilidad nula a horizontal en $i^*$ con movilidad perfecta, y es más plana que la LM con movilidad alta y más inclinada con movilidad baja.

**Estática comparativa**:

▪ *Cambios de posición*: cualquier mejora autónoma de la balanza de pagos, como una depreciación real ($\uparrow e$) o un aumento de $Y^*$, desplaza la BP hacia abajo o a la derecha, y una subida de $i^*$, en cambio, la desplaza hacia arriba. Las medidas que afectan a la cuenta corriente (aranceles, depreciaciones) desplazan la BP y la IS, pues la balanza comercial aparece en ambas.

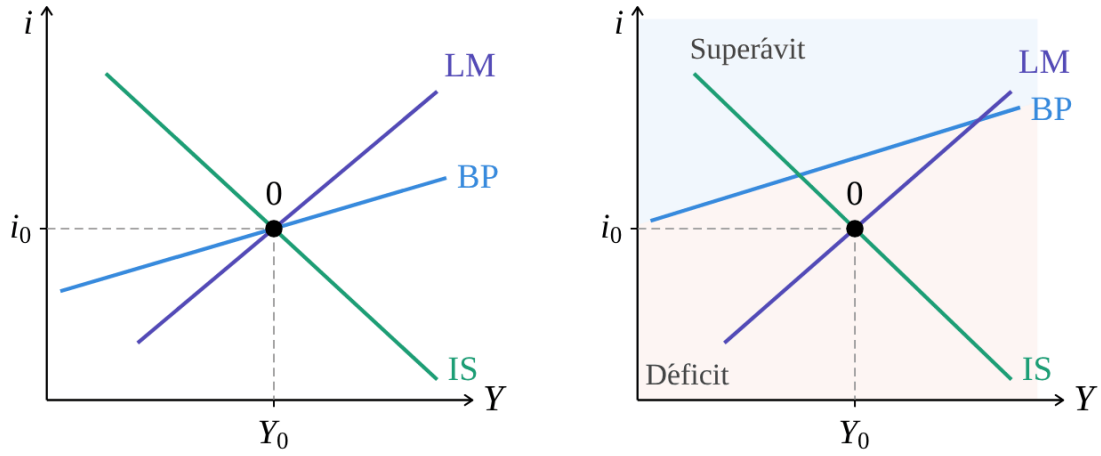
▪ *Cambios de pendiente*: una mayor movilidad del capital ($\uparrow \gamma$) aplana la BP, mientras que una mayor propensión a importar ($\uparrow m$) la hace más rígida.

### 1.2.4. Equilibrio: equilibrio interno y externo

El equilibrio se sitúa donde se cortan las tres curvas y fija tres variables endógenas. Si el tipo de cambio flota, son $Y$, $i$ y el propio tipo de cambio nominal, cuyo lugar ocupa la oferta monetaria si es fijo. En ese punto se cumplen simultáneamente el **equilibrio interno** (mercados de bienes y de dinero) y el equilibrio externo (balanza de pagos).

Gráficamente, en el plano $(Y, i)$ (figura A32.4), las tres curvas se cortan en el punto 0 del panel (a). Puede darse, sin embargo, equilibrio interno sin equilibrio externo. Si la intersección de IS

y LM se sitúa por debajo de la BP, como en el panel (b), hay déficit de balanza de pagos, y si se sitúa por encima, superávit.




(a) Equilibrio interno y externo

(b) Equilibrio interno con déficit

Figura A32.4: Equilibrio del modelo IS-LM-BP. En el panel (a), las tres curvas se cortan en el punto 0 y se cumplen a la vez el equilibrio interno y el externo. En el panel (b), la IS y la LM se cortan por debajo de la BP, de modo que hay equilibrio interno con déficit de balanza de pagos.

El modo en que se corrige el desequilibrio depende del régimen cambiario:

- *Tipo de cambio flexible*: la balanza de pagos está siempre en equilibrio, porque el tipo de cambio se ajusta hasta que la oferta y la demanda de divisas se igualan. Un déficit incipiente deprecia la moneda, lo que mejora la balanza comercial (desplaza IS y BP a la derecha) hasta restablecer el equilibrio.

- *Tipo de cambio fijo*: las autoridades deben evitar la depreciación, así que el ajuste recae sobre la oferta monetaria. Un déficit se financia con venta de reservas (que contrae la base monetaria y desplaza la LM a la izquierda) o con endeudamiento exterior. También puede corregirse con medidas de ajuste sobre la balanza comercial (reducción de la absorción, desviación del gasto) o sobre la cuenta financiera[4].

[4]La financiación mediante endeudamiento exterior tiene un límite en el deterioro de la calificación crediticia y en lo que Eichengreen y Hausmann (1999) llamaron el *pecado original*. El pecado original es la imposibilidad de endeudarse en el exterior en moneda nacional, que expone al país al riesgo de tipo de cambio. Las medidas de ajuste, por su parte, son costosas, porque la reducción de la absorción es contractiva y muchas medidas de desviación del gasto (subvenciones a la exportación) no tienen cabida en la OMC.

Las políticas de financiación y ajuste de un desequilibrio exterior (enfoques de elasticidades, absorción y monetario) se desarrollan en el tema B12; los mecanismos de ajuste propios de cada régimen cambiario, en el tema B15, apartado 1.4.

## 1.3. Políticas monetaria y fiscal con precios fijos

Los resultados de la política económica dependen del régimen cambiario y del grado de movilidad del capital. Dos políticas, dos regímenes y cuatro grados de movilidad (perfecta, alta, baja y nula) dan dieciséis casos. Los cuatro casos de referencia suponen perfecta movilidad del capital, la más próxima a la realidad de las economías avanzadas, y después se generalizan a movilidad imperfecta. Por último, se tratan el país grande y el trilema de Mundell-Fleming.

### 1.3.1. Tipo de cambio flexible

**Política monetaria expansiva** (**eficaz**): en el plano $(Y, i)$ (figura A32.5), el aumento de la oferta monetaria ($\Delta M^s > 0$) tiene un *efecto expansivo interno* que reduce el tipo de interés y estimula la inversión. La LM se desplaza de $\mathrm{LM}_0$ a $\mathrm{LM}_1$ y la economía pasa del punto 0 al $0'$, con $i_{0'} < i^*$. Al efecto interno le sigue un *efecto expansivo externo*. La salida de capital deprecia la moneda ($\Delta E > 0$), y la depreciación mejora la balanza comercial (la IS se desplaza de $\mathrm{IS}_0$ a $\mathrm{IS}_1$) hasta que el tipo de interés vuelve a igualarse con el exterior en el punto 1. En el equilibrio final la inversión recupera su nivel inicial y todo el aumento de la renta, de $Y_0$ a $Y_1$, descansa en las exportaciones netas. La política monetaria es plenamente eficaz por la depreciación que induce.

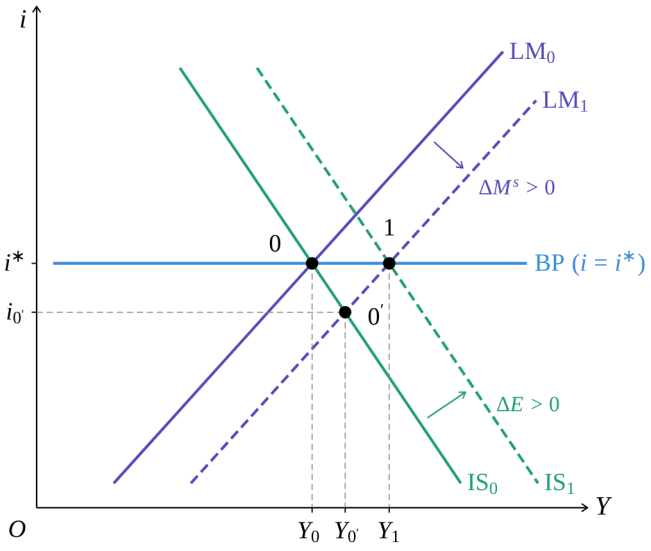



Figura A32.5: Mundell-Fleming con tipo de cambio flexible. La expansión monetaria deprecia la moneda y es plenamente eficaz.

**Política fiscal expansiva** (**ineficaz**): en el plano $(Y, i)$ (figura A32.6), el aumento del gasto público ($\Delta G_0 > 0$) tiene un efecto expansivo interno que eleva la renta y, con ella, la demanda de

dinero y el tipo de interés. La IS se desplaza de $\mathrm{IS}_0$ a $\mathrm{IS}_1$ y la economía pasa del punto 0 al $0'$, con $i_{0'} > i^*$. Al efecto interno le sigue un *efecto contractivo externo*. La entrada de capital aprecia la moneda ($\Delta E < 0$), y la apreciación deteriora la balanza comercial (la IS retrocede) hasta que el tipo de interés vuelve a igualarse con el exterior. Con perfecta movilidad, el retroceso es completo. En el equilibrio final, 1, la renta vuelve a su nivel inicial ($Y_0 = Y_1$) y el gasto público ha desplazado íntegramente a las exportaciones netas, con lo que el efecto expulsión (*crowding out*) exterior es pleno.

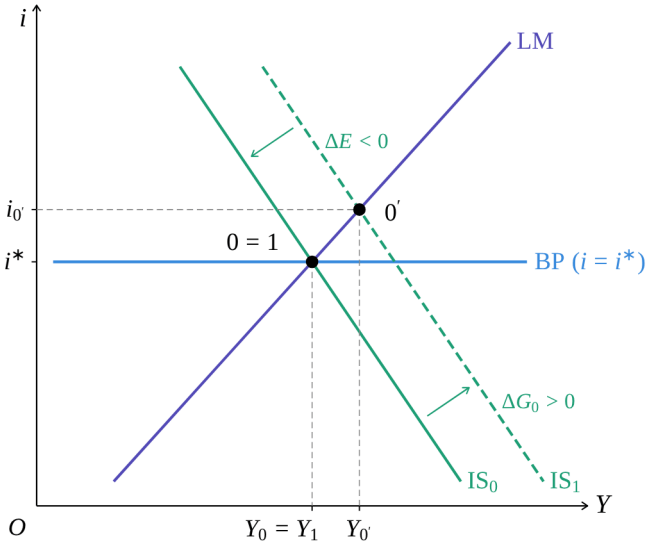



Figura A32.6: Mundell-Fleming con tipo de cambio flexible y perfecta movilidad de capitales. La expansión fiscal aprecia la moneda y es ineficaz sobre la renta.

### 1.3.2. Tipo de cambio fijo

Los resultados se invierten, porque el ajuste de la balanza de pagos recae ahora sobre la oferta monetaria.

**Política monetaria expansiva** (**ineficaz**): en el plano $(Y, i)$ (figura A32.7), la expansión ($\Delta M^s > 0$) desplaza la LM de $\mathrm{LM}_0$ a $\mathrm{LM}_1$ y presiona a la baja el tipo de interés (punto $0'$, con $i_{0'} < i^*$). La bajada provoca una salida de capital y presiona la moneda hacia la depreciación. Para defender la paridad, el banco central vende reservas ($\Delta R < 0$) y contrae la oferta monetaria hasta que la LM vuelve a $\mathrm{LM}_0$. En el equilibrio final, 1, la renta no varía ($Y_0 = Y_1$) y el único efecto es una pérdida de reservas. Por tanto, con tipo de cambio fijo y libre movilidad de capital, el banco central pierde el control de la oferta monetaria.

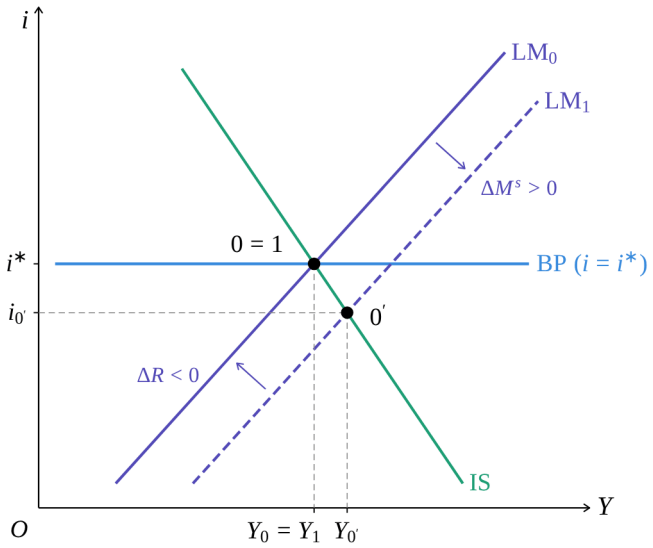



Figura A32.7: Mundell-Fleming con tipo de cambio fijo. La expansión monetaria provoca pérdida de reservas y es ineficaz.

**Política fiscal expansiva** (**eficaz**): en el plano $(Y, i)$ (figura A32.8), el estímulo ($\Delta G_0 > 0$) desplaza la IS de $\text{IS}_0$ a $\text{IS}_1$, eleva la renta y presiona al alza el tipo de interés (punto $0'$, con $i_{0'} > i^*$). La subida atrae capital y presiona la moneda hacia la apreciación. Para defender la paridad, el banco central compra divisas ($\Delta R > 0$) y expande la oferta monetaria (la LM se desplaza de $\text{LM}_0$ a $\text{LM}_1$) hasta que el tipo de interés vuelve al nivel exterior en el punto 1. Como el tipo de interés no sube, no hay expulsión de la inversión y la renta aumenta en todo el efecto multiplicador del gasto público, de $Y_0$ a $Y_1$, y la política fiscal es así plenamente eficaz.

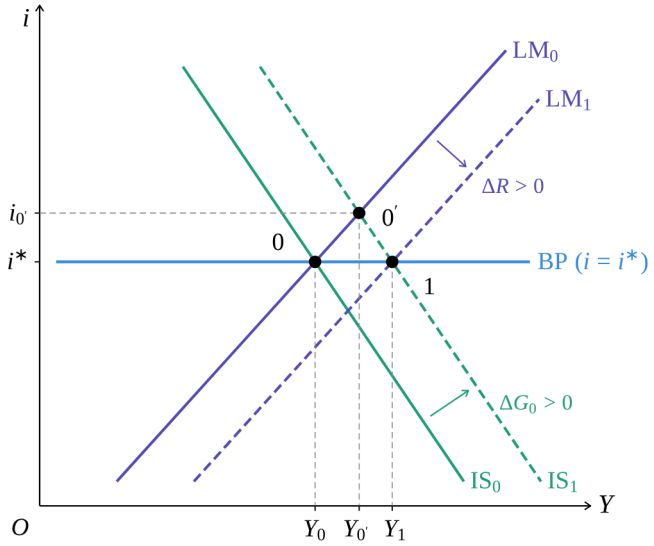



Figura A32.8: Mundell-Fleming con tipo de cambio fijo. La expansión fiscal obliga a una expansión monetaria acomodaticia y es plenamente eficaz.

### 1.3.3. Movilidad imperfecta del capital: los dieciséis casos

Con movilidad imperfecta, el signo del saldo de la balanza de pagos tras una expansión fiscal depende de la pendiente relativa de la BP y la LM. Con alta movilidad domina la entrada de capital y hay superávit, mientras que con baja movilidad domina el deterioro comercial y hay déficit. Una expansión monetaria, por el contrario, genera déficit con cualquier grado de movilidad, pues reduce el tipo de interés y eleva la renta.

- *Tipo de cambio flexible*: la política monetaria es siempre eficaz (con movilidad nula, por el canal de la inversión) y tanto más cuanto mayor es la movilidad del capital, porque mayor es la depreciación que induce. La eficacia de la política fiscal disminuye con la movilidad, porque con movilidad nula o baja el déficit comercial deprecia la moneda y refuerza el estímulo. Con alta movilidad, en cambio, la entrada de capital aprecia la moneda y lo compensa en parte (con movilidad perfecta, por completo).

- *Tipo de cambio fijo*: la política monetaria es ineficaz con cualquier grado de movilidad, salvo esterilización transitoria, porque el déficit que genera obliga a revertirla. La política fiscal, en cambio, es eficaz siempre que exista alguna movilidad del capital, y tanto más cuanto mayor es esta. Con alta movilidad, el superávit expande la oferta monetaria y refuerza el estímulo, y con baja movilidad el déficit la contrae y lo compensa en parte. Con movilidad nula, la renta queda anclada por el equilibrio comercial (la BP es vertical) y la contracción monetaria inducida anula el estímulo.

En suma, el régimen cambiario decide qué política es eficaz y la movilidad del capital decide cuánto. Con perfecta movilidad, además, ambas políticas son alternativas y no complementarias.

Tabla A32.1: Eficacia de las políticas monetaria y fiscal según el régimen cambiario y la movilidad del capital.

| Movilidad del capital | Flexible: política monetaria | Flexible: política fiscal | Fijo: política monetaria | Fijo: política fiscal |
|---|---|---|---|---|
| Nula ($\gamma = 0$) | Eficaz (mín.) | Eficaz (máx.) | Ineficaz | Ineficaz |
| Baja (BP más inclinada que LM) | Eficaz | Eficaz | Ineficaz | Poco eficaz |
| Alta (BP más plana que LM) | Muy eficaz | Poco eficaz | Ineficaz | Muy eficaz |
| Perfecta ($\gamma \to \infty$) | Eficaz (máx.) | Ineficaz | Ineficaz | Eficaz (máx.) |

### 1.3.4. País grande y trilema de Mundell-Fleming

> Los efectos desbordamiento entre economías grandes se desarrollan en el tema B19, apartado 1.2.2, y la coordinación internacional de políticas, en ese mismo tema.

**País grande**: si la actividad de la economía afecta a la del resto del mundo, aparecen efectos desbordamiento. Son positivos en la política fiscal expansiva (*efecto locomotora*, vía demanda de importaciones) y negativos en la monetaria expansiva con tipo flexible y perfecta movilidad (*empobrecimiento del vecino*, vía depreciación competitiva), que puede desembocar en guerras de divisas[5].

**Trilema de Mundell-Fleming** o trinidad imposible: es imposible lograr simultáneamente un tipo de cambio fijo, la perfecta movilidad del capital y una política monetaria independiente. Cada combinación obliga a renunciar al tercer vértice:

- *Tipo fijo y libre movilidad*: fijar un tipo de interés distinto del exterior comprometería la paridad, de modo que se renuncia a la política monetaria (Hong Kong).
- *Política monetaria independiente y libre movilidad*: se renuncia al tipo fijo (Estados Unidos).
- *Tipo fijo y política monetaria independiente*: se renuncia a la libre movilidad mediante controles de capital (China durante décadas).

El trilema está en línea con la regla de Tinbergen, según la cual el número de objetivos no puede superar el de instrumentos independientes.

> El trilema como criterio de elección del régimen cambiario se desarrolla en el tema B15, apartado 1.1, y como aplicación de la regla de Tinbergen, en el tema B17, apartado 1.5.

[5] El empobrecimiento del vecino se invocó contra los programas de expansión cuantitativa (*quantitative easing*, QE) de la Reserva Federal y del BCE en la década de 2010, que las economías emergentes denunciaron como guerras de divisas. Los desbordamientos de la política fiscal, sin embargo, no son siempre positivos. La expansión fiscal alemana tras la reunificación, acompañada del endurecimiento monetario del Bundesbank, se cita como ejemplo de desbordamiento negativo vía tipos de interés.

## 1.4. Conclusiones

**Aspectos positivos**: el modelo IS-LM-BP tiene utilidad pedagógica por su simplicidad y ofrece la intuición fundamental del corto plazo en economía abierta. Asimismo, sus resultados siguen siendo la referencia con la que la Parte B razona la elección de régimen cambiario, las crisis de balanza de pagos y la integración monetaria.

**Limitaciones**: las críticas se agrupan en tres categorías:

- *Metodológicas*: es un análisis de equilibrio general incompleto, que ignora los efectos de la acumulación de activos exteriores y de deuda pública sobre la riqueza. Es, además, estático, sin dinámica del ajuste ni retardos, y está sujeto a la crítica de Lucas, pues mezcla parámetros estructurales con coeficientes de forma reducida que cambian con el régimen de política.

- *De especificación*: prescinde de la oferta agregada, con lo que solo puede explicar el nivel de renta y no el de precios. Prescinde también de las expectativas (de precios y de tipo de cambio) que, como muestra la paridad de intereses, gobiernan los flujos de capital. Por último, hace de los tipos de interés y de cambio los únicos nexos entre el sector real y el nominal.

- *Empíricas*: la evidencia de panel de Ilzetzki, Mendoza y Végh (2013) respalda la predicción del modelo sobre la política fiscal (multiplicador relativamente elevado con tipo fijo y nulo o reducido con tipo flexible). Sin embargo, muestra que la magnitud del multiplicador depende de factores que el modelo ignora, como el nivel de desarrollo o de deuda pública (tema A40, apartado 3.2).

# 2. El modelo OA-DA en economía abierta: determinación de la renta y los precios

**Origen**: el esquema de oferta y demanda agregadas se consolida como herramienta docente en los años setenta, cuando las perturbaciones de oferta hicieron insuficiente el modelo IS-LM, que suponía precios constantes. La oferta agregada se fundamenta en la función de Lucas, de la nueva macroeconomía clásica (NMC), o en el modelo de negociación de Layard y Nickell, de la nueva economía keynesiana (NEK) de primera generación (NEK1G). La versión de economía abierta que aquí se sigue es la de Bajo y Monés (2000).

**Idea**: el modelo de oferta y demanda agregadas levanta el supuesto de precios fijos y determina simultáneamente la renta y el nivel de precios.

## 2.1. Supuestos

El modelo mantiene los supuestos del IS-LM-BP (país pequeño, neutralidad al riesgo, Marshall-Lerner) salvo el de precios fijos y añade otros sobre la formación de precios y expectativas.

**Entorno**: la economía es abierta y pequeña, con horizonte de medio plazo y flexibilidad de precios y salarios. La movilidad del capital es alta, pero no perfecta, de modo que el tipo de interés nacional puede separarse transitoriamente del exterior. Coexisten dos índices de precios. El deflactor del PIB, $P$, recoge solo los bienes de producción nacional. El índice de precios de consumo, $P_c$, incluye además los bienes importados, que se suponen sustitutivos imperfectos de los nacionales, y se define como media geométrica ponderada:

$$P_c = P^{\sigma}\,(E\,P^{*})^{1-\sigma}, \qquad \sigma \in (0,1)$$

donde $\sigma$ es el peso de los bienes nacionales en la cesta de consumo, y el valor esperado del IPC es $P_c^e = (P^e)^{\sigma}(E^e P^{*e})^{1-\sigma}$[6]. Las empresas solo emplean *inputs* nacionales, lo que aísla el efecto del tipo de cambio sobre la oferta de trabajo (el efecto sobre los costes se analiza con el *shock* de precios energéticos).

**Agentes**: el modelo descansa en una asimetría entre empresas y trabajadores. Las empresas contratan según el deflactor, pues venden solo bienes nacionales, mientras que los trabajadores negocian y deciden su oferta de trabajo según el IPC esperado, pues consumen también bienes importados. La formación de expectativas depende de la escuela, con expectativas racionales y problema de extracción de señales en la NMC e inercia nominal en la fijación de precios y salarios en la NEK1G.

## 2.2. Modelo

**Planteamiento**: el modelo combina una demanda agregada común a ambas escuelas y una oferta agregada cuya fundamentación difiere según la escuela.

[6] En una economía importadora de petróleo, por ejemplo, el IPC incluye los carburantes importados, que no forman parte del deflactor del PIB. La media geométrica implica que, en logaritmos ($s = \ln E$), $p_c = \sigma p + (1-\sigma)(s + p^{*})$, de modo que una variación del precio nacional se traslada al IPC solo en la proporción $\sigma$.

### 2.2.1. Demanda agregada

**Definición**: la *demanda agregada* es el lugar geométrico de las combinaciones de renta y nivel de precios ($Y$, $P$) que garantizan el equilibrio simultáneo en los mercados de bienes y de dinero. Se obtiene introduciendo la LM en la IS del modelo IS-LM-BP y haciendo endógeno el nivel de precios, que ya aparecía a través de los saldos reales, del tipo de cambio real y, vía efecto Fisher, del tipo de interés real:

$$Y^d = \text{DA}(\underbrace{A_0^+,\, M_0^{s+},\, P^{e+},\, \tau^-,\, P^-}_{\text{economía cerrada}};\ \underbrace{Y^{*+},\, E^+,\, P^{*+}}_{\text{apertura}})$$

**Determinantes**:

- *Positivos*: el gasto autónomo ($A_0$), la oferta monetaria ($M_0^s$), el nivel de precios esperado ($P^e$) y las variables exteriores ($Y^*$, $E$, $P^*$), pues cualquier mejora de la balanza comercial favorece la demanda de bienes nacionales.

- *Negativos*: el tipo impositivo ($\tau$) y el nivel de precios ($P$). Un aumento de $P^e$, en cambio, estimula la demanda por el efecto Fisher, pues para un tipo nominal dado reduce el tipo de interés real esperado.

**Desarrollo gráfico**: en el plano $(Y, P)$ (figura A32.9), la DA tiene pendiente negativa por dos canales. En el canal del tipo de interés (efecto Keynes), común a la economía cerrada, un mayor $P$ reduce la oferta real de dinero, eleva el tipo de interés y contrae la inversión. En el canal del tipo de cambio real, propio de la economía abierta, un mayor $P$ aprecia el tipo de cambio real ($EP^*/P$ cae) y deteriora la balanza comercial. Por este segundo canal, la DA de economía abierta es más elástica (más plana) que la de economía cerrada. Por eso, la misma subida del nivel de precios, de $P_0$ a $P_1$, lleva la economía del punto 0 al 1 y reduce más la demanda que sobre la DA (cerrada) ($Y_1 < Y_1^c$).

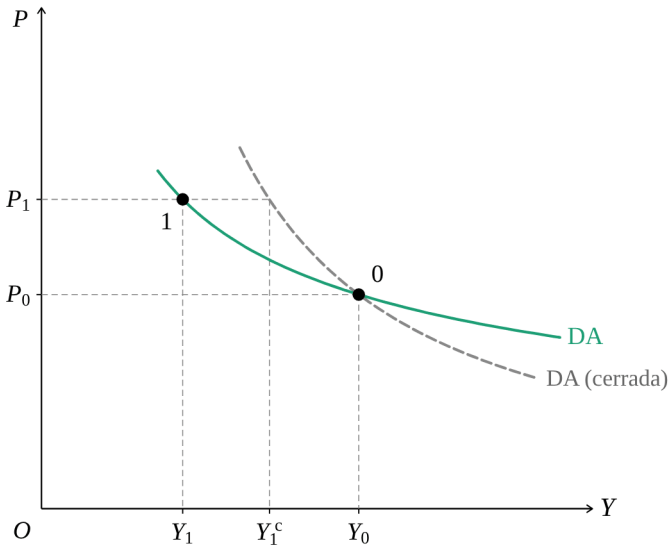



Figura A32.9: Demanda agregada en economía abierta. Al canal del tipo de interés se suma el del tipo de cambio real, de modo que la DA es más plana que la de la economía cerrada y la misma subida del nivel de precios, de $P_0$ a $P_1$, reduce más la demanda ($Y_1 < Y_1^c$).

**Estática comparativa**:

▪ *Cambios de posición*: desplazan la DA los determinantes distintos de $P$, en el sentido de su signo.

▪ *Cambios de pendiente*: dependen de los parámetros de la IS y la LM (multiplicador, $v$, $h$, $k$) y del grado de apertura. Cuanto mayor es el peso de los bienes importados (menor $\sigma$), más sensible es la demanda al tipo de cambio real y más plana la DA.

### 2.2.2. Oferta agregada de la NMC (Lucas, 1972-1975)

**Definición**: en economía cerrada, la función de oferta de Lucas es $y_t = \bar{y}_t + \zeta\,(p_t - p_t^e) + \varepsilon_t^s$, con $\zeta > 0$. Solo las sorpresas de precios afectan al *output*, de modo que la oferta agregada a largo plazo, cuando $P = P^e$, es vertical. En economía abierta la función de oferta incorpora, junto a las variables de la economía cerrada, las que determinan el IPC esperado:

$$Y^s = \mathrm{OA}(K^+,\, P^+,\, P^{e-};\, E^{e-},\, P^{*e-}) \quad \Leftrightarrow \quad Y^s = \mathrm{OA}(K^+,\, (P/P_c^e)^+)$$

**Determinantes**:

▪ *Positivos*: la capacidad productiva ($K$), que amplía la producción para cualquier nivel de empleo, y el nivel de precios ($P$), que reduce el salario real que pagan las empresas.

▪ *Negativos*: el nivel de precios esperado ($P^e$) y el componente importado del IPC esperado ($E^e$, $P^{*e}$). Los tres elevan el IPC esperado y, con él, el salario nominal que exigen los trabajadores para cualquier nivel de empleo, lo que reduce el empleo de equilibrio y la oferta.

**Desarrollo gráfico**: en el plano empleo-salario nominal y en el plano $(Y, P)$ (figura A32.10, paneles a y b), la OA tiene pendiente positiva en cualquier horizonte. Su origen está en el mercado de trabajo del panel (a), donde las empresas demandan trabajo según $w/P$ y los trabajadores lo ofrecen según $w/P_c^e$. Considérese un aumento del nivel de precios nacional perfectamente anticipado ($\Delta p = \Delta p^e$, en logaritmos), con $E^e$ y $P^{*e}$ dados:

1. *Demanda de trabajo* ($L_0^d$): el salario real para las empresas cae en $\Delta p$, lo que eleva la demanda de trabajo (la curva se desplaza hacia arriba en $\Delta p$, hasta $L_1^d$).

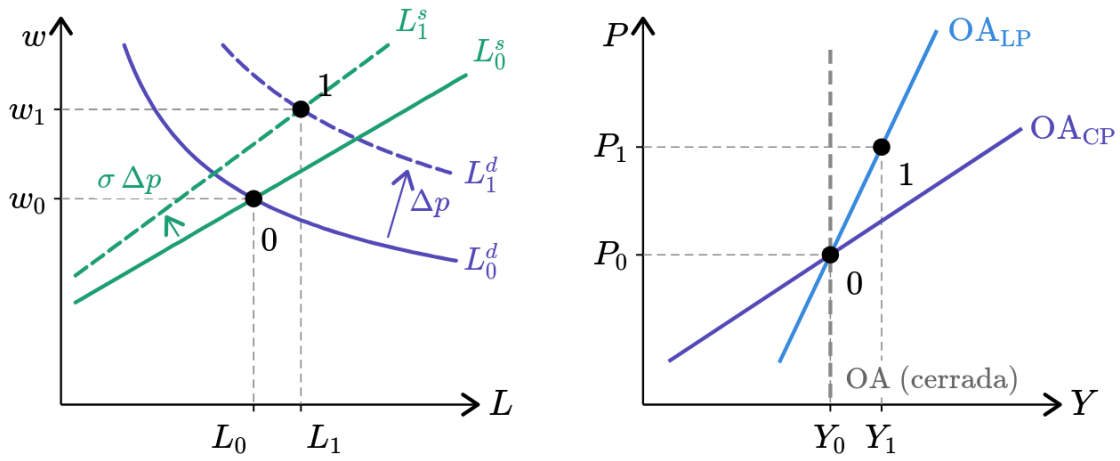
2. *Oferta de trabajo* ($L_0^s$): el IPC esperado solo sube $\Delta p_c^e = \sigma\, \Delta p^e < \Delta p$, ya que los bienes importados, una fracción $1 - \sigma$ de lo que consumen los trabajadores, mantienen su precio. La curva sube solo $\sigma\, \Delta p$, hasta $L_1^s$, y la oferta de trabajo retrocede menos de lo que avanza la demanda.

3. *Equilibrio*: pasa del punto 0 al 1, con un aumento neto del empleo (de $L_0$ a $L_1$) y de la producción (de $Y_0$ a $Y_1$ en el panel b), de modo que la OA tiene pendiente positiva incluso sin sorpresas de precios[7].

- *Largo plazo*: en economía abierta no existe un único nivel de empleo y producción determinado por la tecnología, sino tantos como niveles del tipo de cambio real, y las fluctuaciones del tipo de cambio real alteran la tasa natural de paro. Por eso, la $\text{OA}_{\text{LP}}$ no es vertical como la OA (cerrada).

- *Corto plazo*: con $K$ dado y con errores de expectativas por el problema de extracción de señales, la $\text{OA}_{\text{CP}}$ es más plana que la $\text{OA}_{\text{LP}}$, como en la economía cerrada.

[7] El resultado exige $\sigma < 1$, porque si toda la cesta de consumo fuera nacional ($\sigma = 1$), el IPC esperado y el deflactor variarían en la misma proporción y la OA de largo plazo sería vertical, como en la economía cerrada.




(a) Mercado de trabajo

(b) Oferta agregada

Figura A32.10: Oferta agregada de la NMC en economía abierta. En el panel (a), un aumento anticipado del nivel de precios desplaza la demanda de trabajo en $\Delta p$ y la oferta solo en $\sigma\,\Delta p$, y el empleo aumenta de $L_0$ a $L_1$. En el panel (b), la oferta agregada tiene pendiente positiva también a largo plazo, frente a la vertical de la economía cerrada, y es más plana a corto plazo.

**Estática comparativa**: los aumentos de capacidad productiva ($\uparrow K$) desplazan la OA a la derecha, y los del IPC esperado, que elevan las demandas salariales, a la izquierda.

> La función de oferta de Lucas y el problema de extracción de señales se desarrollan en el tema A06, apartado 2.1, y en el tema A42, apartado 2.1.

### 2.2.3. Oferta agregada de la NEK1G (Layard y Nickell, 1985)

**Definición**: se parte del modelo de negociación con una ecuación de salarios, la del salario real negociado, SRN : $w/P_c^e = f(u^-, z^+)$, que en economía abierta se fija en términos del IPC esperado. La otra ecuación, de precios, es la del salario real pagado, SRP : $w/P = A_L/(1+\mu)$, que las empresas fijan en términos del deflactor. La oferta agregada resultante es:

$$Y^s = \text{OA}(A_L^+,\, P^+,\, P^{e-},\, \mu^-,\, z^-;\; E^{e-},\, P^{*e-})$$

**Determinantes**:

- *Positivos*: la productividad del trabajo ($A_L$), que eleva el salario real que las empresas están dispuestas a pagar para cualquier nivel de empleo, y el nivel de precios ($P$).

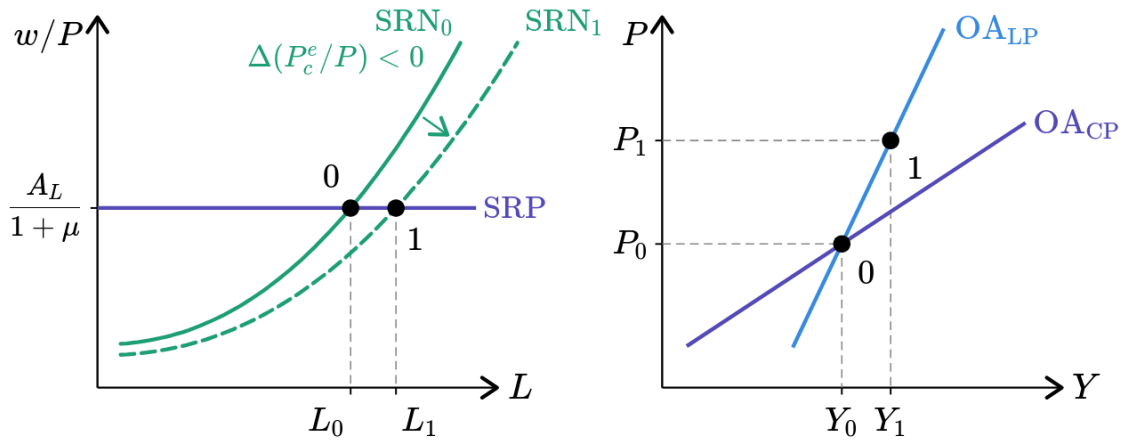
- *Negativos*: el nivel de precios esperado ($P^e$), el margen empresarial ($\mu$), los factores institucionales que elevan el poder de negociación sindical ($z$, como las prestaciones, el salario mínimo o la cuña fiscal) y el componente importado del IPC esperado ($E^e$, $P^{*e}$).

**Desarrollo gráfico**: en el plano empleo-salario real y en el plano $(Y, P)$ (figura A32.11, paneles a y b), la OA tiene pendiente positiva por un mecanismo análogo al de la NMC. La NEK, sin embargo, incorpora poder de mercado ($\mu$, $z$) y rigideces reales, mientras que la NMC supone competencia perfecta. A corto plazo, la inercia nominal en salarios y precios genera pendiente positiva, porque ante un exceso de demanda las empresas aumentan la producción antes de que los precios se ajusten. A largo plazo, la asimetría deflactor/IPC la mantiene positiva:

1. *Salario real pagado* (SRP): es horizontal en $A_L/(1+\mu)$, porque las empresas trasladan íntegramente el salario a los precios, de modo que el salario real en términos del deflactor queda fijado por el margen.

2. *Salario real negociado* ($\text{SRN}_0$): es creciente, porque el aumento del empleo eleva las reivindicaciones salariales. Como el IPC esperado sube menos que el deflactor, la subida de $P$ reduce $P_c^e/P$ y desplaza hacia abajo el SRN expresado en términos del deflactor, hasta $\text{SRN}_1$. El salario fijado por el SRP compra más en términos del IPC, y los trabajadores aceptan un mayor empleo.

3. *Equilibrio*: pasa del punto 0 al 1, de modo que el empleo pasa de $L_0$ a $L_1$ y la producción, de $Y_0$ a $Y_1$ en el panel (b), sobre una $\text{OA}_{\text{LP}}$ de pendiente positiva y más inclinada que la $\text{OA}_{\text{CP}}$.




(a) Salarios y precios (b) Oferta agregada

Figura A32.11: Oferta agregada de la NEK1G en economía abierta. En el panel (a), la subida del nivel de precios reduce el cociente entre el IPC esperado y el deflactor, desplaza hacia abajo el SRN expresado en términos del deflactor y eleva el empleo de equilibrio para un SRP dado. En el panel (b), la oferta agregada tiene, por ello, pendiente positiva también a largo plazo.

**Estática comparativa**: los determinantes distintos de $P$ desplazan la OA en el sentido de su signo. A diferencia de la NMC, aquí los factores institucionales del mercado de trabajo alteran la posición de la oferta, lo que da cabida a las políticas de oferta.

> El modelo de negociación y la tasa de paro que no acelera la inflación (NAIRU) se desarrollan en el tema A28, apartado 2.2.

### 2.2.4. Propiedades de la oferta agregada en economía abierta

Se extraen tres propiedades comunes a ambas escuelas:

1. *Pendiente positiva a largo plazo*: la asimetría deflactor/IPC impide que la OA de largo plazo sea vertical, a diferencia de la economía cerrada, aunque la OA de largo plazo es más inclinada que la de corto plazo.

2. *Efecto filtración*: sea cual sea el horizonte, la apertura aplana la OA frente a la de la economía cerrada, y el aplanamiento aumenta con el peso de los bienes importados en el consumo $(1-\sigma)$[8].

3. *Capacidad productiva*: la oferta se desplaza con $K$ en la NMC y con $A_L$ en la NEK, variables que canalizan tanto las perturbaciones como las políticas de oferta.

### 2.2.5. Equilibrio

El equilibrio determina simultáneamente el nivel de precios y la renta en la intersección de la DA y la OA, y, a través de ellos, el tipo de interés, el empleo, el salario real y el tipo de cambio real.

Como en el IS-LM-BP (apartado 1.2.4), con tipo fijo el saldo de la balanza de pagos altera la oferta monetaria y desplaza la DA hasta que el equilibrio interno es compatible con el externo. Con tipo flexible, en cambio, se ajusta el tipo de cambio nominal, que, a través del IPC esperado, desplaza también la OA. Cada ejercicio de estática comparativa exige, por tanto, seguir dos rondas, la del desplazamiento inicial de la curva afectada y la del desplazamiento inducido por el ajuste exterior.

## 2.3. Políticas monetaria y fiscal

En cada régimen, flexible o fijo, se distingue el corto plazo, donde las dos escuelas difieren, del largo plazo, donde coinciden.

### 2.3.1. Tipo de cambio flexible

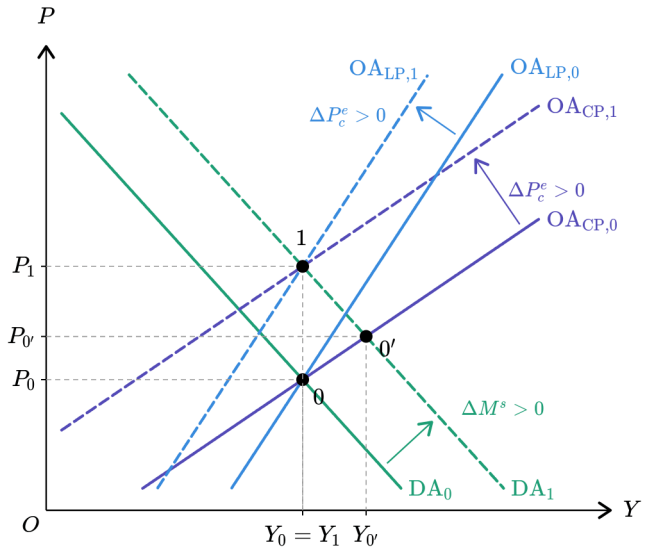
**Política monetaria expansiva**: la DA se desplaza a la derecha y la moneda se deprecia, lo que refuerza el estímulo a corto plazo como en el IS-LM-BP (paso del punto 0 al 0′ en el panel (a) de la figura A32.12). Sin embargo, la depreciación eleva el IPC esperado y desplaza la OA a la izquierda, y el aumento del nivel de precios erosiona los saldos reales.

Con el paso del tiempo, la expansión monetaria se traslada en idéntico porcentaje a precios y tipo de cambio nominal. Ni el tipo de cambio real ni los saldos reales ni la renta acaban variando, y se reproduce así la neutralidad propia de la economía cerrada.

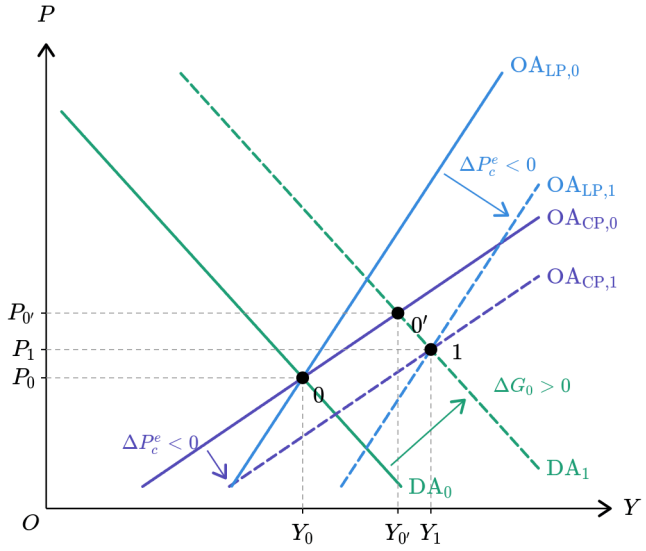
[8] Véase la comparación formal de las pendientes para la NMC y la NEK1G en Bajo y Monés (2000).

**Política fiscal expansiva**: desplaza la DA a la derecha (panel (b) de la figura A32.12) y, si la movilidad del capital es alta (BP más plana que LM, apartado 1.3.3), aprecia la moneda al atraer capital, efecto que en el IS-LM-BP con movilidad perfecta anulaba el estímulo. Aquí operan, además, la subida del nivel de precios, que contrae la demanda, y la apreciación nominal, que reduce el IPC esperado y desplaza la OA a la derecha.

En consecuencia, a largo plazo la política fiscal tiene efectos expansivos y es más eficaz que en la economía cerrada. Frente al IS-LM-BP con movilidad imperfecta, en cambio, no puede afirmarse *a priori* si es más o menos efectiva, pues la subida del nivel de precios resta estímulo, mientras que el desplazamiento favorable de la oferta compensa en parte el efecto expulsión exterior. Incluso con movilidad perfecta, la política fiscal dejaría de ser ineficaz.




(a) Política monetaria expansiva




(b) Política fiscal expansiva con alta movilidad del capital

Figura A32.12: Tipo de cambio flexible en el modelo OA-DA de economía abierta. En el panel (a), la expansión monetaria desplaza la DA a la derecha (de 0 a 0′). La depreciación, al elevar el IPC esperado, desplaza las OA de corto y largo plazo a la izquierda hasta el punto 1, con la renta inicial y precios más altos. En el panel (b), la expansión fiscal desplaza la DA a la derecha, y la apreciación, al reducir el IPC esperado, desplaza las OA a la derecha, de modo que la renta final, $Y_1$, supera a la inicial.

**Aumento de los precios internacionales**: con tipo flexible, la moneda se aprecia nominalmente en la misma proporción y el tipo de cambio real, el IPC y la renta no varían. El régimen flexible aísla así a la economía de la inflación importada.

### 2.3.2. Tipo de cambio fijo

**Política monetaria expansiva**: la incorporación de la oferta agregada no altera el resultado del IS-LM-BP (apartados 1.3.2 y 1.3.3). Así, la política monetaria es ineficaz sobre la renta y

sobre los precios, en cualquier horizonte y para ambas escuelas.

**Política fiscal expansiva**: en el plano $(Y, P)$ (figura A32.13), el aumento del gasto ($\Delta G_0 > 0$) desplaza la DA a la derecha, de $\text{DA}_0$ a $\text{DA}_1$, en dos rondas. En la primera, el estímulo eleva la renta y el tipo de interés real. En la segunda, con alta movilidad, el superávit de balanza de pagos induce una expansión monetaria endógena que desplaza de nuevo la DA. El ajuste de la oferta difiere por escuela:

- *NMC*: si la expansión no es anticipada, la economía se desplaza transitoriamente a lo largo de la OA de corto plazo hasta un punto de mayor renta y precios ($0'$). Después, al revisarse las expectativas ($\Delta P_c^e > 0$), la OA de corto plazo se desplaza a la izquierda hasta el punto 1 sobre la $\text{OA}_{\text{LP}}$. Si la expansión es anticipada, la economía pasa directamente al nuevo equilibrio. En ambos casos el nivel de precios sube, pero, por la asimetría deflactor/IPC, el empleo y la renta aumentan también a largo plazo.

- *NEK1G*: por la inercia nominal, la producción responde antes que los precios. Después, el mayor empleo eleva los salarios, las empresas los trasladan a los precios y la OA se desplaza a la izquierda hasta un equilibrio con renta superior a la inicial.

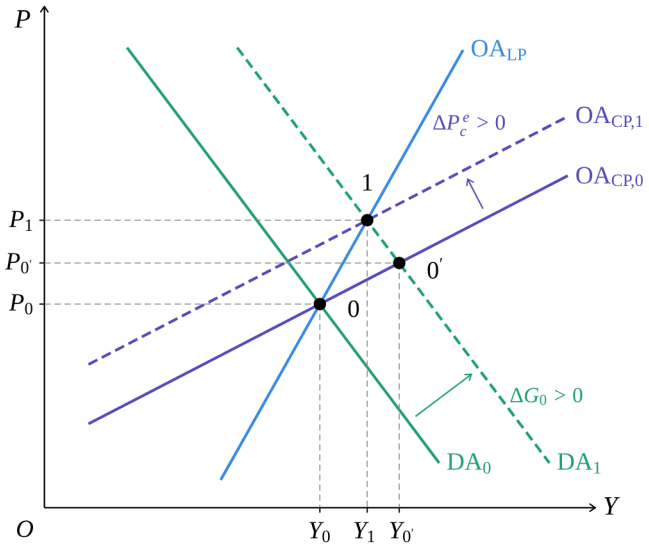



Figura A32.13: Efectos de una expansión fiscal a corto y largo plazo en el modelo OA-DA de economía abierta.

En suma, la expansión fiscal es eficaz para las dos escuelas, también a largo plazo, gracias a la pendiente positiva de la OA de largo plazo, y su eficacia queda en un punto intermedio. Supera la de la economía cerrada, donde esa OA es vertical y la expansión se agota en los precios. No llega, en cambio, a la del IS-LM-BP, porque los precios más altos suponen una apreciación real que deteriora la balanza comercial y resta parte del estímulo.

**Devaluación**: en el plano $(Y, P)$ (figura A32.14), la devaluación ($\Delta E > 0$) desplaza la DA a la derecha, de $\text{DA}_0$ a $\text{DA}_1$, en dos rondas (mejora de la balanza comercial y expansión monetaria endógena por el superávit). Las OA de corto y largo plazo, sin embargo, se desplazan a la izquierda, porque la depreciación eleva el precio de los bienes importados y, con él, el IPC esperado ($\Delta P_c^e > 0$) y las demandas salariales.

- *NMC*: si la devaluación es anticipada, y con ella la subida de precios que provoca, solo suben los precios, mientras que, si no lo es, tiene efectos expansivos transitorios.

- *NEK1G*: la devaluación tiene efectos expansivos a corto plazo (punto $0'$) por la inercia nominal, pero a largo plazo el aumento salarial compensa íntegramente la devaluación.

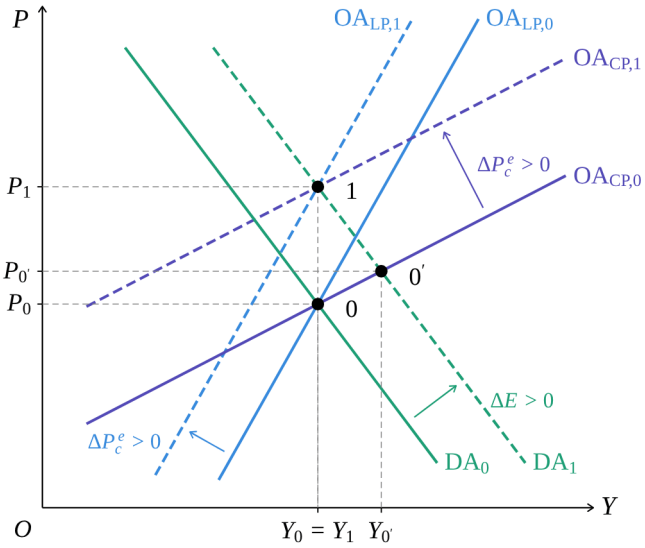



Figura A32.14: Devaluación en el modelo OA-DA de economía abierta. La demanda se expande inicialmente y la oferta se desplaza a largo plazo por el aumento del IPC esperado y de las demandas salariales.

En ninguna de las dos escuelas la devaluación tiene efectos duraderos. Con el tiempo, los precios nacionales suben en la misma medida que la devaluación, lo que deja el tipo de cambio real donde estaba y devuelve la renta al nivel de partida (punto 1, con $Y_0 = Y_1$)[9]. Solo la histéresis (NEK) podría volver permanente el efecto transitorio sobre el empleo. El IS-LM-BP llegaba a la conclusión opuesta porque suponía precios fijos.

**Aumento de los precios internacionales**: con tipo fijo, una subida de $P^*$ equivale a una devaluación y a largo plazo se traslada por completo a los precios nacionales.

> La comparación de ambos regímenes ante perturbaciones de demanda y de oferta, como criterio de elección del régimen cambiario, se desarrolla en el tema B15, apartado 1. El modelo con dos economías ilustra los costes de la integración monetaria en el tema B16, apartado 1.2.

## 2.4. *Shocks* y políticas de oferta

Las perturbaciones de oferta adversas plantean a la política de demanda un dilema que las de demanda no plantean, porque mueven renta y precios en direcciones opuestas.

[9] Los incrementos salariales derivados de la devaluación son de mayor magnitud que los derivados de la política fiscal, porque la devaluación eleva directamente el precio de los bienes importados de la cesta de consumo. Por ello, el aumento de precios es mayor y la corrección de la renta más severa.

### 2.4.1. *Shock* de precios energéticos

**Planteamiento**: para analizar una perturbación en el precio de los *inputs* importados (energía, materias primas) basta incluir un factor importado, $O$, con precio real en moneda nacional $p_o = E\,P_o^*/P$, en la función de producción, $Y = F(K, L, O)$. El programa de la empresa es:

$$\max_{L,\,O}\ \Pi = F(K,L,O) - \frac{w}{P}\,L - \frac{E\,P_o^*}{P}\,O$$

$$\Rightarrow \begin{cases} \dfrac{\partial F}{\partial L} = w/P & \text{(demanda de trabajo)} \\ \dfrac{\partial F}{\partial O} = E\,P_o^*/P = p_o & \text{(demanda de energía)} \end{cases}$$

donde $\Pi$ es el beneficio real, $w/P$ el salario real, $P_o^*$ el precio exterior del factor importado y $\partial F/\partial L$ y $\partial F/\partial O$ las productividades marginales de los factores.

**CPO**: cada factor se remunera según su productividad marginal en términos reales (condición neoclásica de contratación). Si los factores son complementarios en la producción ($F_{LO} > 0$), la demanda de cada uno depende negativamente de su propio precio real y del precio real del otro, $L^d = L(w/P^-, p_o^-)$ y $O^d = O(p_o^-, w/P^-)$. Ese *efecto cruzado* transmite el encarecimiento de la energía al empleo.

**CSO**: con $F$ estrictamente cóncava en $(L, O)$, las CPO son necesarias y suficientes para un máximo único.

**Estática comparativa**: en el plano $(Y, P)$ (figura A32.15), un aumento del precio nominal de la energía ($\Delta P_o^* > 0$), o una depreciación del tipo de cambio que encarezca los *inputs* importados en moneda nacional, reduce la demanda de energía. Por ese efecto cruzado, la menor demanda de energía reduce la productividad marginal del trabajo y, con ella, la demanda de trabajo. En consecuencia, caen el empleo y la producción para cualquier nivel de precios y la OA se desplaza a la izquierda, de $\text{OA}_0$ a $\text{OA}_1$. El resultado es *estanflación*, con menor renta y mayor nivel de precios (paso del punto 0 al 1, con $Y_1 < Y_0$ y $P_1 > P_0$).

Con tipo fijo, el deterioro de la balanza comercial provoca un déficit de balanza de pagos y una contracción monetaria endógena que desplaza también la DA a la izquierda y agrava la recesión, aunque modera la subida de precios. Con tipo flexible, por el contrario, la depreciación provocada por el déficit comercial encarece aún más los *inputs* importados y el IPC, y refuerza el desplazamiento de la OA.

Acomodar la perturbación con una expansión de la DA, p. ej. monetaria con tipo flexible ($\Delta M_0^s > 0$, que desplaza la DA de $\text{DA}_0$ a $\text{DA}_1$ y lleva la economía del punto 1 al 2), estabiliza la renta a corto plazo ($Y_2 = Y_0$), pero consolida la subida de precios ($P_2 > P_1$). No acomodarla, en cambio, modera la subida de precios a costa de una recesión mayor.

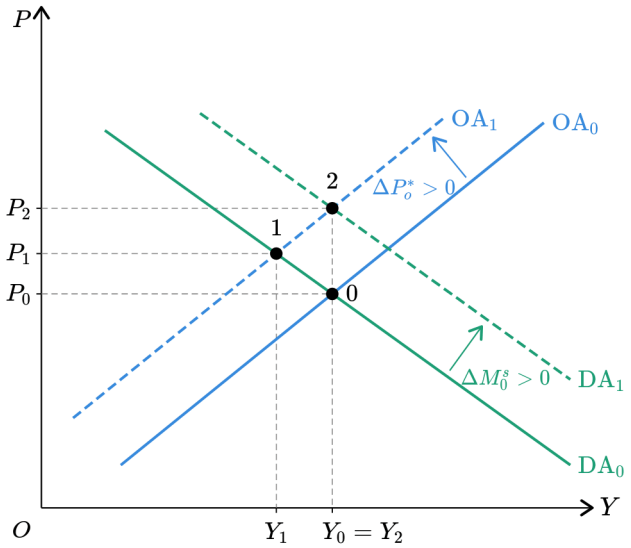



Figura A32.15: Perturbación de precios energéticos en el modelo OA-DA de economía abierta. La subida del precio exterior de la energía desplaza la OA a la izquierda y provoca estanflación (de 0 a 1). Acomodar la perturbación con una expansión de la demanda devuelve la renta a su nivel inicial a costa de un nivel de precios mayor (de 1 a 2).

### 2.4.2. *Shocks* de productividad y salariales

***Shock* de productividad**: un aumento permanente de la productividad ($\uparrow A_L$ en la NEK, $\uparrow K$ o progreso técnico en la NMC) desplaza la OA a la derecha, de modo que aumenta la renta y baja el nivel de precios. Como ambos se mueven en la dirección deseada, la perturbación no plantea dilema. La caída de los precios deprecia el tipo de cambio real ($EP^*/P$ sube) y mejora la balanza comercial. El efecto final depende del régimen cambiario:

- *Tipo fijo*: el superávit resultante induce una expansión monetaria que desplaza la DA a la derecha y amplifica el efecto sobre la renta.

- *Tipo flexible*: la moneda se aprecia nominalmente y absorbe parte de la ganancia de competitividad.

***Shock* salarial**: un aumento autónomo de las reivindicaciones salariales ($\uparrow z$) desplaza el SRN hacia arriba y la OA a la izquierda y provoca estanflación, como en el *shock* energético. En la NMC no existe este canal, pues no hay poder de negociación. El dilema de la política de demanda es el mismo, acomodar (mayor inflación, la espiral precios-salarios) o no acomodar (desempleo).

### 2.4.3. Políticas de oferta

Las **políticas de oferta** actúan sobre los determinantes de la OA para desplazarla a la derecha, con lo que elevan la renta y reducen el nivel de precios de forma permanente. En el marco de la NEK1G (figura A32.11, panel a), se agrupan en tres tipos:

- *Políticas de mercado de trabajo*: las que reducen el poder de negociación o las aspiraciones salariales ($\downarrow z$), como la reforma de las prestaciones por desempleo o la reducción de la cuña fiscal. Desplazan el SRN hacia abajo y reducen la NAIRU.

- *Políticas de competencia*: las que reducen el margen empresarial ($\downarrow \mu$) mediante liberalización de mercados de bienes y servicios, apertura comercial o regulación de la competencia. Desplazan el SRP hacia arriba (mayor salario real compatible con los precios) y elevan el empleo de equilibrio.

- *Políticas de productividad*: las que elevan $A_L$ o $K$ (inversión en capital físico, humano y tecnológico, infraestructuras), que desplazan el SRP hacia arriba y amplían la capacidad productiva.

En economía abierta, al reducir el nivel de precios nacional, estas políticas deprecian el tipo de cambio real. Con tipo fijo, la depreciación real se traduce en superávit exterior y expansión monetaria endógena (la *devaluación interna*, sustituto de la devaluación nominal cuando esta no está disponible, como en una unión monetaria). Frente a la devaluación nominal, ineficaz a largo plazo, las políticas de oferta alteran la NAIRU y la renta de equilibrio.

> Las políticas de empleo se desarrollan en el anexo del tema A27; los efectos del régimen cambiario sobre el mercado de trabajo, en el tema B15, apartado 1.2; la devaluación interna en una unión monetaria, en el tema B16.

## 2.5. Conclusiones

**Aspectos positivos**: el modelo supera dos limitaciones de especificación del IS-LM-BP. Con la oferta agregada determina el nivel de precios junto con la renta y analiza las perturbaciones y políticas de oferta. Con las expectativas, distingue entre políticas anticipadas y no anticipadas y entre corto y largo plazo. Su resultado propio, la pendiente positiva de la oferta agregada a largo plazo, da efectos reales permanentes a la política fiscal y no a la devaluación, y hace que la tasa natural de paro deje de ser única.

**Limitaciones**: el modelo sigue siendo estático, pues distingue puntos de equilibrio, pero no la trayectoria del ajuste. Determina el nivel de precios y no la inflación, la variable que observan y persiguen los bancos centrales. Además, trata la política monetaria como un control de la cantidad de dinero, cuando la práctica moderna es la fijación del tipo de interés según una regla.

Por último, la especificación de la oferta depende de la escuela, y la versión de la NMC atribuye a la información imperfecta unos efectos que la evidencia no respalda.

# 3. Inflación y brecha de producción: el modelo IS-PC-MR en economía abierta (Galí y Monacelli, 2005)

**Origen**: el modelo IS-PC-MR es la exposición gráfica de Carlin y Soskice (2005) del modelo canónico de la nueva economía keynesiana, condensado en tres ecuaciones por Clarida, Galí y Gertler (1999), y su extensión a una economía pequeña y abierta se debe a Galí y Monacelli (2005).

**Idea**: es la versión dinámica del análisis de oferta y demanda agregadas. Sustituye el nivel de precios por la inflación y la renta por la brecha de producción (*output gap*). Además, modela explícitamente el comportamiento del banco central, que fija el tipo de interés según una regla.

> El modelo en economía cerrada, con su microfundamentación, sus extensiones y su análisis gráfico, se desarrolla en el tema A07, apartado 2, y como teoría del ciclo en el tema A42, apartado 2.4.

## 3.1. Supuestos

**Entorno**: el modelo supone competencia monopolística (empresas precio-determinantes) y rigideces nominales *à la* Calvo (en cada periodo solo una fracción de las empresas revisa su precio).

En la versión abierta, la economía es pequeña, con tipo de cambio flexible (puro o dirigido) como caso de referencia y el tipo fijo como caso límite, perfecta movilidad del capital y paridad descubierta de intereses. Los bienes nacionales y extranjeros son sustitutivos imperfectos y una fracción de la cesta de consumo es importada.

**Agentes**: intervienen consumidores-oferentes de trabajo, empresas, sector público y autoridad monetaria, con expectativas racionales. La autoridad monetaria es creíble, tiene un objetivo de inflación $\pi^T$ y fija el tipo de interés nominal a corto plazo, de modo que la oferta monetaria es endógena.

**Tecnología**: la producción utiliza únicamente trabajo, con una tecnología estocástica de rendimientos constantes cuyas perturbaciones definen el *output* natural, $\bar{y}_t$, y el tipo de interés real natural, $\bar{r}_t$, los del equilibrio con precios flexibles.

## 3.2. Modelo en economía cerrada

**Planteamiento**: el modelo se compone de tres ecuaciones, una por cada agente con comportamiento endógeno (consumidores, empresas y autoridad monetaria), con las variables en desviaciones respecto de sus valores naturales.

**Curva IS dinámica** (ecuación de Euler y equilibrio del mercado de bienes, log-linealizados):

$$\tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta}(i_t - \mathbb{E}_t[\pi_{t+1}] - \bar{r}_t) + \varepsilon_t^d$$

La brecha de producción depende positivamente de la esperada y negativamente de la desviación del tipo de interés real respecto del natural, con una sensibilidad $1/\theta$, la elasticidad de sustitución intertemporal del consumo. El término $\varepsilon_t^d$ recoge los *shocks* de demanda (gasto público, preferencias).

**Curva de Phillips de la NEK** (maximización de beneficios con precios *à la* Calvo):

$$\pi_t = \beta\,\mathbb{E}_t[\pi_{t+1}] + \kappa\,\tilde{y}_t + \varepsilon_t^{\pi}, \qquad \kappa > 0,\ \beta \in (0,1)$$

La inflación depende de la inflación esperada y de la brecha de producción, con $\beta$ el factor de descuento y $\kappa$ decreciente en el grado de rigidez. El término $\varepsilon_t^{\pi}$ recoge los *shocks* de oferta o de costes (*cost-push*).

**Regla de Taylor**: el banco central fija el tipo nominal según las desviaciones de la inflación y del *output*[10]:

$$i_t = \bar{i}_t + \alpha_{\pi}\,\tilde{\pi}_t + \alpha_y\,\tilde{y}_t + \varepsilon_t^{m}, \qquad \alpha_{\pi} > 1,\ \alpha_y \geq 0$$

donde $\bar{i}_t = \bar{r}_t + \pi^T$ es el tipo de interés nominal neutral, $\tilde{\pi}_t = \pi_t - \pi^T$ la desviación de la inflación respecto del objetivo y $\varepsilon_t^m$ recoge las desviaciones discrecionales de la regla.

La condición $\alpha_{\pi} > 1$ es el *principio de Taylor*, según el cual, ante un aumento de la inflación, el tipo nominal sube más que la inflación, $\Delta i > \Delta\pi > 0$, de modo que el tipo real sube, la demanda se contrae y la brecha de producción cae.

Ese principio garantiza la estabilidad y unicidad del equilibrio (Blanchard-Kahn). En cambio, si $\alpha_{\pi} < 1$, un aumento de la inflación reduciría el tipo real, estimularía la demanda y validaría la propia inflación[11].

La regla, sustituida en la IS, da una relación decreciente en el plano $(\tilde{y}, \pi)$ que hace las veces de curva MR (en rigor, esta se deriva de la minimización de pérdidas del banco central)[12].

**Equilibrio**: en ausencia de perturbaciones, $\tilde{y}_t = 0$, $\pi_t = \pi^T$ e $i_t = \bar{i}_t$ y la economía se sitúa en su *output* natural con la inflación en el objetivo, que es el equilibrio de precios flexibles.

**Desarrollo gráfico**: el modelo IS-PC-MR se representa en el plano $(\tilde{y}, r)$ y en el plano $(\tilde{y}, \pi)$, que comparten el eje horizontal (figura A32.16, paneles superior e inferior):

1. *Curva IS*: recoge el equilibrio del mercado de bienes y es decreciente en el tipo de interés real, porque un tipo superior al natural contrae la demanda.

2. *Curva de Phillips* ($\mathrm{CP}_0$): recoge la fijación de precios y es creciente, con pendiente $\kappa$, porque una brecha positiva eleva la inflación.

3. *Curva MR*: es decreciente, porque el banco central responde a una mayor inflación con un tipo real más alto y una menor brecha de producción.

4. *Equilibrio*: en el punto 0, la CP y la MR se cruzan en $(0,\ \pi^T)$, y la IS asocia a esa brecha el tipo de interés natural $\bar{r}$.

[10] La regla de Taylor procede del comportamiento observado de los bancos centrales (Taylor, 1993) y puede justificarse como aproximación a la política óptima.

[11] La condición de Blanchard-Kahn exige que el número de raíces inestables del sistema iguale al número de variables no predeterminadas (aquí, $\tilde{y}_t$ y $\pi_t$). Con $\alpha_{\pi} > 1$ se cumple y el equilibrio es único. Con $\alpha_{\pi} < 1$ (en rigor, si $\kappa(\alpha_{\pi} - 1) + (1 - \beta)\alpha_y < 0$) surgen equilibrios múltiples de manchas solares, en los que expectativas de inflación autocumplidas desplazan la economía sin ninguna perturbación fundamental.

[12] El principio de Taylor puede obtenerse también como política óptima. Si el banco central minimiza una función de pérdida cuadrática en la brecha de producción y en la desviación de la inflación, $\mathcal{L} = \tilde{y}_t^2 + b(\pi_t - \pi^T)^2$, sujeto a la curva de Phillips, la condición de primer orden es $\tilde{y}_t = -\kappa b(\pi_t - \pi^T)$ (la curva MR). Sustituyéndola en la IS se obtiene una regla de tipos con coeficiente sobre la inflación superior a la unidad.

La ruptura de la divina coincidencia se estudia en el tema A07, apartado 3.1.2, y el multiplicador fiscal según la respuesta monetaria y en el límite inferior de los tipos de interés (*zero lower bound*), en el tema A40, apartado 2.1.

**Estática comparativa**: la perturbación desplaza la IS (si es de demanda) o la CP (si es de oferta), y la regla de Taylor determina la respuesta del tipo de interés:

- *Shock de demanda* (expansión fiscal, preferencias): la brecha de producción y la inflación suben transitoriamente, tanto menos cuanto mayores son $\alpha_\pi$ y $\alpha_y$, porque el banco central eleva el tipo real, sin conflicto entre ambos objetivos.

- *Shock de productividad*: eleva el *output* natural y, si es transitorio, reduce el tipo natural. Si el banco central acompaña esa caída, la economía converge al nuevo *output* natural con la inflación en el objetivo, de modo que estabilizar la inflación equivale a estabilizar la brecha de producción (la *divina coincidencia*).

- *Shock de costes* ($\varepsilon_t^\pi > 0$: energía, salarios, márgenes): la CP se desplaza hacia arriba, de $\text{CP}_0$ a $\text{CP}_1$, y con una brecha de producción nula la inflación subiría hasta $\pi_{0'}$ (punto $0'$). Para contener la inflación, el banco central eleva el tipo real hasta $r_1$ y abre una brecha de producción negativa, $\tilde{y}_1$, tanto mayor cuanto más pesa $\alpha_\pi$ respecto de $\alpha_y$. La economía pasa al punto 1, con una inflación $\pi_1$ menor que $\pi_{0'}$. Así, la divina coincidencia se rompe, y la disyuntiva entre inflación y *output* hace de los *shocks* de costes el problema central de la política de estabilización.

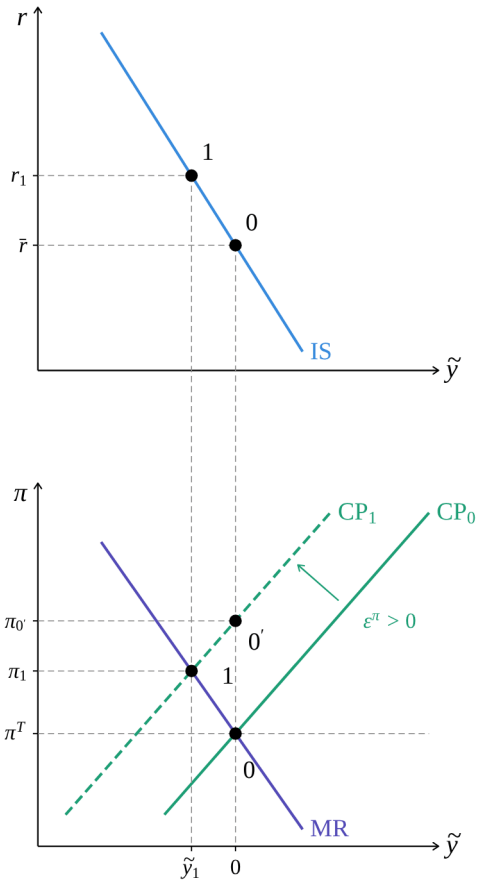



Figura A32.16: Respuesta del banco central a un *shock* inflacionista en el modelo IS-PC-MR de economía cerrada.

## 3.3. Modelo en economía abierta

Galí y Monacelli mantienen la estructura de tres ecuaciones del modelo base, modifican cada una y añaden la paridad de intereses y la noción de tipo de cambio natural.

### 3.3.1. Modificaciones al modelo base

Se denota por $p_{o,t}$ el precio relativo de los bienes importados (inversa de la relación real de intercambio o, para los *inputs* energéticos, su precio real en moneda nacional) y por $\tilde{E}_t$ la desviación del tipo de cambio nominal respecto de su nivel natural.

**Curva IS**: la demanda de bienes nacionales depende, además, del precio relativo de los bienes importados:

$$\tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta_\alpha}(i_t - \mathbb{E}_t[\pi_{t+1}] - \bar{r}_t) + \Gamma_y\, p_{o,t} + \varepsilon^d_t, \qquad \Gamma_y \gtrless 0$$

La sensibilidad al tipo de interés, $1/\theta_\alpha$, depende ahora del grado de apertura, porque los consumidores sustituyen entre bienes nacionales y extranjeros, no solo entre presente y futuro. El

tipo natural $\bar{r}_t$ depende también del *output* exterior, de modo que la economía responde a las perturbaciones de productividad de sus socios[13]. El signo de $\Gamma_y$ es indeterminado *a priori* porque un encarecimiento de las importaciones tiene dos efectos opuestos:

- *Efecto de sustitución* o desviación del gasto: traslada demanda desde los bienes extranjeros hacia los nacionales y es expansivo para la producción interior ($\Gamma_y > 0$).

- *Efecto renta*: el mayor coste de las importaciones reduce la renta real disponible y, con ella, la demanda ($\Gamma_y < 0$).

Cuál domina depende del bien importado. Así, con bienes de consumo sustitutivos de los nacionales prevalece la sustitución, mientras que con *inputs* energéticos, de demanda poco elástica y sin sustitutivo nacional, prevalece el efecto renta.

**Curva de Phillips**: la inflación relevante es la del IPC, que incluye los bienes importados, de modo que el precio de las importaciones entra directamente:

$$\pi_t = \beta\, \mathbb{E}_t[\pi_{t+1}] + \kappa_\alpha\, \tilde{y}_t + \Gamma_\pi\, p_{o,t} + \varepsilon^\pi_t, \qquad \Gamma_\pi > 0, \qquad \beta \in (0,1)$$

La pendiente $\kappa_\alpha$ depende también de la apertura, y el término $\Gamma_\pi p_{o,t}$ es el canal de la *inflación importada*, por el que una depreciación o una subida de los precios exteriores eleva la inflación de consumo aunque la brecha de producción no varíe.

**Regla de Taylor ampliada**: el banco central puede responder, además, a la desviación del tipo de cambio respecto de su nivel natural:

$$i_t = \bar{i}_t + \alpha_\pi\, \tilde{\pi}_t + \alpha_y\, \tilde{y}_t + \alpha_E\, \tilde{E}_t + \varepsilon^m_t, \qquad \alpha_E \geq 0$$

El valor de $\alpha_E$ distingue tres casos:

- *Flotación pura* ($\alpha_E = 0$): el banco central ignora el tipo de cambio.

- *Flotación dirigida* ($\alpha_E > 0$): el banco central resiste las desviaciones del tipo de cambio y añade un objetivo de estabilidad externa a los internos.

- *Tipo fijo* (en el límite, $\alpha_E \to \infty$): el banco central fija el tipo de cambio y renuncia a la regla sobre inflación y *output*.

**Paridad descubierta de intereses**: con perfecta movilidad, se cumple $i_t - i^*_t = \mathbb{E}_t[\Delta s_{t+1}]$, con $s_t = \ln E_t$. Una subida del tipo nacional por encima del exterior aprecia el tipo de cambio de inmediato hasta que la depreciación esperada compensa el diferencial. El *tipo de cambio natural* es el compatible con el *output* y el tipo de interés naturales y con el equilibrio exterior, y la regla ampliada puede intentar corregir sus desviaciones, $\tilde{E}_t$.

[13] En Galí y Monacelli (2005), $\theta_\alpha$ y $\kappa_\alpha$ son funciones del grado de apertura $\alpha \in [0,1]$ (peso de los bienes importados en el consumo), de la aversión al riesgo y de las elasticidades de sustitución entre bienes nacionales y extranjeros. Con $\alpha = 0$ se recupera la economía cerrada. El tipo natural depende del crecimiento esperado de la productividad nacional y del *output* exterior. En el original, la IS y la curva de Phillips se escriben en términos de la inflación interna (de los bienes nacionales) y la relación real de intercambio se sustituye en función de la brecha de producción. Los términos en $p_{o,t}$ de este apartado son, en cambio, una extensión estilizada para analizar perturbaciones en el precio de las importaciones, como las energéticas.

### 3.3.2. Equilibrio y mecanismo de transmisión

En ausencia de perturbaciones, todas las variables están en sus niveles naturales, incluido el tipo de cambio, y la inflación en el objetivo. Una subida del tipo de interés contrae el consumo (canal del tipo de interés) y, vía paridad de intereses, aprecia la moneda. La apreciación reduce las exportaciones netas (canal del tipo de cambio sobre la demanda) y abarata las importaciones (canal del tipo de cambio sobre la inflación del IPC). La política monetaria es, así, más potente sobre la inflación y su efecto se reparte entre el sector interior y el exterior.

Gráficamente, en el plano $(y, r)$ (figura A32.17), para llevar la producción de $\bar{y}$ (punto 0, con el tipo $\bar{r}$) a $y_1$, en economía cerrada el tipo debe subir a lo largo de la IS (cerrada) hasta $r_c$. En cambio, en la abierta basta un tipo $r_a < r_c$, porque la IS (abierta), que recoge también el efecto de la apreciación inducida sobre la demanda, es más plana[14].

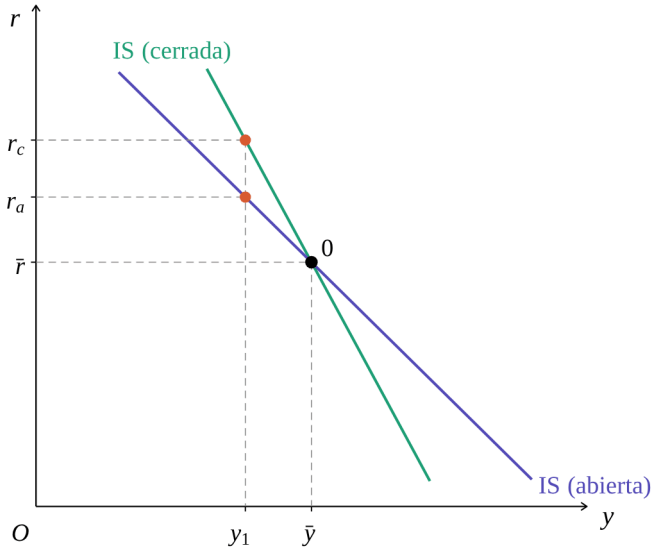



Figura A32.17: Modelo IS-PC-MR en economía abierta. Para cerrar la misma brecha de producción, el banco central necesita mover menos el tipo de interés que en economía cerrada, porque la política monetaria actúa también a través del tipo de cambio.

### 3.3.3. Análisis dinámico de perturbaciones

***Shock* de demanda** (expansión fiscal): la brecha de producción y la inflación suben, el banco central eleva el tipo, la moneda se aprecia y la demanda se contrae por los dos canales, y la apreciación modera además la inflación del IPC. El efecto de la expansión fiscal sobre el *output* es menor que en economía cerrada por el efecto expulsión exterior de Mundell-Fleming, obtenido ahora con precios rígidos, pero no fijos, y con inflación. Si el banco central resiste la apreciación ($\alpha_E > 0$), sube menos el tipo y el estímulo fiscal es mayor, a costa de más inflación.

[14] No es la IS del apartado 1.2.1, trazada para un tipo de cambio dado, que la filtración de importaciones hacía más inclinada.

***Shock* de oferta** (energético): considérese una subida exógena del precio internacional del gas natural (como la de 2022, tras la invasión rusa de Ucrania) en un país importador. Al subir $p_{o,t}$, la inflación aumenta por el canal $\Gamma_\pi$, la demanda de energía importada cae y, si la energía se añade a la tecnología como factor complementario del trabajo (apartado 2.4.1), caen la demanda de trabajo y el *output* natural. Por el lado de la demanda, domina el efecto renta y la brecha de producción se reduce ($\Gamma_y < 0$). Si los hogares sostienen su consumo, el saldo exterior se deteriora, lo que presiona la moneda hacia la depreciación y realimenta la inflación importada.

El banco central se enfrenta a la disyuntiva del *shock* de costes agravada, ya que subir el tipo contiene la inflación y sostiene la moneda, pero profundiza la caída del *output*, y no subirlo tolera la depreciación y la inflación importada. La respuesta óptima depende de si la perturbación es transitoria (acomodar el efecto de primera ronda y evitar los de segunda ronda sobre salarios) o persistente.

***Shock* financiero** (Rey, 2013): la perturbación nace en las condiciones financieras globales (un endurecimiento de la política monetaria de la Reserva Federal, un aumento de la aversión al riesgo global). La perturbación provoca salidas de capital, presión hacia la depreciación y endurecimiento del crédito en la economía pequeña con independencia de su régimen cambiario.

Según Rey, cuando el ciclo financiero global domina los flujos de capital, el trilema se convierte en un *dilema*. Su argumento es que, con libre movilidad del capital, la política monetaria nacional no es independiente ni siquiera con tipo de cambio flexible, porque los tipos a largo, las primas de riesgo y el crédito interno siguen el ciclo global y no el tipo oficial. Sin embargo, Obstfeld, Ostry y Qureshi (2019) matizan la tesis, pues en las economías emergentes la flotación sigue amortiguando las perturbaciones financieras globales y conserva así una autonomía parcial.

En términos del modelo, la perturbación entra como un *shock* a $\bar{r}_t$ y a $\tilde{E}_t$ que la regla de Taylor estándar no puede compensar sin desestabilizar la brecha de producción.

### 3.3.4. Implicaciones de política económica

**Régimen óptimo** (Galí y Monacelli, 2005): la comparación de tres regímenes (regla sobre la inflación interna, regla sobre la del IPC y tipo fijo) muestra que, bajo ciertas condiciones, la política óptima es el objetivo de inflación interna (de los bienes nacionales) con flotación pura ($\alpha_E = 0$). Ese régimen replica la asignación de precios flexibles y deja que el tipo de cambio absorba las perturbaciones. El objetivo de inflación del IPC, que responde implícitamente al tipo de cambio, y el tipo fijo son inferiores en bienestar, porque estabilizan el tipo de cambio a costa de una volatilidad innecesaria de la brecha de producción y de la inflación interna.

Existe, por tanto, una *disyuntiva entre estabilización interna y externa*, ya que cuanto más pesa $\alpha_E$, menos se estabilizan la inflación interna y el *output*. Aun así, la sobrerreacción del tipo de cambio ante perturbaciones monetarias (*overshooting*, Dornbusch, 1976) alimenta la tentación de un $\alpha_E > 0$.

> El modelo de sobrerreacción de Dornbusch se desarrolla en el tema B14, apartado 2.2.2; la elección de régimen cambiario y la intervención en el mercado de divisas, en el tema B15; el diseño de la regla de tipos y la política macroprudencial, en los temas A36 y B27.

## 3.4. Conclusiones

**Aspectos positivos**: sustituye el nivel de precios por la inflación y la cantidad de dinero por una regla de tipos, en línea con la práctica de los objetivos de inflación. Además, está

microfundamentado, es en principio más robusto frente a la crítica de Lucas que los anteriores y analiza las perturbaciones como trayectorias, no como comparaciones de equilibrios.

En economía abierta, muestra que el tipo de cambio es a la vez un canal de transmisión adicional, que hace más potente la política monetaria, y una fuente de perturbaciones y disyuntivas.

**Limitaciones**: la versión básica descansa en un agente representativo, prescinde del capital y del sector financiero y trata la paridad de intereses sin primas de riesgo, lo que deja fuera precisamente los canales por los que operan las perturbaciones financieras. Además, la curva de Phillips puramente *forward-looking* explica mal la persistencia de la inflación, y el límite inferior de los tipos, donde la regla de Taylor deja de operar, exige extensiones específicas.

## Conclusiones

Los tres modelos del tema responden a la misma pregunta, cómo se determinan la renta y los precios en una economía abierta, y cada uno nace de las limitaciones del anterior. El IS-LM-BP explica el corto plazo con precios dados, y en él el régimen cambiario decide qué política de demanda es eficaz, la monetaria con tipo flexible y la fiscal con tipo fijo. El modelo OA-DA añade la oferta y las expectativas, y su oferta agregada con pendiente positiva a largo plazo da efectos reales permanentes a la política fiscal y no a la devaluación. El IS-PC-MR, por último, sigue la trayectoria de la inflación y la brecha de producción bajo una regla de Taylor, con el tipo de cambio como canal adicional de transmisión y como fuente de perturbaciones.

Estos modelos son la base analítica de buena parte de la macroeconomía abierta. El ajuste de la balanza de pagos (tema B12) y la determinación del tipo de cambio, con la paridad de intereses y la sobrerreacción (temas B13 y B14), parten del mismo equilibrio entre el mercado de bienes, el de dinero y el sector exterior. La respuesta de cada régimen cambiario a perturbaciones de demanda y de oferta fundamenta su comparación (tema B15), y la imposibilidad de devaluar dentro de una unión monetaria obliga a la devaluación interna (tema B16). Por su parte, los tres modelos siguen la evolución de las escuelas macroeconómicas (temas A05 a A07) y de la NAIRU (tema A28), y el análisis de las políticas monetaria y fiscal se completa con su diseño y su interrelación (temas A36 y A40).

La existencia de un ciclo financiero global ha puesto a prueba el trilema de Mundell-Fleming. Rey (2013) sostiene que el ciclo reduce el trilema a un dilema, y Obstfeld, Ostry y Qureshi (2019), que la flotación conserva una autonomía monetaria parcial (apartado 3.3.3). El paradigma de la moneda dominante (Gopinath et al., 2020) cuestiona, por su parte, el canal comercial, pues con el comercio facturado en dólares una depreciación apenas eleva las exportaciones a corto plazo y su efecto recae sobre todo en las importaciones. En esa línea, el FMI ha propuesto un marco integrado de políticas (Basu et al., 2020) que combina la política monetaria con la intervención cambiaria, la macroprudencial y las medidas sobre los flujos de capital.

# A33. Teorías de demanda de consumo

Teorías de la demanda de consumo corriente: ciclo vital y renta permanente. La demanda de bienes de consumo duradero. Evidencia empírica e implicaciones de política económica.

## Introducción

Las **teorías de la demanda agregada** estudian los comportamientos que hay detrás de los grandes bloques de los modelos macroeconómicos. El consumo de los hogares forma, con la inversión, el gasto público y el saldo exterior, la demanda de bienes y servicios que recoge la curva IS, y de la forma de la función de consumo dependen el tamaño del multiplicador y, con él, la eficacia de la política fiscal.

El punto de partida es la *Teoría general* de Keynes (1936), que hizo depender el consumo de la renta corriente según una "ley psicológica fundamental". Sin embargo, era una relación entre agregados, sin una derivación explícita a partir de decisiones individuales, y la evidencia de posguerra pronto mostró sus límites. En concreto, Kuznets (1946) observó que la propensión media al consumo era estable a largo plazo, frente a lo que predecía la función keynesiana, y de esa contradicción nacieron la **hipótesis del ciclo vital** (Modigliani y Brumberg, 1954) y la de la **renta permanente** (Friedman, 1957). Desde los años setenta, la crítica de Lucas (1976) y las expectativas racionales llevaron el análisis a hogares que optimizan en un horizonte intertemporal con incertidumbre. Hall (1978) mostró que, con renta permanente y expectativas racionales, el consumo sigue un paseo aleatorio, y la ecuación de Euler del consumo pasó a ser la curva IS dinámica de los modelos de equilibrio general (tema A07).

En economía, el **consumo** se entiende como el uso o adquisición de bienes y servicios por agentes económicos para la satisfacción de sus necesidades, tanto corrientes (p. ej. bienes no duraderos) como futuras (p. ej. bienes duraderos)[1][2].

El estudio del consumo tiene importancia tanto para la microeconomía como para la macroeconomía:

1. *Microeconomía*: permite entender el comportamiento de los individuos y medir, a través de su consumo, su nivel de riqueza y, por extensión, de pobreza.

2. *Macroeconomía*: importa por dos motivos. En primer lugar, el consumo agregado determina el ahorro agregado, que se canaliza a través del sistema financiero hacia la oferta de capital, de modo que plantea la disyuntiva entre el consumo como fuente de bienestar inmediato y el ahorro

[1] Por ejemplo, la compra de un automóvil (bien duradero) genera gasto de consumo en el periodo, pero genera un "servicio de consumo" hasta que se renueva o se retira. Esto es análogo a lo que sucede con la inversión de los hogares, que genera "servicios de vivienda" en el futuro.

[2] El tratamiento del consumo como un único bien compuesto se apoya en el teorema del bien compuesto (Leontief, 1936; Hicks, 1939b), según el cual, si los precios relativos de los bienes de consumo no varían, pueden tratarse como un único bien.

como fuente de crecimiento. En segundo lugar, el consumo privado absorbe cerca del 60 % del producto en las economías desarrolladas, lo que le da un peso central en la demanda agregada y, con la inversión, lo hace imprescindible para explicar las fluctuaciones económicas.

El tema analiza cronológicamente las teorías de la demanda de consumo en tres partes:

- *Antecedentes*: se estudian los antecedentes de las teorías modernas (aportaciones neoclásica, keynesiana y de Duesenberry).

- *Teorías de la renta normal*: se analizan con certidumbre (hipótesis del ciclo vital, hipótesis de la renta permanente) y con incertidumbre (paseo aleatorio de Hall).

- *Desarrollos recientes*: el tema termina con los desarrollos recientes, la demanda de bienes de consumo duradero, la evidencia empírica y las implicaciones de política económica.

# 1. Antecedentes a las teorías modernas sobre el consumo

Los antecedentes de las teorías modernas del consumo son la aportación neoclásica, que trata el consumo como un residuo del ahorro, y la keynesiana.

## 1.1. Aportación neoclásica

**Origen**: la función de consumo agregada neoclásica procede de las aportaciones de diversos autores:

- *Fisher*: Irving Fisher (1930) proporcionó su microfundamentación con su modelo de elección intertemporal de dos periodos.

- *Pigou*: Arthur Pigou (1943) introdujo el efecto saldos reales.

- *Modigliani y Patinkin*: Modigliani (1944) y Patinkin (1956) sistematizaron a partir de ambos la función de consumo agregada neoclásica, como contrapunto a la keynesiana.

**Idea**: en el paradigma neoclásico, el consumo es solo un gasto destinado a satisfacer las necesidades corrientes individuales y se considera apenas un residuo del ahorro[3].

**Razonamiento**: dentro del pensamiento neoclásico existen dos formas de construir la función de consumo agregada:

- *Fisher* (1930): es la forma original. El consumo real agregado es una función creciente de la renta disponible de los agentes y, si el efecto sustitución domina al efecto renta, decreciente del tipo de interés real, $C = f(Y^+, r^-)$. Siguiendo esta formulación, se cumple la dicotomía clásica y la neutralidad del dinero debido a la teoría cuantitativa del dinero.

- *Pigou* (1943): si se considera, además, el *efecto saldos reales* o *efecto Pigou*, el consumo real agregado también dependerá de la riqueza real de los individuos, formada por saldos monetarios reales y *stock* de bonos reales.

Por lo tanto, $C = f(Y^+, r^-, W^+)$, donde $W = M/P + B/P$ es la riqueza real de los agentes, suma de los saldos monetarios y de los bonos reales[4].

Esta formulación rompe la dicotomía clásica (variables monetarias como el dinero o los bonos afectan a variables reales como el consumo), pero no necesariamente la neutralidad del dinero, e influyó en el pensamiento monetarista. Con ella, Pigou buscaba dotar de mayor efectividad al ajuste automático de la economía ante cambios en los precios o los salarios.

> El mercado de fondos prestables y los efectos en él de las políticas fiscal y monetaria se desarrollan en el tema A34, apartado 1.1.2.

**Extensiones**: en el mercado de fondos prestables, las políticas de demanda operan a través del tipo de interés real, cuya flexibilidad garantiza que se cumpla la ley de Say.

[3] No obstante, Smith (1776) ya afirmó que "el consumo es el único fin y propósito de toda producción". Sin embargo, el pensamiento clásico no dio a la demanda de consumo un tratamiento analítico macroeconómico potente.

[4] Metzler (1951) incluye en la formulación anterior como parte de la riqueza el *stock* de capital real: $W = M/P + B/P + K$. Esto hace que, además del efecto Pigou, haya que tener en cuenta el efecto Metzler (tema A04, apartado 3.1.2). Como el ahorro depende de la riqueza, importa también su composición, y una compra de títulos en mercado abierto por el banco central, al reducir la riqueza privada real, altera de forma permanente el tipo de interés real de equilibrio.

▪ *Política monetaria expansiva*: reduce el tipo de interés real y estimula el consumo y la inversión hasta que la subida de precios reequilibra la economía. Sin embargo, con efecto saldos reales y un *stock* nominal de bonos que no crece en la misma proporción que el dinero, se rompe la neutralidad. La razón es que la subida de precios reduce la riqueza real y el consumo, y el tipo de interés real queda permanentemente más bajo para que la inversión compense exactamente esa caída.

▪ *Política fiscal expansiva*: eleva el tipo de interés real y desplaza totalmente el gasto privado, con un efecto expulsión (*crowding out*) total. En el paradigma neoclásico, ese efecto justifica la *regla del equilibrio presupuestario* y su única excepción, el endeudamiento para financiar inversión pública o *regla de oro de las finanzas públicas*[5]. Si la demanda de dinero depende del tipo de interés, sube además el nivel de precios. Con efecto saldos reales, el tipo de interés sube menos y el desplazamiento, que sigue siendo total, recae más sobre el consumo (por la caída de la riqueza real) y menos sobre la inversión.

El paradigma neoclásico explica que el presidente Hoover elevase los impuestos (Revenue Act de 1932) para equilibrar el presupuesto durante la Gran Depresión, con efectos adversos.

**Conclusiones**:

▪ *Aspectos positivos*: fundamenta el consumo en la elección intertemporal e incorpora como determinantes el tipo de interés real y, con Pigou, la riqueza, ideas que recogen después las teorías del ciclo vital y de la renta permanente. El efecto saldos reales es, además, la base de la réplica neoclásica al equilibrio con desempleo keynesiano.

▪ *Limitaciones*: relega el consumo a un papel pasivo y descansa en la flexibilidad de los precios y del tipo de interés, por lo que no explica caídas persistentes de la demanda como la de la Gran Depresión. Además, el efecto saldos reales es cuantitativamente débil y exige caídas de precios intensas, que a su vez elevan la carga real de las deudas.

## 1.2. Aportación keynesiana

Frente a esas limitaciones surge la aportación keynesiana, que hace depender el consumo de la renta corriente, absoluta en Keynes y relativa en Duesenberry.

### 1.2.1. Teoría de la renta corriente absoluta (Keynes, 1936)

**Origen**: la primera función agregada de demanda de consumo se debe a Keynes, que la formula en la *Teoría general* (1936).

**Idea**: el consumo depende de la renta corriente disponible a través de una propensión marginal a consumir estable y menor que la unidad, de la que se sigue el multiplicador.

**Razonamiento**: en su teoría del consumo, Keynes parte de tres ideas fundamentales, de las que se deriva una cuarta sobre el ahorro:

▪ *Naturaleza*: la producción tiene como destino final el consumo. De esta idea parte el proceso multiplicador, y con ella Keynes se aleja del papel residual que los clásicos reservaban al consumo.

[5] Dicho de otro modo, solo es admisible la deuda pública autoliquidable, la que financia inversiones cuyos rendimientos permiten devolverla.

▪ *Renta corriente*: considera que las decisiones de consumo dependen de la renta actual o corriente. Así, el tipo de interés pasa a segundo plano y lo que más pesa en el consumo es la renta, al revés de lo que sostenían los clásicos.

▪ *Propensiones al consumo*: los consumidores tienen planes de consumo estables que se manifiestan en las propensiones al consumo y al ahorro:

- *Propensión marginal a consumir*: indica el consumo realizado por unidad adicional de renta, $c \equiv \partial C/\partial Y \in (0,1)$.

  Para Keynes, esta representa una *ley psicológica fundamental*, según la cual, si los consumidores reciben una unidad adicional de renta, gastan una parte y ahorran otra. Por ello, al aumentar su renta aumentan su consumo, pero en menor medida que el aumento de la renta.

  Además, la propensión marginal a consumir es bastante estable, de manera que se puede considerar un parámetro constante[6].

  El valor de la propensión marginal a consumir fundamenta el *principio del multiplicador* (Kahn, 1931; Keynes, 1936). El principio establece que los aumentos en los componentes autónomos de la demanda agregada dan lugar a aumentos mayores de la renta nacional, $\partial Y/\partial A_0 = 1/(1-c)$, donde $A_0$ es el gasto autónomo.

- *Propensión media al consumo*: fracción de la renta corriente que se dedica al consumo, $C/Y = C_0/Y + c$. Keynes suponía que cae al crecer la renta, $d(C/Y)/dY < 0$, porque veía en el ahorro un bien de lujo y pensaba que, en proporción a lo que ingresan, los hogares acomodados ahorran más que los modestos.

▪ *Propensiones al ahorro*: como, dada la renta disponible, elegir el consumo equivale a elegir el ahorro, de las anteriores se obtienen:

- *Propensión marginal a ahorrar*: cuánto ahorran los hogares por unidad adicional de renta, $s \equiv \partial S/\partial Y = (1-c) \in (0,1)$.

- *Propensión media al ahorro* o *tasa de ahorro* (*savings ratio*): proporción de renta disponible que se mantiene en forma de ahorro. El efecto multiplicador no exige que la tasa de ahorro se mantenga constante (en Keynes, aumenta con la renta), sino que se basa en la constancia de la propensión marginal.

De todo ello resulta una demanda agregada de consumo en términos reales lineal y estable en la renta real disponible del periodo, con un *componente autónomo*, el consumo de subsistencia, y un *componente endógeno*, gobernado por la propensión marginal a consumir:

$$C = C_0 + c(Y - T_0), \quad C_0 > 0,\ c \in (0,1)$$

donde $C_0$ es el consumo autónomo, $Y$ la renta real y $T_0$ los impuestos, de modo que $Y - T_0$ es la renta disponible. El ahorro es, así, un excedente de ingreso.

**Desarrollo gráfico**: la función de consumo keynesiana se representa en el plano renta disponible-consumo (figura A33.1):

[6] En rigor, $c$ puede depender del tipo de interés, de la riqueza o de la distribución de la renta.

1. *Curva de consumo*, $C(Y - T_0)$: recoge el consumo planeado[7] para cada nivel de renta[8]. Corta el eje vertical en el consumo autónomo, $C_0$, y su pendiente es la propensión marginal a consumir, $c$.

2. *Bisectriz* (45°): recoge los puntos en los que el consumo agota la renta disponible, es decir, con propensión media a consumir unitaria.

3. *Equilibrio*: en el punto de corte, punto 0, el plan de consumo se satisface exactamente con la renta disponible, $(Y - T_0)_0$, y el consumo es $C_0/(1 - c)$.

- *Derecha del punto 0*: los hogares ahorran la diferencia vertical entre la bisectriz y la curva de consumo (p. ej. en $(Y - T_0)_1$), y la propensión media al consumo es inferior a la unidad.

- *Izquierda del punto 0*: desahorran o se endeudan por esa diferencia para satisfacer su plan de consumo.

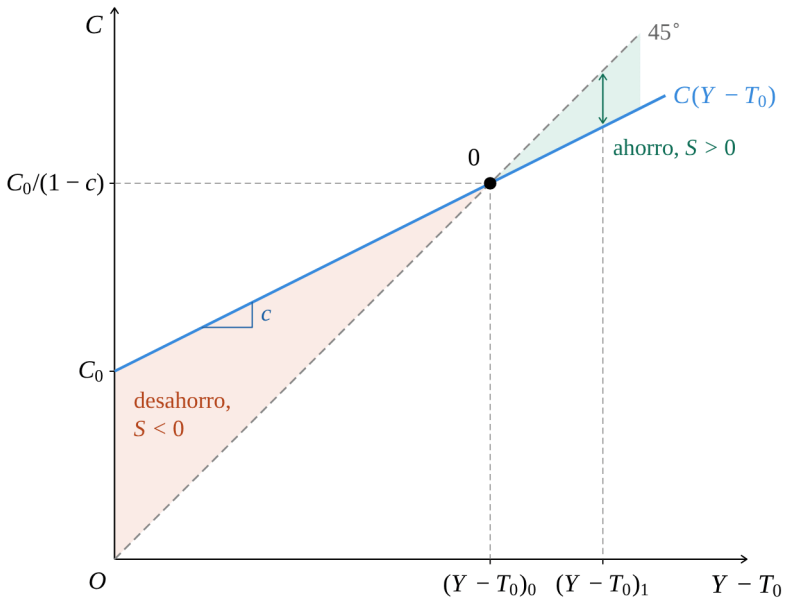



Figura A33.1: Función de consumo keynesiana. El consumo planeado crece con la renta disponible con pendiente $c < 1$. A la derecha del punto de corte con la bisectriz los hogares ahorran y a la izquierda desahorran.

**Extensiones**: del modelo de Keynes se deduce que el Gobierno debe desempeñar un papel activo.

- *Política monetaria*: sí es efectiva con el tipo de interés y, en principio, con la renta. Sin embargo, debido a los *animal spirits* y a la baja sensibilidad de la inversión al tipo de interés (eslabón débil), la política monetaria será poco efectiva con la renta y, por ello, con el consumo.

[7] Es indiferente hablar de consumo planeado y consumo efectivo en Keynes, ya que en una situación de desequilibrio los planes de consumo se realizan ($C = C^p$) y es la inversión la variable que se ajusta vía existencias.

[8] Keynes denomina a esta curva *propensión a consumir* (*propensity to consume*), es decir, la relación funcional entre consumo y renta, que no debe confundirse con la propensión media, $C/Y$, ni con la marginal, su pendiente.

▪ *Política fiscal*: sí es efectiva con la renta y el tipo de interés, porque altera la renta corriente, de la que depende el consumo, a través de los multiplicadores. En el caso de una economía cerrada con impuestos de suma fija:

$$\frac{\partial Y}{\partial G_0} = \frac{1}{1-c} > 1 \qquad \frac{\partial Y}{\partial T_0} = -\frac{c}{1-c} < 0$$

donde $G_0$ es el gasto público autónomo y $T_0$ los impuestos de suma fija. Además, un aumento del gasto público siempre será más efectivo que una disminución impositiva de la misma cuantía, ya que estimula directamente la demanda agregada. La rebaja de impuestos, en cambio, aumenta la renta disponible de los hogares, que pueden dedicar esa cantidad al consumo o al ahorro (analíticamente, porque $c < 1$).

Dentro de la política fiscal, las medidas redistributivas (impuestos y transferencias que trasladan renta hacia los hogares de menor renta) sí serán efectivas para estimular el consumo, ya que estos hogares tienen una propensión marginal a consumir mayor.

Además de sus implicaciones de política económica, de la aportación de Keynes se desprenden otros dos resultados:

▪ *Paradoja del ahorro* (Keynes, 1936): también llamada *paradoja de la frugalidad*, ataca la defensa clásica del ahorro como fuente del crecimiento económico.

La paradoja del ahorro establece que un aumento del ahorro autónomo de los hogares (descenso del consumo de subsistencia), *ceteris paribus*, no provoca necesariamente un aumento del ahorro privado, pero sí una caída de la renta, por lo que el ahorro puede obstaculizar el pleno empleo.

Esto se debe a que la caída del consumo produce una caída de la demanda agregada y, a través del efecto multiplicador, una caída de la producción, la renta nacional y, finalmente, la renta disponible de los hogares[9].

▪ *Teoría del estancamiento secular* (Hansen, 1938): reforzó la defensa del intervencionismo, en concreto de la política fiscal expansiva de inversiones públicas.

Según la teoría keynesiana, si una economía experimentaba un aumento de la renta sostenido, la propensión media al consumo debía disminuir (la propensión media al ahorro debía aumentar), lo que podía llevar a un exceso de ahorro privado.

Hansen sostenía, además, que las oportunidades de inversión privada se estaban agotando por la desaceleración del crecimiento de la población, el fin de la expansión territorial y un progreso técnico menos intensivo en capital. En consecuencia, la ratio inversión privada-renta no aumentaría para absorber el mayor ahorro.

Por todo ello, Hansen apuntaba, a finales de los años treinta, que se llegaría a un *estancamiento secular*, es decir, a una depresión de duración indefinida, un estado estacionario con recursos ociosos. La única manera de garantizar el pleno empleo de recursos y evitar el estancamiento secular vía estímulo de la demanda agregada era la inversión pública, que elevaba el gasto público como porcentaje del PIB[10].

La idea se recuperó tras la Gran Recesión (Summers, 2014), por la dificultad de alcanzar una demanda agregada de pleno empleo con la economía en la trampa de liquidez.

[9] Con la inversión, el gasto público y los impuestos exógenos, la condición de equilibrio $I_0 = S + (T_0 - G_0)$ impide que varíe el ahorro privado, de modo que la renta cae, $\Delta Y = \Delta C_0/(1-c) < 0$, hasta que el ahorro vuelve a su nivel inicial.

[10] Como $1 = C/Y + I/Y + G/Y$, si $C/Y$ cae con la renta e $I/Y$ no aumenta, $G/Y$ debe aumentar para mantener el pleno empleo.

**Conclusiones**:

- *Aspectos positivos*:

  - *Influencia teórica*: sobre la función de consumo construyó Keynes su explicación de las fluctuaciones, y la sencillez de esa función le valió una acogida inmediata entre los economistas.

  - *Evidencia empírica a corto plazo*: los primeros datos parecían darle la razón. Tanto las encuestas de presupuestos familiares (sección cruzada) como las series anuales cortas arrojaban una propensión media al consumo que bajaba al subir la renta. A corto plazo, el ahorro funcionaba, por tanto, como un lujo.

- *Limitaciones*:

  - *Evidencia empírica a largo plazo*: con series largas de EE. UU. (1869-1938), Simon Kuznets (1946) halló una propensión media al consumo estable en el largo plazo[11], sin relación con el nivel de renta, y tampoco la tasa de ahorro variaba. La predicción keynesiana sobre consumo y ahorro fallaba, pues, en ese horizonte.

  - *Estancamiento secular*: la predicción de Hansen también falló. Tras la Segunda Guerra Mundial el gasto público se redujo, pero el consumo subió aunque la renta disponible real cayó entre 1945 y 1947. Detrás de esa subida hubo un efecto riqueza, ya que los hogares liquidaron los bonos públicos que habían acumulado durante el racionamiento. El episodio mostró que la riqueza debía entrar en cualquier teoría del consumo.

### 1.2.2. Teoría de la renta corriente relativa (Duesenberry, 1949)

**Origen**: James Duesenberry (1949) añade fundamentos psicológicos a la teoría keynesiana, en vista de la contradicción entre esta y la evidencia sobre la propensión media al consumo (*enigma de Kuznets*).

**Idea**: cada hogar obtiene utilidad de su consumo actual, pero la mide frente a lo que consume su entorno social (renta relativa) y frente a lo que él mismo consumía antes (hábitos). Lo que gobierna el consumo es, así, la posición relativa del hogar y la renta más alta que haya alcanzado, no su renta corriente absoluta.

**Supuestos**: para desarrollar su teoría, Duesenberry se basa en dos hipótesis:

1. *Efecto demostración*[12]: un hogar de renta relativamente baja procura, aunque ingrese menos, no quedarse por debajo del consumo medio del país, y para ello puede endeudarse o desahorrar. Busca demostrar a sus vecinos que está en condiciones de consumir como ellos. Con este efecto se explicaban los datos de sección cruzada, que retratan a los hogares en un instante concreto.

2. *Efecto trinquete*[13]: a la renta relativa se suma otro determinante del consumo, lo que el hogar ya consumía en el pasado. Los hábitos hacen costoso renunciar a un nivel de vida ya conquistado, de modo que lo consumido antes condiciona lo que se consume hoy. Con este efecto se explicaban las series temporales cortas, es decir, la conducta del consumo a corto plazo.

[11] Formalmente, a corto plazo $(C_t/Y_t)_{\text{CP}} = C_0/Y_t + c \;\rightarrow\; d/dY_t(C_t/Y_t) = -C_0/Y_t^2 < 0$, mientras que a largo plazo $(C_t/Y_t)_{\text{LP}} = k \;\rightarrow\; d/dY_t(C_t/Y_t) = 0$, con $k > c$ constante.

[12] La idea original se encuentra ya en Thorstein Veblen, si bien Duesenberry le da un mayor desarrollo. La consideración del efecto demostración rompe con el supuesto de la microeconomía clásica de que las demandas individuales de los agentes son independientes y plantea la interdependencia en las preferencias de los agentes.

[13] La formalización de los hábitos se recoge en el apartado 3.3.2.

**Razonamiento**: con una renta que crece al ritmo de su tendencia y sin cambios en la posición relativa de cada hogar, los dos efectos frenan la reacción del consumo ante la renta en el corto plazo, de modo que es menor que en el largo. Así surgen dos regularidades sobre la propensión media a consumir, que fluctúa en el corto plazo, con valores más altos cuanto más baja es la renta, y es estable en el largo.

**Desarrollo gráfico**: la hipótesis de la renta relativa se representa en el plano renta disponible-consumo (figura A33.2):

1. *Función de consumo a largo plazo*, $C_{\text{LP}}$: pasa por el origen, porque a largo plazo la propensión media al consumo es constante.

2. *Función de consumo a corto plazo*, $C_{\text{CP}}$: tiene menor pendiente que $C_{\text{LP}}$, porque los hábitos adquiridos (efecto trinquete) frenan la reducción del consumo cuando cae la renta.

3. *Equilibrio*: punto 0, $(Y_0, C(Y_0))$, en el que se cortan ambas funciones, con $Y_0$ la renta máxima alcanzada.

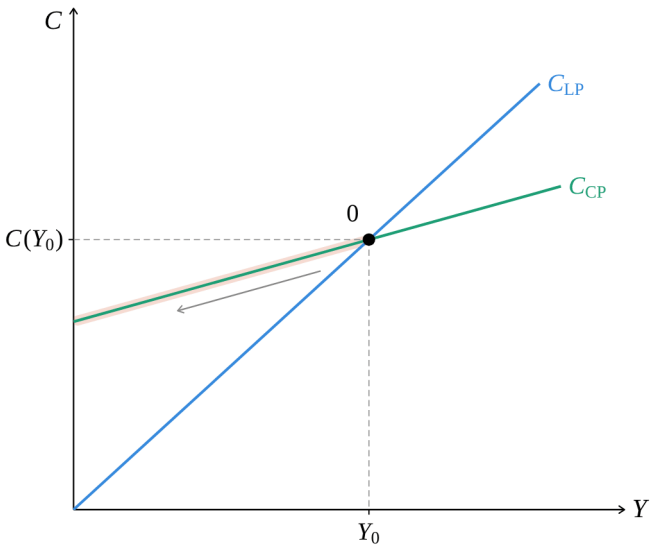



Figura A33.2: Hipótesis de la renta relativa (Duesenberry, 1949). La función de consumo a corto plazo presenta efecto trinquete, mientras que la de largo plazo es proporcional a la renta.

**Estática comparativa**: a partir del punto 0, la respuesta del consumo es asimétrica, porque el efecto trinquete actúa en las caídas de renta y no en las subidas. Se distinguen dos fases:

1. *Caída de la renta*: los hogares se desplazan desde el punto 0 por su curva de corto plazo, $C_{\text{CP}}$ (flecha), y recurren a sus ahorros. Así, el consumo baja menos que si se movieran sobre la curva de largo plazo.

2. *Recuperación*: cuando la recesión queda atrás, el consumo vuelve a subir por $C_{\text{CP}}$ hasta el punto 0 y desde allí retoma su avance sobre $C_{\text{LP}}$.

**Extensiones**: por el efecto trinquete, la propensión media al consumo sube en las recesiones, lo que amortigua la caída de la demanda. Como el ahorro de cada hogar depende de su posición en la distribución de la renta (efecto demostración), la tasa de ahorro agregada no aumenta a medida que la economía se enriquece, mientras esa distribución no cambie.

**Conclusiones**:

- *Aspectos positivos*: aporta dos ideas. Una es la suavización del consumo mediante el ahorro y el desahorro[14]. La otra, la de que el consumo se decide en función de una renta distinta de la del periodo, sirvió de base a los desarrollos posteriores. Además, la teoría de Duesenberry reconcilia la propensión media decreciente a corto plazo con la constante a largo plazo, que Keynes dejaba sin explicar.

- *Limitaciones*: deja sin explicar con rigor el proceso de ajuste por el que, superada la recesión, el consumo vuelve a sus niveles previos y a la curva de tendencia. Tampoco incorpora aspectos dinámicos.

## 1.3. Conclusiones

La aportación neoclásica relaciona el consumo con la renta, el tipo de interés real y, desde Pigou, la riqueza, pero lo trata como un residuo del ahorro y confía en la flexibilidad de los precios para corregir las caídas de la demanda. En cambio, Keynes hace del consumo una función estable de la renta corriente, cuya propensión marginal fundamenta el multiplicador. Su predicción de una propensión media decreciente encaja con los datos de corto plazo, pero no con las series largas de Kuznets. Duesenberry reconcilia ambos horizontes con la renta relativa y los hábitos, aunque sin explicar cómo se ajusta el consumo tras una recesión.

Las funciones keynesianas ligan así el consumo a la renta corriente o pasada, cuando los hogares deciden cuánto consumir y ahorrar teniendo en cuenta también su renta futura, que la sección 2 incorpora a la decisión de consumo.

[14] Esto se debe al efecto trinquete (el consumidor tiene en cuenta inercias y hábitos). Por ello, es una suavización *backward looking* (que mira hacia el pasado).

# 2. El consumo como función intertemporal en horizontes amplios

Ante esta situación, surgen en los años cincuenta las **teorías de la renta normal**, que se apoyan en las aportaciones de Irving Fisher (1930). Todas ellas resuelven el enigma de Kuznets con un mismo mecanismo, hacer depender el consumo de la riqueza y no solo de la renta corriente, y difieren en el horizonte, en cómo se forma la renta relevante y en el tratamiento de la incertidumbre.

## 2.1. Fundamentos del consumo intertemporal (Fisher, 1930)

**Origen**: Irving Fisher (1930) formuló el modelo de elección intertemporal del consumidor del que parten las teorías de la renta normal.

**Idea**: en el modelo de elección intertemporal en dos periodos, el consumidor reparte su riqueza entre consumo presente y futuro, de modo que el consumo depende de la riqueza y no solo de la renta corriente.

> El problema de elección intertemporal del consumidor, la función de utilidad de Samuelson, la restricción de riqueza, la ecuación de Euler y la función de consumo que de ella resulta se desarrollan en el tema A29, apartado 1.

### 2.1.1. Supuestos comunes a las teorías de consumo dinámico

Las teorías del consumo intertemporal comparten estos supuestos:

**Entorno**: un mundo neoclásico con mercados de capitales eficientes y completos, en los que el consumidor puede prestar y endeudarse al mismo tipo de interés. Este supuesto hace posible la suavización del consumo, y su relajación es el objeto de la sección 3.

**Agentes**: consumidores perfectamente racionales y previsores (*forward looking*), que tienen en cuenta sus rentas futuras al decidir el consumo de hoy. Sus preferencias se recogen en una función de utilidad intertemporal aditivamente separable, con impaciencia y homotética[15]. La función de utilidad instantánea es continua, creciente y estrictamente cóncava, lo cual expresa el deseo de suavizar el consumo.

### 2.1.2. Modelo

**Planteamiento**: el consumidor elige la senda de consumo que maximiza la utilidad intertemporal de Samuelson sujeto a su restricción de riqueza:

$$\max_{\{c_t\}} \sum_{t=0}^{T} \beta^t u(c_t) \quad \text{s.a.:} \quad \sum_{t=0}^{T} \frac{c_t}{(1+r)^t} = A_{-1} + \sum_{t=0}^{T} \frac{y_t}{(1+r)^t} \equiv W_0, \qquad \beta \in (0,1)$$

donde $\beta = 1/(1+\rho)$ es el factor de descuento, $\rho$ la tasa de preferencia temporal, $r$ el tipo de interés real, $A_{-1}$ los activos iniciales y $W_0$ la riqueza, suma de los activos y del valor presente de las rentas futuras.

[15] De esta manera, se garantiza la relación proporcional entre la riqueza y el consumo en el modelo HCV y entre la renta permanente y el consumo permanente en el modelo HRP.

**CPO**: con solución interior y la restricción agotada, derivando el lagrangiano respecto al consumo de cada periodo y eliminando el multiplicador, resulta la *ecuación de Euler*, que fija el perfil de la senda:

$$\frac{du}{dc_t} = \frac{1+r}{1+\rho}\,\frac{du}{dc_{t+1}}$$

La ecuación compara lo que el mercado paga por esperar ($r$) con lo que al consumidor le cuesta esperar ($\rho$), y expresa el deseo de mantener una senda de consumo estable, sea constante, creciente o decreciente. Pueden darse tres casos:

- $r = \rho$: el consumo es constante, pues la recompensa por posponer el consumo compensa exactamente la impaciencia.

- $r > \rho$: es creciente. El ahorro rinde más de lo que pesa la impaciencia y conviene consumir menos hoy y más mañana.

- $r < \rho$: es decreciente. La impaciencia domina y el consumidor adelanta consumo, endeudándose o desahorrando.

En todo caso se produce el *alisamiento o suavización del consumo* (*consumption smoothing*), por el que el consumidor iguala entre periodos la utilidad marginal descontada por unidad de precio en valor presente. Gracias a la utilidad marginal decreciente ($u'' < 0$) y a los mercados financieros completos, el consumo no sigue las oscilaciones de la renta, sino una senda regular.

**CSO**: la concavidad estricta de $u$ hace estrictamente cóncava la utilidad intertemporal y la restricción de riqueza es lineal, de modo que la ecuación de Euler, junto con la restricción, es necesaria y suficiente para el máximo.

**Solución**:

- *Función de consumo individual*: la restricción de riqueza fija el nivel de la senda. Sustituyendo el perfil de Euler en la restricción, el consumo de cada periodo resulta ser una fracción de la riqueza del individuo $h$:

$$c_t^h = \kappa^h\, W_t^h, \qquad W_t^h = A_{t-1}^h + \sum_{j=0}^{T-t} \frac{y_{t+j}^{e,h}}{(1+r)^j}$$

donde $c_t^h$ es el consumo del individuo $h$ en el periodo $t$ y $W_t^h$ su riqueza total en $t$. La riqueza total suma los activos acumulados al inicio del periodo, $A_{t-1}^h$ (riqueza no humana), y el valor presente de las rentas laborales que el individuo espera percibir en cada periodo $t+j$, $y_{t+j}^{e,h}$ (riqueza humana). Además, $T$ es el último periodo de vida, de modo que $T-t$ son los periodos que restan, $r$ es el tipo de interés real y $\kappa^h$, la propensión a consumir la riqueza. Esta propensión depende del tipo de interés, de la impaciencia y de los periodos que restan, pero no de la renta corriente, porque la renta del periodo entra solo como un sumando más de la riqueza.

- *Función de consumo agregada*: se obtiene sumando las individuales. Sin embargo, como $\kappa^h$ difiere entre individuos (por edad, preferencias y horizonte), la suma solo es una función estable de la riqueza agregada si la distribución de la población por edades y de la riqueza entre grupos es estable.

Bajo ese supuesto, puede obtenerse una función de consumo agregada canónica como:

$$C_t = \kappa\, W_t$$

con $W_t$ la riqueza agregada (activos más valor presente de las rentas laborales esperadas). Descompuesta en sus componentes observables, la función toma la forma

$$C_t = a\, A_{t-1} + b\, Y_t$$

con $b$ menor que la propensión marginal a consumir keynesiana. Al igual que la función de consumo agregada keynesiana, tiene dos componentes:

- *Componente predeterminado*: $a\, A_{t-1}$, que ocupa el lugar del consumo autónomo keynesiano pero no es una constante. Este componente depende de la riqueza no humana acumulada al inicio del periodo (activos reales y financieros, $A_{t-1}$), dada en $t$ pero variable en el tiempo, a través de la propensión a consumir la riqueza, $a \in (0, 1)$.

- *Componente inducido*: $b\, Y_t$, que depende de la renta corriente, $Y_t$, a través de la propensión marginal a consumir la renta corriente, $b \in (0, 1)$. La renta corriente entra como recurso del periodo y como aproximación a la riqueza humana, que no es observable.

La hipótesis del ciclo vital (HCV) y la de la renta permanente (HRP) son precisamente dos formas de hacer operativa esta función, la primera acotando el horizonte al ciclo vital y la segunda definiendo la renta permanente como la anualidad de la riqueza.

**Desarrollo gráfico**: la elección intertemporal del consumidor se representa en el plano consumo presente-consumo futuro (figura A33.3):

1. *Restricción presupuestaria intertemporal*, RP: recoge las combinaciones de consumo presente y futuro que agotan la riqueza. Pasa por la dotación de rentas, punto D, $(y_t, y_{t+1})$, corta los ejes en $W_0$ y $(1 + r)\, W_0$ y su pendiente es $-(1 + r)$.

2. *Curva de indiferencia intertemporal*, $u_0$: es convexa porque la utilidad instantánea es estrictamente cóncava, lo que expresa el deseo de suavizar el consumo.

3. *Óptimo*: punto 0, $(c_t^0, c_{t+1}^0)$, tangencia entre RP y la curva de indiferencia más alejada del origen. En la figura queda a la derecha de la dotación y el consumidor se endeuda por $c_t^0 - y_t$, mientras que a la izquierda ahorraría.

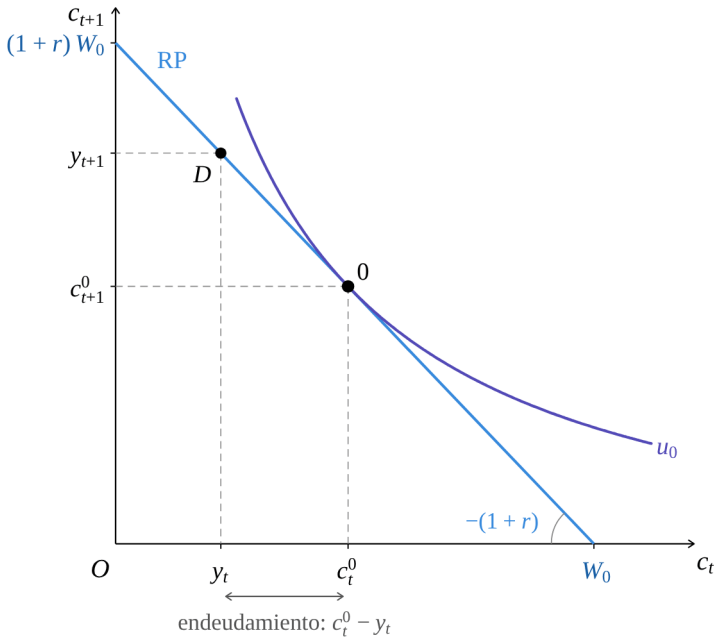



Figura A33.3: Consumo intertemporal en dos periodos, con la restricción presupuestaria y la elección óptima del consumidor.

### 2.1.3. Conclusiones

**Aspectos positivos**: el modelo intertemporal dota a la función de consumo de fundamento microeconómico y sustituye la renta corriente por la riqueza como determinante. Además, explica con un solo mecanismo, la suavización, la evidencia de Kuznets:

- *Corto plazo*: la renta fluctúa más que la riqueza y la propensión media al consumo cae en los auges.
- *Largo plazo*: renta y riqueza crecen a la par y la propensión media es constante.

**Limitaciones**: supone certidumbre sobre las rentas futuras, mercados de capitales perfectos y preferencias separables y homotéticas. Sin esos supuestos, la agregación deja de ser válida y el consumo vuelve a depender de la renta corriente. Las tres teorías que siguen mantienen estos supuestos (salvo Hall, que introduce incertidumbre y utilidad cuadrática), y la sección 3 los levanta.

## 2.2. Hipótesis del ciclo vital (Modigliani y Brumberg, 1954; Ando y Modigliani, 1963)

**Origen**: la hipótesis del ciclo vital se debe a las aportaciones de Franco Modigliani y Richard Brumberg (1954) y de Albert Ando y Franco Modigliani (1963).

**Idea**: esta aportación pone el foco en la función que desempeña el ahorro a la hora de determinar el plan de consumo óptimo.

### 2.2.1. Supuestos

Además de los comunes, el modelo del ciclo vital introduce dos supuestos particulares:

**Entorno**: la vida del individuo es finita, con $T$ periodos conocidos, y el individuo no deja herencias, de modo que agota su riqueza al morir. La renta laboral fluctúa a lo largo de la vida pero es predecible. Concretamente, es baja en la juventud, elevada en la etapa central y nula en la jubilación (perfil de joroba).

### 2.2.2. Modelo

**Planteamiento**: es el problema del modelo común con horizonte igual a la vida del individuo. En cada periodo $t$, el individuo $h$ elige la senda de consumo del periodo actual y de los $T-t$ que le quedan para maximizar su utilidad intertemporal sujeto a su riqueza vital, $W_t^h$, sin dejar herencias.

**CPO**: dan la ecuación de Euler. Con $r=\rho$ el consumo es constante a lo largo de la vida y con $r>\rho$, ligeramente creciente. En ambos casos, su perfil es independiente del de la renta.

**CSO**: son las del modelo común (utilidad instantánea estrictamente cóncava y restricción de riqueza lineal).

**Solución**:

▪ *Función de consumo individual*: en cada periodo, el agente consume una fracción, $\kappa_t^h$, de su riqueza (real) a lo largo de su vida, $W_t^h$:

$$c_t^h = \kappa_t^h W_t^h, \qquad W_t^h = A_{t-1}^h + y_t^h + \sum_{j=1}^{T-t} \frac{y_{t+j}^{e,h}}{(1+r)^j}$$

con la notación del modelo común, donde ahora se separa la renta laboral del periodo, $y_t^h$, de las futuras.

Como en el modelo común, la renta no influye en la fracción $\kappa_t^h$. Para una edad dada, esta fracción es estable, y aumenta con los años a medida que se acorta el tiempo que le queda al agente. El valor actual de la renta futura esperada no es observable y se puede estimar de diferentes formas[16]. En suma, el consumo del agente es proporcional al valor de su riqueza[17].

▪ *Función de consumo agregada*: con una pirámide de edades y un reparto de la riqueza estables, el consumo real agregado es:

$$C_t = aA_{t-1} + bY_t \;\rightarrow\; (C_t/Y_t) = a(A_{t-1}/Y_t) + b$$

con la misma notación.

**Desarrollo gráfico**: la hipótesis del ciclo vital se representa en el plano tiempo-consumo y renta (figura A33.4):

1. *Renta*, $y(t)$: sigue el perfil de joroba supuesto y se anula con la jubilación.

[16] Para hacer operativo el valor actual descontado de la renta futura, Ando y Modigliani introducen la noción de renta esperada media y la aproximan como proporcional a la renta corriente: $\bar{y}_t^{e,h} = \frac{1}{T-t}\sum_{j=1}^{T-t} \frac{y_{t+j}^{e,h}}{(1+r)^j} = \omega\, y_t^h$.

[17] O de manera desagregada: $W_t^h = A_{t-1}^h + y_t^h + (T-t)\,\omega\, y_t^h \;\rightarrow\; c_t^h = \kappa_t^h\,[A_{t-1}^h + [1+\omega(T-t)]\,y_t^h] \;\rightarrow\; c_t^h = a^h A_{t-1}^h + b^h y_t^h$, con $a^h = \kappa_t^h$ y $b^h = \kappa_t^h[1+\omega(T-t)]$.

2. *Consumo*, $c(t)$: es ligeramente creciente, porque se supone $r > \rho$, $\forall t$.

3. *Ahorro y desahorro*: para mantener un consumo regular a lo largo de la vida, el individuo pide prestado en la juventud (área A) y acumula ahorro en los años centrales, de mayor renta (área B). Al final de su vida laboral y durante la jubilación, consume lo acumulado (área C). Por ello, la propensión media al consumo sigue un patrón inverso al de la renta, pues desciende en la etapa central de la vida y aumenta en la juventud y la vejez.

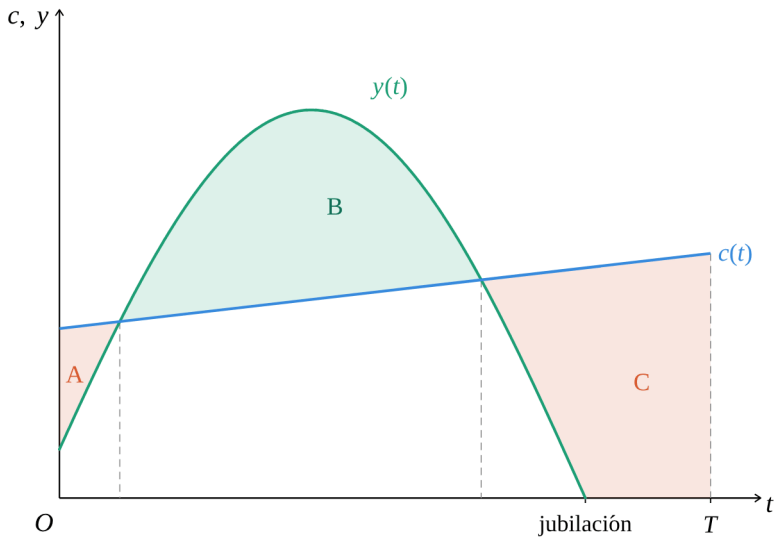



Figura A33.4: Hipótesis del ciclo vital. El perfil de renta y consumo a lo largo de la vida del individuo combina endeudamiento (A), ahorro (B) y desahorro (C).

**Estática comparativa**: a partir de la propensión media al consumo agregada, $C_t/Y_t = a(A_{t-1}/Y_t) + b$, el comportamiento del consumo difiere según el horizonte temporal:

1. *Corto plazo*: los activos o ahorro acumulado, $A_{t-1}$, varían muy poco y pueden considerarse estables. Por ello, a corto plazo, existirá una relación inversa entre la renta y la propensión media al consumo. Así, en fases alcistas disminuye la propensión media al consumo y sube la propensión media a ahorrar y viceversa. De modo intuitivo, un aumento de la renta a corto plazo se trasladará solo en parte al consumo del periodo, pues la maximización de la utilidad a lo largo de la vida exige que el aumento del consumo se reparta entre varios periodos.

2. *Largo plazo*: la acumulación de activos hará que las propensiones media y marginal coincidan y permanezcan estables o constantes.

**Extensiones**: a corto plazo, las políticas activas sí pueden tener efectos sobre el consumo y el ahorro, si bien en el largo plazo, la economía tiene mayor capacidad de ajuste por sí misma. Esta implicación es similar a la de la HRP.

### 2.2.3. Conclusiones

**Aspectos positivos**: la hipótesis del ciclo vital es ampliamente aceptada en la actualidad e incluye tendencias demográficas. Asimismo, es coherente con la evidencia de Kuznets y explica

por qué la propensión media al consumo es decreciente con la renta a corto plazo y en sección cruzada, y constante a largo plazo.

**Limitaciones**: ignora que las restricciones de liquidez pueden limitar el endeudamiento y el desahorro. Además, la evidencia empírica no apoya que exista necesariamente desahorro en la última etapa de la vida.

Otra limitación importante procede de los supuestos. La teoría da al consumidor certeza sobre la renta de todo su ciclo vital y trabaja solo con una renta esperada media estable, sin componentes transitorios.

De modo casi paralelo e independiente, Friedman desarrolló también una teoría del consumo perfectamente compatible con la HCV, si bien dio un tratamiento más formal a la formación de expectativas y la renta, lo cual es un avance.

## 2.3. Hipótesis de la renta permanente (Friedman, 1957)

**Origen**: Milton Friedman (1957) construye una teoría de la demanda de consumo influida por Arthur Pigou e Irving Fisher, en la que el consumo privado no depende solo de la renta del periodo, sino de la riqueza.

**Idea**: los hogares consumen una fracción estable de su renta permanente, la anualidad de su riqueza total, y no de su renta corriente, cuyas desviaciones transitorias se ahorran.

### 2.3.1. Supuestos

**Entorno**:

▪ *Renta*: la renta corriente se divide en dos componentes no correlacionados, $Y_t = Y_t^P + Y_t^T$ con $\text{corr}(Y^P, Y^T) = 0$:

- *Renta permanente* ($Y_t^P$): es la renta que el hogar puede gastar en un periodo sin mermar la riqueza que la genera, de forma que en el periodo siguiente vuelva a disponer del mismo ingreso. Corresponde a la parte prevista y planeada de la renta.
- *Renta transitoria* ($Y_t^T$): es la desviación aleatoria, positiva o negativa, de la renta corriente respecto a la permanente (p. ej. un premio de lotería, un periodo de desempleo)[18]. Corresponde al componente imprevisto de la renta.

▪ *Consumo*: al igual que la renta, el consumo corriente se divide en dos componentes no correlacionados, $C_t = C_t^P + C_t^T$ con $\text{corr}(C^P, C^T) = 0$:

- *Consumo permanente* ($C_t^P$): es el consumo planeado, que refleja el deseo de suavizar el consumo a lo largo de la vida.
- *Consumo transitorio* ($C_t^T$): es la desviación aleatoria, positiva o negativa, del consumo corriente respecto al permanente (p. ej. un gasto por enfermedad). Corresponde al componente imprevisto del consumo.

[18] En el agregado, las rentas positivas y negativas no se cancelan en su totalidad. Por ello, siempre alguna parte de la renta agregada es transitoria.

▪ *Independencia*: además de la ausencia de correlación dentro de cada par, el consumo transitorio y la renta transitoria tampoco están correlacionados entre sí, $\text{corr}(C^T, Y^T) = 0$. Este supuesto es decisivo, puesto que implica que la propensión marginal a consumir la renta transitoria es nula. Así, un aumento transitorio de la renta se ahorra íntegramente y una caída transitoria (p. ej. un despido) se absorbe con desahorro, sin alterar el consumo.

**Agentes**: el consumidor es el del modelo común, con preferencias intertemporales homotéticas (utilidad instantánea isoelástica), de modo que el plan de consumo óptimo es proporcional a los recursos totales. Como los individuos no conocen su renta futura, ni la renta permanente ni la transitoria son observables *ex ante*. Friedman supone que los agentes estiman la renta permanente según la hipótesis de las expectativas adaptativas (HEA), $Y_t^P = \lambda \sum_{j=0}^{J} (1-\lambda)^j Y_{t-j}$, con $\lambda \in (0, 1]$.

Es decir, estiman una media ponderada de las rentas corrientes pasadas (retardo distribuido) que da más peso a las más recientes y que se revisa tanto más deprisa cuanto mayor es $\lambda$.

### 2.3.2. Modelo

**Planteamiento**: es el problema del modelo común con horizonte indefinido. El consumidor maximiza su utilidad intertemporal sujeto a su riqueza real neta total, $W_t = A_t + W_H$, que tiene dos componentes:

▪ *Riqueza no humana* ($A_t$): el valor real neto de los activos acumulados al comienzo del periodo, $A_t = B_t/r + M_t + K_t$, es decir, bonos, dinero y capital real (p. ej. el valor neto de la vivienda[19]).

▪ *Riqueza humana* ($W_H$): el valor descontado de las rentas laborales que el individuo espera percibir a lo largo de su horizonte, $W_H = \sum_{j=0}^{\infty} y_{t+j}^e/(1+r)^{j+1}$, con $r$ el tipo de interés real a largo plazo.

**CPO**: dan la ecuación de Euler, $du/dC_t = \frac{1+r}{1+\rho}\, du/dC_{t+1}$. Con preferencias homotéticas (utilidad isoelástica, $u(C) = C^{1-\theta}/(1-\theta)$, con $\theta$ el coeficiente de aversión relativa al riesgo), la CPO se resuelve para la tasa de crecimiento del consumo, $1 + g_C \equiv \frac{C_{t+1}}{C_t} = \left(\frac{1+r}{1+\rho}\right)^{1/\theta}$ (aproximadamente, $g_C \approx (r-\rho)/\theta$), que no depende del nivel de consumo ni de la riqueza.

El resultado de Friedman se apoya, como en el modelo común, en que la CPO fija la inclinación del perfil de consumo y la riqueza, su nivel. Por eso, dos consumidores con la misma $r$ y los mismos gustos eligen perfiles de la misma forma, a escala de su riqueza.

**CSO**: son las del modelo común.

**Solución**:

▪ *Función de consumo individual*: de la CPO, el plan óptimo es $C_{t+j} = (1+g_C)^j \cdot C_t$. Sustituyendo en la restricción presupuestaria intertemporal con horizonte indefinido, $\sum_{j=0}^{\infty} C_{t+j}/(1+r)^{j+1} = W_t$, y sumando la progresión geométrica, se obtiene el nivel del consumo planeado como proporción de la riqueza total:

$$C_t^P = (r - g_C)\, W_t \equiv \kappa\, W_t$$

En el caso $r = \rho$, $g_C = 0$ y $\kappa = r$, el consumidor consume cada periodo exactamente la *anualidad* de su riqueza, es decir, la renta que esta rinde si se mantiene intacta. Friedman llama a esa

[19] Neto del crédito hipotecario.

anualidad renta permanente, $Y_t^P = r\,W_t$, con $r \in (0,1)$, de modo que la función de consumo individual es

$$C_t^P = \kappa W_t = (\kappa/r)\,Y_t^P \equiv k_P\,Y_t^P$$

Esta función refleja cómo el consumo permanente es proporcional a la renta permanente, con $k_P = 1$ si $r = \rho$ y $k_P < 1$ si $r > \rho$.

Friedman no deriva $k_P$ del programa, sino que postula sus determinantes, $k_P = f(r^-, (W_H/W)^-, z^\pm) \in (0,1]$. Concretamente, $k_P$ depende de:

- *Tipo de interés*: negativamente, por ser el coste de oportunidad de consumir en lugar de ahorrar, pues cuanto mayor es $r$ respecto a $\rho$, más se inclina el perfil hacia el futuro y menor es el consumo presente.

- *Riqueza humana*: cuanto mayor es la parte humana de la riqueza, menor es $k_P$, porque la renta del trabajo es más insegura (puede perderse con un despido) y peor aval para un crédito al consumo.

- *Otras variables*: $z$, que recoge las preferencias, la demografía, la incertidumbre o el estatus social.

Lo que no determina $k_P$ es el nivel de la renta permanente. Por eso la propensión media a consumir la renta permanente es constante.

▪ *Función de consumo agregada*: agregando las funciones individuales, $C_t = k_P\,Y_t^P + C_t^T$ e $Y_t = Y_t^P + Y_t^T$. La forma de la función depende de qué ocurre con los componentes transitorios en cada horizonte.

- *Corto plazo*: la renta permanente se estima como retardo distribuido de la renta corriente (HEA), $Y_t^P = \lambda \sum_j (1-\lambda)^j Y_{t-j}$. Sustituyendo, el consumo agregado se aproxima con una función keynesiana, $C_t = C_0 + bY_t$, con $b = k_P\lambda < k_P$, de modo que de un aumento de la renta corriente solo se consume la parte que se incorpora a la renta permanente y el resto se ahorra (alisamiento).

- *Largo plazo*: los componentes transitorios se cancelan. El consumo transitorio lo hace en el agregado, y la renta transitoria, conforme la renta ahorrada pasa a ser permanente[20]. Queda una función estable, $C_t = k_P\,Y_t^P$, con propensión media constante (véase la figura del anexo).

**Estática comparativa**: el consumo permanente responde a los cambios en sus determinantes, no solo a los de la renta corriente.

▪ *Shock de renta transitoria*: si el agente lo identifica como tal, no altera el consumo y el ahorro actúa como variable de ajuste.

▪ *Shock de renta permanente*: altera el consumo en la proporción $k_P$, de forma gradual, a medida que las expectativas adaptativas lo incorporan a $Y^P$.

▪ *Riqueza*: un aumento de la riqueza, humana o no humana, eleva la renta permanente y con ella el consumo (efecto riqueza).

▪ *Tipo de interés*: una subida de $r$ reduce el consumo, porque reduce $k_P$ y el valor descontado de la riqueza, $W$.

[20] Friedman (1957) estimó para los consumidores estadounidenses un horizonte de unos tres años. Los cambios de la renta medida que persisten al menos ese tiempo se tratan como permanentes.

**Extensiones**: la riqueza entra en la decisión de consumo a través de la renta permanente, lo que hace el consumo privado más estable que en Keynes (alisamiento). Por esa vía se abren también los canales de la riqueza, directos e indirectos, en que se apoya el mecanismo de transmisión monetarista.

- *Política monetaria*: un impulso transitorio no anticipado expande el consumo por el efecto riqueza directo (saldos monetarios) e indirecto (tipo de interés). Uno permanente, en cambio, acaba siendo neutral y se traduce en inflación, conforme a la teoría cuantitativa.

- *Política fiscal*: un impulso transitorio no anticipado afecta solo a la renta transitoria y, si los hogares lo identifican como tal, se ahorra íntegramente, sin efecto sobre el consumo. Uno permanente sí se incorpora a la renta permanente y eleva el consumo. Que lo compense la anticipación de los impuestos futuros necesarios para financiarlo no es un resultado de Friedman, sino de Barro (1974), que lo formaliza como equivalencia ricardiana (apartado 3.2.1).

### 2.3.3. Conclusiones

**Aspectos positivos**: presenta tres:

- *Compatibilidad*: la HRP y la HCV llegan a predicciones similares.

- *Evidencia de Kuznets*: explica la propensión media constante a largo plazo junto con una función keynesiana a corto plazo.

- *Presupuestos familiares*: explica dos regularidades de estos estudios. La propensión media al consumo de los hogares de renta baja es, a corto plazo, mayor que la de los de renta alta y mayor que $k_P$. Además, la propensión media a consumir es anticíclica (la de ahorrar, procíclica), pues en el auge la renta transitoria se ahorra y en la recesión se desahorra.

**Limitaciones**: presenta cuatro:

- *Certidumbre*: el programa del consumidor se plantea con información perfecta sobre los flujos de renta de toda la vida. La incertidumbre solo entra, de forma *ad hoc*, a través de la renta transitoria y de la estimación adaptativa de la renta permanente.

- *Inconsistencia interna*: el consumidor suaviza su consumo mirando hacia el futuro (*forward looking*) pero estima su $Y^P$ mirando hacia atrás (expectativas adaptativas), lo cual es contradictorio.

- *HEA*: la hipótesis de expectativas adaptativas puede conducir a errores sistemáticos, frente a la de expectativas racionales.

- *Contrastes*: son difíciles, pues la renta y el consumo permanentes no son observables.

Robert Hall intentó superar la limitación de la certidumbre perfecta al incorporar el riesgo y las expectativas racionales.

## 2.4. Paseo aleatorio de Hall (1978)

**Origen**: Robert Hall (1978) cambió el modo de estudiar la demanda de consumo intertemporal.

**Idea**: en un marco de incertidumbre, expectativas racionales y utilidad cuadrática, la HCV o HRP llevan a que el consumo (real) agregado siga un paseo aleatorio o martingala (hipótesis del paseo aleatorio, HPA). Con $r \neq \rho$, el consumo sigue un proceso autorregresivo de primer orden.

### 2.4.1. Supuestos

**Entorno**: el tiempo es discreto y hay incertidumbre. El individuo no conoce sus rentas futuras, de modo que el consumo es una variable aleatoria. Con todo, no existe incertidumbre sobre tipos de interés ni sobre el descuento, con $r = \rho = 0$[21].

**Agentes**: incorporan en su patrón de consumo toda la información relevante según la conocen y no cometen errores sistemáticos (hipótesis de expectativas racionales). La función de utilidad instantánea es cuadrática, lo cual implica que la utilidad marginal es lineal, $u'(c_t) = 1 - \eta c_t$.

**Restricciones**: el individuo debe liquidar todas sus deudas pendientes al final de su vida (*condición de no-Ponzi*).

### 2.4.2. Modelo

**Planteamiento**: el consumidor maximiza la utilidad esperada de su senda de consumo sujeto a su riqueza esperada:

$$\max_{c_{t+j}} \ \mathbb{E}_t \left[ \sum_{j=0}^{T-t} \left( c_{t+j} - \frac{\eta}{2} c_{t+j}^2 \right) \right], \ \eta > 0$$

$$\text{s.a.:} \begin{cases} \sum_{j=0}^{T-t} \mathbb{E}_t[c_{t+j}] \leq a_t + \sum_{j=0}^{T-t} \mathbb{E}_t[y_{t+j}] \\ a_{T+1} = 0 \end{cases}$$

donde $c$ es el consumo, $\eta$ el parámetro de la utilidad cuadrática, $a_t$ la riqueza no humana al inicio del periodo, $y$ la renta y $\mathbb{E}_t$ la expectativa condicionada a la información disponible en $t$.

**CPO**: con solución interior y la restricción agotada, derivando el lagrangiano como en el modelo común, la ecuación de Euler esperada iguala la utilidad marginal de hoy a la esperada de mañana, $\underbrace{1 - \eta c_t}_{\text{Umg}_t} = \underbrace{\mathbb{E}_t[1 - \eta c_{t+1}]}_{\mathbb{E}[\text{Umg}_{t+1}]}$. Como la utilidad marginal es lineal, $\mathbb{E}_t[1 - \eta c_{t+1}] = 1 - \eta\, \mathbb{E}_t[c_{t+1}]$, la CPO se reduce a $c_t = \mathbb{E}_t[c_{t+1}]$.

Para prever el consumo de mañana basta, por tanto, con el de hoy, y ninguna otra variable del pasado añade capacidad predictiva. La intuición es la del alisamiento del consumo. Si el consumidor esperase un aumento del consumo, la utilidad marginal actual sería mayor que la esperada futura, y le convendría elevar el consumo actual hasta que la expectativa de cambio desapareciese.

[21] Basta con $r = \rho$ para que el consumo siga un paseo aleatorio. El supuesto $r = \rho = 0$ solo simplifica la restricción presupuestaria.

**CSO**: la utilidad cuadrática es estrictamente cóncava en el tramo relevante ($c_t < 1/\eta$, donde la utilidad marginal es positiva) y la restricción es lineal, de modo que la CPO es necesaria y suficiente.

**Solución**:

▪ *Función de consumo individual*: introduciendo la CPO en la restricción presupuestaria, el consumidor reparte por igual entre los periodos que le quedan su riqueza esperada, activos más valor esperado de las rentas futuras:

$$c_t = \frac{1}{T-t+1}\left(a_t + \sum_{j=0}^{T-t} \mathbb{E}_t[y_{t+j}]\right)$$

La función de consumo de Hall es idéntica a la función de consumo de la HCV, solo que con renta esperada en lugar de cierta. Como la utilidad marginal es lineal, solo importa la esperanza de la renta y no su varianza (*equivalencia cierta*), de modo que la incertidumbre carece de efectos sobre el consumo y la conducta es la misma que en certidumbre. El ahorro precautorio (apartado 3.3.3) revisa precisamente este resultado.

▪ *Función de consumo agregada*: el modelo de Hall es un modelo de agente representativo, por lo que el modelo agregado hereda las propiedades del individual.

**Estática comparativa**: una innovación (cambio no esperado) en la riqueza esperada, $\Delta a_t$ o una revisión de $\sum \mathbb{E}_t[y_{t+j}]$, altera el consumo en $1/(T-t+1)$ de su cuantía, de una vez y para siempre, mientras que una variación anticipada de la renta no lo altera.

**Dinámica**: la CPO y la hipótesis de expectativas racionales implican que el consumo sigue un paseo aleatorio o martingala[22]:

$$c_{t+1} = c_t + \varepsilon_{t+1}, \quad \mathbb{E}_t[\varepsilon_{t+1}] = 0$$

donde $\varepsilon_{t+1}$ es la innovación, es decir, la nueva información sobre la riqueza esperada que llega en $t+1$. La conclusión principal de Hall es que las variaciones del consumo son imprevisibles, porque todo lo previsible quedó recogido en la decisión anterior y el consumo solo se mueve con la información nueva. Como esta opera a través de la riqueza esperada, su efecto persiste y no se deshace después[23].

**Extensiones**: la política económica solo es efectiva a través de su impacto en $\varepsilon_{t+1}$, es decir, solo si no es anticipada.

▪ *Política monetaria*: con $r = \rho = 0$ el tipo de interés no entra en la decisión. En cambio, con $r \neq \rho$, el consumo sigue un proceso autorregresivo de primer orden, AR(1), en lugar de un paseo aleatorio puro[24] y el tipo de interés afecta a la inclinación de la senda.

▪ *Política fiscal*: una rebaja fiscal anunciada no altera el consumo cuando se hace efectiva, pues ya se incorporó a la riqueza esperada en el anuncio. Solo una medida sorpresa lo desplaza, y de forma permanente.

[22] Por definición, con expectativas racionales los agentes no cometen errores sistemáticos: $c_t - \mathbb{E}_{t-1}[c_t] = \varepsilon_t$, con $\mathbb{E}_{t-1}[\varepsilon_t] = 0$. Por la ecuación de Euler, $\mathbb{E}_{t-1}[c_t] = c_{t-1}$. Por lo tanto, $c_t = c_{t-1} + \varepsilon_t \;\rightarrow\; c_t - c_{t-1} = \varepsilon_t$, donde el término aleatorio refleja los cambios no esperados.

[23] Esto contradecía la visión tradicional, según la cual ante una caída de la renta o producción era previsible un aumento posterior del consumo.

[24] Con utilidad cuadrática y $r \neq \rho$, la ecuación de Euler, $1-\eta c_t = \frac{1+r}{1+\rho}\,(1-\eta\,\mathbb{E}_t[c_{t+1}])$, da $\mathbb{E}_t[c_{t+1}] = \frac{r-\rho}{\eta(1+r)} + \frac{1+\rho}{1+r}\,c_t$, es decir, un AR(1) con coeficiente $(1+\rho)/(1+r)$.

### 2.4.3. Conclusiones

**Aspectos positivos**: incorpora la incertidumbre y las expectativas racionales a las teorías de la renta normal, cerrando la inconsistencia de la HRP entre alisamiento hacia adelante y expectativas hacia atrás. Además, ofrece una restricción contrastable (ninguna variable retardada debe predecir $\Delta c_t$), y en el contraste original de Hall la renta retardada no tiene poder predictivo, aunque los precios de las acciones retardados sí, ligeramente.

**Limitaciones**: presenta cuatro:

- *Utilidad cuadrática*: implica que la utilidad marginal es lineal, puede ser negativa a partir de cierto nivel de consumo y el coeficiente absoluto de aversión al riesgo es creciente con la renta.
- *Equivalencia cierta*: excluye el ahorro precautorio.
- *Agente representativo*: la agregación ignora la heterogeneidad de los hogares.
- *Evidencia*: rechaza el paseo aleatorio en dos direcciones, el exceso de sensibilidad y el exceso de suavidad (apartado 3.1).

## 2.5. Conclusiones

Las teorías de la renta normal convergen en la suavización del consumo y difieren en los rasgos que resume la tabla A33.1, que las compara también con las funciones de renta corriente y relativa de la sección 1.

Tabla A33.1: Teorías de la demanda de consumo: comparación.

| Teoría | Renta relevante | Horizonte | Expectativas | Propensión marginal a corto plazo | Propensión media a largo plazo | Eficacia de la política fiscal |
|---|---|---|---|---|---|---|
| Keynes (1936) | Corriente | Un periodo | — | $c \in (0,1)$, alta | Decreciente | Muy alta (multiplicador) |
| Duesenberry (1949) | Relativa y pasada | Un periodo con hábitos | — | Baja en recesión (trinquete) | Constante | Alta, asimétrica |
| HCV (Modigliani y Brumberg, 1954) | Riqueza vital | Vida finita | Certidumbre | Baja ($\kappa$ sobre la riqueza) | Constante | Baja si transitoria |
| HRP (Friedman, 1957) | Permanente | Indefinido | Adaptativas | Baja respecto a la transitoria | Constante ($k_P$) | Nula si se percibe como transitoria, parcial si permanente |
| Hall (1978) | Riqueza esperada | Vida finita | Racionales | Solo ante sorpresas | Constante | Solo si no anticipada |

# 3. Desarrollos recientes y demanda de bienes de consumo duradero

Las teorías de la renta normal y del paseo aleatorio descansan sobre fundamentos microeconómicos sólidos, pero los datos no siempre las respaldan. Para cerrar esa brecha, la literatura posterior ha ido relajando parte de sus supuestos con el fin de explicar regularidades como el exceso de sensibilidad del consumo y su exceso de suavidad, las decisiones que combinan consumo y ocio, los hábitos de consumo o la demanda de bienes de consumo duradero.

## 3.1. Exceso de sensibilidad y exceso de suavidad

Los dos fenómenos que organizan la literatura posterior a Hall son dos violaciones de la hipótesis del paseo aleatorio en direcciones opuestas. Así, el consumo reacciona a lo que ya se sabía, pero poco a lo que no se sabía.

### 3.1.1. Exceso de sensibilidad (Flavin, 1981)

**Origen**: el artículo de Marjorie Flavin (1981) fue el primer contraste sistemático de la HPA frente a la renta corriente.

**Idea**: el consumo responde a cambios anticipados de la renta que, según Hall, no deberían alterarlo, pues ya estaban incorporados en la riqueza esperada (p. ej. una devolución previsible del impuesto sobre la renta eleva el consumo cuando se cobra).

**Razonamiento**: la HPA implica que $\Delta c_t$ es ortogonal a toda la información disponible en $t-1$, incluida la parte predecible de la renta. Flavin modela la renta como un proceso autorregresivo de medias móviles (ARMA), separa su innovación de su componente anticipado y añade este último a la ecuación del paseo aleatorio[25]. Bajo la HPA su coeficiente debe ser nulo, pero Flavin estima que es positivo y significativo. Esta restricción es la parte de la predicción de Hall que la evidencia rechaza con más claridad (tema A29, apartado 1.4.2).

Se han propuesto cuatro explicaciones:

- *Restricciones de liquidez*: impiden adelantar el consumo de una renta futura conocida.
- *Reglas simples*: una parte de los consumidores consume su renta corriente. Campbell y Mankiw (1989) estiman que estos consumidores perciben en torno a la mitad de la renta agregada.
- *Ahorro precautorio*: frena el consumo hasta que la renta se materializa.
- *Bienes duraderos*: su compra se financia con la renta del periodo.

En consecuencia, si el consumo responde a la renta corriente, la política fiscal anticipada recupera parte de su eficacia.

**Conclusiones**:

- *Aspectos positivos*: ofrece un contraste estructural de la HPA. Ese contraste separa la parte anticipada de la renta de su innovación y muestra que el consumo responde a la primera. Este resultado orienta los desarrollos que siguen.

[25] Formalmente, Marjorie Flavin propone: $\Delta c_t = \alpha \Delta y_t + v_t + \phi \varepsilon_t$, donde $\alpha$ mide la sobrerreacción del consumo ante un cambio en la renta actual.

- *Limitaciones*: el contraste se basa en series agregadas y en un modelo estadístico concreto de la renta, por lo que su resultado depende de esa especificación. Además, documenta la anomalía sin discriminar entre las explicaciones propuestas, para lo que hace falta evidencia microeconómica.

### 3.1.2. Exceso de suavidad (Deaton, 1987)

**Origen**: el exceso de suavidad fue documentado por Angus Deaton (1987) y por John Campbell y Angus Deaton (1989).

**Idea**: el consumo reacciona menos de lo debido a las innovaciones (cambios no anticipados) de la renta.

**Razonamiento**: en la HPA, una innovación de la renta altera el consumo en la anualidad de la revisión de la riqueza esperada. Si la renta agregada es un proceso muy persistente (una raíz unitaria con deriva y crecimiento positivamente autocorrelacionado, como indican los datos), una sorpresa de renta de una unidad eleva la renta esperada de todos los periodos futuros en al menos una unidad. La renta permanente sube entonces más que la corriente y el consumo debería variar más que la renta.

La evidencia muestra lo contrario, pues la varianza de $\Delta c_t$ es muy inferior a la de las innovaciones de la renta, resultado que se conoce como *paradoja de Deaton*. Se han propuesto tres explicaciones:

- *Información*: los hogares no distinguen bien las perturbaciones permanentes de las transitorias y reaccionan a una sorpresa como si fuera en parte transitoria.

- *Restricciones y prudencia*: los hogares están restringidos o ahorran por precaución.

- *Preferencias*: incorporan hábitos o el ocio.

Campbell y Deaton muestran que ambos fenómenos son la misma anomalía vista desde dos lados. Si el consumo reacciona a la información pasada (exceso de sensibilidad), necesariamente reacciona menos a la nueva (exceso de suavidad). Para la política económica, la lección es que el ajuste del consumo dista de ser perfecto, lo que devuelve espacio a las políticas estabilizadoras.

**Conclusiones**:

- *Aspectos positivos*: muestra que la HPA no implica por sí misma un consumo suave, pues con una renta muy persistente predice lo contrario. Además, la paradoja de Deaton convierte la persistencia de la renta en un elemento central del contraste de las teorías del consumo.

- *Limitaciones*: la paradoja depende del proceso que se suponga para la renta. Con series cortas es difícil distinguir una raíz unitaria de un proceso estacionario en torno a una tendencia, y en este último caso la HPA sí predice un consumo suave.

## 3.2. Cambios en los supuestos sobre el entorno y los agentes

Un primer grupo de desarrollos cambia los supuestos sobre el horizonte de los agentes, su acceso al crédito y su racionalidad.

### 3.2.1. Horizonte infinito: modelos de generaciones solapadas y la equivalencia ricardiana (Barro, 1974)

**Origen**: Robert Barro (*Are Government Bonds Net Wealth?*, 1974) recuperó la proposición de Ricardo en un modelo de generaciones solapadas con altruismo.

**Idea**: con horizonte infinito, o finito pero con altruismo intergeneracional, los hogares internalizan la restricción presupuestaria del sector público.

> La equivalencia ricardiana se desarrolla en el tema A06, apartado 3.2.1, y su frontera con los modelos de generaciones solapadas, en el tema A29, apartado 3.3. Sus consecuencias para la política fiscal se encuentran en el tema A38.

**Razonamiento**: los agentes anticipan que la deuda emitida hoy exige impuestos futuros de igual valor actual, de modo que no la consideran riqueza neta. Ante una rebaja de impuestos financiada con deuda, ahorran íntegramente la renta adicional para pagar esos impuestos futuros y el consumo no varía. Por tanto, dada la senda del gasto público, su forma de financiación (impuestos o deuda) es irrelevante para el consumo, es decir, se cumple la equivalencia ricardiana.

**Conclusiones**:

- *Aspectos positivos*: obliga a tratar la deuda pública como impuestos diferidos y muestra que el efecto de una rebaja fiscal sobre el consumo depende de cómo perciban los hogares su financiación futura.

- *Limitaciones*: descansa en supuestos exigentes (altruismo operativo entre generaciones, mercados de capitales perfectos, impuestos de cuantía fija). De hecho, basta con que una parte de los hogares sufra restricciones de liquidez para que una rebaja de impuestos financiada con deuda eleve el consumo.

### 3.2.2. Restricciones de liquidez (Deaton, 1991)

**Origen**: Angus Deaton (1991) incorporó la imposibilidad de endeudarse al problema intertemporal con incertidumbre (tema A29, apartado 1.4.3). Campbell y Mankiw (1989) ya habían interpretado la fracción de consumidores que gastan su renta corriente como reflejo, en parte, de estas restricciones.

**Idea**: los modelos de HRP, HCV y HPA se basan en la existencia de mercados financieros perfectos (eficientes y completos). No obstante, en el mundo real existen imperfecciones en estos mercados (p. ej. el tipo de financiación no coincide con el tipo de rendimiento).

**Razonamiento**: las restricciones de liquidez empujan al alza el ahorro de dos maneras:

1. *Límite de gasto efectivo*: cuando la restricción se activa, el consumidor no obtiene todo el crédito que querría y su consumo queda por debajo del deseado. Formalmente, la CPO se cumple con desigualdad, $du/dc_t > (du/dc_{t+1})(1+r)/(1+\rho)$. La relación marginal de sustitución intertemporal supera al coste de oportunidad del consumo presente y el individuo querría consumir más en el periodo $t$ y menos en el periodo $t+1$, pero no puede.

2. *Límite de gasto anticipado*: el consumidor que hoy puede endeudarse sabe que, si su renta cae más adelante, quizá no le presten. Si no tiene ahorros con los que cubrirse, esa perspectiva le lleva a recortar ya su consumo. La mera posibilidad de restricciones futuras introduce, por tanto, una incertidumbre que deprime el consumo actual aunque hoy ninguna restricción sea efectiva.

**Conclusiones**:

- *Aspectos positivos*: las restricciones de liquidez permiten explicar el exceso de sensibilidad del consumo respecto a su renta disponible, en la medida en que dificultan la suavización.

- *Limitaciones*: cuantitativamente se quedan cortas para dar cuenta del ahorro que se observa, y por eso suelen combinarse con el motivo precautorio.

### 3.2.3. Racionalidad limitada (Simon, 1957; Kahneman y Tversky, 1979; Thaler y Sunstein, 2008)

**Origen**: Herbert Simon (1957) acuñó la racionalidad limitada, Daniel Kahneman y Amos Tversky formularon la teoría de la perspectiva (1979), y Richard Thaler y Cass Sunstein (*Nudge*, 2008), la arquitectura de la elección. Su aplicación al consumo intertemporal se debe sobre todo a David Laibson (1997) y a Hersh Shefrin y Richard Thaler (1988).

**Idea**: todos ellos cuestionan el *homo economicus* neoclásico e introducen elementos que ponen de manifiesto la racionalidad limitada de los agentes.

**Razonamiento**: entre sus conceptos destacan:

1. *Teoría del "empujoncito"* (*nudge*): el diseño del contexto de la decisión influye sobre ella, lo que convierte a Administraciones, empresas o vendedores en "arquitectos de la elección". Todos ellos se sirven de sesgos como el efecto marco (*framing effect*), por el que el encuadre de una decisión con connotaciones positivas o negativas influye sobre ella.

2. *Efecto dotación* o de propiedad (*endowment effect*): los consumidores valoran más un bien cuando lo poseen que cuando no lo poseen (Thaler, 1980).

3. *Descuento cuasihiperbólico* (Laibson, 1997): además del factor de descuento exponencial, $\beta$, el consumidor aplica a todo periodo futuro un factor adicional, $b \in (0, 1)$, de modo que pondera el presente más de lo que implica el descuento exponencial (tema A01, apartado 3.2.1). Sus preferencias son *dinámicamente inconsistentes*, porque el plan de ahorro que hoy considera óptimo deja de serlo mañana. De ahí que, si anticipa esa inconsistencia, demande activos ilíquidos (la vivienda o los planes de pensiones) como mecanismo de compromiso, y que su consumo siga a la renta corriente, lo que explica el exceso de sensibilidad.

4. *Ciclo vital conductual* (Shefrin y Thaler, 1988): los hogares reparten su riqueza en *cuentas mentales*, la renta corriente, los activos corrientes y la renta futura, con propensiones marginales al consumo distintas, la mayor en la renta corriente y la menor en la renta futura. Así, una misma renta en valor actual no genera el mismo consumo según la cuenta a la que llegue, frente a la fungibilidad que supone la hipótesis del ciclo vital.

**Conclusiones**:

- *Aspectos positivos*: las limitaciones a la racionalidad pueden explicar tanto el exceso de suavidad como el exceso de sensibilidad, y el descuento cuasihiperbólico y las cuentas mentales dan un mecanismo concreto para el segundo.

- *Limitaciones*: se trata de un conjunto de sesgos más que de un modelo unificado del consumo, de modo que sus predicciones dependen del sesgo que se elija y son difíciles de distinguir empíricamente de las de los modelos con agentes racionales sujetos a restricciones. Además, su traslado del individuo al consumo agregado no es inmediato.

## 3.3. Cambios en los supuestos sobre la función de utilidad

Un segundo grupo de desarrollos modifica la especificación de la función de utilidad, a la que incorpora el ocio, los hábitos o el motivo precautorio.

### 3.3.1. Decisiones de consumo-ocio (Lucas y Rapping, 1969)

**Origen**: la extensión se debe a Robert Lucas y Leonard Rapping (1969), y el modelo se desarrolla en el tema A25, apartado 2.2.

**Idea**: considerar la decisión de trabajo junto a la de consumo endogeniza la renta del individuo en el programa de elección.

**Razonamiento**: la utilidad instantánea pasa a depender del consumo y del ocio, $u(c_t, o_t)$, y la renta deja de ser exógena, pues es el salario real por las horas trabajadas. El resultado depende de la separabilidad de $u$:

- *Utilidad separable*: si $u$ es aditivamente separable en consumo y ocio, la utilidad marginal del consumo no depende del ocio, la ecuación de Euler del consumo es la de Hall y el consumo sigue un paseo aleatorio o un AR(1). La decisión de trabajo se toma aparte.

- *Utilidad no separable*: las horas trabajadas alteran la utilidad marginal del consumo, y por eso en la ecuación de Euler aparecen el salario real de hoy y el de ayer. Consumo, empleo y salario se mueven juntos, y la senda del consumo deja de ajustarse a un paseo aleatorio o a un AR(1).

**Conclusiones**:

- *Aspectos positivos*: la respuesta del consumo a la renta corriente refleja en parte la del empleo, lo que contribuye a explicar el exceso de sensibilidad y, si el ocio amortigua las sorpresas de renta, el exceso de suavidad.

- *Limitaciones*: el resultado descansa en la no separabilidad entre consumo y ocio, cuyo signo y cuantía son difíciles de estimar, de modo que la dirección del efecto queda abierta. Además, si las horas no se eligen libremente (p. ej. con desempleo involuntario), la covariación entre consumo y empleo refleja una restricción y no las preferencias.

### 3.3.2. Hábitos de consumo (Deaton, 1992)

**Origen**: Angus Deaton (1992) recoge la formación de hábitos como explicación de la persistencia del consumo, una idea que procede del efecto trinquete de Duesenberry (apartado 1.2.2).

**Idea**: cuando la utilidad intertemporal no se descompone en sumandos independientes para cada periodo, lo consumido ayer condiciona la satisfacción que da el consumo de hoy. Por ello, cambiar el nivel de consumo se vuelve costoso y aparecen hábitos e inercias.

**Razonamiento**: la utilidad instantánea depende del consumo corriente relativo a un nivel de referencia formado por el consumo pasado, $u(c_t - \gamma\, c_{t-1})$, con $\gamma \in (0,1)$ la intensidad del hábito, o, en forma de cociente, $u(c_t/c_{t-1})$. Elevar el consumo hoy tiene un coste adicional, pues eleva la referencia de mañana, de modo que la ecuación de Euler se cumple para el consumo ajustado por el hábito y no para el consumo en sí.

En consecuencia, el consumidor, ante una sorpresa de riqueza, no salta al nuevo nivel, sino que se acerca a él gradualmente, por lo que cada variación del consumo arrastra parte de la anterior,

$\Delta c_t = \gamma\,\Delta c_{t-1} + \varepsilon_t$. El pasado ayuda así a predecir el presente y el consumo deja de ser un paseo aleatorio.

**Conclusiones**:

▪ *Aspectos positivos*: la respuesta inicial a una innovación es pequeña (exceso de suavidad) y el ajuste posterior es predecible y correlacionado con la renta pasada (exceso de sensibilidad), de modo que los hábitos explican ambos fenómenos a la vez.

▪ *Limitaciones*: el nivel de referencia y la intensidad del hábito se postulan sin una fundamentación propia, y su estimación es sensible a la especificación. Además, la evidencia con datos de hogares tiende a encontrar hábitos más débiles que la de datos agregados, en los que la persistencia puede deberse en parte a la agregación.

### 3.3.3. Ahorro precautorio (Leland, 1968)

**Origen**: Hayne Leland (1968) formalizó el ahorro precautorio.

**Idea**: los hogares ahorran, además de para trasladar renta entre periodos, para hacer frente a la incertidumbre sobre sus rentas futuras.

**Razonamiento**: con utilidad cuadrática, $u'''(c_t) = 0$, rige la equivalencia cierta (apartado 2.4.2). Solo cuenta, pues, la renta esperada y no hay motivo precautorio. Para que lo haya, la utilidad marginal debe ser convexa, $u'''(c_t) > 0$, como ocurre con la exponencial o la isoelástica. Entonces, por la desigualdad de Jensen, la incertidumbre eleva la utilidad marginal esperada del consumo futuro, $\mathbb{E}_t[u'(c_{t+1})] > u'(\mathbb{E}_t[c_{t+1}])$, y la ecuación de Euler solo se restablece con menos consumo presente. El individuo ahorra por prudencia y su consumo crece en valor esperado aunque el tipo de interés coincida con la tasa de descuento.

La intensidad del efecto depende del grado de prudencia, que se mide con el coeficiente de prudencia absoluta o relativa:

$$\text{CPA} = -\frac{u'''(c_t)}{u''(c_t)} \qquad \text{CPR} = -\frac{u'''(c_t)}{u''(c_t)} \cdot c_t$$

donde $u''$ y $u'''$ son la segunda y tercera derivadas de la utilidad instantánea. Estos coeficientes se parecen formalmente a los de aversión al riesgo, aunque miden otra cosa. La aversión al riesgo refleja lo curvada que es la utilidad, y la prudencia, lo curvada que es la utilidad marginal.

**Extensiones**: la impaciencia matiza el resultado. En el modelo de colchón (*buffer stock*) de Christopher Carroll (1997), el consumidor impaciente no llega a endeudarse, pero mantiene un colchón de riqueza reducido. Al comienzo de la vida, el consumo sigue así de cerca a la renta, y el motivo precautorio solo domina a la impaciencia cuando la riqueza cae por debajo de su objetivo.

**Conclusiones**:

▪ *Aspectos positivos*: el ahorro precautorio explica que el consumo responda a la incertidumbre y no solo a la renta esperada, lo que contribuye al exceso de suavidad ante sorpresas y a la sensibilidad ante restricciones anticipadas.

▪ *Limitaciones*: su cuantía depende de la prudencia del consumidor y de la incertidumbre de la renta, ambas difíciles de medir, por lo que las estimaciones de la parte del ahorro debida a este motivo varían mucho. Además, sus predicciones son difíciles de separar empíricamente de las que generan las restricciones de liquidez, que también ligan el consumo a la renta corriente.

## 3.4. La demanda de bienes de consumo duradero (Mankiw, 1982; Bernanke, 1985)

**Origen**: la extensión a los bienes duraderos se debe a Gregory Mankiw (1982) y a Ben Bernanke (1985).

**Idea**: hasta aquí solo se han considerado bienes no duraderos, que proporcionan utilidad únicamente en el periodo en que se consumen. Sin embargo, los consumidores también adquieren bienes duraderos (p. ej. coches).

**Planteamiento**: un bien duradero no proporciona utilidad cuando se compra sino mientras dura. Sea un consumidor que vive dos periodos y consume un bien no duradero, $c_t$, y los servicios, $s(d_t)$, de un *stock* de duraderos, $d_t$, que se deprecia a la tasa $\delta$ y se repone con el gasto en duraderos, $x_t = d_t - (1-\delta)\, d_{t-1}$, con $d_0$ dado. El programa es:

$$\max_{c_1, c_2, d_1, d_2} \; u[c_1, s(d_1)] + \beta\, u[c_2, s(d_2)]$$

$$\text{s.a.:} \quad c_1 + x_1 + \frac{c_2 + x_2}{1+r} = W_0$$

donde $c_1$ y $c_2$ son el consumo del bien no duradero y $d_1$ y $d_2$ el *stock* de duraderos en cada periodo (las variables de decisión). El término $s(d_t)$ recoge los servicios del *stock*, $x_t$ es el gasto en duraderos y el resto de la notación es la del apartado 2.1.2.

**CPO**: con solución interior y la restricción agotada, derivando el lagrangiano respecto a $c_t$ y $d_t$, se obtienen dos. La del no duradero es la ecuación de Euler habitual, $u_c(1) = \beta(1+r)\, u_c(2)$. La del duradero iguala la utilidad marginal de sus servicios al *coste de uso* del *stock* en términos del no duradero:

$$u_s(1)\, \frac{ds}{dd_1} = u_c(1)\, \frac{r+\delta}{1+r}$$

Comprar una unidad más de duradero en 1 cuesta una unidad de consumo hoy, pero devuelve $(1-\delta)$ unidades mañana, que descontadas valen $(1-\delta)/(1+r)$. El coste neto es, por tanto, el interés perdido más la depreciación, $(r+\delta)/(1+r)$, análogo al coste de uso del capital de la empresa (tema A34). Con precio relativo del duradero $p_d$ y revalorización esperada $\pi_d$, el coste de uso es $p_d\,(r + \delta - \pi_d)$.

**CSO**: se cumplen con $u$ y $s$ cóncavas y restricción lineal.

**Solución**: por tanto, la suavización se aplica al *stock*, no al gasto. El gasto es mucho más volátil que el consumo de no duraderos y responde con fuerza al tipo de interés y al crédito. Las dos aportaciones difieren en la función de utilidad y en el ajuste del *stock*:

- *Mankiw (1982)*: función de utilidad aditivamente separable entre duraderos y no duraderos, sin costes de ajuste. Con utilidad cuadrática y expectativas racionales, el *stock* de duraderos sigue un paseo aleatorio, como en Hall, y el gasto, $x_t = d_t - (1-\delta)\, d_{t-1}$, un proceso ARMA(1,1) con raíz unitaria, cuya primera diferencia es de medias móviles de orden 1, $\Delta x_t = \varepsilon_t - (1-\delta)\, \varepsilon_{t-1}$.

- *Bernanke (1985)*: función de utilidad no separable y costes de ajuste cuadráticos, de modo que el *stock* se ajusta solo parcialmente hacia su nivel deseado en cada periodo. Con costes fijos (comprar un coche no admite fracciones), el ajuste sigue una regla $(S, s)$ (Grossman y Laroque, 1990), según la cual el consumidor deja que el *stock* se deprecie hasta un umbral y entonces lo repone de una vez.

Con bienes duraderos, el gasto en consumo no sigue un paseo aleatorio ni un AR(1), sea cual sea la función de utilidad.

**Conclusiones**:

- *Aspectos positivos*: al ligar las compras de un bien que se mantiene en *stock* con la renta del momento, ofrece una explicación parcial del exceso de sensibilidad que documenta buena parte de la literatura empírica.

- *Limitaciones*: la evidencia rechaza ambas versiones. En Mankiw, el coeficiente estimado del término de medias móviles implica $\delta \approx 1$, es decir, el gasto en duraderos se comporta como el de bienes no duraderos. En Bernanke, el gasto en duraderos responde a la renta corriente más de lo que predice el modelo.

## 3.5. Evidencia empírica e implicaciones de política económica

**Balance empírico**: la evidencia acumulada a lo largo del tema dibuja un cuadro estable. A largo plazo, la propensión media al consumo es constante (Kuznets), lo que descarta la función keynesiana como relación estructural y respalda las teorías de la renta normal. En cambio, a corto plazo, la restricción de ortogonalidad de Hall se cumple solo de forma aproximada, pues el consumo responde a la renta anticipada (exceso de sensibilidad) y poco a las sorpresas (exceso de suavidad).

Los desarrollos anteriores explican ambas anomalías, y la evidencia microeconómica reciente sobre transferencias fiscales confirma propensiones marginales al consumo elevadas y muy heterogéneas entre hogares.

**Implicaciones de política**: en términos generales, la investigación que vino después se ha dedicado a modelizar fricciones que impiden al hogar trazar un plan de consumo plenamente eficiente en relación con su renta vital. Cuanto más fuertes sean esas fricciones, más dependerá el consumo de la renta corriente y más eficaces resultarán, en consecuencia, las medidas de política económica, tanto si se anticipan como si no.

Las implicaciones difieren según el instrumento:

- *Política monetaria*: actúa sobre el consumo a través del tipo de interés real, que inclina la senda de consumo (ecuación de Euler, apartado 2.1.2), y del efecto riqueza sobre el valor de los activos (apartado 2.3.2). Los hogares con restricciones de liquidez no responden por la primera vía, sino a través de la renta corriente.

- *Política fiscal*: con hogares previsores, una rebaja impositiva transitoria apenas altera el consumo (HRP), una anunciada no lo altera cuando se hace efectiva (HPA) y una financiada con deuda tampoco lo altera si se cumple la equivalencia ricardiana. Con restricciones de liquidez o reglas simples, en cambio, el consumo responde a la renta corriente, y las transferencias y rebajas temporales, sobre todo las dirigidas a los hogares de menor renta, recuperan eficacia.

## 3.6. Conclusiones

Los desarrollos recientes relajan uno a uno los supuestos de las teorías de la renta normal (mercados perfectos, racionalidad ilimitada, utilidad separable y cuadrática) para explicar el exceso de sensibilidad y de suavidad observados. El resultado común es que una fracción de los hogares consume su renta corriente, por restricciones de liquidez o prudencia, lo que devuelve a la

política fiscal parte de la eficacia que las teorías intertemporales le negaban. Ese resultado fundamenta los modelos neokeynesianos con dos tipos de agentes (TANK) y con agentes heterogéneos (HANK) de la macroeconomía actual.

## Conclusiones

El tema busca una función de consumo compatible con la evidencia de corto y de largo plazo. Keynes liga el consumo a la renta corriente, con una propensión media decreciente que las series largas de Kuznets desmienten, y Duesenberry reconcilia ambos horizontes con la renta relativa y los hábitos, pero sin mirar a la renta futura. Las teorías de la renta normal, en cambio, toman de Fisher el problema intertemporal y resuelven el enigma, pues en el ciclo vital y en la renta permanente el consumo depende de la riqueza y se suaviza. Con expectativas racionales, Hall concluye que solo las sorpresas de renta lo alteran. Sin embargo, la evidencia rechaza esa predicción en dos direcciones, el exceso de sensibilidad y el de suavidad, que los desarrollos recientes explican con restricciones de liquidez, prudencia, hábitos y bienes duraderos. Cuanto más depende el consumo de la renta corriente, más eficaz es la política fiscal.

La teoría del consumo descansa en el problema intertemporal del hogar (tema A29), del que también se deriva la equivalencia ricardiana, según la cual financiar el gasto público con deuda no altera el consumo (tema A06). La propensión marginal a consumir es, a su vez, el puente con la política fiscal, porque fija el multiplicador keynesiano (tema A04) y condiciona la eficacia del estímulo (tema A40), sobre todo cuando una parte de los hogares no puede endeudarse (tema A07). A largo plazo, en cambio, el ahorro determina el nivel de la renta por trabajador del estado estacionario y no su tasa de crecimiento (tema A43).

La evidencia reciente apunta a una propensión marginal a consumir elevada. Parker, Souleles, Johnson y McClelland (2013) estiman una respuesta fuerte del gasto a los pagos de estímulo de 2008 en EE. UU., y Kaplan y Violante (2014) la explican con hogares que tienen riqueza ilíquida (vivienda, pensiones) pero poca liquidez y consumen su renta corriente. Los modelos HANK (Kaplan, Moll y Violante, 2018) trasladan esa heterogeneidad a la política monetaria, que actúa sobre el consumo sobre todo a través de la renta. Sigue abierto también el debate sobre el estancamiento secular. Summers (2014) atribuye la caída del tipo de interés natural a un exceso de ahorro sobre la inversión deseada. Mian, Straub y Sufi (2020) sitúan parte de ese exceso en el ahorro creciente del 1 % más rico, que ha financiado deuda de los demás hogares y del sector público más que inversión.

## Anexo. Análisis gráfico de la hipótesis de la renta permanente

Gráficamente, en el plano renta-consumo (figura A33.5), se puede ilustrar por qué los hogares de renta baja tienen mayor propensión media al consumo y por qué en ciclos recesivos aumenta la propensión media al consumo:

1. *Función de consumo a largo plazo*, $C^{\text{LP}} = k_P Y^P$: es una recta que parte del origen, con pendiente $k_P$, porque el consumo permanente es proporcional a la renta permanente.

2. *Función de consumo a corto plazo*, $C^{\text{CP}}$: estimada con datos de corte transversal por grupos de renta, tiene menor pendiente y ordenada en el origen positiva, porque cada grupo consume según su renta permanente media y no según la observada.

3. *Punto de corte*: en la intersección de ambas funciones se determina la renta permanente media para toda la distribución de renta considerada, $Y^P$.

- *Renta alta*: la renta permanente media de los estratos de renta alta es $Y_A^P$ y la renta observada, $Y_t^A$. Dado que $Y_t^A > Y_A^P$, la renta transitoria media es positiva.

- *Renta baja*: la renta permanente media de los estratos de renta baja es $Y_B^P$ y la renta observada, $Y_t^B$. Dado que $Y_t^B < Y_B^P$, la renta transitoria media es negativa.

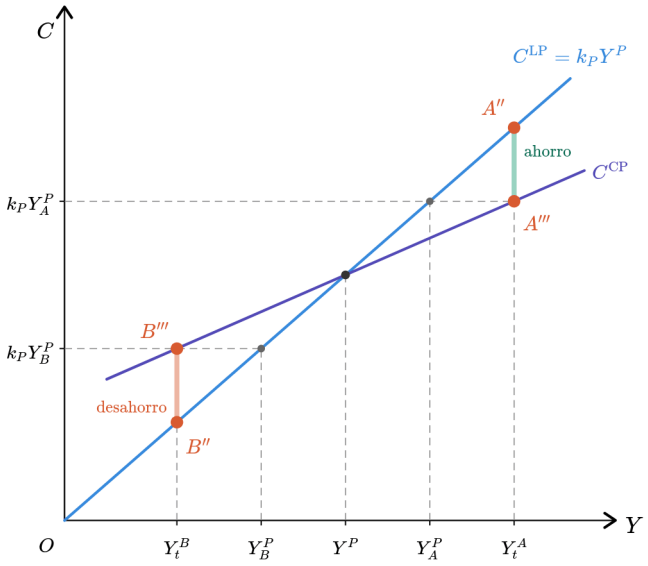



Figura A33.5: Hipótesis de la renta permanente. La función de consumo a largo plazo, $C = k_P Y^P$, parte del origen, mientras que la de corto plazo, obtenida por grupos de renta, es más plana porque los hogares consumen según su renta permanente. El grupo de renta baja, con renta transitoria negativa, consume por encima de lo que indica su renta observada (distancia $B''B'''$).

A partir de aquí, como $\text{corr}(C^T, Y^T) = 0$ (consumo y renta transitorios no guardan correlación), el consumo transitorio medio será nulo en cualquier grupo de renta que se considere. Esto significa que a ambos lados de la distribución de renta, los consumidores demandarán a corto plazo un nivel de consumo acorde con la renta permanente media del grupo ($C^P = k_P Y^P$) para mantener

su consumo permanente. El grupo B tendrá, por tanto, una propensión media a consumir a corto plazo superior a su propensión media a consumir a largo plazo (dada por $k_P$). El exceso de consumo vendrá dado por la distancia vertical $B''B'''$, entre la función a largo plazo ($B''$) y la de corto plazo ($B'''$) a la renta observada $Y_t^B$. Ese exceso se financia con desahorro, de modo que la caída transitoria de la renta no altera el consumo.

# A34. Teorías de demanda de inversión

Teorías de la inversión en bienes de equipo. Incertidumbre e irreversibilidad de la inversión. Implicaciones de política económica.

## Introducción

Las **teorías de la demanda agregada** estudian los comportamientos que hay detrás de los grandes bloques de los modelos macroeconómicos. La inversión de las empresas forma, con el consumo, el gasto público y el saldo exterior, la demanda de bienes y servicios que recoge la curva IS, y de su sensibilidad al tipo de interés depende buena parte de la eficacia relativa de las políticas monetaria y fiscal.

El punto de partida es la *Teoría general* de Keynes (1936), que hizo depender la inversión de la comparación entre la **eficiencia marginal del capital** y el tipo de interés. Sin embargo, era una relación entre agregados, sin una derivación explícita a partir de decisiones individuales. Jorgenson (1963) derivó la inversión de la maximización intertemporal del beneficio, y Tobin (1969) la ligó a la valoración bursátil del capital con su *q*. Desde los años setenta, la crítica de Lucas (1976) y las expectativas racionales llevaron el análisis a empresas que optimizan en un horizonte intertemporal con incertidumbre, y los costes de ajuste y la irreversibilidad renovaron la teoría de la inversión.

El concepto de inversión es uno de los conceptos económicos más difíciles de delimitar, si bien está estrechamente relacionado con el concepto de capital. La **inversión**, variable flujo, mide cómo cambia con el tiempo una variable *stock*, el capital productivo. Invertir es, para una unidad productiva, acercar el capital del que dispone al que desearía tener. En macroeconomía, a diferencia de las finanzas, el término designa la creación de capital, no la compra de activos financieros[1]. El tipo de capital define el tipo de inversión, y este tema considera la inversión económica, entendida como la afectación de bienes económicos a fines productivos.

Para la macroeconomía, la demanda de inversión importa por tres razones. Ningún otro componente de la demanda oscila tanto, así que explica buena parte de las fluctuaciones a corto plazo. Mirando más lejos, la parte del producto que se invierte nace del encuentro entre lo que las empresas quieren invertir y lo que los hogares deciden ahorrar, y de ambas cosas depende la capacidad productiva a largo plazo. Además, en las recesiones y en las fases de actividad débil, la política económica suele orientar sus medidas a tratar de reanimar la inversión.

El tema estudia, en orden cronológico, las teorías que identifican los determinantes de la inversión y de las que se deducen distintas implicaciones de política económica. Se estructura como sigue:

- *Fundamentos*: aportaciones de Fisher, Keynes, teorías del acelerador y Hayek.

[1] La inversión bruta incluye la inversión neta (variación del *stock* de capital) y la inversión de reposición (que compensa la depreciación): $I^b = I^n + \delta K$, donde $I^n = \dot{K}$.

- *Teorías modernas*: modelo de Jorgenson sin costes de ajuste, concepto de costes de ajuste, $q$ de Tobin, modelo con costes de ajuste.

- *Incertidumbre, irreversibilidad y financiación*: inversión como opción real, decisiones de financiación e implicaciones de política económica.

Conviene adelantar ya la principal enseñanza del tema para la política económica. Las medidas que mueven el tipo de interés o los tipos impositivos pueden influir en la inversión, pero lo que más pesa en ella es un entorno macroeconómico estable y previsible.

# 1. Fundamentos de la teoría moderna sobre la demanda de inversión

La **inversión real** es el gasto real destinado a la compra de bienes de capital y, por tanto, es un gasto de naturaleza distinta al alquiler de capital en los mercados de factores. Los servicios de los factores se alquilan y deben suministrarse de nuevo en cada momento del proceso productivo. Los bienes de capital adquiridos, en cambio, generan un flujo de servicios productivos a lo largo del tiempo.

Para entender las teorías dinámicas actuales de la inversión conviene repasar brevemente las aportaciones que dieron lugar a tres conceptos importantes de la teoría moderna. En primer lugar, las aportaciones de Fisher y Keynes, que ponen énfasis en la conducta de los agentes. Como no estudian el proceso de ajuste del *stock* de capital en sí mismo, constituyen más bien teorías de la demanda de capital que de la inversión. En segundo lugar, el principio del acelerador, que enfatiza el papel de la demanda agregada en la determinación de la inversión. En tercer lugar, la inversión como proceso de ajuste según Hayek, que enfatiza el papel de la inversión como fenómeno de desequilibrio.

## 1.1. Aportación neoclásica: la tasa de retorno sobre el coste (Fisher, 1930)

**Origen**: la aportación de Irving Fisher (1930) fue uno de los primeros intentos de microfundamentar intertemporalmente las ideas clásicas sobre la inversión.

**Idea**: Fisher considera que cuando un agente invierte, compra un derecho sobre un flujo de retornos durante la vida útil del bien de capital. Así, desarrolló su "segunda aproximación", un modelo intertemporal de decisión de inversión. Su principal resultado es que lo que llamó *tasa de retorno sobre el coste* (tasa interna de rentabilidad, en términos actuales) debe ser igual al tipo de interés real de mercado, con independencia de las preferencias del inversor.

### 1.1.1. Supuestos

**Entorno**: el modelo tiene dos periodos y adopta los supuestos habituales de la escuela neoclásica. Los mercados financieros son perfectos, de modo que el agente presta o se endeuda al mismo tipo de interés real, $r$, y el ahorro se canaliza eficientemente hacia la inversión.

**Agentes**: un consumidor-inversor racional que dispone de rentas $y_1$ e $y_2$ y maximiza la utilidad de su consumo intertemporal, como en el tema A33, apartado 2.1.2.

**Tecnología**: una oportunidad de producción con rendimientos decrecientes, en la que el capital $k$ invertido en el periodo 1 rinde $f(k)$ en el periodo 2, con $f' > 0$ y $f'' < 0$, y se deprecia a la tasa $\delta$.

### 1.1.2. Modelo

**Planteamiento**: la riqueza del agente depende de cuánto invierta, $W(k) = y_1 - k + [y_2 + f(k) + (1-\delta)k]/(1+r)$, y su programa es $\max u(c_1, c_2)$ sujeto a $c_1 + c_2/(1+r) = W(k)$. Como $k$ solo entra a través de $W$, el problema se separa en dos etapas (*teorema de separación de Fisher*). Primero, la inversión maximiza la riqueza y, después, el consumo se elige sobre esa riqueza como en A33.

El modelo del consumidor-productor de Fisher, con el teorema de separación y su representación gráfica, se desarrolla en el tema A29, apartado 2.4.2.

**CPO**: de la primera etapa, $\partial W/\partial k = 0$:

$$\underbrace{\frac{df}{dk} - \delta}_{\text{Pmg}_k \text{ neta}} = r$$

Se invierte hasta que la tasa de retorno sobre el coste, el rendimiento marginal del capital neto de depreciación, iguala el tipo de interés real. La condición es independiente de las preferencias, de modo que la decisión de inversión es la misma para todos los agentes, sean cuales sean sus gustos, lo que justifica separar la teoría de la inversión de la del consumo.

**CSO**: $\partial^2 W/\partial k^2 = f''(k)/(1+r) < 0$, de modo que la CPO es suficiente.

**Solución**:

- *Función de demanda de inversión individual*: la CPO define el capital deseado, $k^* = k(r)$, decreciente en el tipo de interés. Una subida de $r$ exige una productividad marginal mayor y, por los rendimientos decrecientes, un *stock* de capital menor, que se alcanza con menos inversión.

- *Función de demanda de inversión agregada*: como todos los agentes siguen la misma regla, agregando se obtiene la función de demanda de inversión neoclásica, decreciente en el tipo de interés real:

$$I = I(r), \qquad I'(r) < 0$$

**Desarrollo gráfico**: en el plano fondos prestables (ahorro e inversión reales)-tipo de interés real, la función de inversión se integra en el mercado de fondos prestables (neo)clásico (figura A34.1):

1. *Demanda de fondos prestables*, $I(r)$: recoge la inversión y tiene pendiente negativa, porque, por los rendimientos decrecientes del capital, un tipo de interés más alto reduce el capital deseado. Con sector público, a la inversión se suma el déficit (desahorro o endeudamiento del sector público).

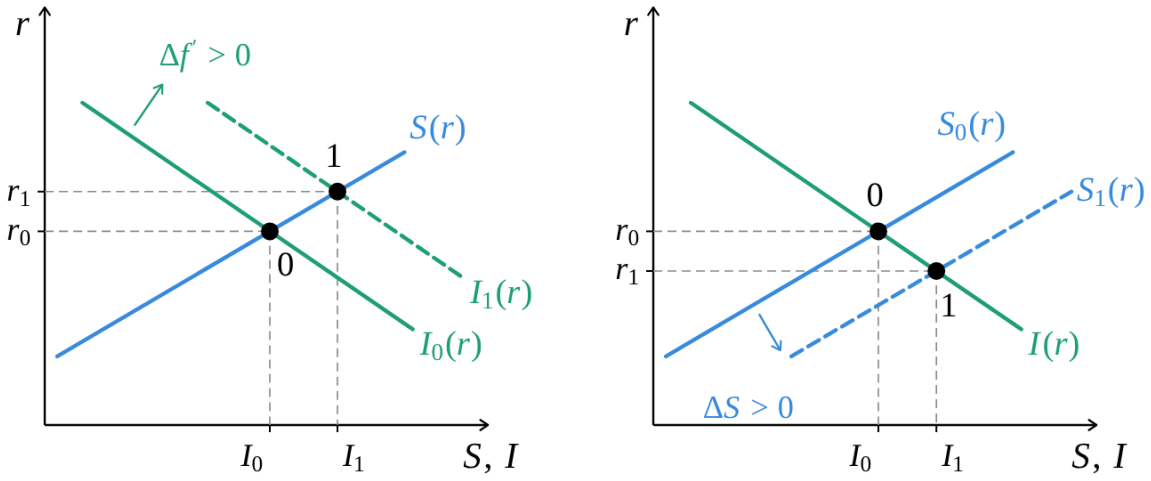
2. *Oferta de fondos prestables*, $S(r)$: recoge el ahorro, que se supone creciente en el tipo de interés real.

3. *Equilibrio*: en el punto 0, el tipo de interés real de equilibrio, $r_0$, iguala el ahorro y la inversión deseados, $I_0$.

**Estática comparativa**: se distinguen dos perturbaciones, una en cada panel de la figura:

- *Mejora tecnológica* ($f'$ mayor para cada $k$): desplaza la demanda de fondos prestables hacia la derecha, de $I_0(r)$ a $I_1(r)$ (panel a). En el paso del punto 0 al 1, suben la inversión y el tipo de interés.

- *Aumento del ahorro*: desplaza la oferta hacia la derecha, de $S_0(r)$ a $S_1(r)$ (panel b). En el paso del punto 0 al 1, el tipo de interés baja y la inversión sube.




(a) Mejora tecnológica

(b) Aumento del ahorro

Figura A34.1: Mercado de fondos prestables neoclásico. La oferta (ahorro), $S(r)$, y la demanda (inversión), $I(r)$, se cortan en el equilibrio 0. (a) Una mejora tecnológica ($\Delta f' > 0$) desplaza la demanda a $I_1(r)$ y eleva el tipo de interés y la inversión (punto 1). (b) Un aumento del ahorro ($\Delta S > 0$) desplaza la oferta a $S_1(r)$, reduce el tipo de interés y eleva la inversión (punto 1).

En ambos casos, la flexibilidad del tipo de interés real garantiza que la renta generada (oferta) sea igual al gasto de la economía (demanda), es decir, que se cumpla la ley de Say.

**Extensiones**: en el mercado de fondos prestables, las políticas de demanda operan a través del tipo de interés real.

- *Política monetaria expansiva*: produce un exceso de oferta de fondos prestables (entra en juego el dinero del banco central puesto en circulación), con la consecuente caída del tipo de interés y, por ello, mayor inversión. Sin embargo, como los precios son flexibles, el nivel general de precios sube hasta restablecer el equilibrio original, de modo que la política monetaria es neutral a largo plazo.

- *Política fiscal expansiva*: desplaza la demanda de fondos prestables hacia la derecha, como la mejora tecnológica del panel (a), y produce un exceso de demanda. En consecuencia, sube el tipo de interés real y se contraen el consumo y la inversión privada (el ahorro privado aumenta) en favor de la financiación del sector público, lo que supone un efecto expulsión (*crowding out*) total.

### 1.1.3. Conclusiones

**Aspectos positivos**: la aportación de Fisher tiene el mérito de microfundamentar intertemporalmente la inversión y de separarla, por el teorema de separación, de las preferencias del

inversor.

**Limitaciones**: Fisher solo representó gráficamente un horizonte de dos periodos (su formulación algebraica abarca un número cualquiera de años), y trabaja en un marco de pleno empleo donde el tipo de interés real iguala el ahorro y la inversión deseados. Keynes cuestiona de manera fundamental este último supuesto a raíz de la Gran Depresión americana.

## 1.2. Aportación keynesiana: la eficiencia marginal del capital (Keynes, 1936)

**Origen**: Keynes (1936) analizó la inversión como magnitud económica fundamental.

**Idea**: Keynes introduce dos ideas muy relevantes a la hora de entender la inversión:

- *Naturaleza de la inversión*: Keynes acepta las ideas de Fisher en cuanto a la tasa de retorno sobre el coste, si bien la bautiza como *eficiencia marginal del capital*. Sin embargo, señala que los retornos son esperados bajo una incertidumbre no reducible a probabilidades, de modo que dependen del estado de la confianza y del horizonte durante el que se esperan obtener. Por tanto, para Keynes lo decisivo no es la productividad física del capital, como lo era para muchos autores neoclásicos (p. ej. Marshall), sino las expectativas sobre sus rendimientos.

- *Causalidad*: para Keynes, la inversión es la variable que impulsa la economía. Movida por las expectativas de los empresarios, la inversión determina el ahorro y no al revés. Esto supone, en línea con el pensamiento de Keynes, la extensión de la "inversión de la ley de Say" al ámbito del ahorro y la inversión y no solo del consumo.

### 1.2.1. Supuestos

**Entorno**: hay desempleo e incertidumbre no reducible a probabilidades sobre los rendimientos futuros del capital. Por ello, la inversión efectiva, que forma parte del producto o renta nacional, puede diferir de la deseada o planeada, que forma parte de la demanda agregada. La brecha entre ambos tipos de inversión da lugar a la acumulación o desacumulación involuntaria de existencias.

**Agentes**: empresarios que comparan los rendimientos futuros con el tipo de interés de mercado y cuyas expectativas sobre esos rendimientos están dominadas por los instintos animales (*animal spirits*), volátiles e impredecibles.

**Tecnología**: rendimientos marginales decrecientes del capital, de modo que los proyectos pueden ordenarse por su rentabilidad esperada, que cae conforme aumenta la inversión.

### 1.2.2. Modelo

**Planteamiento**: no hay programa de optimización explícito. La decisión de inversión se realiza comparando la eficiencia marginal del capital (EMK) con el tipo de interés de mercado:

- *Eficiencia marginal del capital*: se trata del tipo de descuento al que la suma actualizada de los rendimientos previstos de un proyecto de inversión o bien de capital iguala su coste. Ese coste es el de reemplazo o precio de oferta, es decir, lo mínimo que cuesta reproducir el bien, distinto de su cotización en el mercado. Formalmente, $C_K = \sum_{j=1}^{T} Q_j^e/(1+\text{EMK})^j$, donde $C_K$ es el coste de reposición del bien de capital, $Q_j^e$ el rendimiento esperado en el año $j$ y $T$ su vida útil.

▪ *Tipo de interés*: el tipo de interés de mercado, que en Keynes es el monetario, refleja el coste de oportunidad de no realizar una inversión alternativa.

**Solución**:

▪ *Condición de demanda de inversión*: con base en lo anterior, Keynes deduce que debe cumplirse EMK $> r$. Es decir, se invierte si lo que se prevé obtener en el futuro con la inversión considerada es superior a lo que se deja de ganar en una inversión alternativa. En caso contrario (EMK $\leq r$), el proyecto no se acomete. En el agregado, la no reposición del capital depreciado y la desacumulación de existencias pueden hacer negativa la inversión neta.

▪ *Función de demanda de inversión individual*: cada proyecto $p$ tiene un coste $C_p$ y una $\text{EMK}_p$ dada por sus rendimientos esperados. La inversión de la empresa $i$ es la suma de los proyectos aceptados, $I_i = \sum_p I_p$, y la decisión sobre cada uno es discreta, $I_p = C_p > 0$ si $\text{EMK}_p > r$ e $I_p = 0$ si $\text{EMK}_p \leq r$.

La función resultante es escalonada y decreciente en el tipo de interés, pues al subir $r$ se descartan los proyectos con menor EMK, y las expectativas entran a través de los $Q_j^e$. Agregando sobre las empresas, $I = \sum_i I_i(r)$, la función sigue siendo escalonada. Con un continuo de proyectos, en cambio, los escalones se funden en la curva de EMK continua y decreciente del desarrollo gráfico, cuya linealización es la función agregada keynesiana.

▪ *Función de demanda de inversión agregada*: con base en todas estas ideas, Keynes la expresa, en términos reales, como:

$$I = I_0 + I(r), \qquad I'(r) < 0$$

donde $I_0$ es el componente exógeno, que refleja el peso de las expectativas o *animal spirits* y tiene un valor elevado. La pendiente $I'(r)$ mide la sensibilidad de la inversión al tipo de interés, que se considera baja. En suma, Keynes considera, generalmente, que en la función de inversión $I_0$ es grande e $I'(r) \approx 0$.

**Desarrollo gráfico**: la decisión de inversión se representa en el plano inversión-EMK y tipo de interés (figura A34.2):

1. *Curva de EMK*, EMK$(I)$: ordena los proyectos de inversión de mayor a menor EMK y, con un continuo de proyectos, tiene pendiente negativa. Según Keynes, la EMK cae conforme aumenta la inversión porque el rendimiento esperado disminuye a medida que aumenta la oferta de ese tipo de capital y la presión sobre las industrias que lo producen eleva su precio de oferta.

A ello se suma el riesgo creciente del endeudamiento (Kalecki, 1937). Los primeros proyectos pueden financiarse sin recurrir a él, pero, conforme aumenta su número, crecen el endeudamiento y el riesgo y, con ellos, cae la rentabilidad de los proyectos medida por la EMK.

2. *Tipo de interés*, $r$: el tipo de interés de mercado, $r_0$, no depende de la inversión, de modo que es una recta horizontal.

3. *Equilibrio*: por los rendimientos marginales decrecientes del capital, el óptimo se alcanza en el punto 0, donde la EMK iguala el tipo de interés de mercado y la inversión es $I(r_0)$. A su izquierda, EMK $> r$ y los proyectos se acometen, mientras que a su derecha, EMK $< r$ y ya no es rentable llevar a cabo más proyectos de inversión.

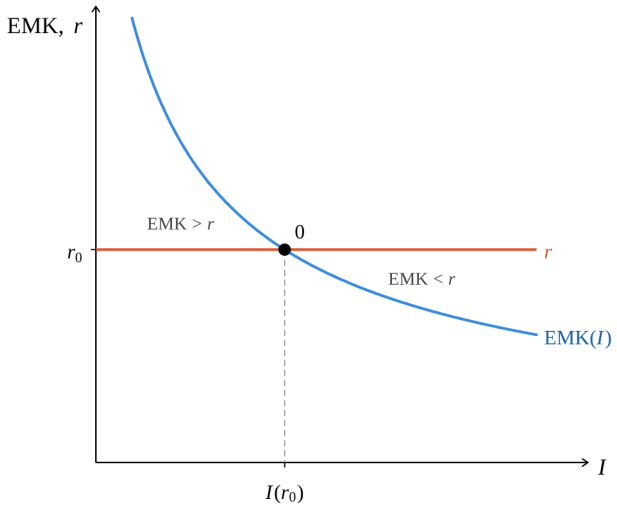



Figura A34.2: Eficiencia marginal del capital (Keynes, 1936). La inversión se realiza hasta el proyecto cuya EMK iguala al tipo de interés.

**Estática comparativa**: se distinguen dos movimientos, de los que domina el segundo:

- *Subida del tipo de interés*: reduce la inversión a lo largo de la curva de EMK, pero poco, dado $I'(r) \approx 0$.

- *Cambios en las expectativas ($I_0$), la tecnología o el capital*: desplazan la curva de EMK. En particular, una mejora de las expectativas o de la tecnología la desplaza hacia la derecha y eleva la inversión para cualquier tipo de interés.

**Extensiones**: del dominio de los instintos animales sobre las expectativas se deduce el carácter errático e inestable de la inversión y, con ello, de la demanda agregada.

La escasa sensibilidad de la inversión al tipo de interés, $I'(r) \approx 0$, constituye el *primer eslabón débil* de la demanda según Keynes. Junto con la trampa de liquidez (segundo eslabón débil), explica la ineficacia de la política monetaria en el modelo keynesiano. Junto con la ilusión monetaria (que causa rigidez salarial nominal a la baja), explica además que la economía no retorne por sí sola al pleno empleo. En consecuencia, la eficacia de las dos políticas de demanda es muy distinta.

- *Política monetaria*: es poco efectiva con la renta y el empleo, y menos que la política fiscal, porque solo actúa a través de una inversión poco sensible al tipo de interés.

- *Política fiscal*: eleva con fuerza la renta y el empleo y apenas reduce la inversión, que, gobernada por los *animal spirits*, responde poco al tipo de interés, y por ello existe un efecto expulsión parcial.

### 1.2.3. Conclusiones

**Aspectos positivos**: amplía las aportaciones de Fisher e introduce conceptos esenciales para la inversión como la importancia de las expectativas.

**Limitaciones**:

- *Demanda de capital*: explica cuánto capital quiere la empresa, no cuánto invierte, pues nada dice de la velocidad con que el capital se acerca a ese objetivo.

- *Estimación*: es difícil estimar en la práctica la EMK y, con ella, la función de inversión, ya que los instintos animales no son observables.

- *EMK decreciente*: la subida del precio de oferta de los bienes de capital, que para Keynes (1936) es la causa principal del descenso de la EMK a corto plazo, pierde fuerza si hay desempleo y capacidad ociosa en las industrias que los producen.

## 1.3. Revitalización del principio del acelerador

Las teorías del acelerador, formuladas a comienzos del s. XX, contribuyeron al avance de las teorías keynesianas del multiplicador y del ciclo económico.

Tras la *Teoría general* volvieron al primer plano, porque la parte endógena de la función de inversión keynesiana, en su formulación original, apenas tenía poder explicativo.

Las versiones de esa etapa se apoyaban más en un ajuste aceptable a los datos que en la microeconomía, y buscaban explicar a qué velocidad se acerca el capital a su nivel óptimo. Su hipótesis común liga la inversión neta a cómo evoluciona la demanda agregada. Cuando el gasto crece y supera lo producido, a las empresas se les abren dos caminos, subir precios o producir más.

El **principio del acelerador** de la inversión supone que las empresas eligen la segunda vía y producen más. Más demanda agregada exige más producción, producir más requiere más equipos, y así crecen la inversión y el capital.

### 1.3.1. Teoría del acelerador simple o rígido (J. M. Clark, 1917)

**Origen**: John Maurice Clark (1917) formuló el principio, que ya contaba con algunos precedentes. Harrod (1936) y Samuelson (1939) lo incorporaron después al análisis keynesiano del ciclo.

**Idea**: la inversión neta no depende del nivel de la demanda sino de su variación, porque, como el capital deseado es proporcional al *output*, solo un aumento del *output* exige capital nuevo.

**Supuestos**:

- *Entorno*: tiempo discreto y mercado de bienes de capital perfecto, sin retrasos en la adaptación del capital. Gracias a ello, la empresa alcanza en cada periodo el *stock* que desea, de modo que el capital efectivo y el óptimo o deseado coinciden siempre, $K_t = K_t^*$.

- *Agentes*: son miopes. Sus expectativas sobre las variables coinciden con sus niveles corrientes (p. ej. $Y_t^e = Y_t$).

- *Tecnología*: con coeficientes fijos, cada unidad de *output* real requiere siempre la misma cantidad de capital, y el capital deseado resulta $K_t^* = vY_t$, con $v = K/Y$ la ratio capital-output deseada. No se considera la depreciación, por lo que la inversión neta coincide con la bruta, $I_t^n = I_t^b$.

**Planteamiento**: no hay programa de optimización como tal, sino que el capital deseado es una proporción técnica del *output*, $K_t^* = vY_t$, y se alcanza en cada periodo.

**Solución**: de lo anterior se deduce la función de demanda de inversión, en la que la inversión depende de cambios en la demanda:

$$I_t = v \cdot \Delta Y_t$$

donde $I_t$ es la inversión neta y $\Delta Y_t$ la variación de la producción.

**Estática comparativa**: se distinguen tres perturbaciones:

- *Aumento permanente del nivel de* $Y$: si se produce un *shock* de $Y$ a partir del periodo $t$, la inversión es positiva solo en el periodo del cambio, $I_t = v\,\Delta Y_t$, y vuelve a cero después, pues el capital ya se ha ajustado al nuevo nivel.

- *Aumento transitorio de* $Y$: si se produce un *shock* de $Y$ exclusivamente en el periodo $t$, la inversión es positiva cuando $Y$ sube y negativa (desinversión) cuando vuelve a su nivel.

- *Crecimiento sostenido de* $Y$: si $Y$ crece en una cuantía constante durante $[t, t+j]$, es decir, con $\Delta Y$ constante, la inversión es positiva y constante. Solo cambia si $Y$ se acelera o desacelera, $\Delta I_t = v\,\Delta^2 Y_t$.

De aquí procede el nombre de "acelerador". Así, no es necesario que la producción caiga para que caiga la inversión, sino que basta con que crezca más despacio, y el efecto es mayor cuanto mayor es $v$.

**Extensiones**: al incluir el acelerador en el modelo macroeconómico, multiplicador y acelerador interactúan (Samuelson, 1939). Con consumo retardado, $C_t = c(1-\tau)\,Y_{t-1}$, e inversión inducida por la variación del consumo, $I_t = v\,(C_t - C_{t-1})$, la renta sigue la ecuación en diferencias

$$Y_t = c(1-\tau)(1+v)\,Y_{t-1} - c(1-\tau)\,v\,Y_{t-2} + G_0$$

donde $c$ es la propensión marginal a consumir, $\tau$ el tipo impositivo, $v$ el acelerador y $G_0$ el gasto público autónomo. Una vez estabilizada la renta, la inversión inducida se anula y el efecto a largo plazo de un aumento permanente de $G_0$ es el multiplicador keynesiano, $1/[1-c(1-\tau)]$. El acelerador amplía, en cambio, la respuesta transitoria, que puede sobrepasar ese nivel y oscilar, y la senda solo converge si $c(1-\tau)\,v < 1$.

Solo si, por simplificación, la inversión inducida depende del nivel de renta, $I = bY$, con $b$ la propensión marginal a invertir (*supermultiplicador*), el multiplicador es permanentemente mayor, $1/[1-c(1-\tau)-b]$, siempre que $c(1-\tau)+b < 1$. De ello se siguen las implicaciones de política económica del acelerador simple.

- *Política monetaria*: pierde todavía más eficacia que en Keynes para influir en la inversión, porque el tipo de interés ni siquiera entra en la función de inversión, que solo responde a la producción.

- *Política fiscal*: a corto plazo mueve la renta y el empleo, y con ellos la inversión, con más fuerza aún que en Keynes. Hay dos razones. Primero, el acelerador amplía la respuesta transitoria de la renta al gasto público (y la hace permanente en la versión del supermultiplicador). Segundo, de acuerdo con la teoría del acelerador, el efecto expulsión sobre la inversión privada es nulo.

**Conclusiones**:

▪ *Aspectos positivos*: el modelo del acelerador se ajusta empíricamente bastante bien para explicar la inversión en existencias.

▪ *Limitaciones*: las principales críticas apuntan a que los supuestos son muy restrictivos:

- *Mercado de bienes de capital*: da por hecho que la empresa consigue todo el equipo que quiere sin alterar su precio. Puede valer para una empresa pequeña, pero no en el agregado, donde una mayor demanda de equipo tiende a encarecerlo.
- *Ritmo de ajuste*: sin retardos y con el *stock* efectivo siempre igual al deseado, no hay un ritmo de ajuste propiamente dicho, porque ante cualquier cambio en la demanda hay que invertir, y de forma inmediata.
- *Agentes*: no considera el efecto de las expectativas, dado que los empresarios son miopes.
- *Tecnología*: exige coeficientes fijos. Sin embargo, si el trabajo puede sustituir al capital, las empresas pueden responder a las oscilaciones de la demanda cambiando el uso que hacen de sus factores, sin tener que invertir.

### 1.3.2. Teoría del acelerador flexible (Goodwin, 1951; Chenery, 1952; Koyck, 1954)

**Origen**: Richard Goodwin (1951), Hollis Chenery (1952) y Leendert Koyck (1954) generalizaron el acelerador simple a raíz de sus limitaciones.

**Idea**: el capital se ajusta a su nivel deseado con retardo, de modo que la inversión responde a las variaciones pasadas de la demanda y no solo a la corriente.

**Supuestos**: se mantienen los del acelerador simple, con una excepción. Ahora existen retardos, de forma que las empresas no pueden llevar de golpe su capital hasta el nivel óptimo.

**Planteamiento**: tampoco hay programa de optimización, pues el retardo se introduce de forma *ad hoc* en la regla de ajuste.

**Solución**: en la función de demanda de inversión, la inversión responde a la variación de la renta del periodo corriente y de todos los precedentes, con retardos de peso decreciente:

$$I_t = v\,(1-\gamma)\sum_{j=0}^{\infty}\gamma^j \Delta Y_{t-j} \quad \Leftrightarrow \quad K_t - K_{t-1} = (1-\gamma)\,(K_t^* - K_{t-1})$$

donde $\gamma \in (0,1)$ es el parámetro de retardo (en cada periodo desaparece la fracción $1-\gamma$ de la distancia que separa el capital deseado del existente). Las variaciones de la producción pesan menos cuanto más antiguas son, y por eso el capital se aproxima al óptimo poco a poco.

**Desarrollo gráfico**: en el plano tiempo-capital (figura A34.3), la diferencia entre ambas teorías se aprecia en la senda del *stock* de capital cuando en $t_0$ el capital deseado pasa de $K_0$ a $K^*$:

1. *Capital inicial y deseado*: se parte de un *stock* de capital previo, $K_0$, y de un *stock* de capital óptimo o deseado, $K^*$, que lo supera desde $t_0$.

2. *Senda del acelerador rígido*, $K_t$: el *stock* de capital salta en $t_0$ hasta $K^*$, porque se ajusta por completo en el periodo del cambio.

3. *Senda del acelerador flexible*, $K_t$: el *stock* de capital se ajusta de forma gradual y se aproxima a $K^*$ cerrando en cada periodo la fracción $1-\gamma$ de la brecha, lo cual resulta más realista.

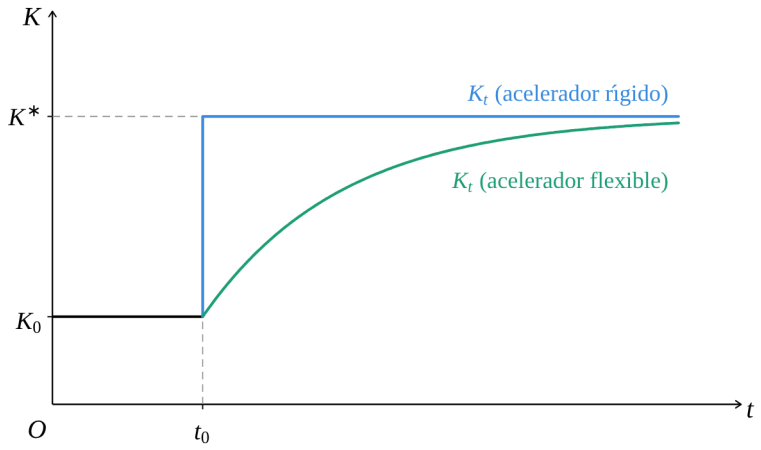



Figura A34.3: Acelerador flexible. El *stock* de capital se ajusta parcialmente hacia su nivel deseado.

**Estática comparativa**: un aumento permanente de la producción eleva el capital de forma gradual. La inversión alcanza su máximo en el periodo del cambio y decrece después a la tasa $1-\gamma$ por periodo. Por su parte, un aumento transitorio produce una inversión menor y también repartida en el tiempo. Cuanto mayor es $\gamma$, más lento es el ajuste.

**Extensiones**: las implicaciones de política son similares a las del acelerador rígido, atenuadas por el retardo.

▪ *Política monetaria*: al igual que en la teoría del acelerador simple, carece por completo de eficacia para influir en la inversión.

▪ *Política fiscal*: pierde algo de fuerza respecto al acelerador rígido, porque con retardos la inversión no responde de golpe, sino de forma escalonada.

**Conclusiones**:

▪ *Aspectos positivos*: en el terreno empírico, el acelerador flexible obtuvo buenos resultados. Las variaciones de la demanda y la inversión muestran una correlación positiva, aunque eso no prueba por sí solo una relación causal.

▪ *Limitaciones*: salvo en lo relativo a los retardos, recibe críticas parecidas a las del acelerador simple:

- *Tipo de interés*: la función de inversión lo ignora.
- *Costes de ajuste*: se introduce el retardo sin explicar qué costes lo originan.
- *Optimización*: ningún agente optimiza intertemporalmente, de ahí que, en rigor, sea una regla de ajuste más que una teoría de la inversión.

### 1.3.3. Conclusiones

**Aspectos positivos**: ambas teorías del acelerador son útiles para explicar la evidencia empírica (p. ej. inversión en existencias).

**Limitaciones**: *a priori*, no existen motivos para suponer que las empresas deseen mantener una ratio capital-output fija con independencia de los precios relativos de los factores. Además, como recuerda Haavelmo (1960), la demanda de inversión no puede deducirse sin más de la demanda de capital, porque un *stock* deseado no determina a qué ritmo se invierte. Para explicar bien la inversión en capital fijo hay que contestar tres preguntas, que las teorías del acelerador solo abordan parcialmente:

- *Stock deseado*: cuánto capital se quiere.
- *Ritmo de ajuste*: con qué rapidez se alcanza ese nivel.
- *Reposición*: cómo se repone en el tiempo el capital gastado.

## 1.4. Perspectiva austriaca: inversión como ajuste óptimo (Hayek, 1941)

**Origen**: en la escuela austriaca, la aportación de Friedrich von Hayek (1941) pone el énfasis en el proceso de ajuste del capital y es un antecedente de las teorías modernas con costes de ajuste.

**Idea**: Hayek consideraba la inversión un fenómeno de desequilibrio, y la decisión de inversión, una decisión sobre la velocidad óptima de ajuste hacia el equilibrio.

**Razonamiento**: una empresa puede decidir que necesita una fábrica (la decisión sobre el *stock* de capital). Sin embargo, a qué velocidad construirla y cuánto gastar mensualmente en ello (la decisión de inversión) es una decisión aparte, ligada a la velocidad de ajuste al *stock* deseado.

**Extensiones**: en materia de política económica, Hayek defendía que existía un "orden espontáneo" derivado de la interacción de agentes descentralizados, el cual no podía ser diseñado por un planificador social. Por tanto, *a priori*, la intervención del Estado debía ser mínima para dejar actuar las velocidades óptimas de ajuste de los empresarios.

**Conclusiones**:

- *Aspectos positivos*: la aportación de Hayek tiene el mérito de poner énfasis en el proceso de ajuste del capital, que posteriormente ha recibido gran atención a través de los costes de ajuste.
- *Limitaciones*: carece de formalismo matemático, como es propio de la escuela, lo que explica que los desarrollos posteriores eclipsaran su aportación.

## 1.5. Conclusiones

Las aportaciones de la sección 1 fijan los determinantes que retoman las teorías modernas. Fisher iguala la tasa de retorno sobre el coste al tipo de interés real y separa así la decisión de invertir de las preferencias del inversor. Keynes conserva esa comparación, pero con rendimientos esperados e inciertos, y por ello la inversión depende más de los *animal spirits* que del tipo de interés. El acelerador traslada el peso a la variación de la producción y, en su versión flexible, introduce un ajuste gradual con un retardo *ad hoc*, sin optimización ni tipo de interés. Hayek separa expresamente la decisión sobre el *stock* de la decisión sobre la velocidad de ajuste, aunque sin formalizarla.

Ninguna de estas aportaciones une una decisión optimizadora y un ritmo de ajuste formalizado, de modo que ninguna responde a la vez a las tres preguntas de Haavelmo (1960). Esa carencia llevó a los desarrollos modernos de la sección 2, que derivan el *stock* deseado del coste de uso y el ritmo de inversión de los costes de ajuste.

# 2. Teorías modernas de la inversión

Los modelos desarrollados a partir de los años sesenta plantean la decisión de inversión como un problema de optimización dinámica, la maximización del valor descontado de los beneficios presentes y futuros de la empresa.

## 2.1. Modelo de demanda de inversión sin costes de ajuste (Jorgenson, 1963; 1967)

**Origen**: el modelo dinámico neoclásico básico se debe a Dale Jorgenson (1963; con Hall, 1967).

**Idea**: la empresa desea el *stock* de capital que iguala la productividad marginal del capital a su coste de uso, y la inversión es el ajuste hacia ese *stock*. De este modo, los precios relativos de los factores, y no solo la demanda, determinan la inversión.

### 2.1.1. Supuestos

**Entorno**: mundo neoclásico, horizonte infinito, tiempo continuo y precios del producto, del trabajo y del capital constantes ($\dot{P} = \dot{w} = \dot{P}_K = 0$), aunque la inflación es relevante para el coste de uso. Desde el punto de vista de la inversión, el mundo neoclásico tiene cuatro implicaciones relevantes:

- *Tipo de interés*: el tipo de interés real es constante y exógeno, sin prima de riesgo, por la perfección de los mercados financieros.

- *Oferta de capital*: la oferta de bienes de capital es perfectamente elástica, de modo que la empresa adquiere todo el capital que desea al precio de mercado.

- *Costes de ajuste*: no existen de ningún tipo, lo que será la principal limitación del modelo.

- *Ausencia de distorsiones*: no existen impuestos distorsionantes (lo que omite los escudos fiscales).

**Tecnología**: una función de producción neoclásica de buen comportamiento, $f(K, L)$, con productividades marginales positivas y decrecientes ($f'_K > 0$, $f''_K < 0$) y que cumple las condiciones de Inada. Como el propio Jorgenson, se toma una Cobb-Douglas, $Y = AK^{\alpha}L^{1-\alpha}$, y el nivel de producción como dado, porque con rendimientos constantes a escala y precio dado, las condiciones de óptimo solo fijarían la ratio $K/L$, y no la escala de la empresa. El capital se deprecia a una tasa constante y exógena $\delta$.

### 2.1.2. Modelo

**Planteamiento**: la empresa maximiza el valor descontado de su corriente de beneficios eligiendo la inversión bruta y el empleo, $I$ y $L$, sujeta a la ley de acumulación del capital, con el capital inicial dado:

$$\max_{\{I,L\}} V = \int_0^{\infty} e^{-rt} \left[ Pf(K, L) - wL - P_K I \right] dt$$

$$\text{s.a.:} \begin{cases} \dot{K} = I - \delta K \\ K_0 \text{ dado} \end{cases}$$

donde $V$ es el valor de la empresa[2], $P$ el precio del producto, $w$ el salario nominal, $P_K$ el precio de los bienes de capital, $r$ el tipo de interés real, $\delta$ la tasa de depreciación y $K_0$ el capital inicial.

▪ *Variables*: $I$ y $L$ son las variables de control, sobre las que la empresa decide libremente en cada instante, mientras que $K$ es la variable de estado, que solo cambia a través de la inversión.

▪ *Función objetivo*: el beneficio de cada instante, ingresos menos salarios menos gasto en bienes de capital, descontado al tipo de interés real de mercado, $e^{-rt}$. La restricción tecnológica, $f(K, L)$, está incorporada en el ingreso.

▪ *Ley de acumulación*: la inversión neta, $\dot{K}$, es la inversión bruta, $I$, menos la depreciación, $\delta K$.

▪ *Condición inicial*: el *stock* de capital de partida, $K_0$, está dado, porque la empresa hereda el capital instalado.

**CPO**: el hamiltoniano en valor presente u origen es

$$H = e^{-rt}\left[Pf(K, L) - wL - P_K I\right] + \lambda\left(I - \delta K\right)$$

donde $\lambda$ es la variable de coestado o multiplicador dinámico, es decir, el *precio sombra del capital* (lo que la empresa pagaría en el momento 0 por una unidad más de capital en el momento $t$).

Resolviendo por el principio del máximo de Pontryagin, se obtiene que la senda óptima cumple las siguientes condiciones necesarias:

1. *Óptimo respecto al empleo* ($\partial H/\partial L = 0$): el valor de la productividad marginal del trabajo iguala al salario nominal, $P\partial f/\partial L = w$.

2. *Óptimo respecto a la inversión* ($\partial H/\partial I = 0$): el precio sombra del capital iguala a su precio de mercado descontado, $\lambda = e^{-rt}P_K$.

3. *Ecuación de la variable de coestado* ($\partial H/\partial K = -\dot{\lambda}$): $e^{-rt}P\partial f/\partial K - \delta\lambda = -\dot{\lambda}$.

4. *Condición de transversalidad*[3]: $\lim_{t\to\infty} \lambda_t K_t = 0$, es decir, el valor descontado del capital que la empresa conserve en el límite es nulo, lo que excluye acumular capital indefinidamente o sendas explosivas del valor de la empresa.

Derivando la segunda condición respecto al tiempo, con $P_K$ constante, $\dot{\lambda} = -re^{-rt}P_K$, y sustituyendo en la tercera condición se obtiene la condición central del modelo:

$$\underbrace{P\frac{\partial f}{\partial K}}_{\text{VPmg}_K} = P_K(r + \delta) \equiv r_K$$

En el óptimo, el *coste de uso del capital* ha de coincidir con el valor de la productividad marginal de ese factor. Esta expresión tiene una gran importancia, ya que, a diferencia de los modelos

[2]En tiempo continuo se omite el argumento temporal: $K \equiv K(t)$, $I \equiv I(t)$ y $L \equiv L(t)$; se escribe solo en las condiciones inicial y terminal.

[3]La condición de transversalidad es la holgura complementaria de la restricción de no negatividad del capital llevada al límite (tema A43, apartado 3.2).

estáticos de alquiler de capital, permite desglosar el coste del capital en su precio de adquisición ($P_K$), su coste de oportunidad ($r$) y su depreciación ($\delta$).

Con la Cobb-Douglas, $\partial f/\partial K = \alpha Y/K$, de modo que el capital óptimo es $K^* = \alpha PY/r_K$, que crece con el valor de la producción y decrece con el coste de uso. Junto con la ley de acumulación, $\dot{K} = I - \delta K$, estas condiciones caracterizan el comportamiento de la empresa.

**CSO**: el hamiltoniano es cóncavo en $(K, L, I)$, pues $f$ es cóncava y el resto de términos son lineales, de modo que las condiciones de Pontryagin, junto con la de transversalidad, son también suficientes (Mangasarian). Sin costes de ajuste, además, el problema se reduce al estático de cada instante.

**Solución**: conocido el capital óptimo, $K^*$, basta contrastarlo con el que ya está instalado, $K$:

- $K^* = K$: la empresa solo repone el desgaste y la inversión neta es cero.
- $K^* < K$: deja que el capital se deprecie sin reponerlo del todo, o vende parte, y la inversión neta resulta negativa.
- $K^* > K$: amplía el capital y la inversión neta es positiva.

Si el ajuste se completa en un periodo, la inversión neta de ese periodo es la diferencia entre el capital óptimo y el instalado, y la inversión bruta añade la reposición del capital depreciado:

$$I = (K^* - K) + \delta K$$

donde $K^*$ es el *stock* de capital óptimo, que iguala $P\,\partial f/\partial K$ al coste de uso, y $K$ el instalado. Como los tres casos solo difieren en el signo de $K^* - K$, una sola expresión los recoge.

**Estática comparativa**: el capital óptimo, y con él la inversión neta, es decreciente en el coste de uso, es decir, en el tipo de interés real, en el precio de los bienes de capital y en la tasa de depreciación. En cambio, crece con el precio del producto y con la productividad del capital. El modelo dice cuánto cambia $K^*$, pero no a qué ritmo se alcanza.

**Extensiones**: el coste de uso puede enriquecerse en dos direcciones:

1. *Inflación*: si los precios del capital no son fijos, el coste de uso debe recoger la inflación de esos precios y pasa a ser $r_K = P_K(i + \delta - \pi_K)$, con $i = r + \pi^e$ el tipo de interés nominal, o, equivalentemente, $r_K = P_K[r + \delta - (\pi_K - \pi^e)]$. Por tanto, una revalorización esperada de los bienes de capital superior a la inflación general reduce el coste de uso del capital, porque la ganancia de capital esperada compensa parcialmente el coste de oportunidad.

2. *Imposición* (Hall y Jorgenson, 1967): para recoger, por ejemplo, el impuesto sobre sociedades o las desgravaciones fiscales, se puede incluir un coeficiente $\chi_t$ que dependa de parámetros fiscales, $r_K = \chi_t P_K(r + \delta)$.

Por ejemplo, aumentos en el impuesto sobre sociedades reducen el ingreso neto de impuestos por unidad vendida y, con ello, el valor después de impuestos del producto marginal del capital, mientras que aumentos en la desgravación fiscal reducen el coste de uso del capital.

Del coste de uso así ampliado se derivan las implicaciones de política económica del modelo:

- *Política monetaria*: es efectiva, dado que afecta al tipo de interés real y este afecta a la inversión vía coste del capital.
- *Política fiscal*: el efecto dependerá de cómo se articule.

- *Basada en el gasto público*: puede tener un efecto adverso sobre el tipo de interés real y, con ello, sobre el coste de capital.

- *Basada en desgravaciones fiscales a la inversión*: tendrá un efecto positivo a través de la reducción del coste de uso del capital.

#### 2.1.3. Conclusiones

**Aspectos positivos**: el modelo dio un paso en la microfundamentación de la inversión y, con ello, enriqueció el estudio de su demanda.

**Limitaciones**: no obstante, presenta tres:

▪ *Capital*: la CPO ofrece una predicción estática y no admite posibilidad de sustitución intertemporal del capital.

▪ *Tasa de inversión*: el modelo fija cuánto capital se desea, pero no a qué ritmo se invierte. Basta pensar en una rebaja del tipo de interés decidida por el banco central, que abarata el coste de uso y desplaza $K^*$. Para que el capital "salte" a ese nuevo nivel haría falta una tasa de inversión infinita, y la inversión dejaría de ser una función continua del tiempo. La teoría recurre, así, a un mecanismo *ad hoc* de ajuste, pues el salto instantáneo es plausible para una empresa individual, pero no en el agregado[4].

▪ *Expectativas*: la inversión solo depende de variables actuales, aunque la información sobre el futuro sea perfecta. Por ello, la teoría omite el papel de las expectativas. Ocurre así porque modificar el capital instalado no cuesta nada y la empresa puede hacerlo en cada periodo. Si, por el contrario, invertir o desinvertir de golpe resultase más caro que hacerlo gradualmente, las empresas atenderían también al valor futuro de las variables relevantes.

### 2.2. Fundamentos de los modelos con costes de ajuste

Incorporar las expectativas exige pasar a los **modelos de inversión con costes de ajuste**, para lo cual hay que entender antes la noción de coste de ajuste y el concepto de $q$ de Tobin.

#### 2.2.1. Concepto de costes de ajuste

> El concepto de coste de ajuste (Eisner y Strotz, 1963; Lucas, 1967), su tipología y sus formas funcionales se desarrollan en el tema A12, apartado 3.3.2, y aquí se retiene lo necesario para la demanda de inversión.

Los costes de ajuste son los costes en los que incurre la empresa al alterar su dotación de factores productivos, tanto mayores, en general, cuanto mayor es la velocidad de ese cambio.

[4] Jorgenson recurrió a una solución *ad hoc* análoga a la del acelerador flexible, un retardo distribuido, $I_t^n = \sum_{j=0}^{\infty} \omega_j (K^*_{t-j} - K^*_{t-j-1})$, donde $\omega_j$ son ponderaciones decrecientes, como alternativa a los costes de ajuste internos de Eisner y Strotz. La solución exige que exista un *stock* de capital óptimo, algo que la evidencia sobre la dimensión óptima de planta cuestiona. En modelos de crecimiento con población creciente, exige además una brecha creciente entre el capital óptimo y el actual para sostener un flujo constante de inversión por trabajador.

Pueden ser internos, pérdidas temporales de eficiencia dentro de la empresa (mermas de producción al reemplazar máquinas, formación de los nuevos trabajadores), o externos, derivados de la interacción con agentes ajenos a ella (consultores, indemnizaciones, costes de contratación).

Se modelan con una **función de costes de ajuste**, $\varphi(I)$, no negativa, creciente en el valor absoluto de la inversión y convexa, con $\varphi(0) = \varphi'(0) = 0$, $\varphi'(I) \gtrless 0$ según $I \gtrless 0$ y $\varphi'' > 0$, de modo que invertir o desinvertir deprisa cuesta más que hacerlo despacio.

Con esta función se obtiene lo que faltaba en Jorgenson, pues la condición de contratación del capital incorpora, además del coste de uso, el coste marginal de ajuste. Por ello, la empresa no salta al capital deseado, sino que elige un ritmo óptimo de ajuste, y la inversión se suaviza en el tiempo.

### 2.2.2. La teoría de la *q* de Tobin (Brainard y Tobin, 1968; Tobin, 1969)

**Origen**: los modelos con costes de ajuste de los años setenta y ochenta se construyen sobre la teoría de la $q$ de James Tobin (con Brainard, 1968; 1969).

**Idea**: la teoría de la $q$ de Tobin (1969) señala que la inversión óptima es una función de la ratio $q$, entendida como el precio sombra relativo del capital. Formalmente, la inversión (real), individual o agregada, depende de la ratio $q$, que mide cuánto vale el capital en el mercado en relación con lo que costaría reponerlo:

$$I = f(q)$$

**Razonamiento**: la $q$ recoge en un solo número las perspectivas futuras de la empresa. Se distinguen dos versiones y tres interpretaciones.

- *q marginal*: lo que se pagaría por una unidad adicional de capital (el valor descontado de los beneficios extra que generará) dividido por lo que cuesta reponerla. Esta versión es la que gobierna la inversión, pero no se puede observar.

- *q media*: valor de mercado del capital instalado dividido por su coste de reposición. Es observable, y coincide con la marginal bajo las condiciones de Hayashi (1982) (rendimientos constantes a escala en la producción y en los costes de ajuste, competencia perfecta y capital homogéneo).

Las tres interpretaciones de la $q$ son:

- *Interpretación financiera* (media): lo que el mercado paga por la empresa, $V$, puesto en relación con lo que costaría reponer su capital instalado, $P_K K$. El valor de mercado es la capitalización bursátil más la deuda y, en un mercado eficiente, coincide con el valor actual de los flujos esperados de la empresa. Formalmente, $q_{\text{me}} = V/(P_K K)$, la definición original de Tobin.

- *Interpretación productiva* (marginal): con un rendimiento neto del capital constante, $\partial f/\partial K - \delta$, y descontando al tipo de interés real, $i - \pi^e$, con $i$ el tipo nominal y $\pi^e$ la inflación esperada, el valor de una unidad adicional de capital es una perpetuidad. Por ello, $q_{\text{mg}} = (\partial f/\partial K - \delta)/(i - \pi^e)$, es decir, la ratio entre la rentabilidad neta del capital y su coste de oportunidad. Si es superior a la unidad, el capital rinde más de lo que cuesta financiarlo, el *stock* es inferior al óptimo y conviene invertir.

- *Interpretación contable*: aproximación empírica de la $q$ media como cociente entre el valor de mercado de acciones y pasivos y su valor en libros. Como los valores en libros no recogen la inflación, infravaloran el coste de reposición y la $q$ contable sobrevalora la económica. Además, no debe confundirse con el *price-to-book ratio*, que solo considera las acciones.

La regla de inversión es la misma en cualquiera de las versiones:

- $q > 1$: el valor de mercado del capital supera su coste de reposición y conviene invertir.
- $q < 1$: conviene desinvertir.
- $q = 1$: el *stock* es el deseado y solo se repone la depreciación.

**Conclusiones**:

- *Aspectos positivos*: la principal ventaja de la $q$, concretamente de la $q$ media, es que, a diferencia de la EMK propuesta por Keynes, es una ratio observable. Permite, además, interpretar fácilmente el incentivo de la empresa para invertir.
- *Limitaciones*:
  - *Ausencia de capital humano*: el coste de reposición omite variables como el capital humano, que el valor de mercado sí recoge, lo que sesga la $q$ al alza, aunque el capital humano puede resultar también difícilmente medible.
  - *Volatilidad*: la $q$ de Tobin puede reflejar comportamientos de rebaño y burbujas especulativas que no respondan a los fundamentos.
  - *Microfundamentación*: en su formulación original carece de ella, y la aportan los modelos con costes de ajuste.

## 2.3. Modelos con costes de ajuste: modelos de inversión basados en la ratio $q$ (Summers, 1981; Abel, 1982; Hayashi, 1982)

**Origen**: Lawrence Summers (1981), Andrew Abel (1982) y Fumio Hayashi (1982) integraron los costes de ajuste en el problema dinámico de la empresa y demostraron que la $q$ marginal de Tobin es el precio sombra del capital.

**Idea**: con costes de ajuste convexos, la empresa no salta al capital deseado sino que elige un ritmo de inversión, que resulta ser una función creciente de la $q$ marginal.

### 2.3.1. Supuestos

**Entorno**: los del modelo de Jorgenson, salvo que existen costes de ajuste internos del capital y no hay depreciación. El precio del bien de capital se normaliza a la unidad.

**Tecnología**: el trabajo se ajusta libremente según las condiciones del mercado, de modo que el beneficio por unidad de capital, $\Pi(K)$, ya incorpora la decisión óptima de empleo. El beneficio por unidad de capital es decreciente en el capital agregado de la industria, $K$, porque más capital en la industria aumenta la oferta y reduce el precio del producto. Además, la función convexa de costes de ajuste, $\varphi(I)$, se incluye de forma aditiva en la función objetivo.

### 2.3.2. Modelo

**Planteamiento**: la empresa representativa elige su senda de inversión para maximizar el valor descontado de sus beneficios, netos del coste de la inversión y de su ajuste, sujeta a la ley de acumulación del capital, con el capital inicial dado:

$$\max_{\{I\}} V = \int_0^\infty e^{-rt} \left[\Pi(K)\, k - I - \varphi(I)\right] dt$$

$$\text{s.a.:} \begin{cases} \dot{k} = I \\ k_0 \text{ dado} \end{cases}$$

donde $V$ es el valor de la empresa, $k$ su *stock* de capital y $K$ el de la industria, que la empresa toma como dado. Además, $\Pi(K)$ es el beneficio por unidad de capital, $I$ la inversión, $\varphi(I)$ el coste de ajuste, $r$ el tipo de interés real y $q$ la variable de coestado.

Como en Jorgenson, $I$ es la variable de control y $k$ la de estado. Como el precio del bien de capital es la unidad, el gasto en inversión es $I$, y, sin depreciación, $\dot{k} = I$. El capital inicial, $k_0$, está dado.

Además del coste de ajuste, que es la diferencia esencial con el modelo de Jorgenson, cambia la formulación, ya que el hamiltoniano se construye en valor corriente. La variable de coestado se capitaliza con el factor de descuento, $q = e^{rt}\lambda$, y mide lo que la empresa pagaría en $t$ por una unidad más de capital en $t$. Este precio sombra del capital en valor corriente es la $q$ marginal de Tobin:

$$H = \Pi(K)\, k - I - \varphi(I) + q\, I$$

**CPO**: por el principio del máximo, la senda óptima cumple cuatro condiciones:

1. *Óptimo respecto a la inversión* ($\partial H/\partial I = 0$): $1 + d\varphi/dI = q$. Esta igualdad es la *condición de demanda de inversión*. Instalar una unidad más de capital cuesta su precio más el coste marginal de ajuste, y la empresa sigue invirtiendo mientras ese coste quede por debajo del valor sombra de esa unidad. Por eso, la rapidez del ajuste depende de lo convexos que sean esos costes.

2. *Ecuación de la variable de coestado* ($\partial H/\partial k = rq - \dot{q}$): $\Pi(K) = rq - \dot{q}$, o bien $\Pi(K) + \dot{q} = rq$. Esta *condición de arbitraje*[5] admite tres lecturas:

- *Arbitraje financiero*: la rentabilidad del capital instalado, dividendo $\Pi(K)$ más ganancia de capital $\dot{q}$, iguala la de invertir su valor al tipo de interés, $rq$.

- *Ecuación de comportamiento*: la valoración del capital se revisa con la bolsa, el tipo de interés y el ciclo.

- *Condición de equilibrio*: la inversión continúa hasta que el ingreso que aporta la última unidad de capital, $\Pi(K)$, se iguala con su coste de oportunidad, $rq - \dot{q}$. El coste de oportunidad son los intereses que no se ganan por no vender el capital, menos la ganancia de capital (o más la pérdida) por conservarlo.

3. *Ley de acumulación*: $\dot{k} = I$.

[5] Con depreciación, la condición es $\Pi(K) + \dot{q} = (r + \delta)q$, de modo que la rentabilidad del capital instalado iguala su coste de uso.

4. *Condición de transversalidad*: $\lim_{t\to\infty} e^{-rt} q\, k = 0$, que excluye las sendas explosivas.

**CSO**: la convexidad del coste de ajuste ($\varphi'' > 0$) y la linealidad de $\Pi(K)k$ en $k$ hacen cóncavo el hamiltoniano, de modo que las condiciones anteriores, junto con la de transversalidad, son también suficientes (Mangasarian).

**Solución**:

▪ *Función de demanda de inversión individual*: invirtiendo la condición de demanda de inversión se obtiene

$$I = (\varphi')^{-1}(q-1) \;\rightarrow\; I = f(q^+)$$

La inversión es creciente en la $q$ marginal, positiva si $q > 1$, negativa si $q < 1$ y nula si $q = 1$. La hipótesis de Tobin queda así derivada del programa de la empresa.

▪ *Función de demanda de inversión agregada*: agregando sobre las $N$ empresas idénticas de la industria, $\dot{K} = Nf(q)$. Junto con la condición de arbitraje, forma un sistema en $(K, q)$ que determina la evolución conjunta del capital y de su valoración, es decir, el ritmo al que el capital se aproxima a su estado estacionario.

**Equilibrio**: el estado estacionario se caracteriza por $\dot{K} = 0$ y $\dot{q} = 0$, es decir, la inversión neta es nula y la valoración del capital no se revisa. De la primera condición se obtiene $q^* = 1$ (con depreciación, $q^* > 1$, pues hay inversión de reposición)[6]. De la segunda resulta $\Pi(K^*) = rq^*$, que determina el capital de la industria. En el equilibrio, las empresas carecen de incentivos para modificar su capital y a los inversores les conviene conservar un capital por el que no prevén ni ganancias ni pérdidas.

**Desarrollo gráfico**: en el plano capital-$q$ marginal (figura A34.4), dos *curvas de demarcación* dividen el espacio en cuatro regiones:

1. *Curva* $\dot{K} = 0$: es horizontal en $q = 1$ y separa la región de inversión ($q > 1$), en la que el capital crece, de la de desinversión ($q < 1$), en la que cae.

2. *Curva* $\dot{q} = 0$: tiene pendiente negativa, pues más capital reduce el beneficio por unidad. Por encima de ella, el ingreso del capital es inferior al rendimiento de invertir su valor al tipo de interés, $rq$, y $q$ se revisa al alza, y por debajo, a la baja.

3. *Sendero de silla*: el sistema en $(K, q)$ tiene un punto de silla, y el sendero es la única senda que converge al equilibrio. Para cada capital heredado, $K(0)$, existe un único $q(0)$ sobre el sendero, compatible con la optimización. Por eso $q$ es una *variable de salto*, que se ajusta de inmediato a la nueva información, y $K$ una *variable predeterminada*, que se ajusta gradualmente.

Las demás sendas son inestables y la condición de transversalidad las descarta. Por encima del sendero, $q > 1$ hace crecer el capital, pero con $K$ elevado los beneficios son pequeños y $q$ solo puede ser alto si se espera que siga subiendo. En consecuencia, $K$ y $q$ divergen, lo que puede leerse como una burbuja especulativa.

4. *Equilibrio*: el estado estacionario, punto 0, se sitúa en el corte de ambas curvas, con $q = 1$ y $K = K^*$.

[6] Con depreciación, la curva $\dot{K} = 0$ deja de ser horizontal y tiene pendiente positiva en el plano $(K, q)$, pues la inversión de reposición, $\delta K$, crece con el capital.

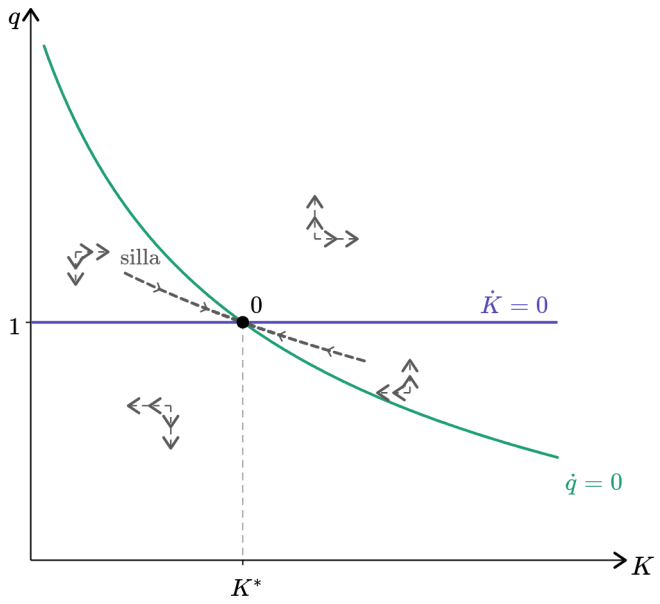



Figura A34.4: Diagrama de fases del modelo $q$. Las curvas $\dot{K} = 0$ y $\dot{q} = 0$ se cortan en el estado estacionario 0, al que conduce el sendero de silla.

**Estática comparativa**: se analizan cuatro perturbaciones (producción, tipo de interés, salario y desgravación fiscal) y, en el caso de la producción y de la desgravación, se distingue entre cambios permanentes y temporales:

▪ *Aumento permanente y no anticipado de la producción* ($\Delta Y > 0$): la curva $\dot{q} = 0$ se desplaza hacia arriba, de $(\dot{q} = 0)_0$ a $(\dot{q} = 0)_1$, igual que con una rebaja del impuesto sobre sociedades. La economía acaba con un capital permanentemente mayor, $K_1^*$, tras un salto inicial de la inversión que luego se diluye hasta cero. Con el mismo *stock*, cada unidad de capital rinde ahora más beneficio, así que a los inversores les basta una ganancia de capital menor para mantener sus títulos. Por eso $q$ sube de golpe del punto 0 al nuevo sendero de silla (punto $0'$), y desde allí $q$ y $K$ lo recorren hasta el punto 1, el nuevo equilibrio (figura A34.5).

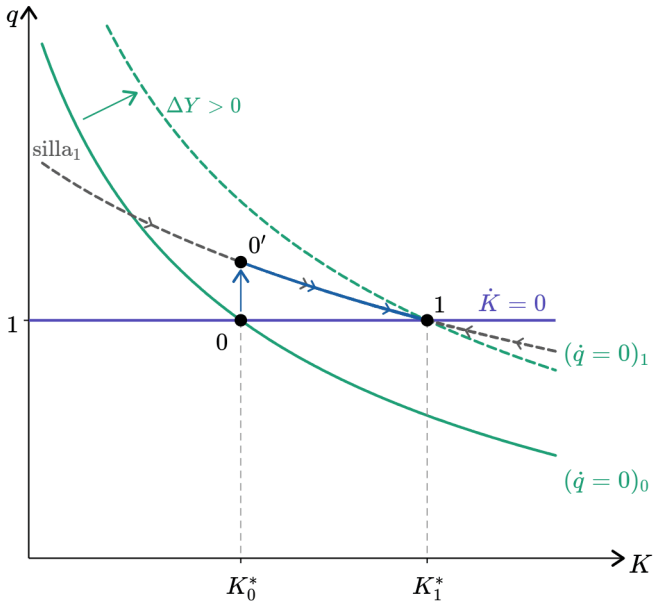



Figura A34.5: Efectos de un aumento permanente de la producción en el diagrama de fases del modelo $q$.

▪ *Aumento temporal y no anticipado de la producción*: las empresas saben que la producción volverá a su nivel en $T$. $q$ sube de golpe (punto $0'$), aunque se queda por debajo del nuevo sendero. El beneficio extra es pasajero, y la empresa prevé que tendrá que desmontar con costes de ajuste parte del capital que instale. Por eso el capital comienza a reducirse cuando la producción todavía no ha vuelto a la normalidad. En $T$ la economía alcanza el sendero de silla inicial (punto $0''$) y por él vuelve al punto 0. La inversión es mayor cuando se espera que la producción futura sea elevada, lo que constituye un efecto acelerador, esta vez microfundamentado (figura A34.6).

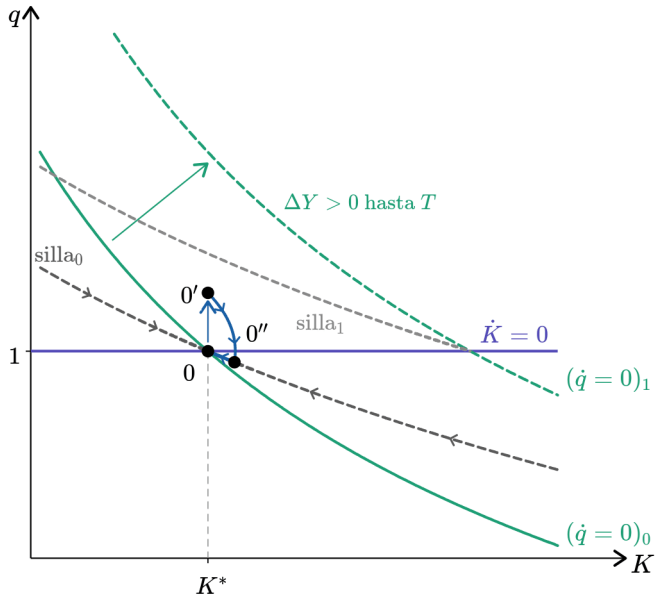



Figura A34.6: Efectos de un aumento transitorio de la producción en el diagrama de fases del modelo $q$.

▪ *Descenso del tipo de interés* (política monetaria expansiva): desplaza $\dot{q} = 0$ hacia arriba y, además, la inclina. Si la bajada es permanente y no anticipada, sus efectos coinciden con los de una subida permanente de la producción. Como $r$ es aquí un tipo instantáneo, es decir, de muy corto plazo, lo que cuenta para invertir es toda la curva de tipos. Por ejemplo, un tipo largo por debajo del corto anticipa recortes de los tipos futuros y anima la inversión.

▪ *Reducción salarial*: en un marco estático eleva el *stock* de capital si el efecto producción cruzado supera al efecto sustitución cruzado. En el marco dinámico, en cambio, como $\Pi(K)$ incorpora el empleo óptimo, la bajada salarial eleva siempre el beneficio por unidad de capital, desplaza hacia arriba la curva $\dot{q} = 0$ y aumenta el capital de estado estacionario. El ajuste es, no obstante, gradual por la convexidad de los costes de ajuste.

▪ *Desgravación permanente sobre el precio del capital* (política fiscal expansiva): desplaza hacia abajo la curva $\dot{K} = 0$, de $q = 1$ a $q = 1 - z$, con $z$ la desgravación (a la compra de capital o por depreciación), que equivale a una subvención por unidad de capital ($\Delta z > 0$). Un impuesto, por el contrario, desplaza la curva hacia arriba. La $q$ reacciona sin demora y baja del punto 0 al nuevo sendero de silla (punto $0'$), que queda más arriba que la nueva curva. Después el capital crece y $q$ sigue bajando por ese sendero hasta el punto 1, el nuevo equilibrio. Se invierte más, pero a costa de unos beneficios menores y de un capital instalado que vale menos (figura A34.7).

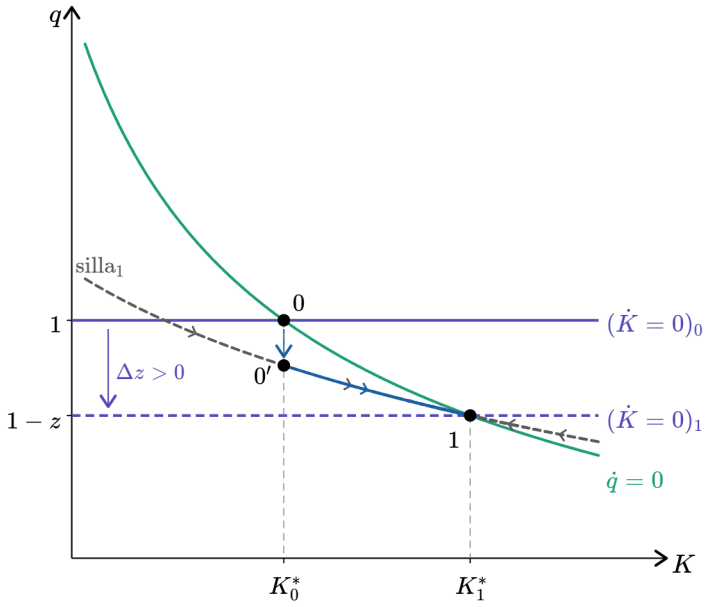



Figura A34.7: Efectos de una desgravación fiscal permanente a la inversión en el diagrama de fases del modelo $q$.

▪ *Desgravación temporal sobre el precio del capital*: $q$ baja menos que en el caso permanente y se queda sin llegar al nuevo sendero (punto $0'$), así que se invierte más desde el principio. Cuanto más cerca está el final del incentivo, más se acelera la inversión, porque las empresas quieren beneficiarse de él mientras dure. En $T$ la economía llega al sendero de silla inicial (punto $0''$) y regresa por él al punto 0. El estímulo a la inversión es, por tanto, mayor con una desgravación temporal que con una permanente (figura A34.8).

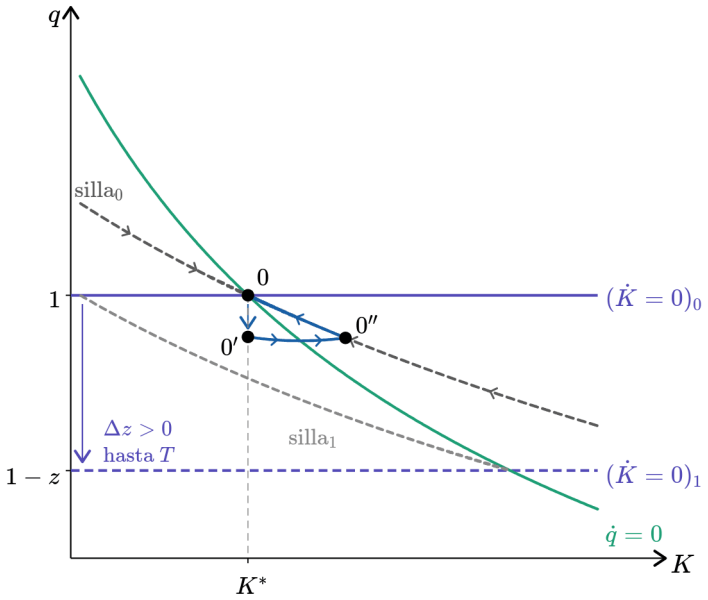



Figura A34.8: Efectos de una desgravación fiscal transitoria a la inversión en el diagrama de fases del modelo $q$.

En suma, la inversión depende de la $q$ marginal, es decir, de toda la senda esperada de beneficios y tipos de interés, y no solo de sus valores corrientes. El ritmo de ajuste es finito y depende de la relevancia de los costes de ajuste.

**Extensiones**:

- *Política monetaria*: eficaz a través del tipo de interés, y más cuanto más permanente se perciba la bajada y más se traslade a los tipos a largo plazo, que son los que incorporan las expectativas de los inversores.

- *Política fiscal*: eficaz a través de los incentivos a la inversión, y más si son temporales, pues las empresas adelantan la inversión para aprovecharlos. Una expansión de la demanda eleva $q$ y la inversión, pero la subida del tipo de interés que acompaña a la expansión contrarresta el aumento de $q$ y de la inversión.

### 2.3.3. Conclusiones

**Aspectos positivos**: es el modelo más sofisticado en términos de microfundamentación y ofrece a la vez un *stock* deseado y un ritmo de ajuste.

**Limitaciones**: la estimación directa es frágil, pues las regresiones de la inversión sobre la $q$ observada arrojan coeficientes muy bajos, que implican costes de ajuste y plazos de convergencia inverosímiles, lo que suele atribuirse a la dificultad de medir la $q$ marginal. Además, los costes de ajuste cuadráticos y simétricos son una simplificación. Con costes asimétricos o irreversibilidad, la respuesta de la inversión es asimétrica y aparece el valor de esperar (apartado 3.1). Por último, la equivalencia entre $q$ marginal y media exige las condiciones de Hayashi.

## 2.4. Conclusiones

Las teorías modernas plantean la inversión como un problema de optimización intertemporal de la empresa. El modelo de Jorgenson determina el *stock* de capital óptimo a partir del coste de uso, pero no el ritmo de ajuste. Los modelos con costes de ajuste añaden ese ritmo y hacen depender la inversión de la $q$ marginal, que resume las expectativas sobre beneficios y tipos futuros. La tabla A34.1 compara las teorías.

Tabla A34.1: Teorías de la demanda de inversión: determinantes, ajuste y eficacia de las políticas.

| Teoría | Determinante de la inversión | Ritmo de ajuste | Política monetaria | Política fiscal |
|---|---|---|---|---|
| Keynes (1936) | Expectativas (EMK frente a $r$) | No modelizado | Poco eficaz | Muy eficaz con la renta; expulsión parcial de $I$ |
| Acelerador (rígido/flexible) | Variación de la producción | Inmediato / retardo distribuido | Ineficaz | Muy eficaz a corto plazo (multiplicador-acelerador) |
| Jorgenson (1963) | Coste de uso del capital, $P_K(r+\delta)$ | Instantáneo (indeterminado) | Eficaz vía $r$ | Eficaz vía desgravaciones; adversa vía $r$ |
| Modelo $q$ (Tobin-Hayashi) | $q$ marginal (senda esperada de $\Pi$ y $r$) | Gradual, según $\varphi''$ | Eficaz; mayor si permanente y vía tipos largos | Eficaz; mayor si temporal |

# 3. Incertidumbre, irreversibilidad y financiación de la inversión

Quedan por ver otras dos aportaciones a la teoría de la inversión, la inversión como opción real bajo incertidumbre e irreversibilidad y la importancia de las decisiones de financiación, y, para terminar, las implicaciones de política económica de las teorías del tema.

## 3.1. Incertidumbre e irreversibilidad: la inversión como opción real (Bernanke, 1983; Dixit y Pindyck, 1994)

**Origen**: el enfoque parte del trabajo de Ben Bernanke (1983) y lo sistematizaron en los noventa Avinash Dixit y Robert Pindyck (1994).

**Idea**: invertir es ejercer una opción de compra sobre un activo irreversible. Esperar tiene valor, porque permite decidir con más información, y ese valor se añade al coste de uso como umbral que la productividad marginal debe superar.

### 3.1.1. Supuestos

El modelo de Jorgenson cambia en tres supuestos a los que antes apenas se prestaba atención:

**Entorno**: se invierte sin certeza sobre variables como los rendimientos futuros, y esa incertidumbre se va despejando en parte con el tiempo.

**Agentes**: la empresa elige cuándo invertir y puede aplazar la decisión, que deja de ser solo de sí o no para ser también de calendario.

**Tecnología**: en su mayoría, las inversiones en capital fijo son irreversibles, del todo o en parte, porque deshacerlas supone perder valor.

Con estos tres rasgos la decisión es más difícil de modelizar, aunque está muy aceptado tratarla como el ejercicio de una opción (real) *call*.

### 3.1.2. Modelo

**Planteamiento**: la empresa dispone de una opción de compra sobre un proyecto irreversible y elige el momento óptimo de ejercerla.

La empresa se enfrenta a un dilema. Si invierte ya, cobra desde hoy los rendimientos del proyecto, pero pierde la posibilidad de aguardar. Si aguarda, decidirá mejor informada, aunque renuncia a esos rendimientos mientras tanto. Invertirá, por tanto, cuando hacerlo valga más que conservar la opción.

**Solución**: el criterio neoclásico de llevar la inversión hasta que la productividad marginal del capital iguale su coste de uso queda sustituido por un listón más alto:

$$P\,\frac{\partial f}{\partial K} \geq r_K + \Theta$$

donde $P\,\partial f/\partial K$ es el valor de la productividad marginal del capital y $r_K$ su coste de uso. $\Theta$ es el *valor de la opción viva*, es decir, lo que aporta esperar (información nueva) neto de lo que cuesta (los rendimientos que se aplazan).

Un ejemplo de dos periodos permite cuantificar ese valor (Dixit y Pindyck, 1994). Un proyecto irreversible cuesta $I = 1600$ y produce cada año una unidad de un bien cuyo precio, hoy de 200, sube a 300 o baja a 100 con igual probabilidad el año próximo y se mantiene después para siempre. El tipo de descuento es del 10 %:

1. *Invertir hoy*: el precio esperado es 200 en todos los años, y el VAN, $-1600 + \sum_{t=0}^{\infty} 200/1{,}1^t = -1600 + 2200 = 600$, es positivo, por lo que el criterio neoclásico aconseja invertir.

2. *Esperar un año*: la empresa solo invierte si el precio sube, y entonces el VAN, descontado a hoy, es $\left(-1600 + \sum_{t=0}^{\infty} 300/1{,}1^t\right)/1{,}1 = (-1600 + 3300)/1{,}1 \approx 1545$. Como la subida del precio ocurre con probabilidad 1/2, esperar vale unos 773.

3. *Valor de la opción*: esperar vale más que invertir hoy, y la diferencia, unos 173, es el valor de la flexibilidad que se pierde al invertir, el coste de oportunidad que el término $\Theta$ del criterio recoge como rendimiento adicional exigido. Con más incertidumbre (precios de 350 o 50) esperar vale unos 1023 y la diferencia sube a 423. Lo que se gana al esperar es no invertir si el precio baja, de modo que la decisión depende de la gravedad del escenario malo.

**Estática comparativa**: cuanto mayor es la incertidumbre, mayor es el valor de la opción viva y mayor la productividad marginal exigida, de modo que la inversión se pospone. El efecto sobre el *stock* de capital esperado a largo plazo es, en cambio, ambiguo, pues la irreversibilidad también impide desinvertir en las fases desfavorables.

Por lo tanto, la respuesta es asimétrica. Según el "principio de las malas noticias" de Bernanke, la decisión de invertir hoy o esperar depende solo de la gravedad de las posibles malas noticias, no de la magnitud de las buenas, y la inversión reacciona con retraso a las mejoras de los fundamentos.

El signo de la relación entre incertidumbre e inversión no es, sin embargo, general. Sin irreversibilidad, con costes de ajuste convexos, competencia perfecta y rendimientos constantes a escala, el beneficio marginal del capital es convexo en el precio del producto, de modo que una mayor incertidumbre eleva la $q$ marginal esperada y estimula la inversión (Hartman, 1972; Abel, 1983). Caballero (1991) reconcilia ambos resultados al mostrar que el efecto negativo de la incertidumbre exige combinar la irreversibilidad con competencia imperfecta o rendimientos decrecientes a escala, que hacen depender el beneficio marginal del capital del *stock* de la propia empresa.

**Extensiones**: la consecuencia de mayor alcance del modelo es que pocas cosas pesan tanto en la inversión como la estabilidad macroeconómica. La autoridad económica puede, por tanto, animar la inversión si reduce la incertidumbre, es decir, si procura, entre otras cosas, que los precios sean estables y que los tipos de cambio y de interés no sufran sobresaltos.

Esa estabilidad tiene un sentido amplio. Abarca las variables económicas, pero también que los derechos de propiedad se respeten en la práctica y que la desigualdad no sea tan grande como para alimentar la tensión social y el conflicto político.

### 3.1.3. Conclusiones

**Aspectos positivos**: según los trabajos empíricos, esperar puede tener un valor elevado, en especial si la incertidumbre es muy alta, de manera que esta frena con fuerza la inversión aunque los inversores sean neutrales al riesgo. Además, el enfoque explica la lentitud de la inversión tras las recesiones y su asimetría.

**Limitaciones**: el valor de la opción exige especificar el proceso estocástico de los rendimientos y supone irreversibilidad total, cuando la mayoría de las inversiones son parcialmente reversibles. Además, como la $q$ marginal, no es directamente observable.

## 3.2. Importancia de las decisiones de financiación

Los modelos anteriores pasan por alto que la empresa decide cómo financiar su inversión prácticamente al mismo tiempo que resuelve invertir. Por lo general puede recurrir a los beneficios no repartidos (resultados y reservas), a la deuda o a la emisión de acciones.

> Las decisiones de financiación de la empresa (estructura del capital, coste medio ponderado, Modigliani-Miller) se desarrollan en los temas B02 y B03, y aquí se retiene su efecto sobre la demanda de inversión.

### 3.2.1. Perspectiva tradicional (Durand, 1952) y perspectiva M&M (Modigliani y Miller, 1958)

**Origen**: la visión tradicional de la estructura del capital se asocia habitualmente a David Durand (1952). Franco Modigliani y Merton Miller (1958) la refutaron bajo mercados perfectos.

**Idea**: en la visión tradicional, existe una estructura óptima del capital que permite minimizar el coste ponderado del capital. Por tanto, las decisiones de inversión no serán independientes de las decisiones de financiación.

**Razonamiento**: para Modigliani-Miller, si los mercados de capitales son perfectos, no existe una estructura óptima del capital. Por ello, la decisión de inversión es independiente de la decisión de financiación (proposición III) y la forma de financiación es irrelevante.

**Conclusiones**:

- *Aspectos positivos*: el teorema de M&M sirve de referencia, porque al precisar los supuestos bajo los que la financiación es irrelevante desplaza la pregunta hacia cuál de ellos falla y con qué efecto sobre la inversión (tema B03). Además, justifica que los modelos de inversión anteriores prescindan de la financiación y descuenten con un único coste del capital.

- *Limitaciones*: la irrelevancia de la financiación que sostiene M&M parece poco plausible, dado que en la realidad existen imperfecciones de mercado (p. ej. costes de transacción, costes de agencia, impuestos distorsionantes o inflación).

### 3.2.2. Racionamiento del crédito (Stiglitz y Weiss, 1981)

**Origen**: el modelo se debe a Joseph Stiglitz y Andrew Weiss (1981). Se desarrolla en el tema A42, apartado 2.3.1, y se aplica a las crisis bancarias en el tema B26, apartado 1.2.3.

**Idea**: una de las imperfecciones de los mercados financieros que puede analizarse de manera más clara es la información asimétrica, que da lugar a selección adversa y, con ello, a racionamiento del crédito.

**Razonamiento**: si el banco sube el tipo de interés, expulsa a los prestatarios más seguros y atrae a los más arriesgados, de modo que el rendimiento esperado del banco deja de crecer con el tipo a partir de cierto nivel. Por ello, prefiere racionar la cantidad antes que subir el precio.

La selección adversa limita la oferta de crédito incluso si los prestatarios aceptarían tipos más altos, lo que restringe la inversión por debajo de su nivel óptimo.

**Conclusiones**:

▪ *Aspectos positivos*: explica que el crédito se racione en equilibrio, sin rigideces de precios, de modo que la inversión depende de la disponibilidad de financiación y no solo de su coste.

▪ *Limitaciones*: el resultado exige que el banco no pueda separar a los prestatarios según su riesgo. Con contratos que combinan tipo de interés y garantías, por ejemplo, los prestatarios se autoseleccionan y el racionamiento puede desaparecer.

### 3.2.3. Teoría de la jerarquía financiera o *pecking order theory* (Myers y Majluf, 1984)

**Origen**: Stewart Myers y Nicholas Majluf (1984) propusieron la teoría a partir de la información asimétrica entre directivos e inversores.

**Idea**: la empresa recurre primero a la financiación más barata y después a las demás. Lo más barato son los beneficios no repartidos (autofinanciación), después viene la deuda y lo más caro es emitir acciones.

**Razonamiento**: la autofinanciación no tiene costes de emisión ni prima de información asimétrica. Por su parte, la deuda tiene costes de transacción pero menores que los de las acciones (los inversores perciben sobrevaloración en emisiones de acciones).

**Desarrollo gráfico**: la jerarquía financiera se representa en el plano inversión-coste de financiación (figura A34.9):

1. *Curva de costes de financiación*, $C(I)$: presenta tres tramos, que reflejan el coste de la autofinanciación, el de la emisión de deuda y el de la emisión de acciones. Es creciente por tramos, porque cada fuente es más cara que la anterior (costes de emisión y prima de información asimétrica).

2. *Curva de demanda de inversión*, $D(I)$: refleja la relación negativa que existe entre el nivel de inversión y el coste, porque al invertir más se acometen proyectos de menor rentabilidad (apartado 1.2.2).

3. *Equilibrio*: en el punto 0, la demanda de inversión corta la curva de costes y fija la inversión, $I^*$, y su coste, $C^*$. En función del tramo en que se produzca el corte, la empresa tendrá mayor o menor holgura para acometer proyectos de inversión, de modo que la inversión depende de la liquidez interna y no solo del coste de uso.

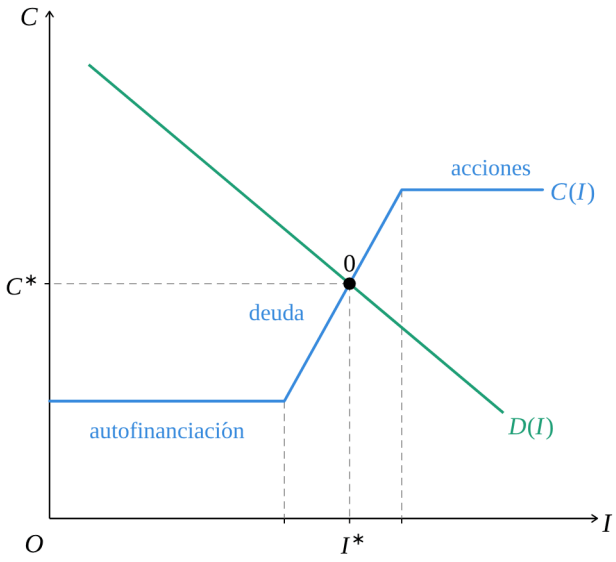



Figura A34.9: Jerarquía financiera (Myers y Majluf, 1984). El coste marginal de los fondos es creciente por tramos (autofinanciación, deuda, acciones) y determina la inversión.

**Conclusiones**:

▪ *Aspectos positivos*: deriva de un solo supuesto (la información asimétrica entre directivos e inversores) el predominio de la autofinanciación y la dependencia de la inversión respecto de la liquidez interna.

▪ *Limitaciones*: ordena las fuentes de financiación pero no fija una estructura del capital objetivo, y prescinde de los impuestos y de los costes de insolvencia que recoge la teoría del equilibrio estático (tema B03, apartado 2.3.5).

### 3.2.4. El acelerador financiero (Bernanke, Gertler y Gilchrist, 1996)

> El acelerador financiero como mecanismo de propagación del ciclo se desarrolla en el tema A42, apartado 2.3.2, y aquí se retiene su efecto sobre la inversión.

**Origen**: Ben Bernanke, Mark Gertler y Simon Gilchrist propusieron el acelerador financiero (*The Financial Accelerator and the Flight to Quality*, 1996).

**Idea**: las imperfecciones financieras amplifican los efectos de los *shocks* reales sobre la inversión.

**Razonamiento**: las restricciones de crédito, vinculadas al valor de los activos de las empresas (colateral), generan un mecanismo de retroalimentación.

Un *shock* positivo eleva el valor de los activos, relaja las restricciones crediticias y amplifica la inversión, mientras que un *shock* negativo reduce el colateral, endurece las condiciones de crédito y contrae la inversión más allá de lo que justificarían los fundamentos.

El mecanismo convierte perturbaciones moderadas en fluctuaciones amplias y persistentes de la inversión, y explica la asimetría de las recesiones con crisis de crédito.

**Conclusiones**:

- *Aspectos positivos*: ofrece un mecanismo de propagación endógeno de los ciclos de inversión, que se ha incorporado a los modelos de equilibrio general con fricciones financieras (Bernanke, Gertler y Gilchrist, 1999).

- *Limitaciones*: la fricción se sitúa en el balance de las empresas prestatarias y deja fuera las restricciones de los propios intermediarios financieros, que la crisis de 2008 situó en primer plano.

#### 3.2.5. Conclusiones

**Aspectos positivos**: las imperfecciones financieras explican por qué la inversión depende de la liquidez interna y del colateral, y no solo del coste de uso o de la $q$, como documentan Fazzari, Hubbard y Petersen (1988) para las empresas con restricciones de acceso al crédito.

**Limitaciones**: la sensibilidad de la inversión al flujo de caja admite otras lecturas (el flujo de caja informa sobre la rentabilidad futura), lo que dificulta identificar las restricciones financieras. Además, la jerarquía financiera no siempre se observa en las decisiones de emisión.

### 3.3. Implicaciones de política económica

Las implicaciones de las teorías de Keynes, del acelerador, de Jorgenson y de los modelos con costes de ajuste se resumen en la tabla A34.1 del apartado 2.4. Las opciones reales y las imperfecciones financieras añaden que la estabilidad macroeconómica, la certidumbre y un marco institucional que minimice las imperfecciones financieras son los principales determinantes de la inversión.

## Conclusiones

El tema recorre las teorías de la inversión desde la comparación entre rentabilidad y coste de Fisher y Keynes hasta la optimización intertemporal de la empresa. El acelerador liga la inversión a la variación de la producción, pero carece de microfundamentos y solo fija el ritmo de ajuste con un retardo *ad hoc*. Jorgenson deriva el *stock* de capital deseado del coste de uso, y los modelos con costes de ajuste añaden ese ritmo y resumen en la $q$ marginal las expectativas sobre beneficios y tipos futuros. Las opciones reales explican el valor de esperar ante la incertidumbre, y las imperfecciones financieras, la dependencia de la inversión respecto de la liquidez interna y del colateral. La política monetaria es más eficaz cuanto más permanente se perciba y los incentivos fiscales, cuanto más temporales, pero los principales determinantes de la inversión son la estabilidad macroeconómica y la certidumbre.

La inversión aplica a la empresa la decisión intertemporal (tema A29), y sus criterios coinciden con los que la economía financiera emplea para valorar proyectos y elegir la estructura del capital (temas B02 y B03). Su volatilidad la convierte en protagonista del ciclo. El acelerador sustenta los modelos de interacción con el multiplicador, y el racionamiento del crédito y el acelerador financiero explican cómo las condiciones financieras amplifican las perturbaciones (tema A42), base del canal del crédito de la política monetaria (tema A37). A largo plazo, en cambio, la acumulación de capital eleva el nivel de la renta per cápita, pero no su tasa de crecimiento (tema A43).

Desde comienzos de la década de 2000, la inversión empresarial de EE. UU. ha sido débil pese a una rentabilidad y una $q$ elevadas. Gutiérrez y Philippon (2017) atribuyen esa debilidad sobre todo a la menor competencia y a una gobernanza orientada al corto plazo, que dedica el flujo de caja a recomprar acciones. Peters y Taylor (2017) muestran, en cambio, que la relación entre inversión y $q$ mejora al incluir el capital intangible, lo que apunta a un problema de medición. También se discute el efecto de la incertidumbre. Bloom (2009) muestra que los *shocks* de incertidumbre, al elevar el valor de esperar, provocan caídas bruscas de la inversión y del empleo seguidas de una recuperación rápida. Asimismo, Baker, Bloom y Davis (2016) asocian un índice de incertidumbre de la política económica con menor inversión en los sectores más expuestos a ella.

# A35. Teorías de la demanda de dinero

Teorías de la demanda de dinero. Implicaciones de política económica.

## Introducción

Las **teorías de la demanda agregada** estudian los comportamientos que hay detrás de los grandes bloques de los modelos macroeconómicos, desde el consumo y la inversión, que forman la demanda de bienes que recoge la curva IS, hasta la demanda de dinero, que está detrás de la curva LM.

El punto de partida es la *Teoría general* de Keynes (1936), que hizo depender la demanda de dinero de la **preferencia por la liquidez**, cuyo motivo especulativo la hacía sensible al tipo de interés. Sin embargo, era una relación entre agregados, sin una derivación explícita a partir de decisiones individuales. Baumol (1952) y Tobin (1956, 1958) dieron fundamento microeconómico a la demanda de dinero por motivo transacción y por motivo cartera, y Friedman (1956) la reformuló como la demanda de un activo más. Desde los años setenta, la crítica de Lucas (1976) y las expectativas racionales llevaron el análisis a agentes que optimizan en un horizonte intertemporal con incertidumbre, y los modelos de equilibrio general con dinero renovaron la teoría de la demanda de dinero.

El **dinero** es el activo que sirve de medio de cambio generalmente aceptado y que, además, cumple las funciones de unidad de cuenta y depósito de valor. La **demanda de dinero** recoge cuántos saldos, nominales o reales, prefieren conservar los agentes como dinero y no en forma de otros activos o de bienes.

Su estudio es central en la macroeconomía por dos razones. Primera, junto con la oferta de dinero determina el tipo de interés (o el nivel de precios, según el paradigma) y, con él, el mecanismo de transmisión de la política monetaria. Segunda, sus dos propiedades principales, la estabilidad de la función y su sensibilidad al tipo de interés, separan a las escuelas y deciden la eficacia relativa de las políticas monetaria y fiscal (apartado 3.2).

Desde los años ochenta, la innovación financiera hizo inestable la función de demanda de dinero, y los bancos centrales dejaron de fijarse objetivos de cantidad de dinero para fijar el tipo de interés (tema A36). Tras 2008, con los tipos de interés cerca de cero, reapareció la **trampa de liquidez**, en la que la demanda de dinero absorbe cualquier aumento de la oferta y la política monetaria convencional pierde eficacia. Hoy el efectivo pierde peso en los pagos, y la emisión de dinero digital por los bancos centrales, como el euro digital, plantea nuevas preguntas sobre la demanda de dinero y sobre la estabilidad de los bancos.

El tema se organiza en tres partes:

- *Fundamentos*: el tema expone las teorías de la demanda de dinero desde la teoría cuantitativa hasta el monetarismo.

- *Desarrollos posteriores*: se presentan los desarrollos en equilibrio general dinámico.

- *Evidencia y política económica*: el tema termina con la evidencia empírica, las implicaciones de política económica y el debate sobre la economía sin efectivo y el dinero digital.

# 1. Fundamentos de la teoría sobre la demanda de dinero

Los fundamentos siguen el orden en que surgieron la teoría cuantitativa, la preferencia por la liquidez, la síntesis neoclásica y el monetarismo.

## 1.1. Aportación clásica y neoclásica: desarrollos basados en la teoría cuantitativa del dinero

El paradigma clásico y neoclásico explica el sector monetario con la **teoría cuantitativa del dinero**, una teoría sobre las causas de los cambios del valor o poder adquisitivo del dinero. De ella se deriva que, a largo plazo, dinero y nivel general de precios guardan una relación proporcional. Esa relación la habían anunciado ya la escuela de Salamanca (M. de Azpilcueta, 1556)[1], J. Bodin (1568) y D. Hume (1752).

La formalizaron después, por separado, dos tradiciones:

- *Enfoque de Fisher* (1911): agregado y de largo plazo, con el acento en los factores institucionales.

- *Enfoque de Cambridge* (1917): individual y de corto plazo, atento a qué lleva a cada agente a mantener dinero.

**Características comunes**: ambos enfoques parten de la identidad contable del gasto, aunque la interpretan de forma distinta, y comparten tres supuestos:

- *Pleno empleo*: aceptan la ley de Say.

- *Ausencia de atesoramiento*: los agentes no atesoran dinero, de modo que $M^s = M^d$ en todo momento.

- *Oferta de dinero exógena*: la fijan las autoridades monetarias y corresponde a un agregado estrecho (efectivo y depósitos a la vista, M1)[2].

### 1.1.1. Enfoque de las transacciones (Newcomb, 1885; Fisher, 1911)

**Origen**: Simon Newcomb (1885) formuló la *ecuación de cambio* e Irving Fisher (1911) la convirtió en la formulación de referencia de la teoría cuantitativa.

**Idea**: el dinero se demanda solo para realizar transacciones, de modo que la cantidad de saldos reales que la economía necesita es una fracción técnica, fijada por las instituciones de pago, del volumen de transacciones.

**Supuestos**:

[1] M. de Azpilcueta desarrolló una teoría del valor-escasez precursora de la teoría cuantitativa del dinero en su *Comentario resolutorio de cambios* (1556), publicado como apéndice de su *Manual de confesores y penitentes*.

[2] Esta es la gran diferencia con la obra de Knut Wicksell, el cual considera un agregado monetario más amplio, que incluye el crédito bancario (en el límite, una economía de crédito puro). Ello hace endógenos tanto la masa monetaria como los estímulos monetarios y da lugar al proceso acumulativo de los precios.

▪ *Identidad*: Fisher parte de la identidad contable del gasto, $MV \equiv PT$[3], que iguala el valor nominal de las compras y el de las ventas. En ella, $M$ es la oferta monetaria (M1), y la *velocidad-transacción*, $V$, cuenta cuántas compras financia por término medio una unidad monetaria en el periodo. A la derecha, $P$ es el nivel general de precios y $T$, el número de transacciones.

▪ *Funciones del dinero*: el dinero solo se utiliza como unidad de cuenta y medio de cambio, de modo que no reporta utilidad.

▪ *Estabilidad*: $V$ y $T$ son constantes o estables[4], la primera por cuestiones institucionales y la segunda por el pleno empleo.

**Planteamiento**: no hay programa de optimización. La demanda de dinero se obtiene de la identidad contable imponiendo el equilibrio del mercado monetario, $M \equiv M^s = M^d$, y la estabilidad de $V$ y $T$.

**Solución**: despejando los saldos reales, la función de demanda de dinero agregada, "mecanicista" y muy estable, es

$$\frac{M^d}{P} = \frac{1}{V_0} T_0$$

**Estática comparativa**: la demanda de saldos reales solo se desplaza si cambian las instituciones de pago ($V_0$) o el volumen de transacciones ($T_0$), y el tipo de interés no la afecta. Por tanto, dada la demanda real de saldos, $M^d/P$, todo aumento de la oferta monetaria ($\Delta M^s > 0$) se traslada íntegramente a los precios.

**Extensiones**: con una demanda de dinero estable, los efectos de la política monetaria son potentes y predecibles.

▪ *Política monetaria*: determina el nivel de precios, pero es neutral sobre las variables reales (*dicotomía clásica*).

▪ *Política fiscal*: ineficaz con la renta, que está en su nivel de pleno empleo.

**Conclusiones**:

▪ *Aspectos positivos*: ofrece la formulación de referencia de la teoría cuantitativa.

▪ *Limitaciones*: presenta dos.

- *Transacciones*: no hay un modo sencillo de calcular su número.
- *Funciones del dinero*: el dinero no solo es unidad de cuenta y medio de cambio, por lo que su demanda podría depender de otros elementos, como el tipo de interés.

[3] $V$ y $T$ son variables reales, que dependen de factores institucionales y tecnológicos, mientras que $M$ y $P$ son nominales.

[4] Las economías desarrolladas han tenido una velocidad de circulación bastante estable a largo plazo, aunque no ha sido constante. Bajó durante la monetización de esas economías y subió más tarde con la sofisticación financiera (Bordo y Jonung, 1987). Donde hubo episodios de inflación alta, en cambio, osciló mucho (Cagan, 1956).

### 1.1.2. Enfoque de los saldos de caja (escuela de Cambridge, 1917)

**Origen**: Alfred Marshall, Arthur Pigou (1917) y el joven Keynes (1923) reformularon la teoría cuantitativa desde la decisión individual de mantener saldos de caja.

**Idea**: los agentes desean mantener una fracción de su renta en dinero, porque este presta servicios útiles, y esa fracción, la $k$ de Cambridge, resume todos los factores que afectan a la velocidad de circulación.

**Supuestos**:

- *Identidad*: se considera la renta real, en lugar del número de transacciones, como medida del volumen de transacciones de la economía, $MV \equiv PY$. Así, la velocidad pasa a ser una *velocidad-renta* del dinero, es decir, las veces que una unidad monetaria entra a formar parte de la renta de alguien en un periodo.

- *Funciones del dinero*: además de unidad de cuenta y medio de cambio, el dinero es depósito de valor, es decir, reporta utilidad. Por ello, la demanda podría depender también de la riqueza y, por extensión, del tipo de interés.

- *Estabilidad*: $V$ e $Y$ se consideran estables. $V$ depende ahora también del coste de oportunidad, la riqueza o las expectativas, pero la escuela de Cambridge la trata como constante.

**Planteamiento**: tampoco hay programa de optimización explícito.

**Solución**: imponiendo el equilibrio del mercado monetario, $M \equiv M^s = M^d$, la función de demanda de dinero agregada es

$$\frac{M^d}{P} = k_0\, Y_0$$

donde $Y_0$ es la renta real, constante por el pleno empleo, y $k_0 \in (0,1)$ la $k$ de Cambridge, la fracción de renta real que los individuos desean mantener en saldos de caja[5]. Se trata como un parámetro constante, "la constante de Cambridge", inversa de la velocidad-renta, $k_0 = 1/V_0$. La renta, y no el número de transacciones, es así el principal determinante de la demanda de dinero[6].

**Estática comparativa**: la demanda de saldos reales solo se desplaza con la renta, y el tipo de interés queda fuera.

**Extensiones**: los resultados de política coinciden con los del enfoque de Fisher.

**Conclusiones**:

- *Aspectos positivos*: permite considerar elementos subjetivos en la demanda de dinero y ofrece la primera teoría genuina de esta.

- *Limitaciones*: aunque considera la función de depósito de valor, no asigna un papel explícito al tipo de interés en la demanda de dinero.

[5] Para Pigou, el dinero se mantiene si el rendimiento que da en comodidad y seguridad supera el ingreso perdido al no invertir en actividades productivas o la satisfacción perdida por no usar el dinero simplemente para comprar bienes de consumo.

[6] Cannan (1921) añadió dos ideas. Cuanto más inflación se anticipa, menos dinero se demanda, y lo que interesa analizar es cuánto dinero se quiere tener, una magnitud *stock*, y no un flujo.

### 1.1.3. Conclusiones

**Aspectos positivos**: ambos enfoques llegan a una demanda de dinero proporcional a una variable de escala, transacciones o renta, con una constante de proporcionalidad institucional, lo que da a la teoría cuantitativa una base sencilla.

**Limitaciones**: la constancia de esa proporción es un supuesto y no un resultado, pues ninguno de los dos enfoques la deriva de un problema de optimización. Si $k$ depende del coste de oportunidad, la demanda de dinero pasa a depender del tipo de interés, y si $k$ depende de expectativas volátiles, la demanda deja de ser estable, que es lo que Keynes explotará.

## 1.2. Teoría keynesiana: la preferencia por la liquidez (Keynes, 1936)

**Origen**: Keynes (1936) concibe por primera vez el dinero como un activo financiero y formula la teoría de la preferencia por la liquidez.

**Idea**: Keynes acepta que la ecuación de Cambridge se cumple a largo plazo, pero a corto plazo la incertidumbre y las expectativas sobre el tipo de interés hacen que el dinero se demande también como depósito de valor. La demanda de dinero depende, así, de la renta y del tipo de interés, entendido no como recompensa a la abstinencia del consumo sino como retribución por renunciar a la liquidez[7], y es inestable.

### 1.2.1. Supuestos

**Entorno**: el horizonte es el corto plazo, con incertidumbre sobre la evolución futura del tipo de interés. Los individuos mantienen su riqueza en dos activos financieros, dinero (activo "estéril" sin rentabilidad) y bonos (con rentabilidad).

El precio de mercado de un bono y su tasa de rentabilidad guardan una relación inversa[8][9], de modo que una disminución de la tasa exigida a los nuevos bonos hace más atractivos los emitidos al cupón anterior, aumenta su demanda y, con ella, su precio.

**Agentes**: cada individuo tiene una expectativa sobre el tipo de interés *normal*, $i^e$, hacia el que cree que tenderá el tipo de interés y que, por simplicidad, se toma como valor crítico[10]. Ese tipo normal es convencional (depende de la experiencia pasada y de la opinión media), cambiante y heterogéneo entre individuos.

### 1.2.2. Modelo

**Planteamiento**: no hay programa de optimización explícito, como es habitual en Keynes. En su lugar, Keynes distingue tres motivos por los que se demanda dinero, cada uno ligado a una función del dinero:

[7] "The rate of interest […] being the reward for parting with liquidity is a measure of the unwillingness of those who possess money to part with their liquid control over it" (Keynes, 1936).

[8] La relación es inversa y convexa para un bono sin opciones con nominal, cupón y vencimiento dados, de modo que, ante una misma variación del tipo, el precio sube más de lo que baja.

[9] Analíticamente: $P_{B,j} = \sum_{t=j+1}^{T} C_t/(1+r_B)^{t-j} + F_T/(1+r_B)^{T-j}$. No obstante, Keynes consideró el caso más sencillo de bonos perpetuos: $P_{B,t} = C/r_{B,t}$.

[10] En rigor, con bonos perpetuos, el umbral por encima del cual el cupón compensa la pérdida de capital esperada es $i^e/(1+i^e)$, algo inferior.

▪ *Transacción*: cubre los pagos previstos, ligados al uso del dinero como medio de cambio. En este motivo Keynes sigue el enfoque de Cambridge y hace la demanda (real) proporcional a la renta real, $M_T^d = f(Y^+)$.

▪ *Precaución*: sirve de reserva ante pagos imprevistos, nacidos de la incertidumbre. *A priori* dependería también del tipo de interés nominal, coste de oportunidad de los saldos líquidos, $M_P^d = f(Y^+, i^-)$, pero Keynes lo omite por simplicidad y escribe $M_P^d = f(Y^+)$.

▪ *Especulación*: deriva de la expectativa de rendimiento futuro. Se debe a la función del dinero como depósito de valor, $M_E^d = f(i^-)$, y es la aportación novedosa de Keynes, pues los neoclásicos ya habían identificado los otros motivos (Pigou contempló el motivo precaución). El motivo especulación establece el nexo entre la demanda de dinero y el tipo de interés y rompe la dicotomía clásica.

**Solución**:

▪ *Función de demanda de dinero individual*: la demanda especulativa individual es discontinua[11], según el tipo de interés efectivo se sitúe por encima o por debajo del valor crítico del individuo:

- *Bonos atractivos* ($i > i^e$): el individuo considera que el tipo de interés es elevado y que caerá, con el consecuente aumento del precio de los bonos. Por eso, mantiene toda su riqueza en bonos y no demanda dinero.

- *Bonos no atractivos* ($i < i^e$): considera que el tipo es bajo y que subirá, con la consecuente bajada del precio de los bonos. Por ello, mantiene toda su riqueza, $W$, en dinero.

$$\frac{M_E^d}{P} = \begin{cases} 0 & \text{si } i > i^e \\ W & \text{si } i < i^e \end{cases}$$

▪ *Función de demanda de dinero agregada*: los valores críticos son heterogéneos y las tenencias individuales insignificantes, de modo que la demanda especulativa agregada es continua, decreciente y convexa[12].

Sumando los tres motivos se obtiene la función de demanda (real) de dinero agregada keynesiana[13]:

$$\frac{M^d}{P} = L\left(Y^+, \underbrace{(r+\pi^e)}_{i}{}^-\right) \to M^d/P = kY - h\,i$$
$$k = \frac{\partial(M^d/P)}{\partial Y} > 0, \qquad h = -\frac{\partial(M^d/P)}{\partial i} > 0$$

donde $k$ y $h$ miden la respuesta de los saldos demandados ante la renta y ante el tipo nominal $i$, respectivamente. La demanda por los motivos transacción y precaución queda recogida en $kY$, y la del motivo especulación[14], núcleo de la preferencia por la liquidez, en $-h\,i$.

[11] Si $i = i^e$, el individuo es indiferente entre ambos activos y su demanda es perfectamente elástica entre 0 y $W$. Keynes no contempla este caso.

[12] Gráficamente, la demanda de dinero especulativa agregada se obtendrá por la suma horizontal de las demandas de dinero especulativas individuales.

[13] Con la forma lineal, $M^d/P = kY - h\,i$, la curva de demanda de dinero es recta. La convexidad keynesiana exige que $h$ aumente al caer el tipo de interés.

[14] Si $h = 0$, la demanda de dinero es vertical en el plano $(M/P, i)$, como en el enfoque neoclásico. Si $h \to \infty$ (trampa de liquidez), es horizontal, al igual que la LM.

**Desarrollo gráfico**: la demanda de dinero keynesiana se representa en el plano dinero (real)-tipo de interés nominal (figura A35.1):

1. *Demanda especulativa individual*, $M_E^d/P$ (panel a): es nula por encima del tipo crítico, $i^e$, donde los bonos son atractivos, y salta a la riqueza, $W$, por debajo de él.

2. *Demanda agregada*, $L(Y_0, i)$ (panel b): parte de la demanda por transacción y precaución, $kY_0$, y es decreciente y convexa, porque cuanto más bajo es el tipo de interés, más agentes lo sitúan por debajo de su valor crítico.

3. *Trampa de liquidez*: al llegar a un tipo mínimo, $i_{\min}$, la curva se hace horizontal. Nadie quiere ya bonos, pues el tipo es tan bajo que todo el mundo cuenta con que subirá.

**Estática comparativa**:

- *Renta*: un aumento de la renta real eleva la demanda de dinero, pues los agentes quieren realizar más transacciones.

- *Tipo de interés*: una subida del tipo de interés nominal eleva el coste de oportunidad de los saldos y reduce la demanda por motivo especulación, con una sensibilidad, $h$, que Keynes supone elevada.

- *Expectativas*: un cambio en las expectativas sobre el tipo normal desplaza toda la curva, de $L(Y_0, i)$ a $L(Y_0, i)'$ en el panel (b), lo que la hace inestable.

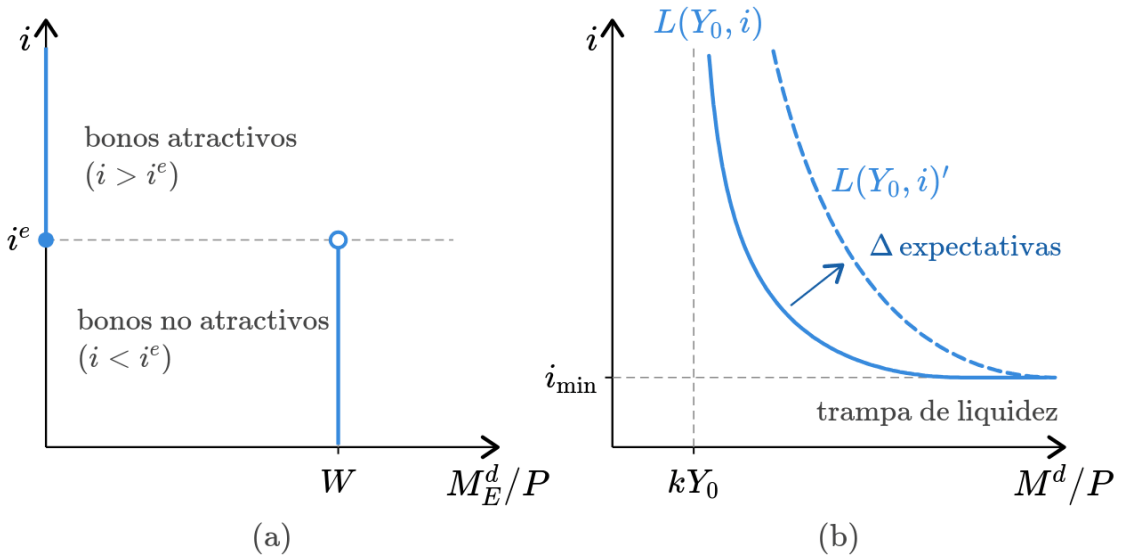



Figura A35.1: Demanda de dinero keynesiana. La demanda especulativa individual es discontinua en el tipo crítico $i^e$ (a), mientras que la agregada es decreciente y convexa y se vuelve horizontal en la trampa de liquidez (b). La curva discontinua muestra su desplazamiento ante un cambio de expectativas.

**Extensiones**: una de las contribuciones de la teoría monetaria keynesiana es la trampa de liquidez, que Keynes planteó solo como posibilidad teórica[15]. En ella, el tipo de interés ha llegado a su mínimo, cada saldo adicional se mantiene como dinero y la demanda es infinitamente elástica respecto al tipo de interés ($h \to \infty$). En la práctica, ese suelo, el límite inferior de los tipos de interés (*zero lower bound*, ZLB), se sitúa en torno a cero (apartado 3.3)[16]. De la demanda de dinero keynesiana se siguen estas implicaciones de política económica.

- *Política monetaria*: poco eficaz con la renta, por la alta sensibilidad de la demanda de dinero al tipo de interés (segundo eslabón débil). En la trampa de liquidez es totalmente ineficaz, porque se rompe el vínculo entre las operaciones de mercado abierto y los tipos de interés.

- *Política fiscal*: muy eficaz con la renta, y especialmente potente en la trampa de liquidez, pues no eleva el tipo de interés.

#### 1.2.3. Conclusiones

**Aspectos positivos**: el abandono de la dicotomía clásica y la noción de trampa de liquidez enriquecieron el análisis de la política económica, al dotar de efectividad diferenciada a la política fiscal y a la monetaria.

**Limitaciones**: cada individuo guarda toda su riqueza en bonos o en dinero, nunca en ambos, así que no hay diversificación. Los tres motivos, además, se postulan sin microfundamentar, tarea que asumirá después la síntesis neoclásica.

### 1.3. Síntesis neoclásica

Los motivos keynesianos de demanda de dinero fueron el punto de partida de las teorías posteriores que buscaban microfundamentarlos.

#### 1.3.1. El motivo transacción: modelos de inventarios (Baumol, 1952; Tobin, 1956)

**Origen**: William Baumol (1952) y James Tobin (1956) aplicaron la teoría de gestión de inventarios a los saldos de transacción.

**Idea**: el dinero se mantiene como un inventario para transacciones del que el agente dispone gradualmente. Existe una disyuntiva entre el coste de transacción de reponerlo y el coste de oportunidad de mantenerlo, de la que resulta una demanda de dinero por motivo transacción que depende también del tipo de interés.

**Supuestos**:

- *Entorno*: información perfecta y marco estático. Todos los flujos son conocidos y se considera un único periodo.

[15] En el modelo IS-LM, la trampa de liquidez corresponde a un equilibrio situado en el tramo horizontal de la LM. El término no es de Keynes, sino de Robertson (1940).

[16] Entre las soluciones propuestas figuran la política fiscal expansiva (Keynes), la elevación de las expectativas de inflación (Krugman, 1998) y, en una línea neofisheriana minoritaria, la subida de los tipos nominales para elevar la inflación esperada.

▪ *Agentes*: al inicio del periodo, el individuo recibe su renta real, $Y$, en una cuenta de ahorro remunerada a un tipo de interés nominal constante y exógeno, $i_0$, que refleja el coste de oportunidad (unitario) de no mantener el dinero en la cuenta. A lo largo del periodo realiza retiradas de efectivo de cuantía real constante, $M$[17], y al final del periodo ha consumido toda su renta.

▪ *Tecnología*: existe un coste fijo unitario o coste de transacción por retirada, $b_0 > 0$, constante, exógeno y real.

**Planteamiento**: el individuo elige la cuantía de cada retirada, $M$, que minimiza su coste total:

$$\min_M \text{CT} = C_T + C_O = b_0 \frac{Y}{M} + i_0 \frac{M}{2}$$

donde CT es el coste total, suma de dos componentes:

▪ *Coste de transacción*: $C_T = b_0 Y/M$ es el producto del coste unitario por el número de retiradas, $Y/M$.

▪ *Coste de oportunidad*: $C_O = i_0 M/2$ es el producto del tipo de interés por el saldo promedio de efectivo, $M/2$.

**CPO**: derivando el coste total respecto a la cuantía de cada retirada,

$$\frac{d\text{CT}}{dM} = -b_0 \frac{Y}{M^2} + \frac{i_0}{2} = 0 \quad \Rightarrow \; M^* = \sqrt{2b_0 Y/i_0}$$
$$\Rightarrow \; n^* = Y/M^*$$

donde $M^*$ es la retirada óptima y $n^*$ el número óptimo de retiradas. La retirada óptima iguala el ahorro marginal en costes de transacción al coste marginal de oportunidad.

**CSO**: $\partial^2 \text{CT}/\partial M^2 = 2b_0 Y/M^3 > 0$, es decir, el coste total es estrictamente convexo en $M$, de modo que la CPO define el mínimo.

**Solución**:

▪ *Función de demanda de dinero individual*: es el saldo promedio mantenido en el periodo:

$$\frac{M^d}{P} \equiv \frac{M^*}{2} = \frac{1}{2}\sqrt{\frac{2b_0 Y}{i_0}} \;\rightarrow\; \frac{M^d}{P} = k(b_0^+, i_0^-) \cdot Y^{1/2}$$

donde $k(\cdot)$ recoge el efecto del coste de transacción y del tipo de interés.

▪ *Función de demanda de dinero agregada*: con agentes idénticos, conserva la forma de la individual y sus elasticidades.

**Desarrollo gráfico**: la minimización del coste total se representa en el plano cuantía de cada retirada-costes (figura A35.2):

1. *Coste de transacción*, $C_T = b_0 Y/M$: decrece con la cuantía de cada retirada, porque hacen falta menos retiradas.

2. *Coste de oportunidad*, $C_O = i_0 M/2$: es una recta que crece con la cuantía retirada, con pendiente $i_0/2$.

[17] El gasto se distribuye uniformemente a lo largo del periodo.

3. *Coste total*, CT: suma vertical de ambos, con forma de U.

4. *Óptimo*: en el punto 0, el mínimo del coste total, la retirada óptima, $M^*$, iguala ambos costes. A su izquierda, $C_T > C_O$ y conviene retirar más, y a su derecha, $C_T < C_O$ y conviene retirar menos.

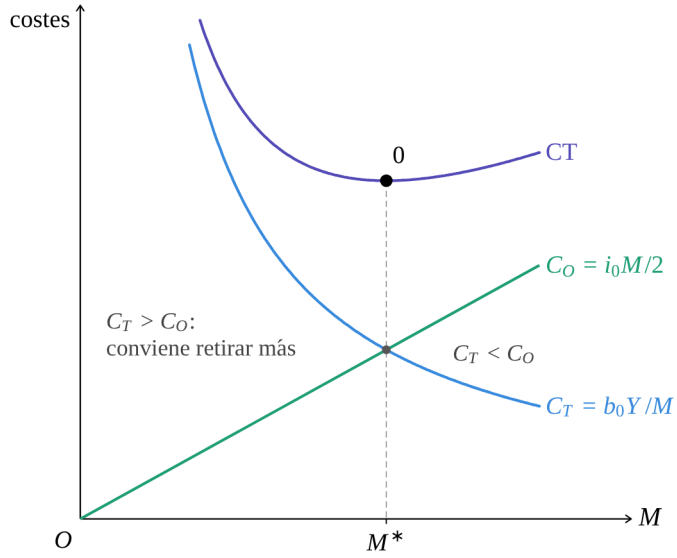



Figura A35.2: Modelo de inventarios (Baumol-Tobin). La retirada óptima, $M^*$, minimiza el coste total, suma del coste de transacción (decreciente) y del coste de oportunidad (creciente), y se sitúa donde ambos se igualan.

**Estática comparativa**: la demanda depende de tres factores:

- *Renta real*: cuanto mayor es la renta, mayor es la demanda de dinero, pero menos que proporcionalmente, pues la elasticidad es $\varepsilon_{M^d Y} = 0{,}5$, lo que refleja economías de escala en la gestión de tesorería.

- *Coste de transacción*: cuanto mayor es el coste unitario de transacción, mayor es la demanda de dinero, para diluir el coste fijo total de transacción.

- *Tipo de interés*: cuanto mayor es el coste de oportunidad del dinero, menor es la demanda, con elasticidad $\varepsilon_{M^d i} = -0{,}5$.

**Extensiones**: con flujos de caja aleatorios, Miller y Orr (1966) proponen una *regla de bandas*. El agente solo ajusta su cartera, hasta un nivel de retorno, cuando el saldo alcanza el límite inferior, que fija él mismo, o el superior[18]. El modelo mantiene las economías de escala y añade que la demanda de dinero aumenta con la incertidumbre de los flujos, lo que enlaza con el motivo precaución.

Cuando hay recursos ociosos, la política monetaria mueve la renta más que en la versión keynesiana del motivo transacción, porque aquí la demanda por ese motivo crece menos que pro-

[18] El límite superior y el nivel de retorno dependen de la varianza de los flujos, del coste de transacción y del tipo de interés.

porcionalmente con la renta. Así, para que el público absorba una misma expansión monetaria manteniéndose fijo el tipo de interés, la renta ha de subir más que con Keynes[19].

**Conclusiones**:

▪ *Aspectos positivos*: microfundamenta el motivo transacción, obtiene elasticidades contrastables y explica las economías de escala en la gestión de tesorería.

▪ *Limitaciones*: supone flujos ciertos y retiradas de cuantía constante y, sobre todo, es una demanda individual, pues con agentes heterogéneos la agregación de raíces cuadradas no es la raíz cuadrada del agregado.

### 1.3.2. El motivo precaución: modelo de costes de ruptura (Whalen, 1966)

**Origen**: Edward Whalen (1966) microfundamentó el motivo precaución con la lógica de inventarios bajo incertidumbre.

**Idea**: se guarda efectivo como reserva porque no se sabe con certeza cuándo entrará y saldrá dinero ni qué gastos imprevistos surgirán. Quedarse sin efectivo tiene un coste, el *coste de ruptura*, y mantenerlo tiene un coste de oportunidad. El saldo precautorio óptimo equilibra ambos.

**Supuestos**:

▪ *Entorno*: marco estático, ahora con incertidumbre. El exceso de gasto sobre la renta de cada mes, $G_t - Y_t$, es una variable aleatoria de media cero y varianza $\sigma^2$.

▪ *Agentes*: en promedio, lo que el individuo ingresa en un número dado de meses, $Y$, coincide con lo que gasta. Reparte esa renta entre efectivo, $M$, con el que cubre los imprevistos sin coste explícito, y bonos o depósitos de ahorro que rinden el tipo nominal $i$.

▪ *Tecnología*: convertir bonos en efectivo conlleva un coste de corretaje unitario, $b > 0$. Si en un mes el exceso de gasto no supera el saldo mantenido, $G_t - Y_t \leq M_t$, no hay que recurrir a los bonos. Si lo supera, $G_t - Y_t > M_t$, se produce una ruptura del *stock* de efectivo y hay que pagar el corretaje.

**Planteamiento**: el individuo elige el saldo precautorio, $M$, que minimiza su coste total. Acotando la probabilidad de ruptura por la desigualdad de Chebyshev, $\Pr(G - Y > M) \leq \sigma^2/M^2$,

$$\min_M \mathrm{CT} = iM + b\,\frac{\sigma^2}{M^2}$$

donde CT es el coste total, suma del coste de oportunidad de mantener el saldo, $iM$, y del coste esperado de ruptura, producto del corretaje, $b$, por la cota de la probabilidad de ruptura.

**CPO**: derivando respecto al saldo precautorio,

$$\frac{\partial \mathrm{CT}}{\partial M} = i - \frac{2b\sigma^2}{M^3} = 0 \;\Rightarrow\; M^* = \left(\frac{2b\sigma^2}{i}\right)^{1/3}$$

[19] Con precios normalizados, la demanda keynesiana por motivo transacción es $M^d = kY$ y la de Baumol-Tobin, $M^d = kY^{1/2}$, de modo que la elasticidad de la renta respecto a la oferta monetaria que la absorbe es 1 con Keynes y 2 con Baumol-Tobin. Por tanto, un aumento del 1 % de la oferta monetaria exige un aumento del 2 % de la renta.

Cada unidad más de efectivo ahorra, en valor esperado, una parte de los gastos de corretaje (beneficio marginal), pero obliga a renunciar al interés que habría rendido en bonos (coste marginal), y el óptimo se alcanza cuando ambos se igualan.

**CSO**: $\partial^2 \mathrm{CT}/\partial M^2 = 6b\sigma^2/M^4 > 0$, por lo que el coste total es convexo y la CPO define el mínimo.

**Solución**:

- *Función de demanda de dinero individual*: el saldo precautorio óptimo es la demanda de dinero por motivo precaución:

$$\frac{M^d}{P} = f(b^+, i^-, \sigma^+) = f(b^+, i^-, Y^+)$$

donde la renta entra a través de la dispersión entre ingresos y gastos, $\sigma$, que crece con el nivel de actividad. Whalen trata el exceso de gasto como la suma de muchos cobros y pagos independientes. Si la renta crece porque aumenta el número de pagos, de igual cuantía, $\sigma^2$ crece en proporción a $Y$ y $M^* \propto Y^{1/3}$. Si, en cambio, crece porque aumenta la cuantía de cada pago, $\sigma$ crece en proporción a $Y$ y $M^* \propto Y^{2/3}$. La elasticidad-renta de la demanda precautoria está, así, entre 1/3 y 2/3, con economías de escala como en el modelo de inventarios.

- *Función de demanda de dinero agregada*: como en el modelo de inventarios, conserva la forma de la individual solo con agentes idénticos.

**Desarrollo gráfico**: la demanda de dinero por motivo precaución se representa en el plano saldo precautorio-coste y beneficio marginales (figura A35.3):

1. *Beneficio marginal*, $\mathrm{Bmg}(M)$: decreciente, porque cada unidad más de efectivo reduce cada vez menos el riesgo de ruptura.

2. *Coste marginal*, $\mathrm{Cmg}(M) = i$: es el interés que se deja de percibir, horizontal porque no depende del saldo.

3. *Óptimo*: en el punto 0, donde ambas curvas se cortan, el saldo precautorio es $M^*$.

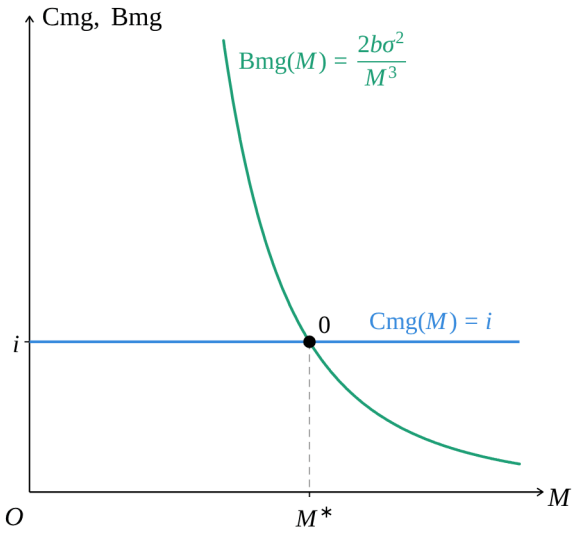



Figura A35.3: Modelo de Whalen (1966). El saldo precautorio óptimo, $M^*$, iguala el beneficio marginal (ahorro esperado en corretaje) al coste marginal, el tipo de interés $i$.

**Estática comparativa**: los signos son los de la función de demanda. Con más renta, los desajustes entre cobros y pagos son mayores, de modo que la curva de beneficio marginal se desplaza hacia arriba y se demanda más dinero.

**Extensiones**: la demanda de dinero puede ser inestable porque depende de la incertidumbre percibida, $\sigma$, que cambia con las expectativas y el entorno. En consecuencia, cualquier medida de política económica que estabilice el entorno reduce la liquidez deseada.

**Conclusiones**:

- *Aspectos positivos*: microfundamenta el motivo precaución y explica que la demanda de dinero aumente con la incertidumbre sobre ingresos y gastos.

- *Limitaciones*: la magnitud del efecto de la renta depende de cómo crezcan los cobros y pagos, lo que resta capacidad predictiva.

### 1.3.3. El motivo especulación: modelo de carteras mixtas (Tobin, 1958)

**Origen**: James Tobin (1958; Premio Nobel 1981) aplicó la teoría de selección de cartera de Markowitz a la demanda de dinero.

**Idea**: la demanda de dinero resulta de un problema de selección de cartera bajo riesgo. Frente a la disyuntiva keynesiana entre bonos o dinero, los individuos mantienen su riqueza simultáneamente en ambos. La aversión al riesgo basta para explicar la preferencia por la liquidez y, si domina el efecto sustitución, la relación negativa entre la demanda de dinero y el tipo de interés.

**Supuestos**:

- *Entorno*: contexto de riesgo, en el que la rentabilidad de los activos es una variable aleatoria. Los agentes cuentan con una riqueza inicial que deben asignar entre dos activos:

- *Bono con riesgo* (p. ej. riesgo de impago o de mercado): con una rentabilidad esperada[20] $R_B^e = r_B + G_B^e$ y un riesgo $\sigma_B > 0$.
- *Activo sin riesgo* (es decir, dinero): con rentabilidad esperada $R_A^e = 0$ y riesgo $\sigma_A = 0$.

▪ *Agentes*: son aversos al riesgo, es decir, la rentabilidad esperada es un bien ($\partial u/\partial R_c^e > 0$) y el riesgo un mal ($\partial u/\partial \sigma_c < 0$), de modo que sus curvas de indiferencia en el plano riesgo-rentabilidad tienen pendiente positiva y son convexas.

**Planteamiento**: el individuo elige la proporción de su riqueza que coloca en bonos, $w_B$, y el resto, $1 - w_B$, en dinero, maximizando la utilidad esperada de la rentabilidad y el riesgo de la cartera resultante. La rentabilidad esperada y el riesgo de una cartera mixta son:

$$R_c^e = w_B r_B + (1 - w_B) \underbrace{R_A^e}_{0} = w_B r_B, \quad w_B \in [0, 1] \tag{1}$$

$$\sigma_c^2 = w_B^2 \sigma_B^2 + (1 - w_B)^2 \underbrace{\sigma_A^2}_{0} + 2 w_B (1 - w_B) \sigma_B \underbrace{\sigma_A}_{0} \rho_{BA} = w_B^2 \sigma_B^2 \;\rightarrow\; \sigma_c = w_B \sigma_B \tag{2}$$

donde $\rho_{BA}$ es la correlación entre ambos activos. De (2), la proporción de bonos es $w_B = \sigma_c/\sigma_B$, de modo que cuanto mayor es el riesgo de los bonos, menor es su peso para un riesgo de cartera dado. Sustituyendo en (1) se obtiene la restricción del individuo, la *recta de oportunidades de inversión* o *frontera eficiente*[21]:

$$R_c^e = \frac{r_B}{\sigma_B}\,\sigma_c, \qquad \frac{\partial R_c^e}{\partial \sigma_c} = \frac{r_B}{\sigma_B} > 0$$

que muestra las combinaciones rentabilidad-riesgo alcanzables, con la máxima rentabilidad esperada para cada riesgo. El programa es:

$$\max_{w_B} u^e = f(R_c^e, \sigma_c) \quad \text{s.a.:} \quad R_c^e = \frac{r_B}{\sigma_B}\,\sigma_c$$

donde $u^e$ es la utilidad esperada de la cartera, creciente en su rentabilidad esperada, $R_c^e$, y decreciente en su riesgo, $\sigma_c$, por la aversión al riesgo.

**CPO**: la pendiente de la curva de indiferencia, la rentabilidad adicional que el agente exige por cada unidad más de riesgo, iguala a la de la recta de oportunidades, la rentabilidad de los bonos por unidad de riesgo:

$$\left.\frac{dR_c^e}{d\sigma_c}\right|_{du^e=0} = -\frac{\partial u^e/\partial \sigma_c}{\partial u^e/\partial R_c^e} = \frac{r_B}{\sigma_B}$$

La condición describe la tangencia entre la curva de indiferencia y la recta de oportunidades, en la que el individuo acepta riesgo adicional hasta que la rentabilidad extra que exige por él iguala a la que el mercado ofrece.

**CSO**: la convexidad estricta de las curvas de indiferencia (aversión al riesgo) y la linealidad de la recta de oportunidades garantizan que la tangencia es un máximo y que la cartera óptima es única.

[20] Se suponen nulas las ganancias o pérdidas de capital esperadas, $G_B^e = (P_{B,t+1}^e - P_{B,t})/P_{B,t} = 0$, aunque son inciertas, $G_B \sim N(0, \sigma_B^2)$. De ellas procede el riesgo del bono, y por eso $R_c^e = w_B r_B$.

[21] No debe confundirse con la recta del mercado de capitales (CML), porque aquí se considera la decisión individual.

**Solución**:

▪ *Función de demanda de dinero individual*: de la CPO se obtiene $w_B^*$ y, con él, la demanda (real) especulativa de dinero, $(1 - w_B^*)W$, que depende del tipo de interés, del riesgo del bono, de la riqueza del individuo y de su grado de aversión al riesgo:

$$\frac{M^d}{P} = f(W^+, \sigma_B^+, r_B^\pm)$$

donde $W$ es la riqueza real actualizada del individuo.

▪ *Función de demanda de dinero agregada*: con agentes homogéneos, como supone Tobin, hereda los signos de la individual.

**Desarrollo gráfico**: la selección de cartera se representa en el plano riesgo-rentabilidad esperada, con la proporción de bonos en el semieje inferior (figura A35.4):

1. *Recta de oportunidades*, $R_c^e = (r_B/\sigma_B)\,\sigma_c$ (cuadro superior): sale del origen, la cartera solo de dinero, con pendiente $r_B/\sigma_B$, y llega a $\sigma_B$, la cartera solo de bonos.

2. *Curva de indiferencia*, $u_0^e$ (cuadro superior): es creciente y convexa por la aversión al riesgo.

3. *Recta de proporciones*, $\sigma_c = w_B\,\sigma_B$ (cuadro inferior): traduce cada nivel de riesgo en la proporción de bonos, $w_B$, y de dinero, $1 - w_B$, que lo genera.

4. *Óptimo*: en el punto 0, la tangencia entre la recta de oportunidades y la curva de indiferencia más alta, el riesgo es $\sigma_c^*$ y la rentabilidad, $R_c^{e*}$. En el cuadro inferior, ese riesgo corresponde a la proporción de bonos $w_B^*$, y el resto, $1 - w_B^*$, se mantiene en dinero.

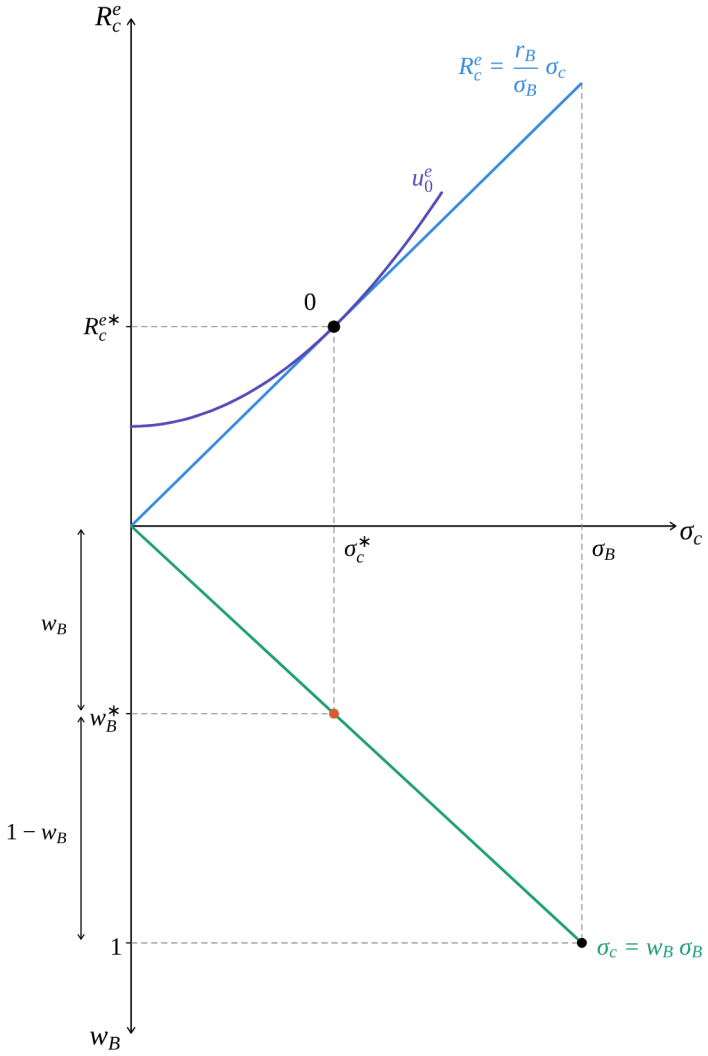



Figura A35.4: Modelo de carteras de Tobin (1958). El inversor averso al riesgo reparte su riqueza entre dinero y bonos sobre la línea de oportunidades rentabilidad-riesgo.

**Estática comparativa**: la demanda de dinero responde a tres variables:

▪ *Riqueza* ($\partial(M^d/P)/\partial W > 0$): si aumenta la riqueza real actualizada, aumenta el límite de gasto del agente y, con ello, su demanda de dinero.

▪ *Riesgo de los bonos* ($\partial(M^d/P)/\partial \sigma_B > 0$): si aumenta el riesgo de los bonos sin aumentar su rentabilidad, la aversión al riesgo lleva a demandar más dinero para evitar la exposición.

▪ *Tipo de interés* ($\partial(M^d/P)/\partial r_B = ?$): el efecto es incierto, pues se descompone en un efecto sustitución y un efecto renta.

- *Efecto sustitución* (ES $< 0$): si sube el tipo de interés, mantener dinero tiene un coste de oportunidad mayor, y a utilidad constante el agente traslada parte de su cartera del dinero a los bonos.

- *Efecto renta* (ER $> 0$): la subida del tipo eleva la rentabilidad esperada de la cartera, de modo que el inversor puede obtener la misma rentabilidad asumiendo menos riesgo, lo que

tiende a aumentar la proporción de dinero. Solo si predomina el efecto sustitución cae la demanda de dinero.

**Extensiones**: el tipo de interés influye en la demanda de dinero a través de la composición de la cartera. Por ello, la política monetaria puede servir para estabilizar la economía. A través del motivo especulación, una variación de la cantidad de dinero mueve el tipo de interés, y este, la inversión, con lo que la dicotomía clásica vuelve a romperse.

**Conclusiones**:

- *Aspectos positivos*: explica la tenencia simultánea de dinero y bonos a partir de la aversión al riesgo y desagrega los efectos renta y sustitución del tipo de interés.

- *Limitaciones*:

  - *Riesgo del dinero*: el dinero no está completamente libre de riesgo, pues la inflación socava su valor[22].

  - *Activos sustitutivos*: en muchas economías hay activos sin riesgo (o casi) pero más rentables que el dinero (p. ej. depósitos a plazo o letras del Tesoro), que lo dominan, de modo que, con idénticos supuestos, nadie mantendría dinero como activo de cartera.

  - *Agregados*: su aplicabilidad se limita a agregados monetarios amplios (M2 y M3) y no al *narrow money* (M1).

#### 1.3.4. Conclusiones

**Aspectos positivos**: la síntesis neoclásica microfundamenta uno a uno los motivos keynesianos y obtiene, en los tres casos, una demanda de dinero que depende negativamente del tipo de interés (en Tobin, si domina el efecto sustitución). Recoge, además, que esa demanda puede ser inestable, porque sus parámetros dependen de expectativas y percepciones del riesgo cambiantes, como subrayaban los keynesianos.

**Limitaciones**: los tres modelos son teorías de la demanda individual que no resuelven la agregación, y cada uno explica un motivo por separado, sin integrarlos en una demanda de dinero conjunta.

### 1.4. Monetarismo

El monetarismo aporta a la teoría de la demanda de dinero la reformulación de la teoría cuantitativa de Friedman y su estimación por Cagan en las hiperinflaciones.

#### 1.4.1. El dinero como bien (Friedman, 1956)

**Origen**: Milton Friedman (1956) reformula la teoría cuantitativa como una teoría de la demanda de dinero.

**Idea**: el dinero es un bien más, cuya demanda, como la de cualquier bien, depende de la renta, de su precio relativo (la rentabilidad de los activos alternativos) y de las preferencias. El resultado es una función de demanda de dinero estable y poco sensible al tipo de interés.

**Supuestos**:

[22] Por ello, un activo financiero más seguro sería un bono indexado a la inflación.

▪ *Entorno*: los agentes eligen entre una amplia gama de activos, financieros y reales. Al ser un bien más, el dinero tiene muchos sustitutivos imperfectos, bonos y acciones, bienes de consumo duradero e incluso activos inmobiliarios[23], a diferencia de Keynes, para quien el dinero tiene pocos pero buenos sustitutivos.

▪ *Agentes*: el dinero es un bien de capital más, demandado por los servicios que presta. Las empresas lo utilizan como factor de producción, y los consumidores como fuente de poder adquisitivo general. Tiene, por tanto, un rendimiento en forma de servicios y, en su caso, el interés de los depósitos. Las decisiones de consumo y de cartera se toman conjuntamente, a diferencia de Keynes y la síntesis neoclásica, donde están separadas. Friedman acepta que puedan existir distintos motivos para demandar dinero, pero no segmenta la demanda en componentes.

**Planteamiento**: como Tobin, Friedman obtiene la demanda de dinero de una decisión de cartera, en la que el agente reparte su riqueza total, humana y no humana, entre dinero y los activos alternativos según sus rentabilidades relativas y sus preferencias. Sin embargo, Friedman no resuelve el programa explícitamente, sino que postula la forma de la función resultante.

**Solución**:

▪ *Función de demanda de dinero individual*[24]:

$$\frac{M^d}{P} = f((Y^P)^+, (W_H/W)^+, r_B^-, r_E^-, \pi^{e-})$$

donde $Y^P$ es la renta permanente, $W_H/W$ la riqueza humana en relación con la riqueza total, $r_B$ y $r_E$ las rentabilidades (nominales) de bonos y acciones, y $\pi^e$ la inflación esperada. Las preferencias, p. ej. por renta fija o variable según el perfil de riesgo, quedan recogidas en la forma de $f$.

▪ *Función de demanda de dinero agregada*: agrupando las rentabilidades de los distintos activos y la riqueza,

$$\frac{M^d}{P} = k(r^-, \pi^{e-}) \cdot Y^P$$

que muestra que la teoría monetarista es una reformulación de la ecuación de Cambridge en la que $k$ depende de la rentabilidad de los activos financieros y reales, y la variable de escala es la renta permanente.

De la función se siguen dos resultados principales. Primero, como el dinero tiene muchos sustitutivos y la demanda depende de la renta permanente, la sensibilidad de la demanda al tipo de interés es baja ($h \approx 0$) y la demanda es muy estable. Segundo, la transmisión monetarista descansa en el reajuste de las carteras entre el dinero y una gama amplia de activos financieros y reales, y la inflación esperada interviene en ella como el rendimiento de los bienes frente al dinero, es decir, como un coste de mantener saldos.

**Estática comparativa**: los determinantes se interpretan como sigue:

▪ *Renta permanente* ($\partial(M^d/P)/\partial Y^P > 0$): la riqueza real anualizada limita la cantidad de dinero que se puede demandar. A mayor riqueza, mayor es la demanda, p. ej. como depósito de valor.

[23] En estos bienes, el rendimiento está implícito y es en especie.

[24] La especificación original de Friedman incluye también la rentabilidad (nominal) del dinero, $r_m$ (p. ej. la remuneración de los depósitos), con $\partial(M^d/P)/\partial r_m > 0$, y un parámetro $u$ que recoge los gustos y preferencias.

- *Riqueza humana* ($\partial(M^d/P)/\partial(W_H/W) > 0$): cuanto mayor es la proporción de riqueza humana, que capitaliza la renta laboral, sobre la riqueza total, mayor es la demanda de dinero, pues la riqueza humana es menos líquida que los activos financieros y su liquidez es más incierta[25].

- *Rentabilidad de los activos* ($\partial(M^d/P)/\partial r_i < 0$): cuanto mayor es la rentabilidad de bonos y acciones, mayor es el coste de oportunidad del dinero.

- *Inflación esperada* ($\partial(M^d/P)/\partial\pi^e < 0$): es la rentabilidad nominal esperada de los activos reales. Por ello, si los individuos esperan inflación, sustituyen dinero (real) por activos reales que no pierden valor (p. ej. oro o inmuebles). La demanda nominal, en cambio, sube con el nivel de precios para mantener el *stock* real deseado ($\partial M^d/\partial P > 0$).

**Extensiones**: en el paradigma monetarista no se dan los "dos eslabones débiles", porque la inversión es muy sensible al tipo de interés, la demanda de dinero poco, y el *efecto liquidez* es elevado.

- *Política monetaria*: sí es efectiva sobre la renta a corto plazo, por el elevado efecto liquidez y el efecto renta[26], lo que constituye la principal diferencia con los neoclásicos. Los monetaristas se oponen a la neutralidad del dinero a corto plazo, y los cambios en la cantidad de dinero explican los ciclos.

No obstante, los monetaristas rechazan el uso discrecional de la política monetaria. Defienden una política monetaria sujeta a reglas[27] y orientada a la inflación, y lo justifican con la ecuación perdida del monetarismo, con unos retardos largos y variables y con su inefectividad a largo plazo (tema A05, apartados 2.2.4 y 2.2.5).

- *Política fiscal*: no es efectiva para cambiar la renta, por el efecto expulsión (*crowding out*) total que provoca una demanda de dinero casi insensible al tipo de interés ($h \approx 0$)[28].

**Conclusiones**:

- *Aspectos positivos*: recupera la estabilidad de la función de demanda de dinero con un fundamento microeconómico, la elección de cartera, y hace de ella el pilar del mecanismo de transmisión monetarista.

- *Limitaciones*: no deriva la función de un programa explícito, sino que postula sus argumentos. Además, como la teoría de carteras de Tobin, se ajusta mejor a agregados monetarios amplios.

### 1.4.2. Modelo de demanda de dinero e hiperinflación (Cagan, 1956)

> El modelo de Cagan y su condición de estabilidad se desarrollan en el tema A41, apartado 3.1.2, y el señoraje óptimo en el tema A39, apartado 2.2.2. Aquí se recoge su aportación a la teoría de la demanda de dinero.

**Origen**: Phillip Cagan (1956), discípulo de Friedman, propuso el modelo.

[25] La riqueza humana solo se convierte en liquidez a medida que se cobra el salario, y con incertidumbre si el contrato no es indefinido. Las acciones, en cambio, pueden liquidarse de inmediato.

[26] $\uparrow M^s \rightarrow\downarrow i \rightarrow\uparrow I \rightarrow\uparrow Y$. Con una demanda de dinero poco sensible al tipo de interés, la expansión monetaria reduce con fuerza el tipo (efecto liquidez elevado) y estimula una inversión muy sensible a él (efecto renta elevado).

[27] Friedman defendió dos reglas: un crecimiento constante de la masa monetaria y la regla de inflación óptima (Friedman, 1969), que anula el coste de oportunidad de mantener dinero ($i = 0$) con una inflación $\pi = -r$.

[28] $\uparrow G \rightarrow\uparrow Y \rightarrow\uparrow M^d \rightarrow\uparrow r \rightarrow\downarrow I$. Con una IS elástica y una LM inelástica, la fuerte subida del tipo de interés expulsa una inversión muy sensible a él.

**Idea**: Cagan estimó para siete hiperinflaciones europeas del siglo XX la función de demanda de dinero monetarista reducida a su único argumento relevante en ese contexto, la inflación esperada.

**Supuestos**:

▪ *Entorno*: hiperinflación, en la que la renta y el tipo de interés real apenas varían frente a las variables nominales[29].

▪ *Agentes*: forman sus expectativas de inflación de forma adaptativa, $\pi_t^e - \pi_{t-1}^e = \lambda(\pi_{t-1} - \pi_{t-1}^e)$, con $\lambda \in (0, 1]$. Su demanda de dinero es exponencial en la inflación esperada.

**Planteamiento**: no hay programa de optimización. Cagan postula la función y la estima:

$$\frac{M^d}{P} = \gamma \cdot e^{-a\pi^e}$$

donde $\gamma > 0$ recoge los factores reales y $a > 0$ es la semielasticidad de la demanda de dinero respecto a la inflación esperada.

**Desarrollo gráfico**: el modelo de Cagan se representa en el plano saldos reales-inflación esperada (panel a) y en el plano inflación-señoraje (panel b) de la figura A35.5:

1. *Demanda de dinero*, $M^d/P = \gamma e^{-a\pi^e}$ (panel a): es decreciente y convexa. Corta el eje horizontal en $\gamma$, con inflación esperada nula, y se acerca al eje vertical a medida que la inflación esperada crece.

2. *Elasticidad unitaria* (panel a): la elasticidad en valor absoluto, $|\varepsilon| = a\pi^e$, vale uno en $\pi^e = 1/a$. Por debajo de ese nivel la demanda es inelástica y, por encima, elástica.

3. *Señoraje de estado estacionario*, $\pi\gamma e^{-a\pi}$ (panel b): con $\pi^e = \pi$, es el producto de la inflación por los saldos reales y tiene forma de curva de Laffer.

4. *Máximo del señoraje* (panel b): se alcanza en $\pi = 1/a$, el mismo nivel en que la demanda tiene elasticidad unitaria.

**Estática comparativa**: con semielasticidad constante, la elasticidad, $a\,\pi^e$, crece con la propia inflación. Una subida de la inflación esperada de $\pi_0^e$ a $\pi_1^e$ lleva la economía del punto 0 al 1 a lo largo de la demanda (panel a). Una vez superado $1/a$, los saldos reales caen más que proporcionalmente, lo que explica la huida del dinero en las fases finales de una hiperinflación, y el señoraje disminuye (panel b).

[29] Se parte de $m_t - p_t = \alpha_0 + \alpha_1 y_t - a\, i_t + \xi_t$, con $y_t$ el logaritmo de la renta real, $\alpha_0, \alpha_1, a > 0$ e $i_t = r_t + \pi_{t+1}^e$, de modo que $m_t - p_t = \alpha_0 + \alpha_1 y_t - a\, r_t - a\, \pi_{t+1}^e + \xi_t$. Con $y_t$ y $r_t$ constantes se obtiene la función de Cagan.

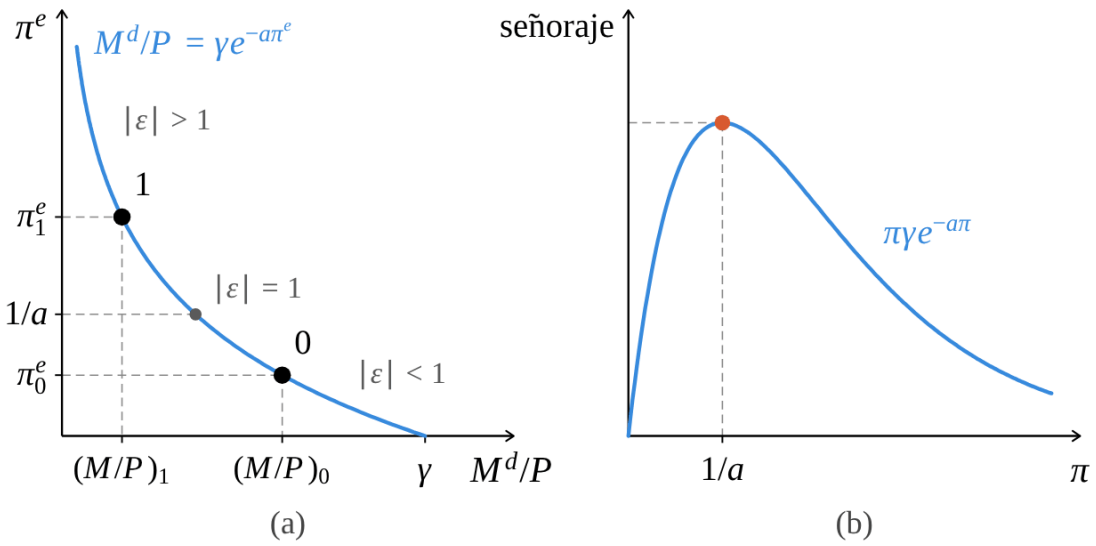



Figura A35.5: Modelo de Cagan (1956). La demanda de saldos reales cae exponencialmente con la inflación esperada y es elástica por encima de $\pi^e = 1/a$ (a). El señoraje de estado estacionario tiene forma de curva de Laffer, con máximo en $\pi = 1/a$ (b).

**Extensiones**: la inflación converge al estado estacionario si $a\lambda < 1$ y se alimenta a sí misma si $a\lambda > 1$. Además, el máximo del señoraje en $\pi = 1/a$ acota lo que el Gobierno puede recaudar de forma permanente emitiendo dinero.

**Conclusiones**:

- *Aspectos positivos*: es la primera estimación econométrica de la semielasticidad de la demanda de dinero respecto a la inflación esperada y prueba que la función es estable incluso en hiperinflación, lo que apoya la tesis monetarista.

- *Limitaciones*: la posible inestabilidad del estado estacionario procede de las expectativas adaptativas, y con expectativas racionales el resultado cambia. Además, la función solo vale cuando los argumentos reales pueden ignorarse.

### 1.4.3. Conclusiones

**Aspectos positivos**: el monetarismo recupera una demanda de dinero estable y poco sensible al tipo de interés con un fundamento de elección de cartera, y hace de ella el pilar de su mecanismo de transmisión. La estimación de Cagan confirma que la función es estable incluso en hiperinflación.

**Limitaciones**: ni Friedman ni Cagan derivan la función de un programa de optimización, sino que postulan sus argumentos. Además, la función de Friedman, como la teoría de carteras de Tobin, se ajusta mejor a agregados monetarios amplios.

## 1.5. Conclusiones

Las teorías de la sección se ordenan en torno a la sensibilidad de la demanda de dinero al tipo de interés ($h$) y su estabilidad, las dos propiedades que determinan la eficacia de las políticas. Junto con la sensibilidad de la inversión al tipo de interés ($v$ en $I = I_0 - v\,r$, que el tema A34 estudia como $I'(r) < 0$), la sensibilidad de la demanda de dinero configura los "dos eslabones" del mecanismo de transmisión.

En el caso keynesiano, los dos eslabones débiles ($v$ baja, $h$ alta) hacen que una política fiscal expansiva eleve la renta sin apenas desplazar la inversión, mientras que una expansión monetaria apenas reduce el tipo de interés y estimula poco la inversión. En el caso monetarista ocurre lo contrario.

Tabla A35.1: Demanda de dinero y eficacia de las políticas según el paradigma.

| Paradigma | Determinantes de $M^d/P$ | $h$ | Estabilidad | Política monetaria | Política fiscal |
|---|---|---|---|---|---|
| Clásicos y neoclásicos | Transacciones o renta ($V$, $k$ constantes) | $\approx 0$ | Muy estable | Potente sobre precios, neutral sobre la renta | Ineficaz (efecto expulsión total) |
| Keynes | Renta y tipo de interés (tres motivos) | Alta | Inestable (expectativas) | Poco eficaz (eslabones débiles, trampa de liquidez) | Muy eficaz con la renta |
| Síntesis neoclásica | Renta, tipo de interés, costes de transacción, riesgo | Media | Inestable (expectativas, incertidumbre) | Eficaz vía tipo de interés | Eficaz |
| Monetarismo | Renta permanente y rentabilidades de los activos | Baja | Estable | Potente a corto plazo, neutral a largo | Ineficaz (efecto expulsión total) |

# 2. Desarrollos posteriores

La investigación reciente ha querido dar a los modelos macroeconómicos una base microeconómica más sólida, con agentes que optimizan a lo largo del tiempo en equilibrio general. Hay tres líneas:

- *Generaciones solapadas*: analizan el dinero como un activo con el que desplazar recursos en el tiempo.
- *Dinero en la utilidad, tiempo de compra y efectivo por adelantado*: toman como dado el motivo para demandar dinero y se centran en la política monetaria.
- *Otras aportaciones*: extienden el análisis a la búsqueda del intercambio y a la economía abierta.

## 2.1. Modelos de generaciones solapadas (Allais, 1947; Samuelson, 1958; Diamond, 1965)

La primera línea añade al modelo de generaciones solapadas un activo sin valor intrínseco, el dinero fiduciario, que sirve de depósito de valor.

> El modelo de generaciones solapadas se desarrolla en el tema A29, apartado 3.2. Aquí se recoge su aportación al estudio del papel del dinero.

**Origen**: Maurice Allais (1947), Paul Samuelson (1958) y Peter Diamond (1965) desarrollaron el modelo.

**Idea**: en una economía de generaciones solapadas sin capital, los jóvenes quieren trasladar consumo a la vejez, pero no hay activo en que hacerlo. El dinero fiduciario, un activo sin valor intrínseco, resuelve el problema si cada generación espera que la siguiente lo acepte. Samuelson concibe así el dinero como depósito de valor y como institución social.

### 2.1.1. Supuestos

**Entorno**: el tiempo es discreto y el horizonte, infinito. En cada periodo nace una generación que vive dos periodos, joven y anciano, y la población crece a la tasa $n$. No hay capital ni otro activo que permita almacenar el bien. Existe un *stock* nominal de dinero fiduciario, $M_t^s$, sin valor intrínseco, que crece a la tasa $g_M$, en manos de los ancianos al inicio de cada periodo. Su precio en bienes es $1/P_t$.

**Agentes**: cada individuo recibe una dotación del bien perecedero, $y_1$ de joven e $y_2$ de anciano, con $y_1 > y_2$, y maximiza $u(c_{1t}, c_{2t+1})$, creciente y cóncava.

### 2.1.2. Modelo

**Planteamiento**: el joven vende parte de su dotación a los ancianos a cambio de dinero, que gastará cuando sea anciano. Elige sus saldos reales, $m_t = M_t^d/P_t$:

$$\max_{m_t} u(c_{1t}, c_{2t+1}) \quad \text{s.a.:} \begin{cases} c_{1t} = y_1 - m_t \\ c_{2t+1} = y_2 + m_t \frac{P_t}{P_{t+1}} \end{cases}$$

donde $P_t/P_{t+1} = 1/(1+\pi_{t+1})$ es la rentabilidad bruta real del dinero, es decir, el bien que se recupera mañana por cada unidad de bien entregada hoy.

**CPO**: derivando respecto a los saldos reales,

$$\frac{\partial u/\partial c_{1t}}{\partial u/\partial c_{2t+1}} = \frac{P_t}{P_{t+1}} = \frac{1}{1+\pi_{t+1}}$$

Se obtiene una ecuación de Euler con dinero como único activo, de modo que la relación marginal de sustitución entre consumo joven y anciano iguala a la rentabilidad real del dinero.

**CSO**: la concavidad de $u$ y la linealidad de las restricciones garantizan que la CPO define el máximo.

**Solución**:

- *Función de demanda de dinero individual*: de la CPO, $m_t^d = f(\pi_{t+1}^{e-}, \rho^-, y_1^+, y_2^-)$, con $\rho$ la tasa de preferencia temporal. Los saldos reales demandados de joven caen con la inflación esperada, que erosiona su rentabilidad (si, como se supone, el efecto sustitución domina al efecto renta), y con la impaciencia. Crecen, en cambio, con la brecha entre la dotación joven y la anciana, que es el motivo para ahorrar.

- *Función de demanda de dinero agregada*: sumando sobre los $N_t = (1+n)^t N_0$ jóvenes, $M_t^d/P_t = N_t\, m_t^d$.

**Equilibrio**: el mercado de dinero se vacía, $M_t^s = N_t\, m_t\, P_t$, de modo que en estado estacionario, con $m$ constante, $1+\pi = (1+g_M)/(1+n)$. La inflación es la diferencia entre el crecimiento del dinero y el de la población, que es el crecimiento de la demanda de dinero.

Pueden darse tres tipos de equilibrio:

- *No monetario*: nadie acepta el dinero y este no tiene valor ($1/P_t = 0$).
- *Monetario estacionario*: los saldos reales por joven son constantes.
- *Continuo de equilibrios no estacionarios*: el valor del dinero decrece hasta anularse.

El dinero tiene, así, valor de equilibrio sin prestar servicios de transacción, solo por su papel intergeneracional, y ese valor es una burbuja racional.

**Estática comparativa**:

- $g_M < n$: el dinero se aprecia y su demanda es alta.
- $g_M > n$: la inflación reduce los saldos reales y, en el límite, la demanda de dinero se anula.
- $g_M = n$: el valor del dinero es constante.

**Extensiones**: en política monetaria, la creación de dinero es un impuesto sobre los saldos de los jóvenes, que reduce su rentabilidad y su demanda. Con un *stock* de dinero constante ($g_M = 0$), la rentabilidad bruta real del dinero es $1+n$ y el dinero sostiene la asignación óptima (regla de oro). Con $g_M > 0$, en cambio, la creación de dinero aleja la economía de la regla de oro, de modo que el dinero no es superneutral en este marco, es decir, su tasa de crecimiento altera la asignación real.

### 2.1.3. Conclusiones

**Aspectos positivos**: explica por qué un activo sin valor intrínseco puede demandarse y tener valor, y fundamenta el dinero como institución social.

**Limitaciones**: no explica por qué se demanda dinero cuando existen otros activos con rendimiento positivo, que es lo que los modelos siguientes abordan, y el equilibrio monetario es frágil, pues coexiste con el no monetario.

## 2.2. Dinero en la utilidad y sus microfundamentos

Los tres modelos de este apartado se ordenan de menor a mayor exigencia sobre el papel del dinero. En el de dinero en la utilidad (*money in the utility*, MIU), el dinero da utilidad. En el de tiempo de compra (*shopping time*, ST), ahorra tiempo que se convierte en ocio, lo que justifica el MIU. En el de efectivo por adelantado (*cash-in-advance*, CIA), por último, sin dinero no hay compra.

### 2.2.1. Modelos MIU (Sidrauski, 1967)

**Origen**: Miguel Sidrauski (1967) introdujo el dinero en la función de utilidad de un modelo de crecimiento óptimo.

**Idea**: Sidrauski quiere saber si la inflación altera la acumulación de capital, y para ello construye un modelo dinámico en el que los saldos reales dan utilidad. La demanda de saldos reales resulta del programa intertemporal del consumidor, y el resultado central es la superneutralidad del dinero.

**Supuestos**:

- *Entorno*: tiempo discreto y horizonte infinito. El consumidor es impaciente y descuenta sus flujos con un factor de descuento constante, $\beta \in (0,1)$.

- *Agentes*: los individuos derivan utilidad directamente de mantener dinero, porque provee servicios de liquidez. El consumo es insaciable y el dinero tiene un nivel de saciedad, ambos con utilidades marginales decrecientes, y la función de utilidad es aditivamente separable.

- *Restricciones*: el agente dedica su renta laboral, más el principal y los intereses de sus bonos, a consumo, dinero y bonos.

**Planteamiento**: el consumidor maximiza la utilidad descontada de su senda de consumo y saldos reales eligiendo dinero y bonos, sujeto a su restricción presupuestaria:

$$\max_{\{m_t, b_t\}} U = \sum_{t=0}^{\infty} \beta^t u(c_t, m_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} c_t = w_t + (1+r)b_{t-1} - b_t + \frac{1}{1+\pi_t} m_{t-1} - m_t \\ b_{-1},\ m_{-1} \text{ dados} \\ \lim_{T\to\infty} b_T/(1+r)^T \geq 0 \end{cases}$$

donde $c_t$ es el consumo, $m_t$ los saldos reales, $w_t$ la renta laboral real, $b_t$ los bonos reales, $r$ el tipo de interés real y $\pi_t$ la inflación, que erosiona los saldos heredados. Los activos iniciales, $b_{-1}$ y $m_{-1}$, están dados, y la condición de no-Ponzi impide endeudarse indefinidamente.

**CPO**: con solución interior, sustituyendo la restricción en la función objetivo y derivando respecto a los saldos reales[30],

$$\underbrace{\frac{\partial u/\partial m_t}{\partial u/\partial c_t}}_{\text{RMS}} = \underbrace{\frac{i}{1+i}}_{\text{coste de oportunidad del dinero}} \tag{3}$$

donde $i$ es el tipo de interés nominal. Según la condición (3), el cociente de utilidades marginales del dinero y del consumo vale $i/(1+i)$, que es lo que cuesta, en consumo de hoy, conservar dinero. Ese coste es el interés nominal, $i$, que se deja de percibir y que se cobraría el periodo siguiente, descontado a hoy por $1+i$.

**CSO**: $u$ es estrictamente cóncava en $(c, m)$ y la restricción, lineal, de modo que, con la condición de transversalidad sobre los activos, la CPO es suficiente.

**Solución**:

- *Función de demanda de dinero individual*: de la CPO se deduce que los saldos reales demandados crecen con el consumo y caen con el tipo de interés nominal:

$$\frac{M^d}{P} = f(i^-, c^+)$$

- *Función de demanda de dinero agregada*: con un agente representativo, hereda la forma de la individual, análoga a la keynesiana con el consumo como variable de escala.

**Equilibrio**: el modelo original de Sidrauski incluye capital físico, y en su estado estacionario el dinero resulta superneutral, es decir, ni el capital per cápita ni la producción dependen del *stock* de dinero ni de su tasa de crecimiento. Tampoco dependen de cuánto valoran los agentes tener saldos. En la transición hacia el estado estacionario, en cambio, la tasa de crecimiento del dinero puede afectar a la acumulación de capital (Fischer, 1979).

**Estática comparativa**: una subida del tipo de interés nominal eleva $i/(1+i)$ y exige una utilidad marginal del dinero mayor, es decir, menos saldos reales. Un aumento del consumo, en cambio, eleva la utilidad marginal del dinero relativa y, con ella, la demanda.

**Extensiones**: en política monetaria, la tasa de crecimiento del dinero solo determina, en el estado estacionario, la inflación y los saldos reales. Su único efecto real es el coste de bienestar de la inflación, que reduce los saldos reales por debajo de su nivel óptimo, el que alcanza la *regla de Friedman* ($i = 0$).

**Conclusiones**:

- *Aspectos positivos*: ofrece una modelización dinámica pionera de la demanda de dinero y el resultado de referencia de superneutralidad, frente al que se miden los modelos posteriores.

- *Limitaciones*: el dinero en la utilidad es un atajo que no explica por qué se demanda, lo que motiva los modelos de tiempo de compra.

[30] Derivando respecto a $m_t$: $-\partial u/\partial c_t + \partial u/\partial m_t + \beta\,(\partial u/\partial c_{t+1})/(1+\pi_{t+1}) = 0$. Con la condición de los bonos, $\partial u/\partial c_t = \beta(1+r)\,\partial u/\partial c_{t+1}$, y $1+i = (1+r)(1+\pi_{t+1})$, se obtiene (3).

### 2.2.2. Modelos ST (McCallum y Goodfriend, 1987)

**Origen**: Bennett McCallum y Marvin Goodfriend (1987) formalizaron el modelo de tiempo de compra (*shopping time*), anticipado por Saving (1971).

**Idea**: el dinero no da utilidad por sí mismo, sino que ahorra tiempo en las transacciones. Como el tiempo ahorrado es ocio, y el ocio sí da utilidad, el dinero entra en la utilidad de forma indirecta, y así el modelo microfundamenta el MIU.

**Supuestos**:

- *Entorno*: tiempo discreto y horizonte infinito, con factor de descuento $\beta \in (0,1)$.

- *Agentes*: la utilidad instantánea depende del consumo y del ocio, $u(c_t, l_t)$, creciente y cóncava en ambos, y el dinero no aparece en ella.

- *Tecnología*: comprar exige tiempo, $s_t = \psi(c_t, m_t)$, creciente en el consumo ($\psi_c > 0$) y decreciente en los saldos reales ($\psi_m < 0$, con $\psi_{mm} > 0$), de modo que más dinero en mano reduce el tiempo de compra, cada vez menos.

- *Restricciones*: la restricción presupuestaria es la del modelo de Sidrauski, salvo que la renta laboral es el salario real, $w_t$, por las horas trabajadas, $h_t$. La dotación de tiempo se reparte entre trabajo, $h_t$, compra y ocio, $l_t = 1 - h_t - \psi(c_t, m_t)$.

**Planteamiento**: el consumidor maximiza la utilidad descontada de consumo y ocio eligiendo consumo, saldos reales y bonos, sujeto a la restricción presupuestaria y a la de tiempo:

$$\max_{\{c_t, m_t, b_t\}} U = \sum_{t=0}^{\infty} \beta^t u(c_t, l_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} c_t = w_t h_t + (1+r) b_{t-1} - b_t + \frac{1}{1+\pi_t} m_{t-1} - m_t \\ l_t = 1 - h_t - \psi(c_t, m_t) \\ b_{-1},\ m_{-1} \text{ dados} \\ \lim_{T\to\infty} b_T/(1+r)^T \geq 0 \end{cases}$$

Sustituyendo la restricción de tiempo en la utilidad se obtiene una utilidad indirecta en consumo y saldos reales, $v(c_t, m_t) \equiv u(c_t, 1 - h_t - \psi(c_t, m_t))$, con $v_m = -u_l\, \psi_m > 0$. Esta utilidad indirecta es la función del modelo MIU, ahora derivada.

**CPO**: derivando respecto a los saldos reales, como en Sidrauski,

$$\frac{v_m}{v_c} = \frac{-u_l\, \psi_m}{u_c - u_l\, \psi_c} = \frac{i}{1+i}$$

El valor del tiempo que ahorra una unidad más de dinero, medido en ocio, $-u_l \psi_m$, relativo al valor neto de una unidad de consumo, iguala al coste de oportunidad del dinero. Es decir, la condición coincide con (3), ahora con contenido, porque la utilidad marginal del dinero es el ocio que compra.

**CSO**: la concavidad de $u$ y la convexidad conjunta de $\psi$ en $(c, m)$ hacen cóncava $v$ en $(c, m)$, y la restricción es lineal, de modo que la CPO es suficiente.

**Solución**:

- *Función de demanda de dinero individual*: de la CPO, $M^d/P = f(i^-, c^+, w^+)$, decreciente en el tipo de interés nominal, creciente en el consumo, que eleva el tiempo de compra, y creciente en el salario real, que es el precio del tiempo ahorrado.

- *Función de demanda de dinero agregada*: con agente representativo, hereda la forma de la individual.

**Estática comparativa**: una subida del tipo de interés reduce los saldos reales, como en Sidrauski. Una mejora de la tecnología de pagos ($\psi_m$ menos negativo), en cambio, reduce el tiempo que ahorra el dinero y, con él, su demanda, lo que ofrece una lectura de la inestabilidad de la función tras la innovación financiera.

**Extensiones**: el modelo da lugar a los dos casos polares. Si $\psi$ no depende de $m$, el dinero no se demanda, y si el consumo exige dinero en proporción fija, se obtiene la restricción de efectivo por adelantado.

En política monetaria, los resultados de superneutralidad del MIU se mantienen si la oferta de trabajo es inelástica (trabajo fijo). Si es endógena, en cambio, la inflación, al reducir los saldos reales y elevar el tiempo de compra, altera el reparto del tiempo entre ocio y trabajo y, con él, la oferta de trabajo. El dinero deja entonces de ser superneutral aunque $u$ sea separable entre consumo y ocio.

**Conclusiones**:

- *Aspectos positivos*: microfundamenta el MIU, conecta los modelos de equilibrio general con los de inventarios de la síntesis neoclásica e introduce el salario como determinante de la demanda de dinero.

- *Limitaciones*: la tecnología de compra $\psi$ es tan *ad hoc* como la utilidad del dinero que sustituye.

### 2.2.3. Modelos CIA (Clower, 1967; Lucas, 1980)

**Origen**: Robert Clower (1967) formuló la restricción de efectivo por adelantado, y Robert Lucas (1980) la integró en un modelo dinámico de equilibrio general.

**Idea**: la crítica a los modelos MIU, que el de tiempo de compra solo atenúa, es que el dinero es intrínsecamente inútil, pues, a diferencia de los bienes, no se consume, y solo tiene sentido como medio para intercambiarlos. Los modelos *cash-in-advance* llevan el tiempo de compra al límite, ya que sin dinero no se compra, de modo que la utilidad del dinero se sustituye por una restricción.

**Supuestos**:

- *Entorno*: tiempo discreto y horizonte infinito, con factor de descuento $\beta \in (0, 1)$. Dentro de cada periodo, el mercado de bienes abre antes que el de activos, de modo que las compras se pagan con el dinero que se tiene al entrar en el periodo.

- *Agentes*: la utilidad instantánea depende solo del consumo, $u(c_t)$, creciente y cóncava, y el dinero no reporta utilidad ni ahorra tiempo.

- *Restricciones*: la restricción presupuestaria es la del modelo de Sidrauski. Además, el consumo del periodo no puede exceder los saldos reales heredados, $c_t \leq m_{t-1}/(1 + \pi_t)$ (*restricción de Clower*): "los bienes compran dinero y el dinero compra bienes, pero los bienes no compran bienes".

**Planteamiento**: el consumidor maximiza la utilidad descontada del consumo eligiendo consumo, saldos reales y bonos, sujeto a la restricción presupuestaria y a la de efectivo por adelantado:

$$\max_{\{c_t, m_t, b_t\}} U = \sum_{t=0}^{\infty} \beta^t u(c_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} c_t + m_t + b_t = w_t + (1+r)b_{t-1} + \frac{m_{t-1}}{1+\pi_t} \\ c_t \leq \frac{m_{t-1}}{1+\pi_t} \\ b_{-1},\ m_{-1} \text{ dados} \\ \lim_{T\to\infty} b_T/(1+r)^T \geq 0 \end{cases}$$

donde las variables y las condiciones inicial y de no-Ponzi son las del modelo de Sidrauski, y la segunda restricción, la de Clower, impide consumir más que los saldos reales heredados.

**CPO**: con $\lambda_t$ el multiplicador de la restricción presupuestaria y $\mu_t$ el de la restricción de efectivo, las condiciones respecto a consumo, bonos y dinero son $du/dc_t = \lambda_t + \mu_t$, $\lambda_t = \beta(1+r)\,\lambda_{t+1}$ y $\lambda_t = \beta\,\dfrac{\lambda_{t+1} + \mu_{t+1}}{1+\pi_{t+1}}$.

Combinándolas, $\mu_{t+1}/\lambda_{t+1} = i_{t+1}$, es decir, el valor de relajar la restricción de efectivo es el tipo de interés nominal, de modo que si $i > 0$ la restricción se cumple con igualdad. El resultado es la contrapartida de la condición (3) del MIU, pues $\mu/(du/dc) = i/(1+i)$. Aquí el dinero no da utilidad, pero mantenerlo cuesta $i$ y es imprescindible para consumir.

**CSO**: la concavidad de $u$ y la linealidad de ambas restricciones garantizan que las condiciones anteriores, con la de transversalidad, son suficientes.

**Solución**:

▪ *Función de demanda de dinero individual*: con la restricción activa, la demanda de dinero es exactamente el valor del consumo del periodo siguiente, $m_t = (1+\pi_{t+1})\,c_{t+1}$, es decir, $M^d/P = c$, con velocidad-renta unitaria y elasticidad-interés nula.

▪ *Función de demanda de dinero agregada*: con agente representativo, $M^d/P = C$, la ecuación cuantitativa con $V = 1$.

**Equilibrio**: en el estado estacionario el dinero es superneutral, como en el MIU.

**Estática comparativa**: la demanda de dinero solo responde al consumo. El tipo de interés nominal no la altera, pero actúa como un impuesto sobre el consumo, pues hay que mantener dinero un periodo para comprar. Una subida de $i$ solo reduce el consumo si existe una alternativa no sujeta a la restricción (ocio, bienes a crédito), y con ocio en la utilidad el impuesto inflacionario distorsiona la elección consumo-ocio.

**Extensiones**: la versión estocástica permite analizar la política monetaria con perturbaciones, y si la restricción se aplica también a la inversión (Stockman, 1981) la inflación reduce el capital y la superneutralidad se rompe.

La política monetaria óptima es, como en el MIU, la regla de Friedman, $i = 0$, con la que la restricción deja de ser costosa.

**Conclusiones**:

▪ *Aspectos positivos*: ofrece una demanda de dinero por motivo transacción sin recurrir a la utilidad del dinero, y una condición de primer orden que se interpreta como coste de oportunidad.

▪ *Limitaciones*: como los MIU, no explica rigurosamente el papel del dinero, pues la restricción se impone en lugar de derivarse. Además, la elasticidad-interés nula contradice la evidencia y el resultado depende del orden de apertura de los mercados dentro del periodo.

#### 2.2.4. Conclusiones

**Aspectos positivos**: los tres modelos derivan del programa intertemporal del agente una función $L(i^-, c^+)$ análoga a la keynesiana (en el CIA, con elasticidad-interés nula), con la que analizan la política monetaria. Su resultado central es la superneutralidad del dinero en el estado estacionario, y la política monetaria óptima, la regla de Friedman ($i = 0$).

**Limitaciones**: los tres modelos toman como dado el motivo para demandar dinero, y la superneutralidad se matiza cuando el dinero entra en la producción, cuando afecta a la oferta de trabajo o cuando existen rigideces nominales. Todos ellos comparten, además, la limitación de la agregación a partir de un agente representativo.

### 2.3. Otras aportaciones

Dos líneas más extienden el análisis a la búsqueda del intercambio y a la economía abierta, con dos monedas en circulación.

#### 2.3.1. Modelos de búsqueda y emparejamiento (Kiyotaki y Wright, 1993)

**Origen**: Nobuhiro Kiyotaki y Randall Wright (1993) trasladaron a la teoría monetaria los modelos de búsqueda y emparejamiento del mercado de trabajo.

**Idea**: se demanda dinero porque abarata la búsqueda de contrapartida. En un equilibrio general sin subastador walrasiano, los agentes se cruzan al azar, y el trueque solo funciona si hay doble coincidencia de deseos. Esa exigencia desaparece con un medio de pago que todos aceptan.

**Razonamiento**: cada individuo cambia lo que produce por bienes o por dinero, que después gastará. Como emparejarse con dinero es más rápido que hacerlo con trueque, que resulta más caro, aparece una demanda de dinero para transacciones y un equilibrio monetario aunque nadie valore el dinero *per se*.

**Extensiones**: Shi (1995) y Trejos y Wright (1995) endogenizaron los precios mediante la negociación bilateral, y Lagos y Wright (2005) hicieron del enfoque un marco manejable para la política monetaria. En su modelo, con dinero divisible, se alternan un mercado descentralizado, en el que el dinero es esencial, y otro centralizado, en el que los agentes reajustan sus saldos. Con preferencias cuasilineales, todos salen del mercado centralizado con los mismos saldos, lo que simplifica la agregación y permite analizar la inflación y la regla de Friedman (apartado 2.2.1). El coste de bienestar de la inflación resulta, además, mayor que en los modelos con dinero en la función de utilidad.

**Conclusiones**:

▪ *Aspectos positivos*: aporta una explicación del dinero como medio de cambio complementaria a la del depósito de valor de los modelos de generaciones solapadas.

▪ *Limitaciones*: en la versión original, como el dinero y los bienes son indivisibles y se cambian uno por uno, los precios quedan fijos y el modelo no permite analizar si el dinero es neutral o superneutral, lo que limitaba su uso para la política económica hasta las extensiones citadas.

### 2.3.2. Modelos de sustitución de monedas en economía abierta

**Origen**: estos modelos adaptan el modelo MIU de Sidrauski a una economía abierta en la que circulan dos monedas.

**Idea**: con dos monedas, la demanda de moneda nacional depende también de la inflación extranjera, y la moneda extranjera puede sustituirla.

**Razonamiento**: el modelo da lugar a una función de demanda (real) de moneda nacional tal que:

$$\frac{M^d}{P} = f(C^+, \pi^{*+}, r^-, \pi^-)$$

El resultado principal es que, cuando la oferta monetaria nacional acelera su crecimiento, la inflación esperada sube y, si supera con holgura a la del exterior, las carteras se inclinan hacia la moneda extranjera aunque la ley lo restrinja.

Con inflación alta, la moneda nacional va perdiendo su papel de depósito de valor, aunque siga sirviendo para pagar y para expresar precios. Si la inflación es muy alta y errática, la divisa puede llegar a sustituirla también como unidad de cuenta.

**Conclusiones**:

- *Aspectos positivos*: los modelos de sustitución de monedas explican fenómenos observados, como la dolarización.
- *Limitaciones*: son las del modelo de Sidrauski como MIU.

## 2.4. Conclusiones

Los modelos de generaciones solapadas y de búsqueda explican por qué un activo sin valor intrínseco se demanda, como depósito de valor intergeneracional o como reductor de costes de búsqueda, pero, salvo los de búsqueda desde Lagos y Wright (2005), son poco operativos para la política. Los modelos MIU, de tiempo de compra y CIA, en cambio, toman como dado el motivo para demandar dinero y analizan con él la política monetaria, con la superneutralidad en el estado estacionario como resultado común.

## 3. Evidencia empírica e implicaciones de política económica

La sección recoge los hechos estilizados sobre la demanda de dinero y las implicaciones de política económica de sus teorías, y termina con el debate sobre la economía sin efectivo y el dinero digital.

### 3.1. Hechos estilizados

1. *Renta y riqueza*: las estimaciones confirman que ambas influyen en los saldos demandados, y Meltzer (1963) encontró que la riqueza no humana explicaba la demanda de dinero mejor que la renta.

2. *Tipos de interés*: la demanda de dinero varía en sentido contrario a los tipos de interés, aunque con poca elasticidad.

3. *Trampa de liquidez*: hay al menos dos episodios que la evidencia relaciona con ella, el estancamiento de Japón desde los últimos años noventa y la etapa de tipos nulos en Estados Unidos (2009-2015).

4. *Velocidad del dinero*: hasta mediados de los setenta, los coeficientes estimados de la demanda de dinero apenas cambiaron, pero desde 1974 la función estimada sobrepredijo sistemáticamente los saldos observados (Goldfeld, 1976). Después, la innovación financiera, que cobró fuerza en los ochenta, hizo inestables tanto la velocidad de circulación como la propia demanda. En EE. UU., por ejemplo, la velocidad de M1 cayó en los ochenta en lugar de seguir subiendo.

### 3.2. Implicaciones de política económica

Las implicaciones de los fundamentos se resumen en la tabla A35.1 del apartado 1.5. Cuanto más estable y menos sensible al tipo de interés es la demanda de dinero, como en los clásicos y los monetaristas, más potente y predecible es la política monetaria. Cuanto más inestable y sensible, en cambio, como en Keynes, más protagonismo gana la fiscal, sobre todo en la trampa de liquidez.

Los modelos de equilibrio general dinámico añaden como implicación propia que el dinero es superneutral en el estado estacionario. Por tanto, la política monetaria solo determina la inflación y los saldos reales, y su único efecto real es el coste de bienestar de la inflación, salvo cuando el dinero afecta a la producción o a la oferta de trabajo.

### 3.3. Debate: economía sin efectivo y dinero digital

> La política monetaria en el límite inferior de los tipos y los instrumentos no convencionales se desarrollan en los temas A36, apartado 3.2, y A37, apartado 3.5, y el euro digital, en el tema B43. Aquí se recoge su relación con la demanda de dinero.

**Origen**: el debate cobra fuerza cuando, tras la crisis financiera global, los tipos oficiales de las principales economías llegan a su límite inferior, y su formulación de referencia es la de Rogoff (2016).

**Idea**: es dinero cualquier cosa con liquidez y credibilidad bastantes para funcionar como medio de cambio, de modo que el *dinero electrónico* (apuntes contables movilizados con tarjetas o

aplicaciones) lo es en igual sentido que los billetes. La cuestión de política es si conviene eliminar el efectivo, y el argumento central es monetario.

**Razonamiento**: mientras exista efectivo, el tipo de interés nominal no puede bajar sensiblemente de cero, porque los agentes atesorarían billetes antes que pagar por prestar. Con precios rígidos, ese suelo impide alcanzar el tipo real negativo[31] que vaciaría el mercado en una recesión profunda[32]. Frente a ese límite caben cuatro respuestas:

- *Política monetaria no convencional* (tipos ligeramente negativos, compras de activos): tiene riesgo de efectos secundarios sobre la estabilidad financiera.

- *Política fiscal expansiva que absorba el exceso de ahorro*: no está disponible para todos los países.

- *Reformas estructurales que flexibilicen los precios*: son costosas y lentas.

- *Eliminación del efectivo*: suprime el ZLB y haría efectivos los tipos negativos oficiales. Rogoff propone una transición gradual, retirando primero los billetes de mayor denominación, lo que además haría aflorar economía sumergida.

**Desarrollo gráfico**: el límite inferior de los tipos de interés se representa en el plano saldos reales-tipo de interés nominal (panel a) y en el plano tiempo-tipo de interés (panel b) de la figura A35.6:

1. *Demanda de dinero*, $L(Y_0, i)$ (panel a): es decreciente y se vuelve horizontal en $i = 0$, porque a un tipo nulo el efectivo, que rinde un 0 % nominal, y los bonos son sustitutivos perfectos.

2. *Oferta de dinero*, $M_0^s/P$ (panel a): es vertical, fijada por el banco central.

3. *Tipo objetivo*, $i^T(t)$ (panel b): es el tipo que prescribe la regla de Taylor, negativo entre $t_1$ y $t_2$. El tipo oficial no puede seguirlo por debajo de cero, y el área sombreada mide el estímulo no aplicado.

4. *Equilibrio*: en el punto 0 del panel (a), el mercado de dinero se vacía sobre el tramo horizontal de la demanda, con $i = 0$, el tipo oficial entre $t_1$ y $t_2$.

**Estática comparativa**: una expansión monetaria desplaza la oferta de $M_0^s/P$ a $M_1^s/P$ y lleva el mercado del punto 0 al 1 sin reducir el tipo de interés, porque los saldos adicionales se mantienen ociosos (panel a). Por tanto, la política monetaria convencional no cierra la brecha del panel (b). Si, en cambio, se elimina el efectivo, desaparece el activo que rinde un 0 % nominal y, con él, el tramo horizontal de la demanda, que pasa a ser la curva discontinua del panel (a). El banco central podría entonces llevar el tipo oficial por debajo de cero, hacia $i^T(t)$.

[31] El tipo de interés real negativo no es creado por las autoridades monetarias, sino por las fuerzas del mercado (exceso de ahorro en China, Alemania, Japón, etc. y escasez de inversión en Estados Unidos y Europa).

[32] Es más, con una demanda agregada muy débil, la economía podría entrar en deflación, lo que agravaría el problema del tipo de interés real excesivamente alto.

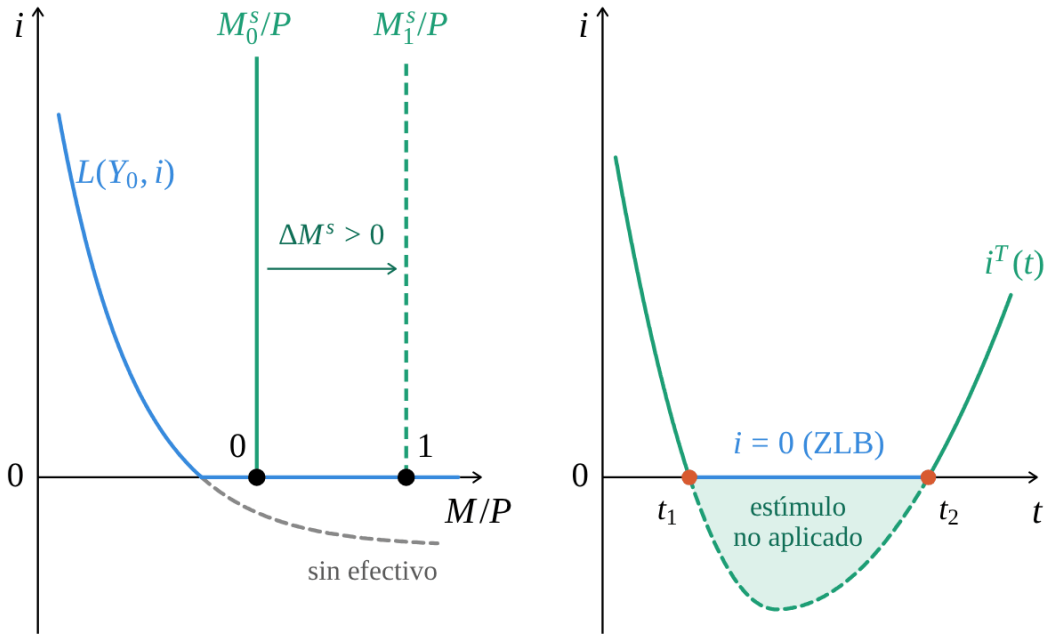



Figura A35.6: Límite inferior de los tipos de interés y demanda de dinero. Con efectivo, la demanda de dinero se vuelve horizontal en $i = 0$, de modo que la expansión monetaria lleva el mercado del punto 0 al 1 sin reducir el tipo de interés (a). Entre $t_1$ y $t_2$ el tipo objetivo, $i^T(t)$, es negativo, pero el tipo oficial no baja de cero, y el área sombreada es el estímulo no aplicado (b).

**Extensiones**: el debate ha evolucionado hacia las *monedas digitales de banco central* (CBDC, euro digital), que buscan preservar el acceso al dinero público, sin eliminar el efectivo, en una economía cada vez menos dependiente de él. Su diseño (remuneración, límites de tenencia) determina su efecto sobre la demanda de depósitos y la transmisión monetaria.

**Conclusiones**:

- *Aspectos positivos*: la eliminación del efectivo ataca directamente la causa del ZLB, sin depender de reformas lentas ni de un margen fiscal del que no todos los países disponen, y suma la ganancia de reducir la economía sumergida.

- *Limitaciones*: supone la pérdida de anonimato y de resiliencia del sistema de pagos, excluye a los colectivos menos digitalizados y concentra el poder de emisión.

## 3.4. Conclusiones

La evidencia confirma los ingredientes de las teorías, pero no zanja el debate sobre la estabilidad. Así, la renta y la riqueza influyen en los saldos demandados, y la demanda varía en sentido contrario al tipo de interés, aunque con poca elasticidad. La trampa de liquidez se asocia al estancamiento de Japón y a los tipos nulos de EE. UU. entre 2009 y 2015. La función fue

relativamente estable hasta mediados de los años setenta y la innovación financiera la desestabilizó, lo que llevó a los bancos centrales a abandonar los agregados monetarios como objetivo intermedio.

De la estabilidad y la sensibilidad al tipo de interés depende la eficacia relativa de las políticas. Con una demanda estable y poco sensible gana potencia la política monetaria, y con una inestable y sensible, la fiscal. El debate sobre el efectivo y el dinero digital traslada la misma pregunta al diseño del propio dinero, porque el efectivo, que rinde un 0 % nominal, impide que el tipo de interés baje sensiblemente de cero.

## Conclusiones

La teoría de la demanda de dinero pasa de la teoría cuantitativa, con saldos proporcionales a las transacciones, al dinero como activo de Keynes, con una demanda especulativa sensible al tipo de interés y potencialmente inestable. La síntesis neoclásica microfundamenta cada motivo, y Friedman trata el dinero como un bien más, con una demanda estable que depende de la renta permanente. Los modelos de equilibrio general dinámico explican por qué se demanda un activo sin valor intrínseco o toman ese motivo como dado, y obtienen la superneutralidad del dinero en el estado estacionario. El recorrido gira en torno a la estabilidad de la función y su sensibilidad al tipo de interés, de las que depende la eficacia relativa de las políticas monetaria y fiscal. La evidencia muestra una función estable hasta mediados de los años setenta, desestabilizada después por la innovación financiera.

La sensibilidad de la demanda de dinero al tipo de interés fija la pendiente de la LM y, con ella, la eficacia relativa de las políticas monetaria y fiscal en los macromodelos clásico y keynesiano (tema A04). La inestabilidad de la demanda de dinero explica que los bancos centrales abandonaran los objetivos de agregados monetarios por los de inflación (tema A36) y condiciona los canales por los que la política monetaria llega a la economía (tema A37). Por último, la función de Cagan determina el señoraje máximo que puede obtener el Estado (tema A39), y la regla de Friedman, que anula el coste de oportunidad de mantener dinero, sirve de referencia para la inflación óptima (tema A41).

Trabajos recientes cuestionan que la función se haya vuelto inestable. Lucas y Nicolini (2015) muestran que la ruptura aparente de la función en EE. UU. desde los años ochenta desaparece cuando el agregado incluye las cuentas de depósito del mercado monetario. Asimismo, Benati, Lucas, Nicolini y Weber (2021) hallan en la mayoría de 38 países una relación estable a largo plazo entre el cociente M1/PIB y el tipo de interés a corto plazo. Sigue abierto también el debate sobre el diseño del dinero digital. Rogoff (2016) propone eliminar el efectivo para suprimir el ZLB. Por su parte, Fernández-Villaverde, Sanches, Schilling y Uhlig (2021) muestran que una moneda digital de banco central abierta a todos puede convertir al banco central, en un pánico, en monopolista de los depósitos. Los límites de tenencia y la ausencia de remuneración previstos para el euro digital buscan contener ese riesgo.

# A36. Política monetaria I: diseño e instrumentos

Política monetaria (I). El diseño y la instrumentación de la política monetaria.

## Introducción

Un banco central hace **política monetaria** cuando altera la cantidad de dinero o su precio para alcanzar ciertas metas macroeconómicas, como la estabilidad de los precios, la estabilidad cambiaria o el empleo, y, como fin último, el mayor bienestar social posible. La política monetaria actúa, por definición, sobre la demanda agregada. Cuando aporta o retira liquidez, el banco central estimula o frena el gasto, y esos cambios en la demanda pueden acabar en la producción y demás variables reales o en los precios y demás variables nominales, y entender cómo y en qué proporción lo hacen es la cuestión central de su estudio.

Su eficacia es materia de una vieja controversia, teórica y empírica. Para la teoría neoclásica, dado que se cumplen la ley de Walras, la ley de Say y la teoría cuantitativa del dinero, rige la **dicotomía clásica**, según la cual las variables nominales no afectan a las reales y la política monetaria es inefectiva[1]. Para la teoría keynesiana, en cambio, la existencia de rigideces de precios y salarios, la posibilidad de una demanda agregada insuficiente y la demanda de dinero por motivo especulativo rompen esa dicotomía. La cantidad de dinero altera entonces el tipo de interés y, a través de él, la inversión, de modo que la política monetaria sí es un instrumento de estabilización[2].

La práctica de los bancos centrales ha seguido la evolución de ese debate. Tras la Segunda Guerra Mundial predominó una política monetaria subordinada a la fiscal y al mantenimiento de los tipos de cambio de Bretton Woods. Más tarde, la inflación de los años setenta dio paso a los objetivos monetarios de inspiración monetarista y, desde principios de los noventa, a la independencia de los bancos centrales y a las estrategias de **objetivos de inflación**, que acompañaron la llamada Gran Moderación. La crisis financiera de 2008 llevó los tipos de interés a su límite inferior y generalizó los **instrumentos no convencionales**, como la compra masiva de activos y el *forward guidance* (orientación prospectiva). El repunte inflacionista de 2021-2023 obligó, en sentido contrario, al endurecimiento monetario más intenso en décadas. Hoy pocos economistas discuten que, a largo plazo, la política monetaria apenas puede alterar el nivel de actividad. Por eso, a los bancos centrales se les encomienda el control de la inflación, la estabilización cíclica a corto plazo y, cada vez más, la estabilidad financiera.

El objeto de este tema es el análisis del diseño y la instrumentación de la política monetaria, que se aborda en tres pasos:

[1] Los monetaristas, herederos de los neoclásicos, sí consideran la política monetaria efectiva a corto plazo, a diferencia de los neoclásicos puros.

[2] Aunque, en la visión keynesiana, menos eficaz que la política fiscal.

- *Fundamentos*: el concepto de dinero, su oferta y demanda y el proceso de creación de dinero.
- *Diseño*: sus elementos, los debates sobre objetivos, variable operativa y discrecionalidad, y la política óptima.
- *Instrumentación*: tanto convencional como no convencional y cualitativa.

# 1. Fundamentos del diseño de la política monetaria

Los fundamentos del diseño son el concepto de dinero, su oferta y su demanda, y el equilibrio entre ambas, que da lugar al proceso de creación de dinero.

## 1.1. Concepto de dinero: definición teórica y operativa

**Definición conceptual**: J. Hicks propone caracterizar el dinero por las funciones que desempeña y no por su naturaleza. El dinero cumple, así, cuatro funciones:

- *Unidad de cuenta*: se utiliza para denominar los bienes, servicios y activos y anotar las transacciones, lo cual reduce drásticamente el número de precios relativos que deben conocerse frente al trueque.

- *Medio de cambio*: se usa y acepta generalmente para liquidar transacciones, lo cual permite superar la doble coincidencia de deseos del trueque.

- *Depósito de valor*: permite transferir poder adquisitivo del presente al futuro.

- *Estándar de pago diferido*: en él se denominan las deudas.

**Definición operativa**: interesa a la política económica y se refiere a los *agregados monetarios*. Estos son pasivos del sistema monetario en manos del sector tenedor de dinero (todos los residentes salvo las propias instituciones monetarias y la Administración central), ordenados por liquidez decreciente y extendidos históricamente en su amplitud por la innovación financiera.

La escala que sigue es la del Eurosistema, si bien el número y la definición de los agregados varían por país[3]:

- *Base monetaria* o *dinero de alta potencia* ($H$): $H = E + R^b$. La emite el banco central y constituye, en consecuencia, un pasivo del banco central. Está compuesta del efectivo en manos del público ($E$), es decir, billetes y monedas fuera del sistema bancario, y de las reservas bancarias ($R^b$), es decir, el efectivo en caja de los bancos y los saldos que mantienen en el banco central, legales o voluntarios.

- *M1*: $M_1 = E + D$, donde $D$ son los depósitos a la vista del sector tenedor de dinero[4]. Constituye el dinero en sentido estricto y el pasivo del sistema monetario. En los modelos de este tema, es el agregado que se identifica con la oferta monetaria, $M^s$.

- *M2*: $M_2 = M_1 + D_2$, donde $D_2$ son los depósitos a plazo de hasta dos años y los disponibles con preaviso de hasta tres meses[5], el "cuasidinero" o *near money*.

- *M3*: $M_3 = M_2 + D_3$, donde $D_3$ incluye instrumentos negociables emitidos por las instituciones monetarias: cesiones temporales, participaciones en fondos del mercado monetario y valores de deuda con vencimiento hasta dos años. Como la sustitución entre sus componentes apenas lo altera, el Eurosistema lo sigue en su análisis monetario.

[3] En EE. UU., la Reserva Federal publica solo $M_1$ y $M_2$ (dejó de publicar $M_3$ en 2006).

[4] Las reservas forman parte de la base monetaria $H$ pero no de $M_1$, porque son un activo de los bancos y no dinero en manos del sector tenedor de dinero.

[5] Una transferencia de una cuenta de ahorro a una cuenta corriente deja $M_2$ inalterado pero eleva $M_1$.

## 1.2. Oferta de dinero: el papel del sistema monetario

La oferta monetaria depende del sistema monetario, es decir, del conjunto de instituciones cuyos pasivos circulan como dinero. Lo integran el banco central y la banca comercial.

**Banco central**: es el organismo público con la exclusiva de emitir moneda de curso legal[6]. Además, diseña y ejecuta la política monetaria, vigila a las entidades financieras y ejerce de prestamista de última instancia.

El pasivo del banco central es la base monetaria, $H = E + R^b$, y su activo lo forman los activos netos frente al exterior o reservas internacionales (ANFE, $R$, es decir, divisas, oro, derechos especiales de giro y bonos extranjeros) y el crédito interno (CI). Este comprende los activos frente al Gobierno (AFG) y frente al sector privado, esencialmente la financiación a la banca (AFSP). El patrimonio neto del banco central se considera nulo[7], de modo que toda adquisición de activos se financia con una obligación. Por eso, desde el activo se ve cómo crece la base monetaria cuando el banco central compra activos exteriores o da crédito interno.

Tabla A36.1: Balance simplificado del banco central.

| **ACTIVO** | | **PASIVO** | |
|---|---|---|---|
| Activos netos frente al exterior (ANFE) | $R$ | Efectivo en manos del público | $E$ |
| **Crédito interno (CI)** | | Reservas bancarias | $R^b$ |
| Activos frente al Gobierno | AFG | | |
| Activos frente al sector privado | AFSP | | |
| **TOTAL ACTIVO** | $R + \mathbf{CI}$ | **TOTAL PASIVO** (base monetaria, $H$) | $E + R^b$ |

**Banca comercial**: como cualquier intermediario financiero, traslada fondos desde quienes tienen excedentes hacia quienes necesitan financiación, pero, a diferencia de los demás, crea dinero mediante la *expansión múltiple de los depósitos*. Su pasivo son los depósitos ($D$) y la financiación recibida del banco central (AFSP, que aparece en el activo de aquel). Su activo lo forman las reservas ($R^b$, que aparecen en el pasivo del banco central) y el crédito al público.

Las reservas pueden ser legales ($R^b_L$), la cantidad mínima fijada por el banco central a través del coeficiente de caja[8], y voluntarias ($R^b_V$), que dependen de factores de mercado como el tipo de interés y la incertidumbre[9]. Sus coeficientes respecto de los depósitos, $q_L$ y $q_V$, suman el encaje bancario, $q = q_L + q_V$, de modo que $R^b = qD$. La razón de ser de las reservas es garantizar

[6] En la Unión Económica y Monetaria (UEM), el art. 128 del Tratado de Funcionamiento de la Unión Europea (TFUE) reserva al BCE autorizar la emisión de billetes. Las monedas las emiten los Estados miembros, y el BCE aprueba su volumen (art. 128.2 TFUE).

[7] El supuesto simplifica la exposición, porque, si el banco central retiene como beneficio parte de los intereses de sus activos, esos intereses pasan a formar parte de su patrimonio neto.

[8] La reserva mínima se calcula como producto de la base de reservas (pasivos del banco con vencimiento o preaviso de hasta dos años, excluidas las cesiones temporales y deducidos los contraídos con otras entidades sujetas al sistema y con el Eurosistema) por el coeficiente de reservas. Se mantiene en la cuenta corriente en el banco central nacional.

[9] Las reservas voluntarias dependen negativamente de la rentabilidad de las inversiones alternativas (coste de oportunidad de mantener dinero ocioso) y positivamente del grado de incertidumbre y del coste de recurrir al mercado interbancario.

la estabilidad y credibilidad del sistema bancario, pues nada obliga a que entradas y salidas de efectivo coincidan[10].

Tabla A36.2: Balance simplificado de un banco comercial.

| **ACTIVO** | | **PASIVO** | |
|---|---|---|---|
| **Reservas** ($R^b$) | | Depósitos | $D$ |
| Reservas legales | $R_L^b$ | Financiación del banco central | AFSP |
| Reservas voluntarias | $R_V^b$ | | |
| Crédito al público | Crédito | | |
| **TOTAL ACTIVO** | $R^b + \textbf{Crédito}$ | **TOTAL PASIVO** | $D + \textbf{AFSP}$ |

**Sistema monetario**: al consolidar ambos balances, las partidas cruzadas entre ellos ($R^b$ y AFSP) se cancelan. Queda así un balance cuyo pasivo es la oferta monetaria, $M_1 = E + D$, y cuyo activo recoge los activos exteriores netos y el crédito interno, tanto al sector público (AFG) como al privado (crédito bancario).

En el balance consolidado se ve que el dinero nace, sobre todo, de la compra de activos exteriores y del crédito interno, objeto del multiplicador monetario.

Tabla A36.3: Balance consolidado del sistema monetario.

| **ACTIVO** | | **PASIVO** | |
|---|---|---|---|
| Activos netos frente al exterior (ANFE) | $R$ | Efectivo en manos del público | $E$ |
| **Crédito interno (CI)** | | Depósitos | $D$ |
| Activos frente al Gobierno | AFG | | |
| Crédito al público | Crédito | | |
| **TOTAL ACTIVO** | $R + \textbf{CI}$ | **TOTAL PASIVO** (oferta monetaria, $M_1$) | $E + D$ |

## 1.3. Demanda de dinero

Para comprender el control imperfecto de la cantidad de dinero hay que entender la demanda. La razón es que la autoridad monetaria emite base monetaria, pero la cantidad de dinero que circula, $M_1$, la determinan el público, que decide cuánto dinero mantener y en qué forma, y los bancos, que deciden qué reservas mantener frente a los depósitos.

**Enfoque teórico**: existe un amplio número de hipótesis sobre la demanda de saldos reales, $M^d/P$, que comparten las variables que la relacionan con la economía real, si bien difieren en el papel asignado a cada una. Las dos concepciones clásicas son:

- *Keynesiana* ($M^d/P = kY - hi$): suma un componente transacción, proporcional a la renta, y un componente especulación, decreciente con el tipo de interés. Este enfoque considera que la demanda de dinero es muy inestable, ya que el segundo componente depende de las expectativas sobre los tipos, en gran medida convencionales y cambiantes. Además, en él los bonos son los

[10] En el balance de la banca no se consideran los activos frente al exterior ni frente al sector público, que no alteran la lógica de la creación de dinero.

únicos sustitutivos del dinero. Estos rasgos dificultan la efectividad de la política monetaria. También la dificultan los dos eslabones débiles (la elevada sensibilidad de la demanda de dinero al tipo de interés, con la trampa de liquidez en el límite, y la escasa sensibilidad de la inversión al tipo de interés).

- *Monetarista* ($M^d/P = f(Y^{P^+}, r_B^-, r_E^-, \pi^{e^-}, (W_H/W)^+)$): considera que el dinero es un bien más, cuya demanda depende de múltiples factores. Da lugar a una demanda de dinero estable y dependiente fundamentalmente de la renta permanente, ya que todos los activos, financieros (bonos, acciones) y reales (cuyo rendimiento es la inflación esperada), son sustitutivos imperfectos del dinero[11].

**Enfoque empírico**: ambas concepciones admiten una forma reducida común, $M^d/P = L(Y^+, i^-)$, donde $Y$ es la renta disponible del público[12] e $i = r + \pi^e$ el tipo de interés nominal del activo sustitutivo.

Las estimaciones de esta función para $M_1$ (Goldfeld, 1973) obtuvieron una elasticidad-renta a largo plazo inferior a la unidad (en torno a 0,7) y una elasticidad-interés reducida y estable, lo que sostuvo el control de los agregados como estrategia. Sin embargo, su ruptura a partir de 1974, cuando la función sobrepredijo sistemáticamente la demanda (*the case of the missing money*, Goldfeld, 1976), es el antecedente del abandono de esa estrategia que se estudia en el apartado 2.2.

**De la demanda de dinero a la demanda de base monetaria**: cualquiera que sea la teoría, la demanda de dinero del público se satisface con dos activos, efectivo y depósitos a la vista. El público decide en qué proporción mantiene cada uno (el coeficiente de efectivo, $c_M = E/M^s$, que depende de los costes de transacción, la bancarización y la confianza en los bancos).

Los depósitos, a su vez, obligan a los bancos a mantener reservas en el banco central en una proporción $q$ (el coeficiente de reservas), de modo que la demanda de depósitos del público genera una demanda de reservas de los bancos.

Efectivo y reservas son los dos componentes de la base monetaria, el pasivo monetario del banco central. Su suma es la demanda de base monetaria, $H^d = E^d + R^{b,d}$, y es frente a ella, y no frente a la demanda de dinero, donde la autoridad monetaria fija su oferta. Esta cadena se formaliza a continuación.

## 1.4. Equilibrio en el mercado monetario: el proceso de creación de dinero

En el mundo real, el mercado de reservas forma parte del mercado interbancario, y en él el tipo de interés se ajusta hasta igualar la oferta y la demanda de reservas. No obstante, es un mercado intervenido, puesto que los bancos centrales actúan sobre él para influir en el tipo de interés con el que se relacionan todos los tipos de la economía.

La cadena anterior se formaliza con tres modelos de complejidad creciente que comparten un mismo entorno (economía cerrada sin imperfecciones) y tres agentes:

- *Banco central*: emite la base monetaria, $H$.
- *Público*: decide la composición de su demanda de dinero.

[11] La ausencia de los dos eslabones débiles hace, para los monetaristas, la política monetaria efectiva a corto plazo, mientras que la teoría cuantitativa la hace inefectiva a largo. Las teorías de la demanda de dinero se desarrollan en el tema A35.

[12] A pesar de que desde la teoría resulta más atractivo considerar la riqueza como variable, en la práctica resulta difícil estimarla, por lo que muchos modelos de los bancos centrales se apoyan en la renta disponible de los hogares.

- *Bancos comerciales*: fijan sus reservas en una proporción $q \in (0,1)$ de los depósitos.

**Modelo con depósito íntegro**: los agentes ingresan entera en un banco la base $H$ que emite el banco central, $D^{(1)} = H$. Ese banco guarda $qH$ como reserva y presta el resto, $(1-q)H$, que acaba también entero en otro banco, $D^{(2)} = (1-q)H$. Este reserva $q(1-q)H$ y presta $(1-q)^2H$, y así *ad infinitum*. Sumando todas las rondas, los depósitos de la economía son:

$$D = H + (1-q)H + (1-q)^2H + \cdots = \sum_{n=0}^{\infty}(1-q)^n H = \underbrace{\frac{1}{q}}_{b} H \rightarrow \ M^s = \frac{1}{q}H$$

Sin efectivo ($M^s = D$), el *multiplicador bancario* es $b = 1/q > 1$. Dada una base $H$, la economía termina con una cantidad de dinero mayor que $H$, tanto mayor cuanto menor sea $q$ (y, por extensión, cuanto mayor sea $b$), y los bancos crean así dinero.

**Modelo con ratio efectivo-depósitos**: el banco central pone en circulación de nuevo $H$ y el público en este caso mantiene efectivo en proporción a sus depósitos, $E^d = c_D D^d$, con $c_D = E/D > 0$. Los bancos mantienen en reservas parte de los depósitos de sus clientes, $R^{b,d} = qD^d$.

La condición de equilibrio del mercado de base monetaria establece que la demanda de efectivo y reservas debe igualar a la base monetaria emitida inicialmente:

$$H^d \equiv E^d + R^{b,d} = H$$

Con la definición $M^s = E + D$, se obtiene:

$$(c_D + q)D = H, \quad M^s = (1+c_D)D \ \rightarrow \ M^s = \underbrace{\frac{1+c_D}{c_D+q}}_{m=f(c_D^-,\, q^-)} H$$

De nuevo $m > 1$, pero ahora el multiplicador depende no solo del coeficiente de reservas de los bancos sino también de la ratio efectivo-depósitos que decide el público.

Cuanto más efectivo mantenga el público fuera del sistema bancario ($c_D \uparrow$, "dinero debajo del colchón" en situaciones de incertidumbre), menor es $m$, que tiende a la unidad cuando todo el dinero es efectivo. Solo si las reservas superan a los depósitos ($q > 1$, caso que el supuesto $q \in (0,1)$ excluye) el multiplicador cae por debajo de uno.

**Modelo con ratio efectivo-dinero**: el público mantiene efectivo en proporción a su dinero, $E^d = c_M M^d$ y $D^d = (1-c_M)M^d$, con $c_M = E/M^s \in (0,1)$, el coeficiente de efectivo ya definido. Con $R^{b,d} = qD^d$ y la misma condición de equilibrio:

$$c_M M^s + q(1-c_M)M^s = H \ \rightarrow \ M^s = \underbrace{\frac{1}{c_M + q(1-c_M)}}_{m=f(c_M^-,\, q^-)} H$$

Ambas parametrizaciones son el mismo multiplicador, pues $c_D = c_M/(1-c_M)$. La segunda es la habitual porque $c_M$ es la fracción de $M_1$ que el público mantiene en efectivo, directamente observable en los agregados.

**Multiplicador y control de la oferta monetaria**: en suma, cada unidad de base genera $m \geq 1$ unidades de dinero, tanto más cuanto menores sean $q$ y $c_M$.

La autoridad monetaria fija $H$ mediante operaciones de mercado abierto y $q_L$ mediante los requerimientos de reservas, pero no controla las reservas voluntarias ni el coeficiente de efectivo.

Su control de la cantidad de dinero es, por tanto, indirecto e imperfecto, y será tanto mejor cuanto más estable sea $m$[13]. El control de la propia base, además, exige gestionar a discreción los activos exteriores y el crédito interno, lo que un tipo de cambio fijo o la financiación del déficit público pueden impedir.

Gráficamente, en el plano base monetaria-cantidad de dinero (figura A36.1), la relación $M^s = m \cdot H$ es un rayo desde el origen con pendiente $m > 1$, más inclinado que la bisectriz (45°).

1. *Multiplicador estable*: $c_M$ y $q$ no varían y una expansión de la base de $H_0$ a $H_1$ eleva la cantidad de dinero de $M_0^s$ a $M_1^s$ a lo largo del rayo (del punto 0 al 1). Por ello, el banco central controla $M^s$ a través de $H$ con precisión.

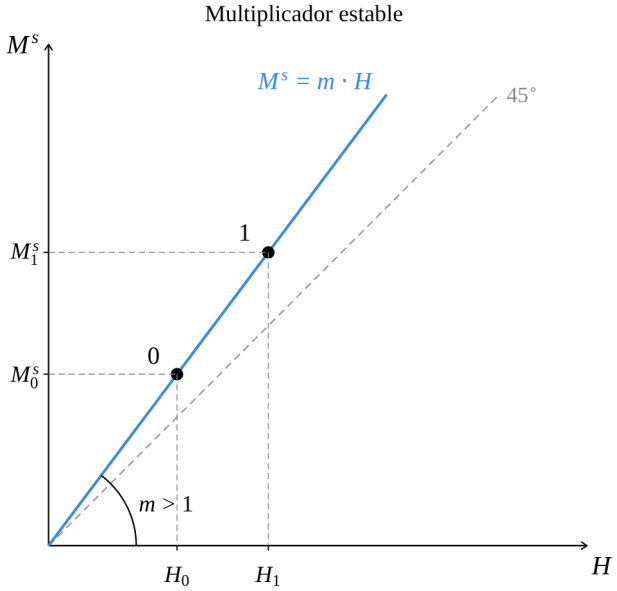



Figura A36.1: Multiplicador monetario estable. La relación $M^s = mH$ es un rayo desde el origen más inclinado que la bisectriz ($m > 1$). Una expansión de la base de $H_0$ a $H_1$ eleva la cantidad de dinero de $M_0^s$ a $M_1^s$.

Gráficamente, en el plano saldos reales-tipo de interés y en el plano $(Y, i)$ (figura A36.2), esa expansión ($\Delta H > 0$) desplaza la oferta de saldos reales de $M_0^s/P$ a $M_1^s/P$ y reduce el tipo de interés de $i_0$ a $i_{0'}$ (punto 0′). En el modelo IS-LM, la LM se desplaza de $\text{LM}_0$ a $\text{LM}_1$ ($\Delta M^s > 0$) y la renta sube de $Y_0$ a $Y_1$, lo que eleva la demanda de dinero de $L(Y_0)$ a $L(Y_1)$ y sitúa el tipo de interés en $i_1$ (punto 1).

[13] Cuando el banco central fija el tipo de interés, como hacen hoy los principales, la causalidad va en sentido inverso. Los bancos crean depósitos al conceder préstamos y el banco central suministra al tipo fijado las reservas que demandan, de modo que el multiplicador describe una relación contable entre la base y el dinero, no un mecanismo de control (McLeay, Radia y Thomas, 2014).

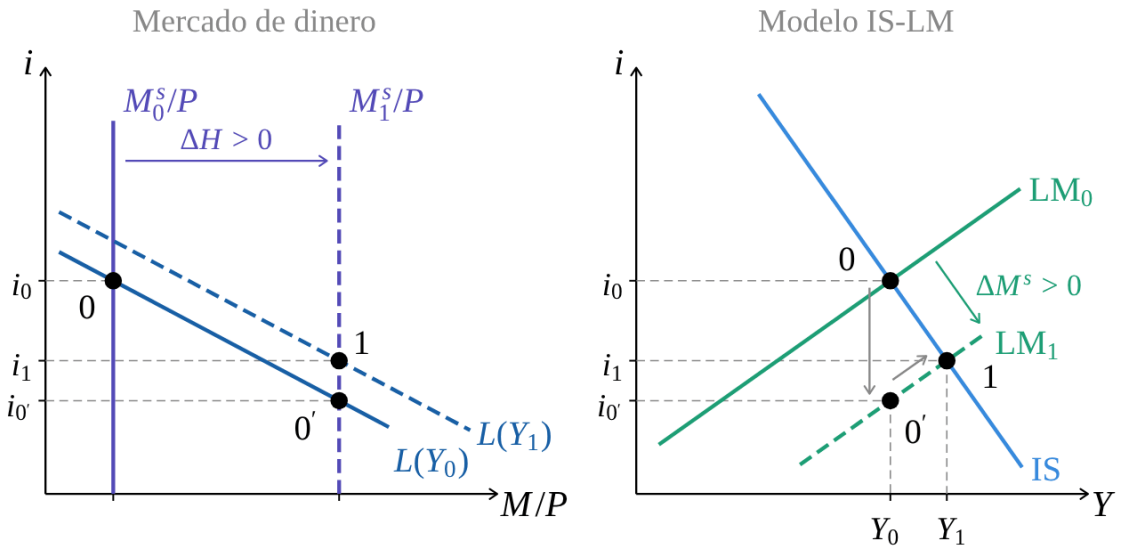



Figura A36.2: Expansión de la base monetaria. La oferta de dinero se desplaza y el tipo de interés desciende.

2. *Multiplicador inestable*: se representa en el plano base monetaria-cantidad de dinero (figura A36.3). Un aumento de $c_M$ o de $q$, como en una crisis de confianza, en la que el público retira efectivo y los bancos acumulan reservas, hace girar el rayo hacia abajo, de $m_0$ a $m_1$. Por ello, con la misma base, $H_0$, la cantidad de dinero cae hasta $M^s_{0'}$ (punto $0'$), y para mantenerla en $M^s_1 = M^s_0$ el banco central debe expandir la base hasta $H_1$ (punto 1).

Así ocurrió tras 2008, cuando los balances de los bancos centrales se multiplicaron con los programas de compras y la acumulación de reservas excedentarias mientras los agregados monetarios apenas crecían, con $m_1 < 1$ en el caso de la Reserva Federal, cuyas reservas bancarias llegaron a superar a los depósitos a la vista ($q > 1$). La inestabilidad de $m$ es, por tanto, la razón por la que el control de la base es un instrumento poco fiable para controlar la cantidad de dinero, argumento que reaparece en el debate sobre la variable operativa (apartado 2.2).

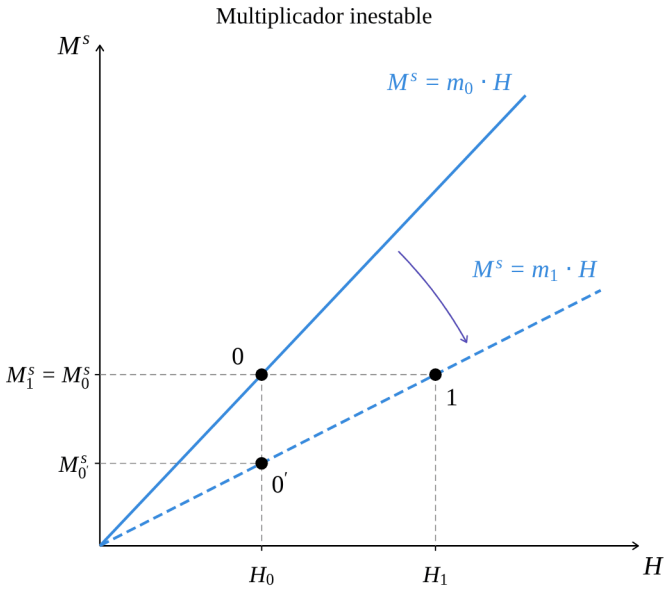



Figura A36.3: Multiplicador inestable. Un aumento de $c_M$ o de $q$ hace girar el rayo hacia abajo, de $m_0$ a $m_1$. Con la misma base, la cantidad de dinero cae (punto $0'$) y, para mantener $M_1^s = M_0^s$, el banco central debe ampliar la base de $H_0$ a $H_1$.

## 1.5. Conclusiones

**Aspectos positivos**: los modelos planteados permiten entender las intuiciones en torno al funcionamiento del mercado monetario, base de la determinación de los tipos de interés de toda la economía. Explican, además, por qué cada unidad de $H$ genera una cantidad proporcional de $M^s$ y por qué el control de la cantidad de dinero es imperfecto. Fenómenos como los vividos durante la Gran Recesión (2008-2014) o el escándalo del Libor (manipulaciones de 2005-2009, destapadas en 2012) ponen de manifiesto la importancia de los mercados monetarios para el sistema económico en su conjunto.

**Limitaciones**: se ha hecho abstracción de la dimensión internacional, pero las reservas de los bancos centrales se ven afectadas por el sector exterior.

Las autoridades monetarias también financian la balanza de pagos e intervienen en el mercado de divisas, de dos formas:

- *No esterilizada*: deja fluctuar la oferta de moneda nacional, lo que puede ser efectivo para mantener el tipo de cambio, pero genera problemas de credibilidad por su incompatibilidad con una política monetaria orientada a objetivos internos.

- *Esterilizada*: mantiene constante la oferta monetaria, y su efectividad depende de la sustituibilidad imperfecta entre bonos nacionales y extranjeros.

Ambas formas de intervención alteran los activos exteriores netos del banco central, pero la esterilizada compensa su efecto sobre la base monetaria con una variación de signo contrario del crédito interno.

## 2. Diseño de la política monetaria

El diseño se articula en torno a cuatro elementos, cuya combinación ha cambiado con el tiempo. Dos debates ordenan esa evolución, el de los objetivos y la variable operativa y el de la discrecionalidad y la credibilidad. El modelo IS-PC-MR los ilustra y permite determinar la política óptima.

### 2.1. Definición y elementos del diseño

El **diseño de la política monetaria**, o forma en que la autoridad monetaria plantea su intervención sobre la cantidad y el precio del dinero, se articula en torno a cuatro elementos:

**Objetivo final**: es la variable económica sobre la que se tiene un escaso control directo y que constituye el objetivo último de la política monetaria. Puede abarcar la estabilidad de precios, la estabilidad financiera, el pleno empleo o la redistribución de la renta.

El énfasis en cada uno de ellos, no obstante, cambia. Por ejemplo, el BCE busca la estabilidad de precios, mientras que la Reserva Federal busca no solo la estabilidad de precios, sino también el máximo empleo y unos tipos de interés a largo plazo moderados[14].

**Objetivo intermedio**: variable económica que se puede controlar a través de la variable operativa con un retardo razonable y un grado de precisión relativamente alto, y que además mantiene una relación estable con el objetivo final y es un indicador representativo de él.

Tradicionalmente, se incluyen la cantidad de dinero ($M_1$, $M_3$), el tipo de cambio y los tipos de interés a medio y largo plazo. Su popularidad, sin embargo, ha caído y en la actualidad se utilizan como variables indicador.

**Variable operativa**: variable económica sobre la que el banco central sí tiene un control directo, en el día a día, a través de sus instrumentos.

Incluye la base monetaria, $H$, o el tipo de interés interbancario a corto plazo, al que los bancos se prestan de un día para otro. Para el BCE es el tipo de interés a corto plazo del euro (€STR), sucesor desde 2022 del índice medio del tipo de interés del euro a un día (EONIA), y para la Reserva Federal, el *fed funds rate*. En la actualidad, existe consenso en considerar el tipo de interés como la variable operativa apropiada.

**Instrumentos**: son las herramientas con las que el banco central maneja la variable operativa. Los convencionales son las operaciones de mercado abierto (OMA), las facilidades permanentes y los requerimientos de reservas[15], a los que desde 2008 se han sumado los instrumentos no convencionales.

### 2.2. Debate sobre los objetivos y la variable operativa

En torno a los objetivos y la variable operativa se han sucedido dos doctrinas, la de reservas y la de seguimiento directo de la inflación. Como alternativas a esta última se han propuesto el seguimiento del PIB nominal y el del nivel de precios.

[14] Fuente: mandato dual de la Reserva Federal (Federal Reserve Reform Act, 1977).

[15] Durante 1930-1980 también había un cuarto instrumento, el "método directo", que incluía requerimientos de márgenes y techos al interés de los depósitos.

### 2.2.1. Doctrina de reservas (1920-1990)

El diseño tradicional arranca de la doctrina de la posición de reservas de la Reserva Federal en los años veinte y culmina en los objetivos de agregados monetarios de los años setenta y ochenta. Enfatiza el papel de la variable intermedia y se apoya en el esquema de objetivos e instrumentos de Jan Tinbergen (1952). De ahí que la estrategia tenga dos niveles:

- *Primer nivel*: se determina la relación entre el objetivo final y el objetivo intermedio.
- *Segundo nivel*: se determina la relación entre el objetivo intermedio y la variable operativa.

**Objetivo final**: la estabilidad de la inflación promedio, si bien en la mayoría de países no había un objetivo claro en cuanto a su magnitud ni una política de comunicación.

**Objetivo intermedio**: durante los años setenta y ochenta hubo un importante debate entre fijar un tipo de interés a largo plazo o una cantidad de agregado monetario[16]. Desde el punto de vista teórico predominó la visión monetarista, que situaba como variable intermedia los agregados monetarios.

Por su parte, William Poole (1970), en un modelo IS-LM estocástico de carácter teórico (figura A36.4), formuló la regla de elección del instrumento según el origen de los *shocks* (en rigor, según sus varianzas relativas y las pendientes de la IS y la LM):

- *Sector monetario más inestable* ($\sigma_M^2 > \sigma_R^2$): si los *shocks* de demanda de dinero que desplazan la LM dominan, hay que controlar el tipo de interés. Con el tipo de interés fijo en $i_0$, la LM es horizontal y la renta no se aparta de $Y_0$, mientras que con objetivo de agregado oscilaría entre $Y'$ e $Y''$ (panel a).

- *Sector real más inestable* ($\sigma_R^2 > \sigma_M^2$): los *shocks* reales que desplazan la IS aconsejan controlar los agregados. Con el agregado fijo, la LM creciente amortigua el efecto de los desplazamientos de la IS, y la renta oscila entre $Y'$ e $Y''$ en lugar de entre $Y'''$ e $Y''''$ (panel b). En los años setenta, los *shocks* reales (de gasto y, en la práctica, los del petróleo) hicieron que la evidencia también apoyara el control de agregados, aunque los del petróleo fueron sobre todo de oferta, que el modelo IS-LM de Poole no recoge.

[16] Durante la década anterior predominaban las ideas keynesianas, y el banco central orientaba su política fijando los tipos de interés. El razonamiento monetarista parte de la ecuación dinámica de la teoría cuantitativa con velocidad estable, supuesta constante ($g_V = 0$): $g_M + g_V = \pi + g_Y \rightarrow g_M = \pi + g_Y$. A corto plazo no puede determinarse qué proporción del efecto recae en precios y cuál en renta real (ecuación perdida del monetarismo), y a largo plazo la economía tiende a la tasa natural. Por ello, Friedman propone una regla de crecimiento constante de la masa monetaria.

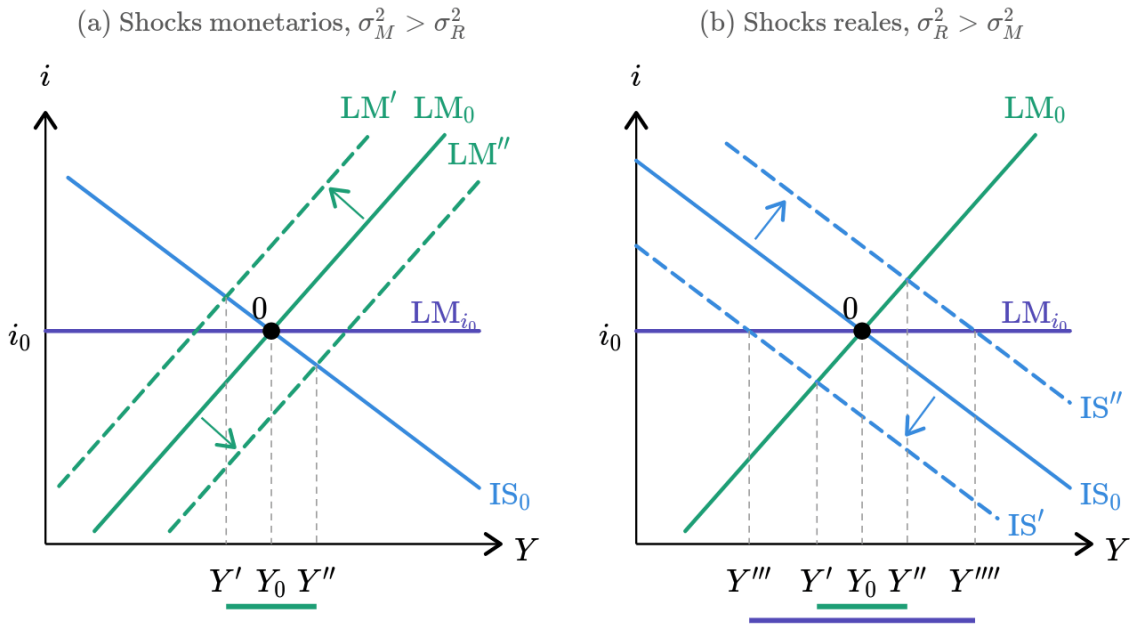



Figura A36.4: Regla de Poole (1970). Con *shocks* monetarios (a), la LM oscila entre LM$'$ y LM$''$ y, con objetivo de agregado, la renta fluctúa entre $Y'$ e $Y''$, mientras que con objetivo de tipo de interés la LM es horizontal en $i_0$ (LM$_{i_0}$) y la renta queda en $Y_0$. Con *shocks* reales (b), la IS oscila entre IS$'$ e IS$''$, y el objetivo de agregado (LM$_0$) limita la renta al intervalo entre $Y'$ e $Y''$, más estrecho que el que deja el objetivo de tipo de interés, entre $Y'''$ e $Y''''$.

**Variable operativa**: el banco central controla la base monetaria a corto plazo.

**Instrumentos**: emplea los convencionales.

**Conclusiones**:

▪ *Aspectos positivos*: la estrategia se aplicó en la mayoría de economías avanzadas desde mediados de los setenta y alcanzó su mayor relevancia entre 1979 y 1982, bajo Paul Volcker en la Reserva Federal. Sus resultados fueron buenos en la reducción de la inflación, pero dañinos para el empleo y la producción.

▪ *Limitaciones*: el objetivo intermedio se controla de forma deficiente, pues hay retardos tanto de detección como de transmisión, el mecanismo por el que actúa la política se conoce de forma imperfecta y el vínculo del objetivo intermedio con el final no es estable. A ello se suman los incentivos a usar la política con miras cortoplacistas, que desembocan en inconsistencia dinámica (*time inconsistency*).

### 2.2.2. Doctrina de seguimiento directo de la inflación (1990-hoy)

A partir de los años ochenta, la innovación y la liberalización financiera volvieron más inestables la demanda de dinero y su velocidad de circulación. Tras una transición a finales de los ochenta, ello derivó en las estrategias de **objetivo directo de inflación** (*inflation targeting*), adoptadas por primera vez en Nueva Zelanda (1990). Siguieron Canadá (1991), Reino Unido (1992), Suecia y Australia (1993), el BCE desde 1999, con una estrategia afín (sus dos pilares incluían el análisis

monetario), y, más tarde, la Reserva Federal (2012) y el Banco de Japón (BoJ, 2013). También lo adoptaron numerosas economías emergentes (Brasil, Chile, México, Polonia, Sudáfrica).

Este régimen supone el paso al diseño de estrategias con un único nivel, que se centra en el objetivo final (inflación) y en la variable operativa (tipo de interés a corto plazo), sin un esquema de relaciones con un objetivo intermedio.

**Objetivo final**: persigue la estabilidad de precios en el medio plazo[17], cuantificada como una variación de un índice de precios al consumo (IPC, índice armonizado de precios de consumo en la UEM, deflactor del consumo en EE. UU.).

Esta cuantificación puede articularse como *objetivo puntual*, como *banda* o como *punto con banda de tolerancia*. El objetivo puntual se sitúa generalmente en torno al 2 %, como en el BCE, el Reino Unido o la Reserva Federal. El punto con banda de tolerancia es el de la mayoría de economías emergentes, y también el de Canadá y Nueva Zelanda, con el 2 % en el centro de un intervalo del 1 al 3 %.

El objetivo debe cumplir tres *requisitos*:

- *Verificable*: mediante el anuncio público de objetivos cuantificados.
- *Flexible* (*flexible inflation targeting*): con discrecionalidad limitada, a medio plazo y con tolerancia a desviaciones transitorias, para atender *shocks* inesperados.
- *Creíble*: para anclar las expectativas, lo que remite al debate sobre la discrecionalidad.

Existe, en todo caso, consenso sobre las bondades de la estabilidad de precios. Esta evita los costes de la inflación y la deflación y los efectos redistributivos arbitrarios, mejora la señal informativa de los precios relativos, reduce las primas de riesgo y las coberturas innecesarias y previene distorsiones en los sistemas impositivos y de Seguridad Social.

**Objetivo intermedio**: no existe como tal. Su papel lo desempeña la *previsión de inflación* del propio banco central al horizonte de la política (en torno a dos años). Si la previsión se desvía del objetivo, el banco central ajusta el tipo de interés (*inflation-forecast targeting*, Svensson, 1997a).

La previsión se apoya en un conjunto amplio de indicadores reales, financieros y monetarios. El BCE lo formalizó en dos pilares, el análisis económico a corto y medio plazo y el análisis monetario a medio y largo plazo, una herencia de la doctrina de reservas que ningún otro banco central mantuvo. El pilar monetario incluyó inicialmente un valor de referencia para $M_3$, $g_{M_3} \approx 4{,}5\%$, cuya revisión anual se abandonó en 2003. Por último, la revisión estratégica de 2021 sustituyó los dos pilares por un marco integrado de análisis económico y análisis monetario y financiero.

**Variable operativa**: el banco central controla el tipo de interés a corto plazo.

**Instrumentos**: emplea los convencionales junto a los no convencionales.

**Conclusiones**:

[17] El TFUE no aclara qué ha de entenderse exactamente por estabilidad de precios. La primera definición cuantitativa la adoptó el Consejo de Gobierno del BCE en 1998, y se revisó en 2003 ("por debajo, pero cerca, del 2 %") y en 2021 (2 % simétrico).

- *Aspectos positivos*: su transparencia facilita la evaluación pública y el anclaje de expectativas. Si los anuncios son creíbles, la estrategia reduce las ratios de sacrificio, porque la desinflación no tiene por qué ser costosa en empleo. Además, coordina a los agentes que fijan precios y salarios y es robusta frente a la innovación financiera, pues no depende de la estabilidad de ningún agregado.

- *Limitaciones*: los retardos exigen un buen seguimiento de los indicadores y hacen que la política dependa de la calidad de la previsión. Ante *shocks* de oferta puede ser procíclica, porque un *shock* positivo reduce la inflación y la expansión monetaria resultante puede alimentar burbujas, mientras que uno negativo lleva a una contracción que agrava la recesión. Por último, conserva cierta discrecionalidad a corto plazo y relega otros objetivos como el crecimiento o la estabilidad financiera, cuya omisión antes de 2008 es la crítica más repetida al régimen.

### 2.2.3. Doctrina de seguimiento del PIB nominal

A raíz de la crisis financiera global se ha debatido el control de la senda de crecimiento del PIB nominal (*NGDP targeting*)[18], que reparte los *shocks* de oferta entre precios y producción.

**Objetivo final**: la estabilidad de la senda de crecimiento del PIB nominal, con un objetivo del orden del 4-5 % anual (2 % de inflación más 2-3 % de crecimiento real), de modo que en recesión la inflación puede superar el 2 %.

**Objetivo intermedio**: no existe, porque el seguimiento del PIB nominal es, como el seguimiento directo de la inflación, una estrategia en un nivel.

**Variable operativa**: el banco central controla el tipo de interés a corto plazo.

**Instrumentos**: emplea los convencionales junto a los no convencionales.

**Conclusiones**:

- *Aspectos positivos*: habría permitido una acción más temprana en 2008, cuando el PIB nominal caía con fuerza. Además, estabiliza mejor el *output* ante *shocks* de oferta, pues reparte el efecto del *shock* entre precios y producción y exige una respuesta monetaria más moderada que un objetivo estricto de inflación.

- *Limitaciones*: el PIB se mide con menor frecuencia y mayor revisión que la inflación y la estrategia es más difícil de evaluar por el público.

### 2.2.4. Doctrina de seguimiento del nivel de precios

Alternativamente, el objetivo puede fijarse sobre una senda del nivel de precios (*price-level targeting*) y no sobre su tasa de variación. Su atractivo aparece en el límite inferior efectivo de los tipos (ELB), porque el compromiso de compensar la inflación perdida eleva las expectativas de inflación y reduce el tipo real cuando el nominal no puede bajar (Eggertsson y Woodford, 2003).

**Objetivo final**: una senda objetivo del nivel de precios, de modo que los periodos de inflación por debajo del objetivo deben compensarse con periodos por encima.

**Objetivo intermedio**: no existe, porque el seguimiento del nivel de precios también es una estrategia en un nivel.

[18] La *senda* no es un mero objetivo de tasa de crecimiento, sino que busca compensar los años malos, con un alisamiento de la variable durante el periodo considerado.

**Variable operativa**: el tipo de interés a corto plazo, complementado por el *forward guidance* (orientación prospectiva), pues la estrategia descansa en el compromiso sobre la senda futura de tipos.

**Instrumentos**: emplea los convencionales junto a los no convencionales.

**Conclusiones**:

- *Aspectos positivos*: ancla las expectativas de inflación a largo plazo y proporciona estímulo adicional en el ELB sin necesidad de bajar los tipos. Las revisiones estratégicas de la Reserva Federal (2020, *average inflation targeting*) y del BCE (2021, objetivo simétrico del 2 % con tolerancia a desviaciones transitorias al alza tras periodos en el ELB) incorporaron versiones atenuadas de esta idea.

- *Limitaciones*: exige que el público entienda y crea el compromiso de compensación, y obliga a tolerar inflación por encima del objetivo tras una recesión, lo que resulta difícil de sostener cuando esa inflación llega. De hecho, tras la inflación de 2021-2023, ambos bancos centrales retiraron esos elementos de compensación en sus revisiones de 2025. La Reserva Federal abandonó la compensación de la inflación promedio y volvió a un objetivo de inflación flexible. El BCE, por su parte, suprimió la tolerancia a desviaciones transitorias al alza tras periodos en el ELB y la sustituyó por una respuesta enérgica o persistente ante desviaciones amplias y sostenidas de la inflación en cualquier dirección.

## 2.3. Debate sobre la discrecionalidad y la credibilidad

El debate anterior puede entenderse a través de la minimización de una **función de pérdida** del banco central sujeta a la estructura de la economía. La cuestión es si esa minimización debe hacerse periodo a periodo o con arreglo a una regla y, en el segundo caso, cómo hacer creíble el compromiso.

**Política monetaria discrecional**: el banco central optimiza en cada periodo dada la situación. Su beneficio es la flexibilidad frente a *shocks* imprevistos (p. ej. la pandemia de COVID-19), y su coste, el problema de la *inconsistencia dinámica* (Kydland y Prescott, 1977). Si la autoridad monetaria tiene un objetivo de *output* superior al natural, tiene un incentivo a generar inflación sorpresa una vez que los agentes han fijado sus expectativas. Con expectativas racionales, sin embargo, los agentes lo anticipan, y el resultado es un *sesgo inflacionista* sin ganancia de *output* (Barro y Gordon, 1983a).

> El modelo de Barro-Gordon y el sesgo inflacionista se desarrollan en el tema A06, apartado 3.1.2.

**Política monetaria reglada**: la solución es plantear reglas o compromisos creíbles que aten las manos de la autoridad, ya sea el crecimiento constante de la oferta monetaria (Friedman), una inflación constante o una regla de tipos de interés. Entre las reglas de tipos, la de referencia es la *regla de Taylor* (1993)[19]:

$$i_t = \underbrace{\bar{r}_t + \pi^T}_{\bar{i}_t} + \alpha_\pi \underbrace{(\pi_t - \pi^T)}_{\tilde{\pi}_t} + \alpha_y \underbrace{(y_t - \bar{y}_t)}_{\tilde{y}_t} \;\rightarrow\; \tilde{i}_t = \alpha_\pi \tilde{\pi}_t + \alpha_y \tilde{y}_t$$

[19] No es la única especificación. Existen versiones en función del desempleo cíclico ($i_t = \bar{i}_t + \alpha_\pi \tilde{\pi}_t - \alpha_u \tilde{u}_t$) y otras como la de Clarida, Galí y Gertler (2000), que incluyen las expectativas de inflación en lugar de la inflación contemporánea (*forward looking*).

donde $\bar{i}_t$ es el tipo de interés nominal neutral (tipo real natural, $\bar{r}_t$, más objetivo de inflación, $\pi^T$), $\tilde{i}_t \equiv i_t - \bar{i}_t$ la desviación del tipo nominal respecto a ese tipo neutral, $\tilde{\pi}_t$ la desviación de la inflación y $\tilde{y}_t$ la brecha de producción (*output gap*). Si la inflación se desvía por encima de su objetivo, el tipo nominal debe subir más que proporcionalmente, $\alpha_\pi > 1$ (*principio de Taylor*), para que el tipo real suba[20].

Los beneficios de las reglas son la comprensión pública de la política, la gestión de las expectativas, indispensable dado el control imperfecto de la cantidad de dinero, y la posibilidad de delegar el trabajo en una institución evaluable. La regla de Taylor añade sencillez y buen ajuste a los datos. Los costes de las reglas son la rigidez frente a *shocks* no contemplados y, en el caso de la regla de Taylor, su carácter *backward looking*, la ausencia de un componente estocástico y la falta de microfundamentos.

**Mecanismos de credibilidad**: entre la discrecionalidad pura y la regla rígida, la literatura ha propuesto tres:

- *Prácticas reputacionales* (Barro y Gordon, 1983b): en un juego repetido indefinidamente entre la autoridad monetaria y los agentes, la creación de reputación puede resolver la inconsistencia dinámica sin necesidad de una regla formal. Para ello, hacen falta estrategias de castigo (los agentes elevan sus expectativas de inflación tras una sorpresa) y que el banco central valore suficientemente el futuro.

- *Delegación* (Rogoff, 1985): delegar la política monetaria en una autoridad independiente y más aversa a la inflación que la sociedad (un banquero central *conservador*) reduce el sesgo inflacionista, a costa de una estabilización subóptima del *output*. Los países de la OCDE han adoptado esta solución mediante la independencia del banco central, que la evidencia empírica correlaciona negativamente con la inflación media sin coste en términos de *output* (Alesina y Summers, 1993).

- *Contratos* (Walsh, 1995; Svensson, 1997b): los firmados entre el Gobierno y la autoridad monetaria modifican la función de pérdida del banco central, penalizando linealmente la inflación (Walsh) o asignándole un objetivo de inflación inferior al socialmente óptimo (Svensson). Eliminan así el sesgo inflacionista sin sacrificar la estabilización y alcanzan, en el modelo de Barro-Gordon, el resultado óptimo bajo compromiso, a diferencia del banquero conservador de Rogoff, que es una solución de *second best*. El *inflation targeting* con rendición de cuentas es la versión práctica de este contrato.

## 2.4. Política óptima: el modelo IS-PC-MR

Los debates anteriores y la determinación de la política monetaria óptima pueden ilustrarse con el modelo canónico de la nueva economía keynesiana.

> El modelo IS-PC-MR se desarrolla en el tema A07, apartado 2.1, y su relación con la política fiscal y el límite inferior de los tipos de interés (*zero lower bound*, ZLB), en el tema A40, apartado 2.1. Aquí se retiene la derivación de la regla óptima.

**Planteamiento**: el banco central elige la senda del tipo de interés real que minimiza su función de pérdida intertemporal sujeto a la estructura de la economía:

[20] El principio de Taylor es también la condición de determinación del equilibrio en el modelo neokeynesiano y la de política monetaria activa en el modelo de Leeper (tema A39, apartado 3.1).

$$\min_{r_t} \ \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[(\pi_t - \pi^T)^2 + \omega \tilde{y}_t^2\right], \quad \omega > 0, \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} \tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta}(r_t - \bar{r}_t) \\ \pi_t - \pi^T = \beta \mathbb{E}_t[\pi_{t+1} - \pi^T] + \kappa \tilde{y}_t + \varepsilon_t^\pi \end{cases}$$

donde $r_t$ es el tipo de interés real (el instrumento), $\pi_t$ la inflación, $\pi^T$ el objetivo de inflación, $\tilde{y}_t$ la brecha de producción, $\omega$ el peso relativo de la brecha en la función de pérdida y $\beta$ el factor de descuento. En las restricciones, $\bar{r}_t$ es el tipo real natural, $1/\theta$ la sensibilidad de la brecha al tipo de interés en la IS dinámica, $\kappa$ la pendiente de la curva de Phillips y $\varepsilon_t^\pi$ un *shock* de costes.

**CPO**: bajo discrecionalidad, el banco central reoptimiza en cada periodo tomando las expectativas como dadas, y, derivando el lagrangiano respecto a $\tilde{y}_t$ y $\pi_t$, se obtiene la *regla monetaria* (MR), $\tilde{y}_t = -\frac{\kappa}{\omega}(\pi_t - \pi^T)$. Esta relación negativa entre la brecha de producción y la desviación de la inflación señala, ante cada *shock*, el punto de la curva de Phillips que minimiza la pérdida. La IS traduce después ese punto en el tipo de interés que hay que fijar, lo que da lugar a una regla de tipos de la forma de Taylor.

Bajo compromiso, en cambio, la condición depende también de la brecha pasada, $\tilde{y}_t - \tilde{y}_{t-1} = -\frac{\kappa}{\omega}(\pi_t - \pi^T)$, como recoge el tercer resultado de la solución.

**CSO**: la función de pérdida es estrictamente convexa (cuadrática) y las restricciones lineales, de modo que la CPO es necesaria y suficiente y la política óptima es única.

**Solución**: del programa se desprenden tres resultados:

1. *Variable de decisión*: el instrumento es el tipo de interés real, a través del nominal y de las expectativas de inflación, y no la cantidad de dinero, que queda determinada endógenamente por la demanda. La doctrina actual se deriva, por tanto, del modelo.

2. *Aversión a la inflación* (inversa de $\omega$): el peso relativo de la brecha de producción define el tipo de banco central. Si $\omega$ es alto, la autoridad es "paloma" (*dove*) y prioriza la estabilización del *output*. En el plano $(\tilde{y}, \pi)$ la MR es entonces muy inclinada, pues la autoridad acepta más inflación a cambio de menos brecha. Si $\omega$ es bajo, es "halcón" (*hawk* o *hard-nosed*) y prioriza la inflación (la MR es plana).

3. *Sesgo inflacionista*: si la autoridad monetaria tiene un objetivo de brecha positivo ($\tilde{y}^T > 0$), la función de pérdida es $(\pi_t - \pi^T)^2 + \omega(\tilde{y}_t - \tilde{y}^T)^2$ y aparece un incentivo a generar inflación sorpresa que, con expectativas racionales, solo produce inflación (el resultado de Barro-Gordon). La solución bajo compromiso, además, difiere de la discrecional incluso sin sesgo, porque el banco central óptimo promete responder a los *shocks* de hoy también mañana (*dependencia de la historia*). Esa promesa ancla las expectativas y reduce el coste de estabilización (Clarida, Galí y Gertler, 1999), resultado que fundamenta el *forward guidance* como instrumento (apartado 3.3.1).

**Desarrollo gráfico**: la política monetaria óptima se representa en el plano $(\tilde{y}, \pi)$ (figura A36.5):

1. *Curva de Phillips* ($\text{CP}_0$): recoge la inflación asociada a cada brecha de producción, dadas las expectativas. Es creciente, con pendiente $\kappa$, porque una brecha positiva presiona al alza los precios, y sin perturbaciones pasa por el objetivo de inflación con brecha nula.

2. *Curva de isopérdida* $L(\tilde{y}, \pi) = L_1$: recoge las combinaciones de brecha e inflación con la misma pérdida, que crece al alejarse del punto $(0,\ \pi^T)$.

3. *Regla monetaria* (MR): es decreciente, con pendiente $-\omega/\kappa$, porque une las tangencias entre las curvas de isopérdida y las curvas de Phillips.

4. *Equilibrio*: sin perturbaciones, en el punto 0 la brecha es nula y la inflación está en el objetivo ($\tilde{y}_0 = 0$, $\pi_0 = \pi^T$).

**Estática comparativa**: un *shock* de costes ($\varepsilon^\pi > 0$) desplaza la curva de Phillips hacia arriba, de $\text{CP}_0$ a $\text{CP}_1$. La respuesta óptima es el punto 1 de la MR, donde $\text{CP}_1$ es tangente a la curva de isopérdida $L_1$, y reparte el ajuste entre inflación ($\pi_1 > \pi^T$) y *output* ($\tilde{y}_1 < 0$) en función de $\omega$. La doctrina de seguimiento directo de la inflación resuelve esa disyuntiva con bandas y un horizonte de medio plazo.

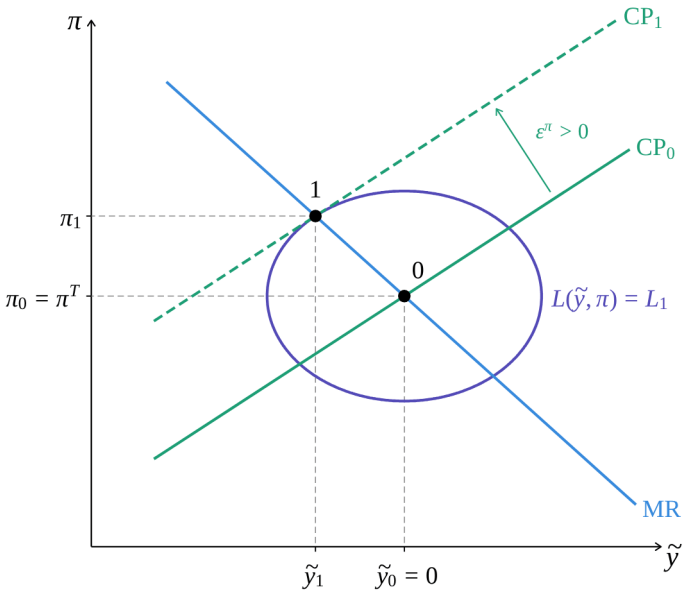



Figura A36.5: Regla monetaria óptima en el modelo IS-PC-MR. El banco central elige sobre la curva de Phillips el punto que minimiza su función de pérdida.

## 2.5. Conclusiones

El diseño moderno de la política monetaria descansa sobre un objetivo final de estabilidad de precios cuantificado y sobre el tipo de interés a corto plazo como variable operativa, dado el control imperfecto de la cantidad de dinero. Lo completan compromisos creíbles (reglas, delegación, transparencia) que resuelven la inconsistencia dinámica. La interrelación de este diseño con la política fiscal, y los regímenes de dominancia que resultan cuando la restricción presupuestaria del sector público se impone a la autoridad monetaria, se desarrollan en los temas A39 y A40.

# 3. Instrumentación de la política monetaria

Para alcanzar sus objetivos, los bancos centrales utilizan un conjunto de instrumentos y procedimientos que constituyen su **marco operativo**. Si el diseño fija qué objetivo alcanzar, el marco operativo determina cómo.

Los marcos operativos de los principales bancos centrales han evolucionado en las últimas décadas. Por ello, las medidas se agrupan en instrumentos convencionales, instrumentos no convencionales, generalizados tras la crisis financiera global de 2008 y la pandemia de COVID-19, e instrumentos cualitativos y medidas transversales, y la sección se cierra con los debates recientes sobre la instrumentación.

## 3.1. Instrumentos cuantitativos convencionales

Los instrumentos convencionales son las operaciones de mercado abierto, las facilidades permanentes y los requerimientos de reservas.

### 3.1.1. Operaciones de mercado abierto

**Definición**: son operaciones de inyección o drenaje de liquidez a las entidades en el mercado monetario, generalmente mediante subasta[21], a cambio de un activo de garantía o colateral[22]. Estas operaciones son el principal instrumento de política monetaria en la mayoría de países a través de la fijación de sus condiciones, concretamente del tipo de interés oficial.

**Objetivo**: sirven para señalizar la posición del banco central, gestionar la liquidez del sistema y controlar los tipos de interés.

**Instrumentación**: abarca cinco tipos de operaciones principales:

1. *Transacciones revertidas*: es el caso más habitual, en el que se provee o drena liquidez temporalmente contra colateral que se recompra o revende al vencimiento (*repos* y *reverse repos*). En el Eurosistema, lo son las operaciones principales de financiación semanales (MRO), las de más largo plazo (LTRO) y las de ajuste y estructurales.

2. *Transacciones en firme*: compras o ventas simples de activos, generalmente deuda pública, en las que el título cambia de manos de forma definitiva y entra o sale del balance del banco central indefinidamente. Alteran la liquidez de forma permanente, por lo que se reservan a las operaciones estructurales. Los programas de compras de activos son transacciones en firme a gran escala.

3. *Certificados de deuda*: emisión por el banco central de títulos propios a corto plazo que las entidades suscriben, drenando liquidez del sistema durante la vida del título.

4. *Swaps de tipos de cambio* (*FX swaps*): compra (venta) al contado de divisa contra moneda nacional con venta (compra) simultánea a plazo. Así, la liquidez en moneda nacional se inyecta (drena) durante el plazo del *swap* sin alterar la posición cambiaria del banco central. Se emplean en las operaciones de ajuste y no deben confundirse con las líneas de *swap* entre bancos centrales, que proveen divisa extranjera.

[21] El Eurosistema recurre a subastas estándar (semanales) o rápidas, a las que solo acuden las entidades de contrapartida autorizadas. Desde octubre de 2008, las MRO se adjudican a tipo fijo con adjudicación plena.

[22] Generalmente, activos financieros cuyo emisor tiene una elevada solvencia, como bonos del Estado o papel comercial. El Eurosistema publica la lista de activos admisibles y aplica recortes de valoración (*haircuts*).

5. *Depósitos a plazo fijo*: el banco central ofrece a las entidades la constitución de depósitos remunerados a un plazo determinado, generalmente mediante subasta rápida, drenando así liquidez por ese periodo. Estos depósitos son la operación de ajuste típica para absorber excesos puntuales, y el BCE la utilizó semanalmente entre 2010 y 2014 para esterilizar las compras del programa para los mercados de valores (SMP).

El diseño del *sistema de garantías* (colaterales) es, además, un "subinstrumento", en la medida en que exigir menos garantías o admitir distintos activos altera cualitativamente las condiciones.

**Efectos**: las OMA tienen dos:

- *Efecto cantidad*: alteran la base monetaria y, vía multiplicador, la oferta monetaria.

- *Efecto precio* (o liquidez): si la compraventa tiene por objeto valores de renta fija, influyen en su cotización y en su tipo de interés efectivo[23].

**Conclusiones**:

- *Aspectos positivos*: son un instrumento muy popular por su flexibilidad en magnitud (grandes cantidades o pequeñas dosis) y en dirección (pueden revertirse sin dificultades para las entidades).

- *Limitaciones*: su efecto sobre la oferta monetaria depende de la estabilidad del multiplicador, que el banco central no controla (apartado 1.4), y exigen mercados de deuda y de colateral profundos y líquidos. Además, con los tipos en su límite inferior pierden eficacia, porque el dinero y los activos a corto plazo pasan a ser casi sustitutos perfectos.

### 3.1.2. Facilidades permanentes

**Definición**: sistema que permite a las entidades de contrapartida acceder a liquidez, o depositarla, en condiciones fijas, es decir, a precios determinados por el banco central que no son de mercado y que generalmente son desfavorables. Este sistema es descentralizado, accesible por iniciativa de las entidades y de uso puntual.

**Objetivo**: las facilidades permanentes sirven para acotar las fluctuaciones del tipo de interés interbancario a un día y proveer o absorber liquidez de emergencia.

**Instrumentación**: adoptan dos formas generales, con nombres distintos en cada banco central:

1. *Facilidad de crédito* (*lending facility*): permite a las entidades obtener liquidez a un día contra activos de garantía elegibles, a un tipo superior al de mercado. En el Eurosistema se conoce como la facilidad marginal de crédito y, en la Reserva Federal, como la ventanilla de descuento (*discount window*), instrumentada de forma descentralizada por los doce bancos federales[24]. Otros bancos centrales siguen otras denominaciones (p. ej. facilidad de crédito complementaria en el caso del BoJ).

2. *Facilidad de depósito* (*deposit facility*): permite depositar liquidez a un día a un tipo inferior al de mercado. En el Eurosistema se conoce como facilidad de depósito y, en el BoJ, como facilidad de depósito complementaria (2008). La Reserva Federal, en cambio, no tiene una facilidad de depósito propiamente dicha. En su lugar, el suelo lo forman la remuneración de las reservas (desde 2008; *interest on reserve balances*, IORB, desde 2021) y, para las entidades sin acceso a ella, la facilidad de *reverse repo* a un día (ON RRP, 2013).

[23] $r_B = (\text{cup}/P_B) \cdot 100$, con $r_B$ el tipo de interés efectivo del bono (rendimiento corriente), cup el cupón y $P_B$ el precio del bono.

[24] Regla de Bagehot (1873): el banco central debe prestar libremente, a tipos penalizadores y contra buen colateral.

**Efectos**: la facilidad de crédito fija un *techo* a los tipos interbancarios y la de depósito un *suelo*, y entre ambos delimitan el *pasillo* (*corridor*) dentro del cual se sitúa el tipo de las operaciones principales.

Cuando el sistema opera con liquidez abundante, el tipo de mercado se pega al suelo y el pasillo degenera en un *sistema de suelo* (*floor system*). La Reserva Federal opera con él desde 2008, y el Eurosistema, de hecho desde 2015 y de derecho desde su nuevo marco operativo de 2024[25].

**Ejemplos**: el BoJ introdujo su facilidad de crédito complementaria en 2001; la ventanilla de descuento de la Reserva Federal presta con una prima sobre el tipo de los fondos federales.

**Conclusiones**:

- *Aspectos positivos*: acotan el tipo a un día sin que el banco central tenga que intervenir, porque las entidades acceden a ellas por iniciativa propia. Las facilidades ofrecen, además, una red de seguridad ante necesidades puntuales de liquidez.

- *Limitaciones*: acudir a la facilidad de crédito puede estigmatizar a la entidad, lo que resta eficacia a esa red. Con liquidez abundante, además, el pasillo pierde su función y el control del tipo pasa a depender de la remuneración de las reservas, cuyo coste para el banco central crece con su balance.

### 3.1.3. Requerimientos de reservas: coeficiente de caja

**Definición**: la autoridad monetaria exige a cada banco comercial mantener en el banco central, como saldo en efectivo, una fracción de sus depósitos, $q_L = R_L^b/D$. Ese dinero queda inmovilizado, pues la entidad no puede prestarlo ni invertirlo.

**Objetivo**: en la actualidad cumple dos funciones:

- *Estabilización*: como basta con cumplir el coeficiente en media durante el periodo de mantenimiento, las entidades usan sus reservas como amortiguador y no necesitan salir al interbancario ante cada desajuste. Por ello, su demanda de reservas se vuelve más estable y los tipos a corto plazo sufren menos tensiones.

- *Refuerzo*: deja a la banca con un déficit estructural de liquidez que la obliga a recurrir al banco central, y esa mayor dependencia aumenta la capacidad de este para controlar los tipos.

**Instrumentación**: el banco central fija el coeficiente y decide si remunera las reservas mínimas.

**Efectos**: un incremento del coeficiente reduce el multiplicador y, con él, la capacidad de conceder crédito y la oferta monetaria.

**Ejemplos**: el Eurosistema lo sitúa en el 1 % desde 2012 y no lo remunera desde 2023; la Reserva Federal lo eliminó en 2020.

**Conclusiones**:

- *Aspectos positivos*: su coactividad y su precisión lo hacen poderoso, y el coeficiente afecta a la vez a todas las entidades sujetas, sin depender de que acudan a las operaciones del banco central.

[25] El nuevo marco operativo del Eurosistema (2024) estrecha el pasillo y fija el tipo de las MRO 15 pb por encima del de la facilidad de depósito. Este último pasa a ser el tipo que guía la orientación de la política monetaria en un sistema de liquidez abundante suministrada a demanda.

- *Limitaciones*: las dificultades que genera a las entidades impiden cambiarlo con frecuencia, por lo que se considera de acción discontinua, y, si no se remunera, grava a la banca frente a los intermediarios no sujetos. Su popularidad ha decrecido, también por la competencia regulatoria entre países para atraer fondos.

## 3.2. Instrumentos cuantitativos no convencionales

Los canales habituales de transmisión quedaron dañados con la crisis financiera global, y los instrumentos convencionales ya no bastaban para cumplir los objetivos de inflación. Para reactivar esos canales, los bancos centrales recurrieron a medidas no convencionales[26]. Estas medidas son difíciles de sistematizar, y aquí se agrupan en cinco tipos.

Tres causas las justifican:

- *Confianza*: la crisis de confianza, con el aumento de las primas de riesgo de los países periféricos y el riesgo de contagio.

- *Transmisión*: la ruptura del mecanismo de transmisión, tanto en el canal interbancario como en el crédito a la economía real.

- *ZLB*: el límite inferior de los tipos de interés, que impidió seguir bajando los tipos oficiales en una situación próxima a la trampa de liquidez.

### 3.2.1. Compras masivas de activos

**Definición**: consiste en la compra masiva de activos financieros (deuda pública a largo plazo, cédulas hipotecarias, deuda privada…) por parte de la autoridad monetaria a través de programas de compras, la llamada *expansión cuantitativa* (*quantitative easing*, QE). La diferencia esencial con las OMA es que los programas de compras ponen el foco en un grupo escogido de activos, generalmente de gran volumen, y no en las condiciones ofrecidas al sistema financiero.

**Objetivo**: sortear el ZLB actuando sobre los tipos a largo plazo y las primas de riesgo cuando el tipo oficial ya no puede bajar, y restablecer la transmisión en segmentos dañados.

**Instrumentación**: adquisiciones en firme, generalmente en el mercado secundario, con un volumen mensual y una duración anunciados, y reinversión de los vencimientos mientras dura el programa.

**Efectos**: operan a través de tres canales:

- *Transmisión directa*: la demanda de los activos comprados eleva su precio, reduce su rendimiento y anima a los bancos a prestar.

- *Reequilibrio de carteras*: los vendedores reinvierten en otros activos, cuyos precios suben y rendimientos caen, lo que reduce los costes de financiación en los mercados de capitales.

- *Señalización*: las compras señalan que los tipos oficiales se mantendrán bajos durante un periodo prolongado, lo que reduce la incertidumbre.

**Ejemplos**: destacan los programas de tres bancos centrales:

- *BoJ*: llevó a cabo el QE por primera vez en 2001 en respuesta a la deflación.

[26] Aunque algunas de estas medidas se han mantenido durante años, todas fueron pensadas como transitorias, con una retirada gradual ligada a la normalización de la situación económica.

▪ *Reserva Federal*: el QE1 (2008-2010) compró 1,25 billones de dólares en títulos respaldados por hipotecas (MBS), unos 175.000 M$ en deuda de agencias y 300.000 M$ en deuda del Tesoro en respuesta a la crisis *subprime*. El QE2 (2010) añadió 600.000 M$ en deuda del Tesoro, y el QE3 (2012) fue abierto, con compras mensuales de 85.000 M$ en deuda del Tesoro y MBS hasta 2014, cuando el balance alcanzó los 4,5 billones de dólares. La respuesta a la pandemia (2020) lo llevó por encima de los 8,9 billones.

▪ *BCE*: el programa de compras de activos (APP), creado en 2014, agrupa los programas de compras de bonos garantizados, de bonos de titulización y de valores de los sectores público y corporativo. El programa de compras de emergencia frente a la pandemia (PEPP), en cambio, es independiente. Los programas se sucedieron en tres etapas:

1. *SMP y CBPP1-2*: el SMP (2010)[27] y los programas de adquisiciones de bonos garantizados (CBPP1 y CBPP2, 2009 y 2011)[28] para restablecer la transmisión.

2. *CBPP3 y ABSPP*: el CBPP3 y el programa de compras de bonos de titulización de activos (ABSPP, 2014), con operativa en los mercados primario y secundario[29].

3. *PSPP, CSPP y PEPP*: el QE europeo propiamente dicho comienza en 2015 con el programa de compras de valores públicos (PSPP), de compra exclusivamente en el mercado secundario de deuda del sector público de los Estados miembros de la zona del euro[30]. Le siguieron en 2016 el programa de compras de bonos corporativos (CSPP) y en marzo de 2020 el PEPP, dotado inicialmente con 750.000 M€ y ampliado hasta 1,85 billones, con flexibilidad entre países y clases de activos. Las compras netas cesaron en 2022 y desde 2023 el balance se reduce (*quantitative tightening*).

**Conclusiones**:

▪ *Aspectos positivos*: permiten seguir relajando las condiciones financieras cuando el tipo oficial ya no puede bajar, y la evidencia les atribuye reducciones sustanciales de los tipos a largo plazo (tema A37, apartado 3.5.3).

▪ *Limitaciones*: amplían el balance del banco central y lo exponen a pérdidas cuando suben los tipos, porque la remuneración de las reservas creadas puede superar el rendimiento de los bonos comprados. Además, la compra masiva de deuda pública acerca la política monetaria a la financiación del sector público.

### 3.2.2. Operaciones de mercado abierto no convencionales

**Definición**: OMA que no destacan por su volumen sino por características diferentes a las habituales, como plazos mucho más largos, condicionalidad o carácter de emergencia.

[27] Conllevó la compra de títulos de deuda soberana en el mercado secundario ante las tensiones en los mercados de deuda de los países periféricos.

[28] Los bonos garantizados o *covered bonds* son instrumentos de renta fija asegurados por una cartera de préstamos hipotecarios o de deuda pública contra la que se tiene un derecho preferente en caso de impago. Constituyen una fuente de financiación importante de las entidades de crédito.

[29] Los bonos garantizados pueden ser retenidos por sus emisores para ofrecerlos como garantía ante el banco central o estar en manos de inversores. Si son retenidos, la compra es una mera sustitución de liquidez. Si no, impacta sobre la financiación de la economía real, en función de en qué se invierta la liquidez obtenida.

[30] A los países bajo programa de rescate se les impusieron condiciones adicionales: la deuda chipriota fue elegible, sujeta a las revisiones de su programa. La griega, en cambio, quedó fuera del PSPP por no cumplir el requisito de calidad crediticia (sí se admitió después en el PEPP).

**Objetivo**: sustituir temporalmente al mercado interbancario cuando este deja de funcionar y, en su versión condicionada, dirigir la liquidez hacia el crédito a la economía real.

**Instrumentación**: el BCE ha empleado tres tipos:

1. *LTRO a tres años* (2011-2012): operaciones de refinanciación a mucho más largo plazo que las subastas semanales (vencimiento a tres años frente a siete días), con un tipo atractivo (alrededor del 1 %), una amplia gama de colaterales y sin estigmatización de las entidades.

2. *TLTRO* (Targeted, 2014-2021): operaciones a plazo amplio condicionadas a la concesión de crédito bancario (excluidas las hipotecas), con un tipo tanto más favorable cuanto más crédito se concede a empresas y hogares[31].

3. *PELTRO* (operaciones de financiación a plazo más largo sin objetivo específico de emergencia frente a la pandemia, 2020): siete operaciones nuevas anunciadas en abril de 2020, a un tipo 25 pb inferior al de las MRO.

**Efectos**: las LTRO a tres años inyectaron más de un billón de euros en dos subastas ("barra libre de liquidez") y aliviaron las tensiones de financiación de la banca europea, pero la liquidez se empleó para sanear balances y no para reactivar el crédito. De ahí la condicionalidad de las TLTRO, vinculada al crédito a empresas no financieras y hogares. Las PELTRO, por su parte, aportaron liquidez al sistema financiero durante la pandemia.

**Conclusiones**:

- *Aspectos positivos*: sostienen la financiación de la banca a plazos que el mercado ya no ofrece, y la condicionalidad de las TLTRO vincula la liquidez al crédito a hogares y empresas.

- *Limitaciones*: su efecto sobre el crédito depende de la demanda de préstamos y del estado de los balances bancarios, como mostraron las LTRO a tres años. La financiación prolongada a tipos favorables mantiene, además, la dependencia de la banca respecto del banco central.

### 3.2.3. Políticas innovadoras de tipos de interés

**Definición**: medidas que actúan sobre la remuneración de las reservas para llevar el tipo de interés por debajo de cero o para diferenciar el tipo según la operación.

**Objetivo**: buscan eliminar, al menos virtualmente, el límite inferior de los tipos y estimular la concesión de crédito.

**Instrumentación**: abarca dos medidas:

1. *Tipos de interés negativos* (2009): cobrar a los bancos por el mantenimiento de su exceso de reservas en lugar de remunerarlo.

[31] Hubo tres rondas. La TLTRO I (2014) constó de ocho operaciones trimestrales. En las dos primeras cada entidad podía pedir hasta el 7 % de su cartera viva de préstamos elegibles, y en las seis siguientes, hasta el triple del crédito neto concedido por encima de una referencia. La TLTRO II (2016) se compuso de cuatro operaciones a cuatro años por hasta el 30 % del saldo vivo. Su tipo, comprendido entre el de la facilidad de depósito y el aplicado en las MRO, podía resultar negativo, y en ese caso era el BCE quien pagaba a las entidades por el crédito concedido. La TLTRO III (2019-2021), de operaciones a tres años, ofreció durante la pandemia un tipo de hasta 50 pb por debajo de la facilidad de depósito media.

2. *Sistema de tipos dual* (2019): remuneración de las reservas en dos tramos (*tiering*), según el cual un múltiplo de las reservas mínimas queda exento del tipo negativo de la facilidad de depósito[32]. Combinado con las TLTRO, permite fijar tipos distintos para prestamistas y prestatarios.

**Efectos**: los tipos negativos muestran que el límite efectivo está por debajo de cero. El sistema dual, por su parte, genera un efecto renta positivo sobre acreedores y deudores a la vez.

**Ejemplos**: los tipos negativos fueron implementados por primera vez por el Riksbank (2009) sobre la facilidad de depósito, y seguidos por el Banco de Dinamarca (2012), el BCE (2014-2022), el Banco Nacional Suizo (BNS)[33] y el BoJ. El sistema dual lo introdujo el BCE en 2019, después de que el BNS (2015) y el BoJ (2016) eximieran ya del tipo negativo parte de las reservas.

**Conclusiones**:

- *Aspectos positivos*: amplían el margen de bajada del tipo oficial cuando ya ha llegado a cero, y el sistema dual reduce lo que esos tipos cuestan a la banca.

- *Limitaciones*: tienen rendimientos decrecientes y efectos adversos sobre la rentabilidad bancaria, y por debajo de cierto umbral (*reversal rate*) la bajada contrae el crédito (tema A37, apartado 3.5.2). El sistema dual, además, desdibuja la línea entre política monetaria y fiscal.

### 3.2.4. Medidas de recomposición del balance

**Definición**: operaciones que actúan sobre la composición del activo del banco central y no sobre su tamaño, por lo que son formas menos agresivas de actuación que el QE.

**Objetivo**: buscan alterar las condiciones del crédito en segmentos concretos o la pendiente de la estructura temporal de los tipos de interés (ETTI).

**Instrumentación**: abarca dos medidas:

1. *Relajación del crédito* (*credit easing*, 2009): sustitución de unos activos por otros (p. ej. deuda pública por papel comercial o préstamos). Bernanke la define por la composición del activo, aunque en 2008-2009 vino acompañada de una fuerte expansión del balance de la Reserva Federal[34].

2. *Operación twist* (2011)[35]: compraventa simultánea de bonos del Tesoro, vendiendo a corto plazo y comprando a largo.

**Efectos**: la relajación del crédito mejora las condiciones del segmento comprado. La operación *twist*, por su parte, aplana la ETTI al reducir los tipos a largo sin ampliar el balance (los tipos a corto siguen anclados en el ZLB por el tipo oficial), para recuperar la inversión y el consumo a crédito.

**Ejemplos**: el término *credit easing* fue acuñado por Ben Bernanke para diferenciar la política de la Reserva Federal de la del BoJ (2001-2006), y la operación *twist* la llevó a cabo la Reserva Federal entre 2011 y 2012.

[32] Solo para los saldos mantenidos en la cuenta corriente en el banco central nacional, y no para los mantenidos en la facilidad de depósito. El múltiplo se fijó en seis veces las reservas mínimas.

[33] En el caso de Suiza la motivación fue distinta, porque el franco, por su estatus de activo refugio, sufre presiones apreciatorias. Por ello, el BNS fijó sus tipos en el −0,75 % entre 2015 y 2022.

[34] Se diferencia de la expansión cuantitativa en que se centra en la relación activo-activo y no en la relación activo-pasivo. Uno de los programas de la Reserva Federal, la Commercial Paper Funding Facility, afectó de forma positiva a los prestatarios del mercado de papel comercial.

[35] La operación *twist* original se llevó a cabo en 1961, cuando el Comité Federal de Mercado Abierto buscaba sostener el dólar sin frenar la recuperación.

**Conclusiones**:

- *Aspectos positivos*: actúan sobre segmentos o plazos concretos sin ampliar necesariamente el balance ni la base monetaria.

- *Limitaciones*: su alcance está acotado por los activos que el banco central tiene para vender, y la relajación del crédito genera ganadores y perdedores, lo que la acerca a la política fiscal.

### 3.2.5. Líneas de *swap*

**Definición**: acuerdos de permuta de divisas entre bancos centrales, que permiten al banco central de un área obtener la divisa de otra para suministrarla a sus bancos.

**Objetivo**: atender la demanda de divisa extranjera, en particular de dólares, de los bancos de un área cuando los mercados de financiación en esa divisa se cierran.

**Instrumentación**: el banco central extranjero entrega su moneda a la Reserva Federal a cambio de dólares al tipo de cambio de mercado y se compromete a recomprarla al mismo tipo en una fecha futura (de un día a tres meses), más un interés. El banco central extranjero presta después esos dólares a sus bancos y asume el riesgo de crédito, mientras que la Reserva Federal no soporta riesgo cambiario.

**Efectos**: abaratan la financiación en dólares fuera de Estados Unidos y reducen las tensiones en los mercados monetarios (Goldberg, Kennedy y Miu, 2011).

**Ejemplos**: tradicionalmente de uso puntual, han ganado peso desde la crisis financiera. Entre diciembre de 2007 y octubre de 2008, la Reserva Federal abrió líneas temporales con catorce bancos centrales, vigentes hasta febrero de 2010 (Goldberg, Kennedy y Miu, 2011). Las del BCE, el BoJ, el Banco de Inglaterra, el BNS y el Banco de Canadá, reabiertas en mayo de 2010, son permanentes desde octubre de 2013. En marzo de 2020 volvieron a ser decisivas, y la Reserva Federal abrió además líneas temporales con otros nueve bancos centrales.

**Conclusiones**:

- *Aspectos positivos*: contienen las crisis de financiación en divisa sin que la Reserva Federal asuma riesgo de crédito frente a bancos extranjeros ni riesgo cambiario.

- *Limitaciones*: solo alcanzan a los bancos centrales con los que existe acuerdo, que elige la Reserva Federal, por lo que las demás economías han de recurrir a sus propias reservas de divisas.

## 3.3. Instrumentos cualitativos y medidas transversales

A los instrumentos cuantitativos se suman dos instrumentos cualitativos, la comunicación y el control del crédito, y dos medidas transversales, la rebaja de las calificaciones crediticias exigidas al colateral y los programas de compras condicionadas.

### 3.3.1. Comunicación y gestión de expectativas

**Definición**: consiste en la provisión de información por parte del banco central acerca del curso futuro de la política monetaria (*forward guidance*), en la línea de la dependencia de la historia de la política óptima bajo compromiso (apartado 2.4). Junto a ella figura la *persuasión moral* (*jawboning*, *open mouth operations*), que consiste en persuadir a los agentes para alcanzar un objetivo sin recurrir a operaciones.

**Objetivo**: la comunicación busca gestionar las expectativas de los agentes sobre los tipos futuros y, con ellas, los tipos a largo plazo.

**Instrumentación**: la orientación puede ser *cualitativa* ("durante un periodo prolongado"), *basada en el calendario* (hasta una fecha) o *dependiente del estado* de la economía (hasta alcanzar umbrales de desempleo o inflación). Se transmite en los comunicados, las conferencias de prensa y las actas, y la refuerzan las comparecencias de la presidencia del BCE ante la Comisión de Asuntos Económicos y Monetarios del Parlamento Europeo.

**Efectos**: si es creíble, reduce los tipos esperados y los tipos a largo plazo sin mover el tipo oficial. La orientación *délfica* (*Delphic*), que solo anuncia la previsión del banco central, puede leerse como señal de debilidad económica, a diferencia de la *odiseica* (*Odyssean*), que compromete su actuación futura (Campbell et al., 2012).

**Ejemplos**: la Reserva Federal pasó de la orientación cualitativa (2008) a la basada en el calendario (2011) y a la dependiente de umbrales (2012, mientras el desempleo superase el 6,5 %). El BCE, por su parte, la incorporó formalmente en julio de 2013, un giro notable respecto de su rechazo previo a prefijar los tipos. La declaración de Mario Draghi en 2012, "dentro de nuestro mandato, el BCE está dispuesto a hacer todo lo necesario para preservar el euro. Y créanme, será suficiente"[36], no fue *forward guidance* sobre los tipos futuros, sino un compromiso verbal que precedió al anuncio de las operaciones monetarias de compraventa (OMT).

El ejemplo paradigmático de persuasión moral es el rescate privado del fondo Long-Term Capital Management orquestado por la Reserva Federal de Nueva York en 1998 sin aportar financiación.

**Conclusiones**:

- *Aspectos positivos*: actúa sobre los tipos a largo plazo sin coste para el balance y sigue disponible cuando el tipo oficial está en su límite inferior.

- *Limitaciones*: su eficacia depende de la credibilidad, porque el compromiso es dinámicamente inconsistente (apartado 2.3), y ata la reputación del banco central a anuncios que las circunstancias pueden obligar a revisar.

### 3.3.2. Control y supervisión del crédito

**Definición**: medidas de control selectivo del crédito (coeficientes específicos, crédito oficial) y *acción directa* sobre las entidades que no colaboran con el banco central (condiciones más restrictivas, negativa a operar, sanciones).

**Objetivo**: busca dirigir fondos hacia sectores prioritarios de la economía y limitar la expansión del crédito sin subir los tipos de interés.

[36] Discurso en la *Global Investment Conference* de Londres, 26 de julio de 2012.

**Instrumentación**: se instrumenta con topes al crecimiento del crédito de cada entidad, coeficientes de inversión obligatoria en activos o sectores determinados y redescuento preferente para el crédito a esos sectores.

**Efectos**: limitan el volumen del crédito de las entidades y orientan su destino, pero alientan la desintermediación hacia canales no controlados.

**Ejemplo**: el *corset* del Banco de Inglaterra (1973-1980) obligaba a constituir depósitos no remunerados en el banco central a las entidades cuyos pasivos remunerados crecieran por encima de un límite. Sin embargo, la desintermediación a través de letras bancarias fuera del balance minó su eficacia.

**Conclusiones**:

- *Aspectos positivos*: actúan con rapidez y de forma selectiva, algo que no permite el tipo de interés, que afecta a toda la economía.

- *Limitaciones*: distorsionan la asignación del crédito, y la desintermediación erosiona su eficacia, como mostró el *corset*. Por ello, las economías avanzadas los abandonaron con la liberalización financiera de los años ochenta.

### 3.3.3. Rebaja de las calificaciones crediticias

**Definición**: consiste en reducir las exigencias de calificación crediticia de los activos utilizados como colateral en las operaciones del Eurosistema.

**Objetivo**: evitar que las rebajas de calificación restrinjan la liquidez del sistema.

**Instrumentación**: rebaja del umbral mínimo de calidad crediticia, suspensión del umbral para un emisor o mantenimiento de la admisibilidad de los activos que pierden la calificación exigida, con recortes de valoración (*haircuts*) mayores para compensar el riesgo.

**Efectos**: amplía el colateral disponible y evita que una rebaja de calificación expulse de golpe a las entidades de la financiación del banco central, a cambio de más riesgo de crédito en su balance.

**Ejemplos**: el Eurosistema rebajó el umbral de A- a BBB- en octubre de 2008 y lo suspendió para la deuda pública griega en mayo de 2010. En abril de 2020 admitió como colateral los activos que perdieran el grado de inversión (*fallen angels*) hasta BB, ante las rebajas derivadas de la pandemia.

**Conclusiones**:

- *Aspectos positivos*: evita que una decisión de las agencias de calificación, ajena al banco central, endurezca de forma automática las condiciones de liquidez del sistema.

- *Limitaciones*: los recortes de valoración solo compensan en parte el mayor riesgo de crédito que asume el banco central. Además, al suspender el umbral para un emisor concreto, como Grecia en 2010, el banco central pasa a juzgar el riesgo de un soberano, una decisión de carácter casi fiscal.

### 3.3.4. Programas de compras condicionadas

**Definición**: consiste en la disposición anunciada del banco central a adquirir deuda soberana en el mercado secundario sujeta a condiciones.

**Objetivo**: busca afrontar las crisis de confianza y evitar la fragmentación financiera de la UEM sin desactivar la transmisión de la política monetaria.

**Instrumentación**: compras en firme en el mercado secundario, sin límite cuantitativo *ex ante*, para países que cumplan condiciones de política económica.

**Efectos**: su eficacia descansa en el anuncio, que descarta una crisis de deuda autocumplida (tema B16, apartado 3.1) y reduce las primas de riesgo sin necesidad de comprar.

**Ejemplos**: el BCE dispone de dos programas:

- *OMT* (operaciones monetarias de compraventa, 2012): bonos de uno a tres años de países sujetos a un programa del Mecanismo Europeo de Estabilidad (MEDE). Nunca se han activado, pero contribuyeron a la senda descendente de las primas de riesgo[37].

- *TPI* (Instrumento para la Protección de la Transmisión, 2022): valores públicos de uno a diez años de países cuya prima de riesgo se aleje de los fundamentos y que cumplan los criterios fiscales y macroeconómicos de la UE. Anunciado al inicio del ciclo de subidas de tipos, cumplió la misma función disuasoria.

**Conclusiones**:

- *Aspectos positivos*: si el anuncio es creíble, protegen la transmisión con un coste mínimo para el balance, pues pueden no llegar a activarse, y la condicionalidad limita el riesgo moral de los Gobiernos.

- *Limitaciones*: la condicionalidad, ligada a un programa del MEDE o a criterios fiscales, puede retrasar la ayuda a los países que más la necesitan. La compra de deuda de Estados concretos expone, además, al banco central a impugnaciones por financiación monetaria, como la que afrontaron las OMT ante el TJUE.

## 3.4. Debates recientes sobre la instrumentación

**Dinero helicóptero** (Friedman, 1969): el término se debe a su ensayo y ha sido rebautizado como *QE for the people*. La idea es que, en lugar de comprar bonos, los bancos centrales inyecten liquidez directamente a los consumidores. No obstante, ningún banco central la ha puesto en marcha, porque el dinero helicóptero difumina la línea divisoria entre política fiscal y monetaria, y se contempla como un último recurso frente a la deflación sistémica.

**Monedas digitales de banco central** (CBDC): la eliminación del efectivo de alta denominación que propone Rogoff (2016) suprimiría el ZLB y permitiría imponer tipos nominales negativos, a costa de la pérdida de anonimato y de la exclusión de los colectivos menos digitalizados. Las CBDC minoristas, como el euro digital, cuya fase de preparación concluyó en 2025 y cuyo reglamento negocian desde julio de 2026 el Parlamento Europeo y el Consejo, se plantean, en cambio, como complemento y no sustituto del efectivo. Su diseño (límites de tenencia, no

[37] Las OMT fueron impugnadas ante el Tribunal Constitucional alemán, que planteó al Tribunal de Justicia de la Unión Europea (TJUE) una cuestión prejudicial sobre la compatibilidad del programa con la prohibición de financiación monetaria. En 2015 (asunto *Gauweiler*), el TJUE confirmó que los Tratados autorizan al Sistema Europeo de Bancos Centrales a adoptar un programa de las características de las OMT.

remuneración) busca evitar la desintermediación bancaria. Al no estar remunerada, sin embargo, una CBDC rinde, como el efectivo, un 0 % nominal y mantiene el límite inferior de los tipos.

> La eliminación del efectivo, las monedas digitales de banco central y su efecto sobre la demanda de dinero en el ZLB, con su representación gráfica, se desarrollan en el tema A35, apartado 3.3.

## 3.5. Conclusiones

A los instrumentos convencionales, que operan sobre el tipo de interés a corto plazo, se han sumado desde 2008 otros que actúan sobre el tamaño y la composición del balance y sobre las expectativas. Las operaciones de mercado abierto sitúan ese tipo dentro del pasillo que delimitan las facilidades permanentes, y con liquidez abundante el pasillo se convierte en un sistema de suelo. Cuando la crisis dañó la transmisión y los tipos llegaron a su límite inferior, los bancos centrales compraron activos a gran escala, prestaron a largo plazo con condiciones ligadas al crédito y llevaron los tipos por debajo de cero. La comunicación y los programas de compras condicionadas actúan, en cambio, sobre todo a través del anuncio, y su eficacia depende de la credibilidad del banco central.

La lección del periodo es que el límite inferior de los tipos no agota la política monetaria, pero que los instrumentos que lo sortean acercan al banco central a la política fiscal. El sistema de tipos dual y el dinero helicóptero desdibujan esa frontera.

## Conclusiones

El diseño de la política monetaria ha pasado del control de agregados monetarios al seguimiento directo de la inflación, con el tipo de interés a corto plazo como variable operativa. Justifican el cambio el control imperfecto de la cantidad de dinero y la inestabilidad de su demanda. La inconsistencia dinámica se resuelve con compromisos creíbles (reglas como la de Taylor, reputación, delegación o contratos). El modelo IS-PC-MR muestra que la política óptima es una regla de tipos que reparte el ajuste ante los *shocks* según la aversión a la inflación del banco central. Desde 2008, la instrumentación se ha ampliado con compras de activos, operaciones condicionadas, tipos negativos y *forward guidance*. El límite inferior de los tipos no agota la política monetaria, pero los instrumentos que lo sortean la acercan a la fiscal.

El diseño de la política monetaria se apoya en la teoría de la demanda de dinero, cuya inestabilidad motivó el abandono de los agregados (tema A35), y se completa con el estudio de su transmisión, convencional y no convencional (tema A37). Sus dos marcos analíticos proceden de las escuelas macroeconómicas, la inconsistencia dinámica de Barro y Gordon (tema A06) y el modelo de tres ecuaciones (tema A07). Los instrumentos que sortean el límite inferior, como el dinero helicóptero, plantean el riesgo de dominancia fiscal (tema A39) y obligan a coordinar las políticas monetaria y fiscal (tema A40). Por último, el Eurosistema aplica estos principios en una unión monetaria sin un Tesoro común (tema B43).

Holston, Laubach y Williams (2017) estiman una caída persistente del tipo de interés natural en las principales economías avanzadas, que reduce el margen hasta el límite inferior. Ante ese menor margen, Bernanke, Kiley y Roberts (2019) muestran que mantener los tipos bajos más tiempo tras tocar el límite conserva parte de sus beneficios aunque solo los mercados financieros anticipen el compromiso. La Reserva Federal (2020) y el BCE (2021) recogieron esa estrategia, atenuada, y retiraron sus elementos de compensación en 2025 (apartado 2.2.4). También se discute el tamaño del balance. Borio (2023) defiende un balance tan pequeño y poco arriesgado como permita el mandato y la vuelta a sistemas de pasillo con reservas escasas, porque los costes de los sistemas de suelo son tardíos y poco visibles. El nuevo marco operativo del Eurosistema (2024) opta, en cambio, por liquidez abundante a demanda (apartado 3.1.2).

# A37. Política monetaria II: transmisión y efectos

Política monetaria (II). Los mecanismos de transmisión de la política monetaria convencional. Rigideces de precios, rigideces de salarios y fricciones financieras. Los mecanismos de transmisión de la política monetaria no convencional.

## Introducción

> **Introducción común.** Los temas A36 y A37 comparten una introducción, que se desarrolla al comienzo del tema **A36**. Define la política monetaria y resume cómo ha cambiado el consenso sobre sus efectos, hasta el actual, que encomienda a los bancos centrales el control de la inflación, la estabilización a corto plazo y, cada vez más, la estabilidad financiera.

Que la **política monetaria** influye en la economía real a corto plazo es hoy poco discutido, pero cómo lo hace, con qué intensidad y con qué retardos sigue siendo objeto de investigación. Friedman describió esos retardos como "largos y variables", y la evidencia confirma la neutralidad del dinero a largo plazo y efectos reales a corto plazo, asimétricos y difíciles de medir. Este tema completa así el estudio del diseño y los instrumentos de la política monetaria (tema A36) con el de sus efectos.

El **mecanismo de transmisión** es la cadena que lleva de una decisión del banco central sobre los tipos de interés oficiales a la demanda agregada, la producción y los precios. El canal tradicional opera a través del coste del capital y del consumo, pero la crisis financiera global puso en primer plano el **canal del crédito** (Bernanke y Gertler, 1995), en el que la situación financiera de bancos, empresas y hogares amplifica los impulsos monetarios, y el de las expectativas, que depende de la credibilidad del banco central. Cada escuela, de la clásica a la nueva economía keynesiana, ha dado una respuesta distinta a la pregunta por la neutralidad del dinero, según el papel que atribuye a las rigideces y a las expectativas.

Con los tipos de interés en su límite inferior, al que la Reserva Federal llegó a finales de 2008 y el BCE entre 2012 y 2014, los bancos centrales recurrieron a **instrumentos no convencionales**, como la expansión cuantitativa (QE) y el *forward guidance* (orientación prospectiva), que actúan sobre las primas de plazo, los precios de los activos y las expectativas y cuya eficacia sigue en debate. El endurecimiento monetario de 2022-2023 ha ofrecido, por su parte, una nueva prueba de la intensidad y la rapidez de la transmisión.

El tema analiza la política monetaria y sus efectos sobre la economía en tres partes:

- *Canales de transmisión*: se presentan los siete principales[1] y los efectos microeconómicos y sectoriales que tienen lugar en ellos.

- *Debates teóricos*: se analizan las aportaciones de cada escuela sobre los efectos de la política monetaria, hasta las rigideces de precios y de salarios y las fricciones financieras de la nueva economía keynesiana.

- *Evidencia empírica y política no convencional*: se recogen los efectos observados de la política monetaria y los mecanismos de transmisión de las medidas no convencionales y su eficacia.

[1] Bajo los canales subyace la infraestructura financiera (T2, que sustituyó a TARGET2 en marzo de 2023, y T2S en Europa) que conecta el banco central con las instituciones de crédito.

# 1. Mecanismos de transmisión de la política monetaria convencional: los canales

El **mecanismo de transmisión** de la política monetaria es el sistema de canales por los que los impulsos monetarios del banco central pasan de las variables instrumentales a los objetivos finales de la autoridad monetaria.

Aunque las escuelas han puesto el énfasis en canales y efectos distintos y no hay consenso sobre el funcionamiento del mecanismo, la cadena de eventos tiene, en general, tres etapas:

1. *Impulso*: el banco central inyecta o drena liquidez mediante operaciones de mercado abierto o altera las facilidades permanentes de depósito o de crédito o los coeficientes de caja.

2. *Transmisión*: el impulso recorre los diferentes canales, comenzando por el del tipo de interés a corto plazo.

3. *Efectos*: la inflación y la actividad económica responden de un modo que la autoridad monetaria no puede predecir ni controlar del todo.

Se distinguen siete canales de transmisión, los de los tipos de interés, del crédito, de la riqueza, del tipo de cambio, de la $q$ de Tobin, de las expectativas y de la redistribución.

## 1.1. Canal de los tipos de interés

Este canal tiene mucha importancia porque, en las economías con sistemas financieros desarrollados, el tipo de interés a corto plazo, que el banco central guía con sus tipos oficiales, es la variable operativa con la que las autoridades monetarias persiguen sus objetivos finales. Como el coste de financiarse condiciona lo que gastan empresas y familias, un recorte del tipo oficial estimula la inversión y el consumo a crédito y, a través de ellos, la economía real.

A corto plazo, la relación entre el sector monetario y el sector real se puede explicar a través de la ecuación de Fisher:

$$i = r + \pi^e$$

Si existen retardos o rigideces que impiden que los precios se ajusten inmediatamente ($\pi^e$ se ajusta lentamente), un descenso del tipo de interés nominal a corto plazo se reflejará en una caída del tipo de interés real a corto plazo.

Si el cambio del tipo de interés nominal persiste y los tipos a largo plazo son una media geométrica de los tipos a corto actuales y esperados, cambiarán también los tipos nominales a largo plazo.

La relación entre los tipos de interés a corto y a largo plazo dependerá de las expectativas de los agentes. Estas expectativas se reflejan en la estructura temporal de los tipos de interés (ETTI), que recoge la relación funcional entre los tipos de interés nominales de préstamos sin riesgo y el tiempo hasta vencimiento. Si presenta una forma normal (pendiente positiva), indica, según la hipótesis pura de las expectativas, que los mercados esperan que los tipos de interés suban en el futuro. Si existe, además, una prima por plazo, una pendiente positiva moderada es compatible con tipos esperados constantes (apartado 3.5.1).

Independientemente de esta relación, los cambios en los tipos de interés oficiales se transmitirán por tres vías:

1. *Efecto sustitución* del ahorro a corto plazo por consumo debido al descenso en el precio relativo del consumo presente. Según la ecuación de Euler, el consumo crece a una tasa aproximada de $(r-\rho)/\theta$ (tema A33, apartados 2.1.2 y 2.3.2), de modo que un menor tipo real adelanta consumo.

2. *Efecto renta* sobre la renta disponible de los deudores, que aumenta al disminuir la carga de intereses de su deuda y, por ello, lleva a un aumento del consumo, y sobre la de los acreedores, que disminuye al reducirse sus ingresos por intereses. En este caso prevalecerá el efecto sobre los deudores, dado que tienen una mayor propensión marginal al consumo.

3. *Efecto sobre los componentes de la demanda agregada con carácter permanente* (p. ej. vivienda, formación bruta de capital fijo). Para las empresas, suponiendo que su función de producción es de buen comportamiento, la disminución del tipo de interés real reduce el coste de uso del capital, $P_K(r+\delta)$ (tema A34, apartado 2.1.2). Por ello, aumenta el *stock* de capital deseado (aquel en que la productividad marginal del capital iguala el coste de uso) y, con él, la inversión. Para los hogares, aumenta el gasto en consumo de bienes duraderos financiados a crédito.

Todos estos efectos dependerán de cuatro factores: la magnitud de las elasticidades de las decisiones de gasto real al tipo de interés nominal, el peso relativo de los componentes del gasto en las decisiones de los agentes económicos, la posición financiera neta y la existencia de tipos variables o fijos. A largo plazo, en cambio, conforme los precios se ajustan, el efecto sobre $r$ se disipa y las variaciones solo afectan a variables nominales. En el modelo IS-PC-MR (apartado 2.4.2), estos efectos se resumen en la curva IS.

## 1.2. Canal del crédito (Bernanke y Gertler, 1995)

En este canal, la política monetaria incide no solo mediante los tipos de interés oficiales, sino también mediante la base monetaria. La idea es que la política monetaria puede aumentar el crédito concedido por los bancos comerciales, al aumentar el volumen de su pasivo o el valor de los activos que los prestatarios aportan como colateral. Esto se traduce en un incremento de la producción.

Bernanke y Gertler (1995) distinguen dos componentes, el canal en sentido estricto y el canal en sentido amplio. Un modelo sencillo de riesgo moral recoge ambos.

### 1.2.1. Canal en sentido estricto: el crédito bancario

El **canal en sentido estricto** (*bank lending channel*) actúa sobre la oferta de crédito. Si el banco central adopta una política expansiva, crecen los pasivos de los bancos comerciales, que disponen así de más fondos para prestar y engrosan los préstamos de su activo. Ese crédito adicional financia inversión y consumo en el sector real y eleva la renta nominal:

$$\uparrow M^s \rightarrow \uparrow \text{ pasivos bancarios} \rightarrow \uparrow \text{ préstamos} \rightarrow \uparrow I, C \rightarrow \uparrow PY$$

El BCE ha utilizado este canal con las operaciones de financiación a plazo más largo (LTRO), incluidas las de tres años de 2011-2012, y con las operaciones de financiación a plazo más largo con objetivo específico (TLTRO I, II y III). Estas últimas se vinculan a la concesión de préstamos a la economía real. La última TLTRO III venció en diciembre de 2024, y su mecanismo de transmisión se analiza en el apartado 3.5.2.

### 1.2.2. Canal en sentido amplio: el balance de los prestatarios

El **canal en sentido amplio** (*balance sheet channel* o acelerador financiero) recoge el efecto del precio de los activos y los efectos adicionales de la política monetaria debidos a las imperfecciones del mercado del crédito relacionadas con los costes de agencia y las asimetrías informativas. Opera por dos vías, la reducción de las tasas de descuento y la compra de activos.

Al bajar el tipo de interés, los rendimientos futuros se descuentan a una tasa menor. Los flujos de caja previstos ganan así valor actual y, con ellos, lo que vale la empresa:

$$\downarrow i \rightarrow\uparrow \text{ valor presente de los flujos de caja } \rightarrow\uparrow \text{ precio de los activos}$$
$$\rightarrow\uparrow \text{ colaterales } \rightarrow\uparrow \text{ crédito } \rightarrow\uparrow PY$$

Los tipos de interés nulos e incluso negativos del BCE (2014-2022) actuaron también por esta vía, y su transmisión y sus límites se analizan en los apartados 3.5.1 y 3.5.2.

Las compras de la autoridad monetaria elevan la cotización de los activos adquiridos:

$$\uparrow \text{ demanda activos } \rightarrow\uparrow \text{ precio } \rightarrow\uparrow \text{ colaterales } \rightarrow\uparrow \text{ crédito } \rightarrow\uparrow PY$$

El BCE ha utilizado este canal con las compras masivas del programa de compras de activos (APP)[2] y del programa de compras de emergencia frente a la pandemia (PEPP). Las reinversiones de ambos programas cesaron en julio de 2023 y a finales de 2024, respectivamente. Sus canales de transmisión y su eficacia se analizan en el apartado 3.5.

El canal del crédito tiene un efecto asimétrico sobre los distintos tipos de empresas y sectores. Por tipo de empresa, las pymes lo sufren más que las grandes compañías, puesto que estas últimas disponen de fuentes de financiación alternativas, como emitir acciones o deuda. Por sectores, aquellos cuya rentabilidad está garantizada (por lo general, al tratarse de oligopolios) obtienen financiación a menor coste, mientras que a las empresas de actividades con más riesgo (como las innovadoras) les resulta más difícil acceder a otras fuentes de financiación.

Donde predomina la tradición anglosajona, las empresas acuden sobre todo a los mercados de capitales para financiarse y los hogares mantienen carteras de bonos y acciones, de modo que el efecto riqueza gana peso en la transmisión de la política monetaria. En la tradición continental, sin embargo, la financiación empresarial se canaliza principalmente a través de la banca comercial, por lo que la transmisión depende en mayor medida del canal bancario.

En septiembre de 2015, la Comisión Europea presentó el Plan de Acción para la Unión de los Mercados de Capitales (UMC). Su objetivo es reducir la dependencia de la financiación bancaria (cuyo canal resultó disfuncional durante la crisis de 2008) y diversificar las fuentes de financiación, lo que refuerza de paso los efectos riqueza. La iniciativa aproxima así la tradición continental a la anglosajona. En marzo de 2025 la Comisión la relanzó como Unión de Ahorros e Inversiones (UAI).

[2] El APP agrupa los programas de compras de bonos de titulización, de bonos garantizados y de valores de los sectores público y corporativo.

### 1.2.3. Modelo del canal del crédito (Holmström y Tirole, 1997)

**Origen**: Bernanke y Gertler (1995) exponen verbalmente los dos canales, sobre su idea anterior (1989) de que el patrimonio neto de los prestatarios reduce los costes de agencia (tema A42, apartado 2.3.2). Holmström y Tirole (1997) los formalizan después en un modelo de riesgo moral con empresas y bancos sujetos a restricciones de capital. Aquí se presenta su versión con inversión variable (Tirole, 2006, cap. 3).

**Idea**: si el empresario puede desviar parte del rendimiento, los inversores solo financian lo que el proyecto puede comprometerles, de modo que la inversión depende del patrimonio neto y del coste de los fondos, dos variables que mueve la política monetaria.

**Supuestos**:

▪ *Entorno*: riesgo moral. Si el empresario se esfuerza, el éxito tiene probabilidad $p_H$. Si no lo hace, tiene probabilidad $p_L = p_H - \Delta p$, y el empresario obtiene un beneficio privado $BI$.

▪ *Agentes*: un empresario con responsabilidad limitada aporta su patrimonio neto, PN, e invierte $I \geq \text{PN}$, que rinde $RI$ si el proyecto tiene éxito y cero si fracasa. Los inversores son neutrales al riesgo y prestan $I - \text{PN}$ a cambio de un rendimiento esperado bruto $1 + r$, donde $r$ es el tipo real que fija la política monetaria.

**Planteamiento**: para que se esfuerce, el empresario ha de recibir en caso de éxito al menos $BI/\Delta p$ (restricción de incentivos), de modo que a los inversores solo puede prometerse la *renta comprometible*, $\rho_0 I \equiv p_H(R - B/\Delta p)\, I$. El empresario elige la escala que maximiza el valor actual neto del proyecto:

$$\max_{I} \; [\rho_1 - (1+r)]\, I \quad \text{s.a.:} \quad \rho_0\, I \geq (1+r)(I - \text{PN})$$

donde $\rho_1 = p_H R$ es el rendimiento esperado por unidad invertida y la restricción exige que la renta comprometible cubra lo que piden los inversores. Se supone $\rho_1 > 1 + r > \rho_0$, es decir, el proyecto es rentable, pero cada unidad invertida compromete menos de lo que cuesta financiarla. El lagrangiano, con multiplicador $\lambda$, es $\mathcal{L} = [\rho_1 - (1+r)]\, I + \lambda\, [(1+r)\,\text{PN} - (1 + r - \rho_0)\, I]$.

**CPO**: la derivada del lagrangiano respecto a la escala se anula:

$$\frac{\partial \mathcal{L}}{\partial I} = \rho_1 - (1+r) - \lambda\,(1 + r - \rho_0) = 0 \quad \Longrightarrow \quad \lambda = \frac{\rho_1 - (1+r)}{1 + r - \rho_0} > 0$$

de modo que la restricción se satura y el empresario invierte cuanto puede financiar. Por el teorema de la envolvente, una unidad de patrimonio invertida en el proyecto rinde $(1+r)(1+\lambda)$, más que el $1+r$ que obtendría prestada. Esa diferencia es la *prima de financiación externa*, que aquí adopta la forma de racionamiento y no de precio.

**CSO**: el objetivo y la restricción son lineales en $I$, por lo que la condición de Kuhn-Tucker (tema A08, apartado 1.4.1) es suficiente.

**Solución**: la inversión es proporcional al patrimonio neto:

$$I^* = k(r)\,\text{PN}, \qquad k(r) = \frac{1}{1 - \rho_0/(1+r)} > 1$$

donde $k(r)$ es el *multiplicador de los fondos propios* (*equity multiplier*), la inversión por unidad de patrimonio, que decrece con $r$ porque un tipo mayor reduce la financiación externa respaldada por la renta comprometible.

**Desarrollo gráfico**: el canal del crédito se representa en el plano patrimonio neto-inversión (figura A37.1):

1. *Recta de autofinanciación* ($I = \mathrm{PN}$): recoge la inversión financiable solo con el patrimonio. La distancia vertical hasta la capacidad de inversión es la financiación externa ($I_0 - \mathrm{PN}_0$ en el punto 0).

2. *Capacidad de inversión* ($I = k(r_0)\,\mathrm{PN}$): recoge la inversión máxima financiable con cada patrimonio. Sale del origen con pendiente $k(r_0) > 1$, tanto mayor cuanto menor es el tipo de interés.

3. *Óptimo*: con el patrimonio $\mathrm{PN}_0$, el empresario invierte $I_0 = k(r_0)\,\mathrm{PN}_0$ (punto 0).

**Estática comparativa**: una política monetaria expansiva ($\Delta r < 0$) eleva la inversión por dos vías:

- *Coste de los fondos*: al bajar el tipo de $r_0$ a $r_1$, crece el multiplicador y la recta gira hacia arriba. Con el mismo patrimonio, la inversión pasa de $I_0$ a $I_{0'}$ (de 0 a 0′).

- *Patrimonio neto*: el menor tipo revaloriza los activos y los flujos de caja, y el patrimonio pasa de $\mathrm{PN}_0$ a $\mathrm{PN}_1$. La inversión crece en $k(r_1)\,\Delta\mathrm{PN}$, más que el propio patrimonio, hasta $I_1$ (de 0′ a 1), amplificación que constituye el canal en sentido amplio.

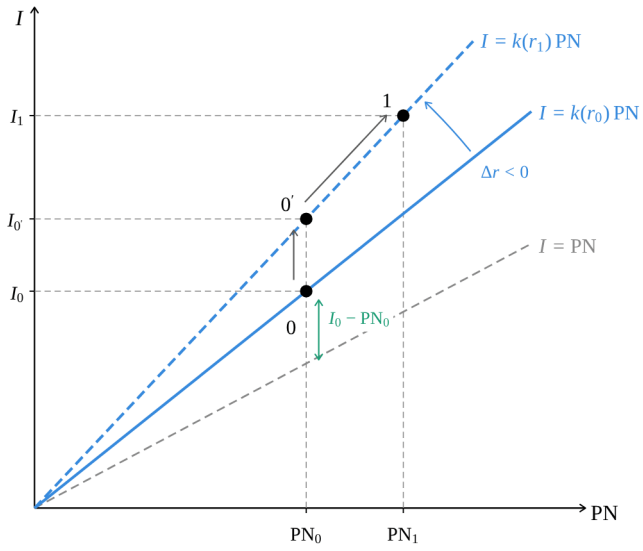



Figura A37.1: Canal del crédito en el modelo de Holmström y Tirole. Una bajada del tipo de interés ($\Delta r < 0$) gira hacia arriba la capacidad de inversión (0→0′), y la revalorización del patrimonio eleva la inversión a lo largo de la nueva recta (0′→1). El tramo verde es la financiación externa en 0.

**Extensiones**: Holmström y Tirole (1997) añaden bancos que supervisan a los prestatarios, lo que reduce el beneficio privado de estos, y que deben arriesgar capital propio para tener incentivos

a hacerlo. Las empresas con patrimonio intermedio solo obtienen financiación si un banco las supervisa, y la cantidad que así se financia depende del capital de los bancos, que es el canal en sentido estricto. Así, cualquier endurecimiento, sea del capital bancario (*credit crunch*), del colateral o del ahorro disponible, recae sobre todo en las empresas peor capitalizadas.

**Conclusiones**:

- *Aspectos positivos*: a partir de un único supuesto, el riesgo moral, explica por qué la inversión depende del patrimonio de las empresas y del capital de los bancos y por qué la política monetaria tiene efectos amplificados y asimétricos.

- *Limitaciones*: es estático y de equilibrio parcial, y toma como dado el efecto de la política monetaria sobre el precio de los activos.

## 1.3. Canales de los precios de los activos: riqueza, tipo de cambio y *q* de Tobin

Los precios de los activos transmiten la política monetaria por tres canales, el de la riqueza, el del tipo de cambio y el de la $q$ de Tobin.

### 1.3.1. Canal de la riqueza

La idea es que la política monetaria puede alterar la riqueza de los individuos vía cambios en su volumen o en la valoración de sus activos. El sustrato teórico de estos efectos son las teorías del consumo basadas en la riqueza, sobre todo la hipótesis del ciclo vital (Modigliani y Brumberg, 1954; Ando y Modigliani, 1963), además de la renta permanente de Friedman (1957). Las compras de activos del BCE (APP y PEPP) actúan también por este canal. Se distinguen dos efectos:

1. *Efecto riqueza directo*: el dinero es uno de los componentes del patrimonio de los agentes. Con precios dados, más liquidez significa más saldos reales y más riqueza, y esta estimula el consumo: $\uparrow M^s/P \rightarrow \uparrow W \rightarrow \uparrow C$.

2. *Efecto riqueza indirecto*, que, a su vez, se desglosa en dos subefectos de magnitud ambigua:

- *Efecto Pigou*: la expansión monetaria eleva el nivel general de precios, que erosiona el valor real de los activos denominados en términos nominales y empobrece a los individuos: $\uparrow M^s \rightarrow \uparrow P \rightarrow \downarrow A/P \rightarrow \downarrow W \rightarrow \downarrow C$ (de signo contrario al efecto directo).

- *Efecto reajuste de cartera*: un exceso de liquidez en la cartera de los individuos hará que estos se quieran deshacer de él mediante la adquisición de nuevos activos financieros y reales. La mayor demanda produce una revalorización de estos activos que, además, pueden ser utilizados como colaterales, lo que genera un aumento del crédito y del gasto.

### 1.3.2. Canal del tipo de cambio

Este canal ha ganado peso a medida que la economía se ha globalizado y los tipos de cambio se han vuelto flexibles. En una economía pequeña y abierta, con tipo de cambio flexible y plena movilidad del capital, una política monetaria expansiva rebaja el tipo de interés. Esa rebaja origina una salida neta de capitales y deprecia el tipo de cambio.

Según la paridad descubierta de intereses, $i = i^* + g_E^e$, con $g_E^e$ la depreciación esperada (tema B14, apartado 1.3.4), un tipo nacional inferior al exterior exige una apreciación esperada. En consecuencia, la moneda se deprecia hoy por encima de su nivel de largo plazo (*overshooting* de Dornbusch, apartado 2.2.2 de ese tema). Esta depreciación tiene dos efectos:

1. *Aumento de la competitividad-precio*: si rige la condición de Marshall-Lerner, el abaratamiento de las exportaciones y el encarecimiento de las importaciones elevan la demanda agregada.

2. *Aumento de la inflación*: los *inputs* importados se encarecen en términos relativos y, con costes más altos, las empresas recortan la oferta agregada, lo que acaba elevando el nivel general de precios.

No obstante, el efecto neto sobre la cuenta corriente es ambiguo, porque la reducción de los tipos de interés eleva también la demanda nacional de importaciones, y su signo es una cuestión empírica. En el modelo IS-PC-MR de economía abierta, este canal permite al banco central mover menos el tipo de interés (tema A32, apartado 3.3.2)[3].

### 1.3.3. Canal de la $q$ de Tobin

La $q$ de Tobin (1969) es el cociente entre el valor de mercado de las empresas y el coste de reposición de su capital (tema A34, apartado 2.2.2). Una política monetaria expansiva rebaja los tipos de interés y hace los bonos menos atractivos que las acciones, cuya demanda y cuya cotización suben. Con ello aumenta la $q$, y cuando supera la unidad, a las empresas les basta emitir pocas acciones para financiar una inversión grande, porque el capital vale en bolsa más de lo que cuesta reponerlo. Es decir, $\downarrow i \rightarrow\uparrow V \rightarrow\uparrow q \rightarrow\uparrow I$, con $V$ el valor de mercado de las empresas. El canal actúa sobre la inversión, mientras que el de la riqueza actúa sobre el consumo.

## 1.4. Canal de la señalización: el papel de las expectativas

Las expectativas sobre los futuros tipos de interés oficiales tienen gran peso, ya que los tipos de interés a medio y largo plazo dependen, en parte, de las expectativas del mercado sobre la evolución de los tipos a corto plazo. La idea es que el banco central puede influir sobre las expectativas de inflación y, por tanto, sobre el nivel de precios de la economía, mediante la señalización del curso de la política monetaria a futuro[4].

Formalmente, al iterar hacia delante la IS dinámica del modelo neokeynesiano (tema A07, apartado 2.1.2), la brecha de producción (*output gap*) actual depende de toda la senda esperada de tipos reales:

$$\tilde{y}_t = -\frac{1}{\theta} \sum_{k=0}^{\infty} \mathbb{E}_t \left[ r_{t+k} - \bar{r}_{t+k} \right]$$

donde $\tilde{y}_t$ es la brecha de producción, $\theta$ la inversa de la elasticidad de sustitución intertemporal, $r_{t+k}$ el tipo real del periodo $t+k$ y $\bar{r}_{t+k}$ el tipo natural. Por ello, el anuncio de tipos futuros más bajos estimula el gasto actual aunque el tipo oficial corriente no cambie.

[3] Más allá del tipo de cambio, la política monetaria puede tener efectos sobre la balanza de pagos a través de los flujos de capital aunque no exista riesgo cambiario. Ocurre, por ejemplo, con los inversores de la zona del euro que quieren invertir en otros países de la zona del euro. En el límite, la reversión brusca de esos flujos da lugar a *sudden stops*.

[4] El canal fue fundamental durante la crisis del euro, cuando el *whatever it takes* de Draghi (2012) estabilizó las expectativas y redujo las primas de riesgo.

Estas expectativas se canalizan a la economía real a través de las decisiones de inversión y consumo de bienes duraderos. El instrumento principal es el *forward guidance* (orientación prospectiva), cuya transmisión en el límite inferior de los tipos de interés (*zero lower bound*, ZLB) se analiza en el apartado 3.5. No obstante, resulta difícil prever qué efectos tendrán las expectativas. Por ejemplo, el aumento del tipo de interés podría interpretarse de dos maneras:

- *Señal de crecimiento*: los agentes pueden leerlo como indicio de que el banco central prevé un crecimiento más robusto de la economía a medio/largo plazo y de que, en consecuencia, interviene ya para anticiparse a la inflación futura.

- *Señal de freno*: pueden entender que la subida de los tipos obstaculizará el crecimiento y, en consecuencia, tendrá un efecto negativo.

## 1.5. Canal de la redistribución

Por último, el **canal redistributivo** es hoy un campo de investigación muy activo. La visión tradicional considera que las propensiones marginales al consumo no varían mucho entre la población, de modo que los efectos mencionados se reparten de manera relativamente homogénea.

Sin embargo, las fricciones financieras y los diferentes niveles de liquidez hacen que las propensiones marginales al consumo sean heterogéneas. Por tanto, al repartir ganancias y pérdidas entre hogares, la política monetaria altera el consumo agregado aunque la renta total no cambie. Auclert (2019) distingue tres vías, que, según su estimación para Italia y Estados Unidos, probablemente amplifican el efecto de la política monetaria:

1. *Heterogeneidad de las rentas*: una expansión eleva más el gasto si las rentas que más responden al ciclo son las de los hogares con mayor propensión al consumo.

2. *Canal de Fisher*: la inflación no anticipada traslada riqueza de los acreedores nominales a los deudores, que gastan una mayor parte.

3. *Exposición a los tipos de interés*: una bajada del tipo real beneficia a los hogares cuyas deudas se revisan antes que sus activos (p. ej. con hipotecas a tipo variable), que es el efecto renta del canal de los tipos de interés.

## 1.6. Conclusiones

Con precios rígidos, una bajada del tipo oficial reduce el tipo real y estimula el gasto por sustitución, por renta y por el abaratamiento del coste de uso del capital. El canal del crédito amplifica ese efecto, pues la política monetaria mueve la oferta de préstamos y el colateral de los prestatarios, sobre todo de los peor capitalizados. Los precios de los activos transmiten el impulso por la riqueza, el tipo de cambio y la $q$ de Tobin, y las expectativas lo extienden a toda la senda esperada de tipos. La redistribución, por último, altera el consumo agregado cuando las propensiones al consumo difieren entre hogares. La sección 2 examina, escuela por escuela, qué efectos reales tiene ese impulso y cuánto duran.

# 2. Efectos de la política monetaria: de la neutralidad a las rigideces y fricciones financieras

Hasta llegar al consenso actual sobre la neutralidad del dinero a largo plazo, las escuelas de pensamiento discutieron durante décadas si la política monetaria es efectiva y qué efectos produce sobre la economía real.

> La teoría cuantitativa y la demanda de dinero se desarrollan en el tema A35, y el pensamiento de cada escuela, en los temas A04 a A07. Aquí se recogen sus implicaciones para la efectividad de la política monetaria.

## 2.1. Teoría clásica: la neutralidad del dinero

La teoría clásica parte de la teoría cuantitativa del dinero y extrae de ella su consecuencia para la efectividad de la política monetaria.

### 2.1.1. La teoría cuantitativa del dinero

**Origen**: la teoría descansa en la *ecuación de cambio*, identidad contable formulada por Newcomb (1885) y popularizada por Fisher (1911) sobre ideas que se remontan a Hume (1752).

**Idea**: el pensamiento clásico se apoya en la *teoría cuantitativa del dinero* (TCD) para explicar el sector monetario. La TCD se pregunta por qué cambia el valor o poder adquisitivo del dinero, y su conclusión principal es que el nivel general de precios guarda una proporción constante con la cantidad de dinero.

**Razonamiento**: esta idea queda recogida en la lectura causal de la ecuación de cambio. Por un lado, el valor monetario del dinero empleado, o valor de las compras, es igual a la oferta monetaria ($M$) por la velocidad-transacción del dinero ($V$), es decir, el ritmo al que circula. Por otro, el valor monetario del conjunto de transacciones, o valor de las ventas, es igual al número de veces que bienes y servicios se cambian por dinero multiplicado por el precio de una transacción representativa ($P \cdot T$):

$$MV \equiv PT$$

En la versión de Cambridge (Pigou, 1917), formulada como demanda de dinero, $M^d = kPY$, con $k = 1/V$ constante, lo que, en equilibrio ($M^s = M^d = M$), equivale a $MV = PY$. Ambas versiones implican, dado que $V$ y $T$ (o $Y$) se consideran estables, que los cambios en $M$ se trasladan proporcionalmente a $P$.

**Conclusiones**:

- *Aspectos positivos*: explica la relación a largo plazo entre el crecimiento monetario y la inflación, que la evidencia respalda (apartado 3.4.2), y fundamenta la tesis de que la inflación es un fenómeno monetario.

- *Limitaciones*: la proporcionalidad descansa en la estabilidad de $V$ y de la producción, que la demanda especulativa de dinero de Keynes y la inestabilidad de la velocidad desde los años ochenta (apartado 3.3) ponen en cuestión. La TCD es, además, una teoría del largo plazo que no explica los efectos reales a corto.

### 2.1.2. Efectividad de la política monetaria

En el paradigma clásico, el "dinero es un velo" y, con ello, se produce la llamada **dicotomía clásica**, según la cual las variables nominales no afectan a las variables reales. Por tanto, la política monetaria es inefectiva.

## 2.2. El debate entre el keynesianismo, la síntesis y el monetarismo

Las aportaciones de Keynes rompen con esta dicotomía clásica. Su formalización posterior en el modelo IS-LM de la síntesis neoclásica permite ilustrar sus principales implicaciones monetarias, que el monetarismo criticaría después.

### 2.2.1. Keynesianismo: la ruptura de la dicotomía clásica por la demanda especulativa de dinero

**Origen**: a raíz de la Gran Depresión (1929-1933), Keynes (1936) formula la teoría de la preferencia por la liquidez, en la que el dinero se demanda también por motivo especulación.

**Idea**: para Keynes, la demanda de dinero por motivo especulación deja a la política monetaria un cauce, el tipo de interés, para actuar sobre las variables reales, aunque dos eslabones débiles le restan eficacia:

1. *Inversión poco sensible*: los *animal spirits* dominan las decisiones de inversión, de modo que el tipo de interés apenas la mueve ($v \approx 0$).

2. *Demanda de dinero muy sensible*: responde mucho al tipo de interés ($h$ alto) y poco a la renta. Dado que el único activo sustitutivo del dinero son los bonos, cambios en la demanda de bonos debidos a una variación del tipo de interés afectan en gran medida a la demanda de dinero[5]. En el límite, si los tipos de interés son excepcionalmente bajos, la demanda de dinero se vuelve infinitamente elástica y se produce una trampa de liquidez.

**Razonamiento**: en el plano $(Y, i)$ (figura A37.2, panel a), una política monetaria expansiva ($\Delta M^s > 0$) desplaza la LM de $\text{LM}_0$ a $\text{LM}_1$ y va a tener tres efectos reales:

- *Efecto liquidez* ($0 \rightarrow 0'$): un aumento de la oferta monetaria hace que los individuos eliminen el exceso de dinero comprando bonos, cuyo precio sube, de modo que el tipo de interés baja de $i_0$ a $i_{0'}$. No obstante, este efecto es escaso.

- *Efecto renta* ($0' \rightarrow 1$): al bajar el tipo de interés, crecen la inversión y el consumo que se financia con crédito. Por tanto, se produce un aumento de la renta, de $Y_0$ a $Y_1$ (movimiento a lo largo de la $\text{LM}_1$ hasta cortar la IS). No obstante, este efecto es reducido debido a la baja sensibilidad de la inversión a los tipos.

- *Efecto expulsión* (*crowding out*): el aumento de la renta provoca un aumento de la demanda de dinero por motivo transacción, que genera un efecto rebote sobre el tipo de interés, hasta $i_1$. En consecuencia, la inversión disminuye en parte y se modera el aumento de la renta. Sin embargo, este último efecto será reducido, porque, con una demanda de dinero muy sensible al tipo, basta una pequeña subida de este para absorber la mayor demanda por motivo transacción, y predomina el efecto renta.

[5] Keynes (1936): el tipo de interés es "la recompensa por renunciar a la liquidez".

En el panel a, la IS (equilibrio del mercado de bienes) es muy inclinada, porque la inversión apenas responde al tipo de interés, y la LM (equilibrio del mercado de dinero) es casi horizontal, porque la demanda de dinero es muy sensible a él. Por eso, la política monetaria será poco eficaz con las variables reales (la renta y la inversión).

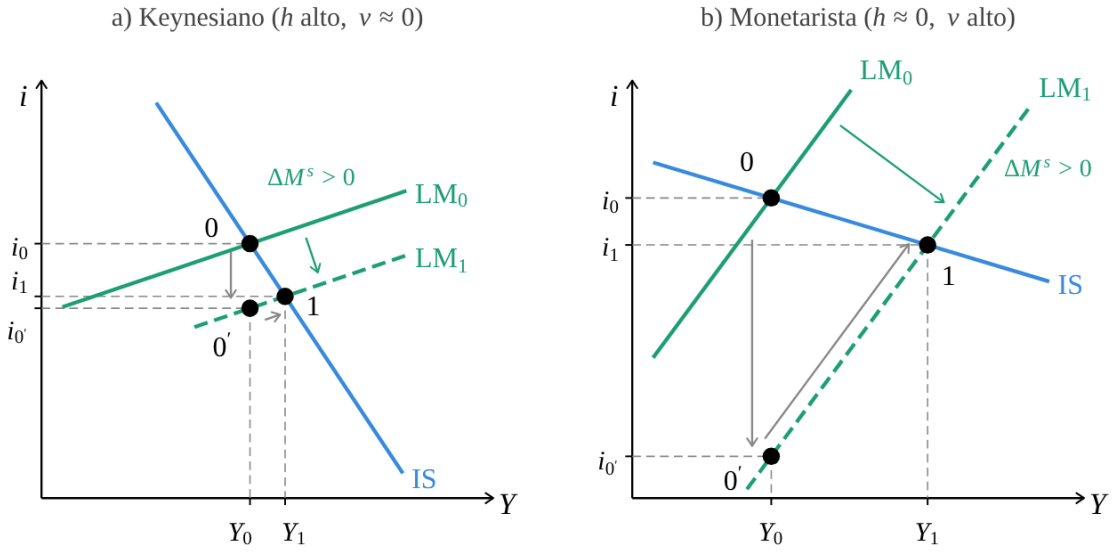



Figura A37.2: Política monetaria expansiva en el IS-LM. El gráfico separa el efecto liquidez ($0 \to 0'$) y el efecto renta y expulsión ($0' \to 1$) en el caso keynesiano (a), con demanda de dinero muy sensible al tipo e inversión insensible, y en el monetarista (b), con los supuestos opuestos.

**Conclusiones**:

- *Aspectos positivos*: rompe la dicotomía clásica al dar al tipo de interés un papel en la transmisión de la política monetaria, el canal principal de la sección 1, y la trampa de liquidez anticipa el problema del límite inferior de los tipos (apartado 3.5.1).

- *Limitaciones*: la escasa eficacia descansa en dos supuestos sobre elasticidades que el monetarismo rechaza, y el análisis, con precios fijos y sin expectativas, no distingue entre los efectos a corto y a largo plazo.

### 2.2.2. Síntesis neoclásica: la curva de Phillips como menú de política económica

**Origen**: la síntesis neoclásica adopta la curva de Phillips en la lectura de Samuelson y Solow (1960)[6].

**Idea**: la síntesis concede algo más de eficacia a la política monetaria porque cuenta con la curva de Phillips, que recoge una disyuntiva entre inflación y desempleo. La curva funciona así

[6] Samuelson y Solow (1960) presentaron la curva de Phillips como un menú para el "ajuste fino" de la economía, aunque advirtieron de que ese menú podía no mantener su forma a más largo plazo.

como un menú, en el que las autoridades pueden comprar un paro más bajo pagándolo con más inflación.

**Razonamiento**: la curva liga de forma inversa la inflación y el desempleo (véase la figura del tema A41, apartado 2.2.1):

$$\pi_t = f(u_t), \quad f' < 0$$

De ahí que, en este esquema, la política monetaria sea eficaz en cualquier horizonte. Si mantiene en el tiempo una inflación estable y más alta, el desempleo queda de forma permanente en un nivel más bajo.

**Conclusiones**:

▪ *Aspectos positivos*: ofrece una relación estimable entre inflación y desempleo, con la que la política monetaria puede elegir la combinación de ambos que prefiera.

▪ *Limitaciones*: la disyuntiva permanente exige que los trabajadores sufran ilusión monetaria y no incorporen la inflación esperada a sus demandas salariales, que es la crítica monetarista. Además, la estanflación de los años setenta, con inflación y desempleo crecientes a la vez, mostró que la curva se había desplazado.

### 2.2.3. Monetarismo: la importancia de las expectativas y la renta permanente

**Origen**: la defensa del intervencionismo suscitó pronto la crítica monetarista. Milton Friedman (1968) y, de forma independiente, Edmund Phelps (1967) introdujeron las expectativas en la curva de Phillips.

**Idea**: la política monetaria tiene efectos reales elevados a corto plazo, porque la inversión responde con fuerza al tipo de interés y la demanda de dinero apenas lo hace. A largo plazo, en cambio, cuando las expectativas se ajustan, se cumple la neutralidad del dinero.

**Razonamiento**: la crítica monetarista se articula en dos vertientes:

▪ *Por el lado de la oferta*: según los monetaristas, la curva de Phillips, tal como la usaba la síntesis, carecía de coherencia, porque exigía un mercado de trabajo que se comportase de forma irracional. Para que la disyuntiva durase en cualquier horizonte, los trabajadores tendrían que sufrir ilusión monetaria, tomando el salario nominal por el real, y desatender la inflación futura, es decir, ser miopes o tener expectativas estáticas sobre los precios. Como alternativa, los monetaristas propusieron una hipótesis sobre la formación de las expectativas y, a partir de ella, una curva de Phillips ampliada:

- *Hipótesis de expectativas adaptativas* (HEA): cada periodo se revisa la inflación esperada en una fracción del error cometido en el anterior, $\pi_t^e = \pi_{t-1}^e + \lambda(\pi_{t-1} - \pi_{t-1}^e)$, con $0 < \lambda \leq 1$, que, en su versión más sencilla ($\lambda = 1$), se reduce a $\pi_t^e = \pi_{t-1}$.

- *Curva de Phillips ampliada por las expectativas*: $\pi_t = \pi_t^e - \kappa_u(u_t - \bar{u})$, donde $\kappa_u$ refleja la sensibilidad de la inflación al desempleo cíclico y $\bar{u}$ la tasa natural de paro. A largo plazo, cuando $\pi_t = \pi_t^e$, la curva es vertical sobre $\bar{u}$ en el plano $(u, \pi)$ (véase la figura del tema A28, apartado 1.3.1), es decir, ninguna tasa de inflación sostenida altera el paro (superneutralidad del dinero).

▪ *Por el lado de la demanda*: los monetaristas niegan que se den los dos eslabones débiles de los que hablaba Keynes, y para ellos la demanda de dinero depende de la renta permanente y no solo de la corriente. A diferencia de Keynes, sostienen que a corto plazo los agentes colocan el dinero sobrante en activos de todo tipo, financieros y reales, y que ese gasto eleva la demanda agregada. En el IS-LM, los tres efectos keynesianos son elevados (figura A37.2, panel b), y el aumento de la renta es considerable por dos razones:

- *Inversión*: responde con fuerza al tipo de interés ($v$ alto), porque las decisiones se toman atendiendo a la rentabilidad y no a los *animal spirits* (IS elástica).
- *Demanda de dinero*: apenas reacciona al tipo de interés ($h \approx 0$, LM inelástica). El dinero tiene sustitutos imperfectos en todos los activos, no solo en los bonos. Por eso, cuando el tipo de interés cambia y se recompone la demanda de bonos, la de dinero apenas se mueve.

**Conclusiones**:

▪ *Aspectos positivos*: explica, antes de que la estanflación lo hiciera evidente, por qué la disyuntiva entre inflación y desempleo solo es transitoria, y devuelve a la política monetaria efectos reales elevados a corto plazo frente a los eslabones débiles keynesianos.

▪ *Limitaciones*: el efecto de corto plazo descansa en expectativas adaptativas, que suponen errores sistemáticos de los agentes. Con expectativas racionales, en cambio, una política monetaria anticipada no tiene efectos reales ni siquiera a corto plazo.

### 2.2.4. Efectividad de la política monetaria

Para Keynes, la política monetaria es poco eficaz por los dos eslabones débiles, y la síntesis la considera eficaz en cualquier horizonte. Por su parte, los monetaristas reconocen su eficacia a corto plazo, pero no quieren que se use de forma activa. Prefieren atarla a reglas y dedicarla a mantener la inflación bajo control, y lo justifican con tres argumentos:

1. *Ecuación perdida del monetarismo*: Friedman deja sin determinar qué parte de una variación del dinero acaba en los precios (variable nominal) y qué parte en la producción (variable real).

2. *Retardos*: largos y variables según Friedman (1961), pueden provocar que los efectos de la política monetaria resulten procíclicos.

3. *Inefectividad en el largo plazo*: se debe a la HEA y a que entonces sí se cumple la neutralidad del dinero.

## 2.3. Nueva macroeconomía clásica: vuelta a la no intervención monetaria

Las aportaciones de la nueva macroeconomía clásica (NMC) suponen un cambio metodológico que permite la vuelta a la defensa de la no intervención monetaria.

### 2.3.1. Consideraciones generales

1. *Apoyo en la fundamentación microeconómica dinámica, estocástica y de equilibrio general*: se da racionalidad dinámica.

2. *Competencia perfecta*: rige en el mercado de bienes y factores, con plena flexibilidad de precios y salarios, de manera que se garantiza el vaciado continuo de los mercados.

3. *Expectativas racionales*: se forman según la hipótesis de expectativas racionales (HER). Los individuos no las ajustan gradualmente, sino que, al predecir la evolución de las variables económicas de su interés, utilizan en todo momento toda la información relevante disponible y la procesan de forma eficiente.

4. *Información incompleta*: los agentes no distinguen con certeza si una variación del precio de su producto responde a un *shock* específico, ante el que convendría variar la producción, o a uno general, ante el que no (problema de extracción de señales, temas A06, apartado 2.1, y A41, apartado 2.3.1).

5. *Oferta agregada*: de estos postulados se deriva la oferta de Lucas, función de los *shocks* de la economía.

$$y_t = \bar{y}_t + \zeta(p_t - p_t^e) + \varepsilon_t^s, \quad \zeta > 0$$

donde $\zeta$ es el coeficiente de Lucas, $\bar{y}_t$ el *output* natural y $\varepsilon_t^s$ el término aleatorio (con esperanza nula) que recoge los *shocks* exógenos de oferta. Ese coeficiente combina la proporción de la variación del precio que los agentes atribuyen a precios relativos (mayor cuanto más probable consideran que el *shock* observado sea específico) y la respuesta de la oferta de los productores ante un *shock* en su mercado.

Los componentes monetarios, que son la parte no anticipada de la política monetaria y las perturbaciones de la velocidad de circulación del dinero, actúan a través de la sorpresa de precios $p_t - p_t^e$. Por tanto, los *shocks* monetarios solo afectan a la producción en la medida en que son sorpresivos.

La NMC supone una revitalización del pensamiento clásico, y en materia de política monetaria se traduce en las aportaciones de Sargent y Wallace (1975, 1976) y Kydland y Prescott (1977).

### 2.3.2. Inefectividad de la política monetaria: la crítica de Sargent y Wallace (1975, 1976)

**Origen**: Thomas Sargent y Neil Wallace (1975, 1976) estudian la elección de la regla monetaria óptima con expectativas racionales y la oferta de Lucas.

**Idea**: la política monetaria sistemática, la que responde a información al alcance de todos (como los grandes indicadores macroeconómicos), no sirve para estabilizar la producción.

**Razonamiento**: la inefectividad se apoya en dos resultados:

1. *Política sistemática*: si la política monetaria sigue una regla conocida (fácilmente anticipable), los agentes la incorporan a sus expectativas. La regla actúa entonces solo sobre los precios, sin consecuencias reales.

2. *Sorpresas*: solo la parte no anticipada de la política monetaria tiene efectos reales, y solo transitorios.

**Desarrollo gráfico**: la crítica de Sargent y Wallace se representa en el plano renta-precios (figura A37.3):

1. *Oferta agregada a largo plazo* ($\mathrm{OA_{LP}}$): es vertical sobre el *output* natural $\bar{Y}$, porque con expectativas cumplidas los precios no afectan a la producción.

2. *Oferta agregada a corto plazo* ($\mathrm{OA_{CP,0}}$): es creciente, porque solo las sorpresas de precios alteran la producción, y hay una por cada nivel de precios esperado.

3. *Demanda agregada* ($\mathrm{DA_0}$): recoge el gasto que corresponde a cada nivel de precios, dada la oferta monetaria, y es decreciente, porque unos precios más altos reducen los saldos reales.

4. *Equilibrio*: en el punto 0, con el *output* $\bar{Y}$ y los precios $P_0$, las expectativas se cumplen.

**Estática comparativa**: supóngase que la autoridad monetaria comunica que aumentará $M^s$ ($\Delta M^s > 0$, que desplaza la DA de $\mathrm{DA_0}$ a $\mathrm{DA_1}$) para que la inflación suba, por ejemplo, del 0 % al 5 %. El resultado depende de cómo reciban el anuncio los agentes:

- *Anuncio creíble*: si los agentes lo anticipan, la oferta a corto plazo pasa a $\mathrm{OA_{CP,1}}$ y solo suben los precios, de $P_0$ a $P_1$ (de 0 a 1).

- *Anuncio no creíble*: los agentes pueden anticipar una expansión de la masa monetaria superior o inferior a la anunciada, de modo que, si la política se ejecuta tal como se anunció, su parte no anticipada tendrá efectos reales transitorios. Esos efectos son expansivos si los agentes esperan precios menores, con la oferta $\mathrm{OA_{CP}}(P^e < P_1)$ (punto $0'$), y contractivos si los esperan mayores, con $\mathrm{OA_{CP}}(P^e > P_1)$ (punto $0''$).

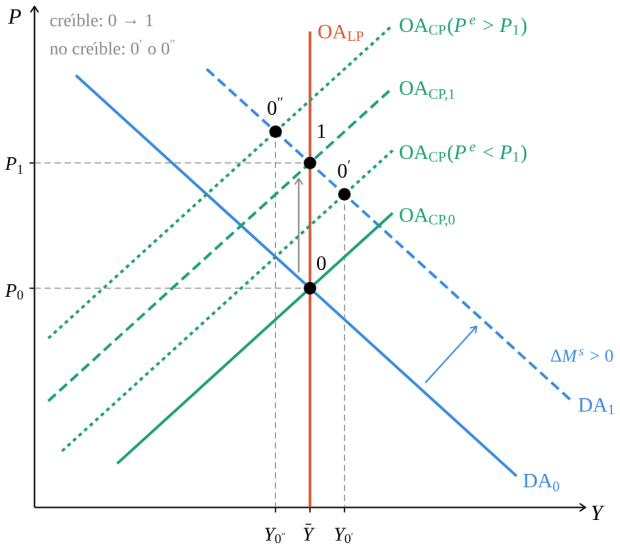



Figura A37.3: Crítica de Sargent y Wallace en el plano OA-DA (en niveles). Una expansión monetaria creíble y anticipada solo eleva los precios (0→1). Si no es creíble, las expectativas erróneas generan efectos reales transitorios, expansivos ($0'$) o contractivos ($0''$).

La falta de credibilidad actúa aquí solo como fuente de sorpresa, y su análisis como problema de política corresponde a Kydland y Prescott. Con todo, la NMC desaconseja la política sorpresiva, pues resta credibilidad a la política monetaria de los periodos siguientes y añade incertidumbre.

**Conclusiones**:

- *Aspectos positivos*: separa los efectos de la política anticipada y de la no anticipada, una separación que respaldaron los primeros contrastes empíricos (Barro, 1977), y refuerza el argumento a favor de reglas sencillas y predecibles.

- *Limitaciones*: exige precios y salarios plenamente flexibles. Con contratos nominales plurianuales (Fischer, 1977), la política anticipada tiene efectos reales, como recoge la NEK, y Mishkin (1982) halla que el dinero anticipado también afecta a la producción.

### 2.3.3. Inconsistencia dinámica de la política monetaria (Kydland y Prescott, 1977; Barro y Gordon, 1983a)

> El modelo de Barro y Gordon se desarrolla en el tema A06, apartado 3.1.2. Aquí se recoge su implicación para la efectividad de la política monetaria.

**Origen**: Kydland y Prescott y Barro y Gordon se preguntan si una política monetaria sorpresiva sostenida en el tiempo es compatible con las expectativas racionales.

**Idea**: una política anunciada es *dinámicamente inconsistente* (*time inconsistent*) cuando, una vez que los agentes han fijado sus contratos confiando en ella, a la autoridad le conviene incumplirla. Como los agentes lo anticipan, la política no es creíble.

**Razonamiento**: la política monetaria discrecional, si la autoridad persigue un paro inferior a la tasa natural, produce un *sesgo inflacionista* sin ganancia de empleo, pues la economía acaba en la tasa natural con más inflación que la prometida. La efectividad de la política monetaria pasa así a depender de su credibilidad. De ahí las soluciones institucionales que la NMC propone: reglas frente a discrecionalidad, independencia del banco central, banquero conservador (Rogoff, 1985) y contratos de desempeño (Walsh, 1995), que se desarrollan en el tema A36.

**Conclusiones**:

- *Aspectos positivos*: explica el sesgo inflacionista como resultado de la falta de compromiso, sin suponer agentes ni autoridades irracionales, y fundamenta las reglas, la independencia del banco central y los objetivos de inflación.

- *Limitaciones*: el sesgo descansa en que la autoridad persiga un paro inferior a la tasa natural, supuesto discutible en un banco central independiente con mandato de estabilidad de precios. El marco es, además, estático y de información completa.

### 2.3.4. Efectividad de la política monetaria

Una política monetaria discrecional, no basada en reglas o que no es creíble, puede tener efectos reales transitorios por sorpresa, pero, una vez anticipada, da lugar a una pérdida de bienestar derivada de un mayor nivel de inflación con el mismo nivel de empleo. De ahí proceden las soluciones institucionales a la inconsistencia dinámica.

## 2.4. Nueva economía keynesiana: vuelta a la defensa del intervencionismo

La respuesta a las aportaciones de la NMC vino de los herederos del keynesianismo, los cuales retoman el poder de la política monetaria.

### 2.4.1. Rigideces de precios, de salarios y fricciones financieras

> La fundamentación microeconómica de las rigideces nominales y reales, incluida la obtención de la curva de Phillips neokeynesiana, se desarrolla en el tema A07, apartados 1.2 a 1.4. El acelerador financiero se estudia en los temas A34, apartado 3.2.4, y A42, apartado 2.3.2, y el racionamiento del crédito, en los temas B26, apartado 1.2.3, y A42, apartado 2.3.1. Aquí se recoge cómo cada fricción hace que la política monetaria no sea neutral y condiciona su transmisión.

Para la nueva economía keynesiana (NEK), la no neutralidad del dinero no exige que la política monetaria sea sorpresiva. Aunque los agentes formen, en general, sus expectativas según la HER, basta con que los mercados de bienes y de factores funcionen en competencia imperfecta (las empresas fijan sus precios con un margen sobre el coste marginal), con información imperfecta y contratos duraderos. Algunos modelos admiten, además, cuasirracionalidad (Akerlof y Yellen, 1985). Tres tipos de fricciones (de precios, de salarios y financieras), amplificadas por las rigideces reales, explican entonces que un impulso monetario altere la producción.

**Rigideces nominales de precios**: una *rigidez nominal* es la resistencia de los precios (o salarios) nominales a ajustarse. Con contratos escalonados (Taylor, 1980), los salarios, y con ellos los precios, se fijan para varios periodos con revisiones solapadas entre empresas, y el nivel de precios hereda inercia. Calvo (1983) simplifica el escalonamiento al suponer que, cada periodo, una empresa puede cambiar su precio con probabilidad $1-\xi$ y ha de conservarlo con probabilidad $\xi$, sin que cuente el tiempo que lleve fijado. Si la empresa lo cambia, aplica un margen a una combinación ponderada de sus costes marginales de hoy y de los que espera. Al agregar se obtiene la *curva de Phillips de la NEK*:

$$\pi_t = \beta\,\mathbb{E}_t[\pi_{t+1}] + \kappa\,\tilde{y}_t, \qquad \kappa > 0,\ \beta \in (0,1)$$

donde $\pi_t$ es la inflación del periodo $t$, $\mathbb{E}_t[\pi_{t+1}]$ la inflación esperada para el periodo siguiente con la información disponible en $t$ y $\beta \in (0,1)$ el factor de descuento. La brecha de producción, $\tilde{y}_t = y_t - \bar{y}_t$, es la diferencia entre el logaritmo de la producción y el de su nivel natural, el de precios flexibles, y $\kappa > 0$ es la pendiente de la curva en el plano $(\tilde{y}_t, \pi_t)$, que mide cuánto responde la inflación a la brecha.

La pendiente decrece con el parámetro de Calvo. Cuanto mayor es $\xi$ (los precios duran por término medio $1/(1-\xi)$ periodos), menos empresas ajustan, más plana es la curva en ese plano y más recae el impulso monetario sobre la producción y menos sobre los precios. Con precios flexibles ($\xi \to 0$), $\kappa$ tiende a infinito, la curva se vuelve vertical y se recupera la neutralidad, mientras que, con $\xi \to 1$, $\kappa$ tiende a cero. Además, la curva es prospectiva, porque la inflación actual depende de las brechas futuras esperadas, por lo que la política monetaria actúa también a través de las expectativas (apartado 1.4).

Los *costes de menú* (Mankiw, 1985) completan el cuadro. Si cambiar el precio tiene un coste fijo, aunque sea pequeño, a la empresa con poder de mercado le compensa no ajustarlo, pues su pérdida privada es de segundo orden. La suma de esas inacciones tiene, sin embargo, efectos de primer orden sobre la producción y el bienestar.

**Rigideces nominales de salarios**: los salarios nominales se revisan a intervalos discretos, por los costes de negociación y la duración de los convenios, y esa inercia se refuerza con las rigideces reales del mercado de trabajo. En el modelo de Christiano, Eichenbaum y Evans (2005), los contratos salariales escalonados son, junto con la utilización variable del capital, lo que más contribuye a la persistencia de los efectos de un *shock* monetario. Además, Erceg, Henderson y Levin (2000) añaden al modelo neokeynesiano salarios fijados escalonadamente *à la* Calvo por hogares con poder de mercado, y muestran que ello cambia la disyuntiva del banco central.

En el modelo básico, con rigidez solo en los precios y sin perturbaciones de costes ni rigidez salarial real, estabilizar la inflación equivale a cerrar la brecha (la "divina coincidencia" del tema A07, apartado 3.1.2). Con salarios nominales también rígidos, en cambio, un *shock* de productividad exige ajustar el salario real moviendo precios o salarios nominales, ambos costosos, de modo que el banco central ya no puede estabilizar a la vez la brecha, la inflación de precios y la salarial. Perseguir solo la de precios resulta claramente subóptimo frente a reglas que respondan también a la brecha o a la inflación salarial.

**Rigideces reales**: una *rigidez real* es la escasa sensibilidad del precio o del salario real deseado a la actividad. Nace, en el mercado de trabajo, de los contratos implícitos, los salarios de eficiencia o los sindicatos y, en el de bienes, de los mercados de clientela, la colusión contracíclica, la información imperfecta sobre la calidad o el exceso de capacidad. Por sí sola no rompe la neutralidad, pero, combinada con fricciones nominales pequeñas, genera una rigidez nominal sustancial (Ball y Romer, 1990), ya que, si el precio óptimo apenas se mueve, no ajustarlo cuesta poco. En la curva de Phillips, la rigidez real reduce $\kappa$ para un $\xi$ dado.

**Fricciones financieras**: con información asimétrica entre prestamistas y prestatarios, la financiación externa es más cara que la interna, y su sobrecoste, la prima de financiación externa, es tanto mayor cuanto menores son el patrimonio neto y el colateral del prestatario. La política monetaria mueve la prima en el mismo sentido que el tipo oficial y amplifica así su efecto, que es el canal del crédito del apartado 1.2. En un modelo neokeynesiano, esa retroalimentación da lugar al acelerador financiero (Bernanke, Gertler y Gilchrist, 1999). Cuando el sistema financiero está dañado, el banco central puede actuar directamente sobre las primas y la oferta de crédito, que es la lógica de la política no convencional (apartado 3.5).

**Consecuencias para la transmisión**: estas fricciones hacen que el ajuste ante un impulso monetario sea lento y vía cantidades, con posibles fallos de coordinación y equilibrios múltiples subóptimos, y refuerzan los canales de la sección 1:

- *Precios rígidos*: como la inflación esperada se ajusta lentamente, una bajada del tipo nominal reduce el tipo real (apartado 1.1).

- *Salarios rígidos*: el salario real se ajusta despacio, lo que prolonga los efectos sobre el empleo y enfrenta la estabilidad de la inflación con la de la actividad.

- *Fricciones financieras*: la política monetaria altera la prima de financiación externa y la oferta de crédito, con mayor efecto sobre los prestatarios con menos patrimonio, como las pymes.

### 2.4.2. Modelo IS-PC-MR y desinflación

> El modelo IS-PC-MR se desarrolla en el tema A07, apartado 2.1.2, y la regla monetaria óptima, en el tema A36, apartado 2.4. Aquí se recoge lo que el modelo enseña sobre la transmisión de una desinflación y el papel de la credibilidad.

**Origen**: el modelo canónico de la NEK, condensado en tres ecuaciones por Clarida, Galí y Gertler (1999), permite seguir la trayectoria de una desinflación deliberada.

**Idea**: con precios rígidos, el banco central mueve el tipo de interés real respecto del natural y, con él, la brecha de producción. La inflación responde a esa brecha y a la inflación esperada, de modo que el coste de reducirla depende de cuánto pueda apoyarse el banco central en las expectativas, es decir, de su credibilidad.

**Planteamiento**: con la notación del tema A07, apartado 2.1.2, el modelo se compone de tres ecuaciones:

- *Curva IS dinámica*: $\tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta}(r_t - \bar{r}_t)$, con $r_t = i_t - \mathbb{E}_t[\pi_{t+1}]$. Recoge el canal de los tipos de interés.

- *Curva de Phillips de la NEK 2G* (CP): $\pi_t = \beta\,\mathbb{E}_t[\pi_{t+1}] + \kappa\,\tilde{y}_t$, con $\mathbb{E}_t[\cdot]$ la esperanza condicionada a la información disponible en $t$.

- *Regla de política monetaria*: $\tilde{i}_t = \alpha_\pi \tilde{\pi}_t + \alpha_y \tilde{y}_t$, con $\alpha_\pi > 1$. Llevada a la IS, da la curva MR, decreciente en el plano $(\tilde{y}, \pi)$, cuya versión óptima bajo discrecionalidad, $\tilde{y}_t = -\frac{\kappa}{\omega}(\pi_t - \pi^T)$, con $\omega$ el peso de la brecha en la pérdida del banco central y $\pi^T$ el objetivo de inflación, se deriva en el tema A36, apartado 2.4.

**Desarrollo gráfico**: el modelo IS-PC-MR se representa en el plano brecha de producción-inflación (figura A37.4):

1. *Curva de Phillips* ($\mathrm{CP}_0$): es creciente, con pendiente $\kappa$, y su posición depende de la inflación esperada. Con $\beta$ próximo a la unidad, corta la vertical de brecha nula a la altura de $\mathbb{E}_t[\pi_{t+1}]$.

2. *Curva MR* ($\mathrm{MR}_0$): es decreciente y pasa por el objetivo de inflación con brecha nula, porque ante más inflación el banco central sube el tipo real y contrae la demanda.

3. *Equilibrio*: en el punto 0, la inflación y la esperada coinciden con el objetivo, $\pi_0^T$, la brecha es nula y el tipo real iguala al natural.

**Estática comparativa**: el banco central reduce de forma permanente su objetivo, de $\pi_0^T$ a $\pi_1^T$ ($\Delta\pi^T < 0$), y la MR se desplaza hacia abajo, de $\mathrm{MR}_0$ a $\mathrm{MR}_1$. La trayectoria depende de la credibilidad del anuncio:

1. *Anuncio creíble*: la inflación esperada baja de inmediato a $\pi_1^T$, la CP se desplaza a la vez hasta $\mathrm{CP}_1$ y la economía pasa directamente de 0 a 1, sin abrir la brecha. El tipo real no se mueve y el nominal baja con la inflación esperada, de modo que la desinflación se transmite solo por el canal de las expectativas.

2. *Credibilidad imperfecta*: si la inflación esperada se mantiene en $\pi_0^T$, la CP sigue en $\mathrm{CP}_0$ y el banco central elige el punto $0'$, donde $\mathrm{CP}_0$ corta a $\mathrm{MR}_1$. Para llegar a él eleva el tipo real por encima del natural ($r_t > \bar{r}_t$), la IS traduce esa subida en una brecha negativa y la CP, en una inflación inferior a la esperada.

3. *Revisión de expectativas*: al observar una inflación menor, los agentes rebajan la esperada ($\Delta\mathbb{E}_t[\pi_{t+1}] < 0$), la CP se desplaza hacia abajo, hasta $\mathrm{CP}_{0'}$, y la economía pasa de $0'$ a $0''$ a lo largo de $\mathrm{MR}_1$.

4. *Convergencia*: el proceso se repite hasta $\text{CP}_1$ y la economía llega al punto 1, con la inflación en $\pi_1^T$, la brecha cerrada y el tipo real igual al natural. Los puntos de producción perdidos en el camino por cada punto de inflación reducida miden la ratio de sacrificio, que es nula con un anuncio plenamente creíble.

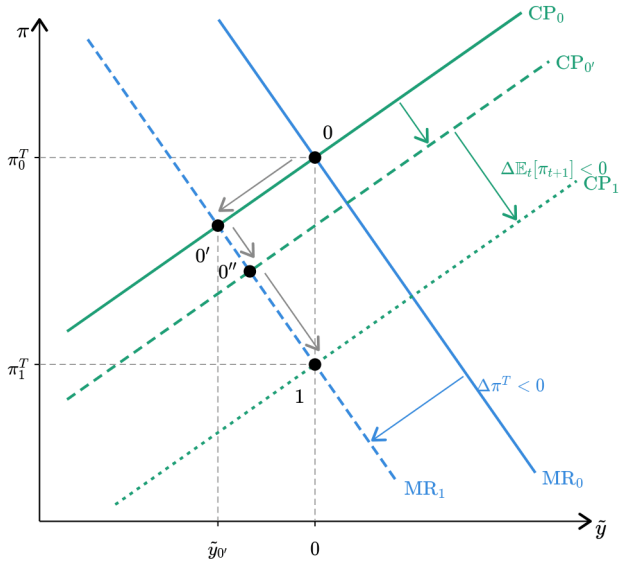



Figura A37.4: Reducción del objetivo de inflación en el modelo IS-PC-MR con credibilidad imperfecta. La MR se desplaza hacia abajo ($0 \to 0'$) y la revisión gradual de las expectativas desplaza la CP hacia abajo ($0' \to 0'' \to 1$) hasta el nuevo equilibrio con $\pi = \pi_1^T$ y brecha nula.

La comparación muestra que la política monetaria puede llegar a la inflación por la vía de la brecha, a través del tipo real y la IS, que cuesta producción, o por la de las expectativas, que no la cuesta, pero exige credibilidad. Por ello, la falta de credibilidad que genera la discrecionalidad encarece la desinflación.

**Extensiones**:

- *Inercia de la inflación*: con una curva de Phillips híbrida, que añade la inflación pasada (tema A41, apartado 2.4.3), incluso con un anuncio creíble la economía sigue una trayectoria gradual como la de $0'$, $0''$ y 1, y la ratio de sacrificio crece con el peso de la inercia.

- *Compromiso*: un banco central que se compromete a responder también en el futuro ancla las expectativas y reduce el coste de estabilizar (tema A36, apartado 2.4).

- *Otros canales*: el modelo básico solo recoge el de los tipos de interés. La economía abierta añade el del tipo de cambio (tema A32, apartado 3.3.2), las fricciones financieras añaden el del crédito (apartado 1.2.3) y los agentes heterogéneos, el de la redistribución (apartado 1.5).

**Conclusiones**:

- *Aspectos positivos*: con las tres ecuaciones del modelo canónico, separa los dos canales por los que la política monetaria llega a la inflación y explica por qué el coste de una desinflación depende de la credibilidad del banco central.

- *Limitaciones*: con la CP puramente prospectiva, una desinflación creíble no cuesta producción, en contra de la evidencia de desinflaciones graduales (tema A41, apartado 2.4.3), y la credibilidad imperfecta se supone, sin explicar cómo se revisan las expectativas. Además, el modelo prescinde del crédito y del límite inferior de los tipos (apartado 3.5.1).

### 2.4.3. Efectividad de la política monetaria

La política monetaria puede evitar que pérdidas microeconómicas de segundo orden se conviertan en pérdidas macroeconómicas de primer orden, y ayudar a estabilizar el *output*. No obstante, en el largo plazo estos efectos se diluyen en la medida en que los agentes ajustan sus expectativas.

## 2.5. Conclusiones

En la sección 2, la eficacia de la política monetaria depende de la flexibilidad de precios y salarios y de cómo se forman las expectativas. Con precios flexibles, la teoría cuantitativa traslada el dinero a los precios y se cumple la dicotomía clásica. Keynes la rompe a través del tipo de interés, aunque los dos eslabones débiles limitan el efecto, y la síntesis convierte la curva de Phillips en un menú permanente. Las expectativas adaptativas del monetarismo reducen ese menú al corto plazo, y las racionales de la NMC lo limitan a las sorpresas. Una autoridad discrecional que persigue un paro inferior al natural solo obtiene, además, más inflación. En cambio, la NEK devuelve a la política anticipada efectos reales transitorios mediante las rigideces nominales y las fricciones financieras. La sección 3 contrasta estas predicciones con la evidencia.

# 3. Evidencia empírica y transmisión de la política monetaria no convencional

Además de los efectos que la teoría atribuye a la política monetaria, la autoridad monetaria considera la evidencia empírica. Tras una nota de metodología, los hechos que recoge esa evidencia se clasifican en:

- *Limitaciones generales*: las que no dependen del marco temporal.
- *Limitaciones específicas*: las ligadas a rasgos concretos de la política monetaria y de su entorno.
- *Hechos temporales*: los condicionados por el horizonte temporal.
- *Política no convencional*: canales de transmisión y efectos observados de las medidas adoptadas cuando los tipos alcanzaron su límite inferior.

## 3.1. Metodología

En los modelos con rigideces nominales, el dinero no es neutral, al menos a corto plazo (apartado 2.4), y gran parte de la literatura empírica se dirige a contrastarlo.

Sin embargo, la identificación de esta relación no es tarea fácil, porque los instrumentos de política monetaria (p. ej. el tipo de interés nominal a corto plazo) son, generalmente, endógenos (el banco central los ajusta en respuesta a la evolución de la economía). Por tanto, la simple correlación entre tipos de interés o masa monetaria y el *output* u otras variables reales no puede usarse como evidencia de la no neutralidad monetaria. Se recurre a técnicas como el enfoque de vectores autorregresivos estructurales (SVAR) y el enfoque narrativo[7]. La evidencia SVAR sobre los efectos de la política monetaria se resume en el tema A40, apartado 3.3.

## 3.2. Limitaciones de carácter general

1. *Existencia de incertidumbre*: por mucho que se hayan modelizado, los mecanismos de transmisión y el funcionamiento mismo de la economía solo se conocen en parte, y las decisiones de los bancos centrales pierden eficacia por ello.

2. *Retardos temporales*: si los retardos son largos, una medida pensada para amortiguar el ciclo puede terminar amplificándolo. Según la evidencia empírica, los desfases son largos y variables. Tras una subida de tipos, los precios alcanzan su mínimo, de media, a los 29 meses (entre 25 y 50 meses en las economías desarrolladas, según Havranek y Rusnak, 2013). Se distinguen dos tipos de desfases:

- *Desfase interno*: comprende dos retardos:
  - *De reconocimiento*: media entre la perturbación y la toma de conciencia, por parte del banco central, de que ha de intervenir. Lo alargan la llegada tardía de la información estadística y la duda sobre si la perturbación es transitoria o permanente.
  - *De acción*: va del reconocimiento de esa necesidad a la puesta en marcha de las medidas. En política monetaria es corto, porque sus decisiones se adoptan con rapidez, a diferencia de las fiscales.

[7] El enfoque narrativo identifica los *shocks* a partir de registros históricos (Romer y Romer, 1989).

- *Desfase externo*: abarca desde que las medidas se aplican hasta que surten efecto en la economía.

3. *Globalización e interdependencia económica*: cuando una economía pequeña, que no pesa lo bastante para mover las condiciones financieras mundiales, se abre al capital exterior, sus tipos de interés pasan a seguir sobre todo a los internacionales. Para las políticas nacionales el reto es considerable, porque los efectos desbordamiento (*spillover effects*) ligan la política óptima de cada país a la que siguen los demás. Esa interdependencia explica la aparición de foros de coordinación internacional como el G20 y encuentros de banqueros centrales como el simposio anual de Jackson Hole, que organiza la Reserva Federal de Kansas City[8].

### 3.3. Limitaciones de carácter específico

1. *Asimetría de la política monetaria* ("metáfora de la cuerda"): la política monetaria frena mejor de lo que impulsa (Cover, 1992)[9]. Retirar dinero de la economía recorta de forma directa la capacidad de gasto del sector privado, mientras que inyectarlo no garantiza que aumente el gasto privado. El resultado de una expansión depende, entre otros factores, del multiplicador bancario. En la primera etapa de la provisión ilimitada de liquidez del BCE (adjudicación plena en sus operaciones), por ejemplo, los bancos comerciales emplearon los fondos en sanear sus balances y no en dar crédito.

2. *Límites institucionales de los tipos de interés*: la política monetaria restrictiva, aunque sea más eficaz que la expansiva, tropieza también con dificultades por la oposición social que despiertan varios de sus efectos:

- *Distribución de la renta*: cuando los tenedores de deuda pública son sobre todo hogares de renta alta y el grueso de la recaudación procede de las rentas medias y bajas, subir los tipos traslada renta de las clases medias y bajas a las altas, porque los intereses adicionales se pagan con impuestos.

- *Crecimiento*: los tipos altos frenan la actividad al deprimir la inversión. Además, si el tipo de cambio flota, atraen capitales que aprecian la moneda, y esa apreciación resta competitividad y reduce las exportaciones netas.

- *Pymes*: con tipos más altos, su financiación se complica más que la de otras empresas, porque las pymes apenas cuentan con fuentes alternativas de fondos.

3. *Cambios en la velocidad de circulación del dinero*: en la visión neoclásica y en la monetarista, $V$ era una magnitud estable, fijada por factores institucionales (apartado 2.1.1). No obstante, desde los años ochenta la innovación financiera y la creciente liberalización de capitales han hecho más inestable el sector monetario, y la relación a corto y medio plazo entre la oferta monetaria y el nivel de precios se ha debilitado. Por ello, la mayoría de los bancos centrales de economías avanzadas dejó de usar los agregados monetarios como objetivo intermedio y pasó a objetivos directos de inflación o a estrategias mixtas.

[8] Tras 2008, la expansión cuantitativa (*quantitative easing*, QE) de la Fed generó efectos desbordamiento en las economías emergentes. El gobernador del Banco de la Reserva de la India, R. Rajan, la criticó (2014) como una expansión monetaria competitiva, con beneficios sobre todo internos y costes sobre todo en el exterior.

[9] La asimetría obedece a dos motivos, el comportamiento asimétrico de prestamistas y prestatarios y la rigidez nominal de los precios a la baja, que hace que el ajuste se produzca vía cantidades y no vía precios.

4. *Interrelación entre la política monetaria y la política fiscal: aritmética monetarista desagradable* (Sargent y Wallace, 1981): con una política fiscal sistemáticamente expansiva y un tipo de interés real superior al crecimiento de la economía, la deuda crece más rápido que la renta. Por ello, acabará teniendo que monetizarse. Como los agentes lo anticipan, la autoridad monetaria no controla la inflación ni siquiera en el presente, y financiar el déficit con deuda puede resultar más inflacionista que monetizarlo (tema A39, apartado 3.2.1).

5. *Tipo de cambio fijo*: si los capitales circulan libremente, una economía que quiera mantener una paridad fija deja a su banco central sin margen para decidir la política monetaria, como establece el trilema de Mundell-Fleming (trinidad imposible).

## 3.4. Hechos observados vinculados al horizonte temporal

Las correlaciones entre dinero, inflación y PIB, que no prueban por sí solas una relación causal, están bien documentadas y se ordenan aquí según el horizonte.

### 3.4.1. Relaciones a corto plazo

1. *Dinero y PIB real*: crecimiento monetario ↔ crecimiento del PIB real (+). Las fases de expansión monetaria rápida tienden a ir por delante de las de mayor crecimiento del PIB.

2. *Inflación y PIB real*: inflación ↔ crecimiento del PIB real (+). Los periodos de inflación más alta coinciden, por lo general, con un crecimiento real más intenso.

3. *Tipo de interés y PIB*: tipo de interés ↔ PIB (−). Los recortes de tipos se anticipan normalmente a una expansión del *output*[10].

### 3.4.2. Relaciones a largo plazo

1. *Dinero e inflación*: crecimiento monetario ↔ inflación (+). La correlación entre ambas tasas roza la unidad (0,92–0,96, según el agregado monetario elegido). Este resultado avala la tesis central de la teoría cuantitativa, el origen monetario de la inflación[11].

2. *Dinero y PIB real*: crecimiento monetario ↔ crecimiento del PIB real. En la muestra completa las dos tasas no están correlacionadas, aunque en la submuestra de la OCDE la correlación es positiva. Ello parece desmentir la disyuntiva entre paro e inflación que defendía la síntesis neoclásica, y respaldar la dicotomía clásica.

[10] Estas correlaciones no identifican por sí solas el efecto de la política monetaria (apartado 3.1). La evidencia identificada (tema A40, apartado 3.1) confirma su signo, porque con el VAR de Christiano, Eichenbaum y Evans (1999) una subida inesperada del tipo de interés reduce el producto de forma gradual y los precios reaccionan con retraso, y con la medida narrativa de Romer y Romer (2004) los efectos son más intensos y rápidos.

[11] McCandless y Weber (1995), para 110 países en 1960-1990. Friedman (1963, *Inflation: Causes and Consequences*): "inflation is always and everywhere a monetary phenomenon".

## 3.5. Transmisión de la política monetaria no convencional

La Reserva Federal llevó su tipo oficial al límite inferior a finales de 2008 y volvió a él en 2020, mientras que el BCE lo alcanzó entre 2012 y 2014 y no lo abandonó hasta 2022. Para seguir estimulando la economía, los bancos centrales recurrieron a instrumentos no convencionales: compras masivas de activos, *forward guidance*, tipos negativos y financiación a largo plazo condicionada al crédito.

> Los instrumentos no convencionales se describen en el tema A36, apartados 3.2 y 3.3.1, y su empleo por el BCE desde 2009, en el tema B43, apartado 3.3.2. Su papel cuando el límite inferior restringe la política monetaria y su relación con la política fiscal se analizan en el tema A40, apartado 2.1, y una síntesis de la evidencia, en el apartado 3.3 de ese tema. Aquí se recogen su mecanismo de transmisión y la evidencia principal sobre sus efectos.

### 3.5.1. El límite inferior efectivo y el debilitamiento de la transmisión convencional

Cuando el tipo oficial alcanza su límite inferior, el canal de los tipos de interés (apartado 1.1) se bloquea aunque la regla prescriba nuevos recortes. Ese límite es el cero (ZLB) o, en la práctica, un **límite inferior efectivo** (*effective lower bound*, ELB) algo inferior. La razón es que, por debajo de cierto tipo negativo, bancos y público prefieren mantener efectivo, que rinde un 0 % nominal pese a sus costes de almacenamiento y seguridad. El BCE, por ejemplo, llegó a situar la facilidad de depósito en $-0{,}5$ % en 2019-2022.

Gráficamente, en el plano tiempo-tipo de interés (figura A37.5), entre $t_1$ y $t_2$ el tipo objetivo que prescribe la regla de Taylor, $i^T(t)$, es negativo, pero el tipo oficial no puede bajar del límite $i = 0$ (ZLB). La diferencia entre ambos es el estímulo no aplicado.

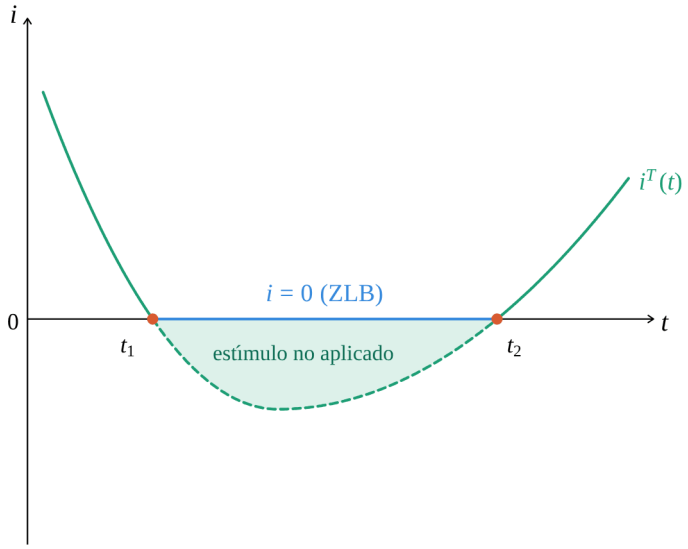



Figura A37.5: Límite inferior de los tipos de interés nominales. Entre $t_1$ y $t_2$ el tipo objetivo que desearía fijar el banco central (el que prescribe la regla de Taylor), $i^T(t)$, es negativo, pero el tipo oficial no puede bajar de cero (en la práctica, de un límite inferior efectivo ligeramente negativo). El área sombreada es el estímulo que la política monetaria convencional no puede aplicar.

Si además cae la inflación esperada, el tipo real sube y puede desencadenarse una espiral deflacionista. Las operaciones de mercado abierto ordinarias pierden también eficacia, porque con tipos nulos el dinero y las letras del Tesoro son casi sustitutos perfectos (trampa de liquidez, apartado 2.2.1). En 2008-2012 se sumó un canal del crédito dañado, pues los bancos destinaron la liquidez a sanear sus balances más que a prestar.

Las medidas no convencionales actúan, por ello, sobre los demás determinantes de las condiciones financieras. En aproximación logarítmica, un tipo a largo plazo se descompone en:

$$i_t^{(n)} \approx \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{E}_t[i_{t+k}] + \rho_t^{(n)}$$

donde $i_t^{(n)}$ es el tipo en $t$ de un bono a $n$ periodos, $i_{t+k}$ el tipo a corto plazo del periodo $t+k$, $\mathbb{E}_t[\cdot]$ la esperanza condicionada a la información disponible en $t$ y $\rho_t^{(n)}$ la prima por plazo. El *forward guidance* actúa sobre el primer sumando y las compras de activos, sobre ambos. Los tipos negativos rebajan el suelo del tipo corto, y la financiación condicionada abarata el pasivo de los bancos.

### 3.5.2. Canales de transmisión de las medidas no convencionales

**Compras de activos a gran escala**: las compras masivas de bonos (QE) operan por cuatro canales:

- *Reequilibrio de carteras y duración*: al retirar bonos a largo plazo del mercado, el banco central reduce el riesgo de duración en manos de los inversores y, con él, la prima por plazo. Los vendedores se desplazan hacia otros activos, cuyos rendimientos también bajan. Este canal exige que los activos no sean sustitutos perfectos, porque, si lo fueran, sin otras fricciones, desaparecería y las compras solo actuarían en la medida en que alterasen las expectativas sobre los tipos futuros (teorema de Wallace, tema A39, apartado 3.4.1).

- *Señalización*: las compras anuncian una política expansiva prolongada y reducen los tipos cortos esperados.

- *Liquidez y funcionamiento del mercado*: en episodios de tensión (crisis de deuda soberana, marzo de 2020), el banco central actúa como comprador de último recurso y reduce las primas de liquidez y la fragmentación.

- *Tipo de cambio*: los menores rendimientos deprecian la moneda, lo que estimula las exportaciones netas (apartado 1.3.2), pero también genera efectos desbordamiento sobre otras economías.

Al elevar el precio de los activos, las compras refuerzan, además, los canales de la riqueza y del colateral (apartados 1.2.2 y 1.3.1).

***Forward guidance***: al comprometerse a mantener el tipo en el límite inferior más tiempo del que prescribiría su regla (Eggertsson y Woodford, 2003), el banco central rebaja los tipos esperados y los largos. Estimula así el gasto actual, porque la brecha de producción depende de toda la senda esperada de tipos reales (apartado 1.4 y tema A07, apartado 3.1.3).

El *forward guidance* exige credibilidad, pues el compromiso es dinámicamente inconsistente (apartado 2.3.3), y los modelos estándar exageran su potencia. De hecho, en los anuncios de la Reserva Federal, sus efectos sobre las expectativas de producción e inflación fueron muy inferiores a los que esos modelos predicen, el llamado *forward guidance puzzle* (Del Negro, Giannoni y Patterson, 2023).

**Tipos negativos**: se transmiten como un recorte convencional, pues abaratan el interbancario, desplazan a la baja la ETTI e incentivan a cada banco a prestar o invertir su exceso de reservas, pero chocan con la rentabilidad bancaria. Como los depósitos minoristas apenas se remuneran por debajo de cero, cada recorte estrecha el margen de intermediación. Por debajo del *tipo de reversión* (*reversal rate*), esa pérdida reduce el capital de los bancos más de lo que lo elevan las plusvalías de su cartera de renta fija, y el recorte pasa a contraer el crédito (Brunnermeier y Koby, 2018).

Ese umbral es más alto cuanto menores son la cartera de renta fija y el capital de los bancos, más estrictos sus requisitos de capital y menor la traslación a los depósitos. Además, sube con el tiempo, al vencer los bonos revalorizados, y con las compras de activos, por lo que conviene agotar antes los recortes. Para aliviar ese coste, el BCE introdujo en 2019 la remuneración de las reservas en dos tramos (tema A36, apartado 3.2.3).

**Financiación condicionada**: las TLTRO actúan sobre el canal del crédito en sentido estricto (apartado 1.2.1). Prestan a los bancos a varios años a un tipo tanto menor cuanto más crédito conceden a empresas no financieras y hogares (salvo para la compra de vivienda). En la TLTRO III, ese tipo llegó a situarse, entre junio de 2020 y junio de 2022, 50 puntos básicos (pb) por debajo del tipo medio de la facilidad de depósito. Las TLTRO abaratan así el pasivo bancario y condicionan la liquidez a que llegue a la economía real, algo que no aseguraron las LTRO a tres años de 2011-2012.

### 3.5.3. Evidencia sobre su eficacia

La evidencia apunta a un efecto sustancial sobre los precios de los activos y más moderado sobre la actividad y la inflación:

- *Compras de la Reserva Federal*: Gagnon, Raskin, Remache y Sack (2011) estiman que las compras de 2008-2010, por más de 1,7 billones de dólares, redujeron la prima por plazo del bono del Tesoro a diez años entre 30 y 100 pb. La caída de los tipos largos se debió sobre todo a menores primas de riesgo, entre ellas la prima por plazo, y no a menores tipos cortos esperados. Krishnamurthy y Vissing-Jorgensen (2011), sin embargo, muestran que el efecto depende de los activos comprados, y hallan apoyo para los canales de señalización, de demanda de activos seguros y de inflación, pero poco para el de duración.

- *Compras del BCE*: Altavilla, Carboni y Motto (2015) estudian los 17 anuncios relacionados con el APP entre septiembre de 2014 y marzo de 2015. En conjunto, esos anuncios redujeron el rendimiento a diez años de la deuda soberana de la zona del euro entre 30 y 50 pb, según el método, y en torno al doble en Italia y España. El anuncio de enero de 2015, que añadió la deuda pública, estaba en gran parte descontado.

- *Conjunto de medidas del BCE*: la evaluación del Eurosistema para la revisión de su estrategia (Altavilla et al., 2021) estima el efecto de las medidas no convencionales adoptadas desde 2014. Esas medidas elevaron el crecimiento anual del PIB de la zona del euro en unos 0,6 puntos porcentuales de media en 2015-2020, y sin ellas la inflación anual habría sido unos 0,4 puntos más baja. Además, combinar instrumentos resultó más eficiente que recurrir a uno solo.

En suma, el límite inferior no agota la política monetaria, pero hace su transmisión más incierta, al depender de la credibilidad del banco central, de la segmentación de los mercados y de la salud de los bancos. Por ello, la estrategia del BCE de 2021 exigió, cerca de ese límite, medidas especialmente enérgicas o persistentes (tema B43, apartado 3.2.2). La evaluación de 2025 reformuló el principio de forma simétrica, como una respuesta adecuadamente enérgica o persistente ante desviaciones amplias y duraderas de la inflación respecto del objetivo, en cualquiera de los dos sentidos.

## Conclusiones

El impulso monetario llega a la economía real sobre todo por el tipo de interés, y los canales del crédito, la riqueza, el tipo de cambio, las expectativas y la redistribución lo refuerzan o lo debilitan según la estructura financiera. Los efectos reales de la política monetaria dependen del marco teórico. Para la teoría clásica, el dinero es neutral y la política monetaria no afecta a la producción. Keynes le abre un cauce real por el tipo de interés, aunque la trampa de liquidez y una inversión poco sensible al tipo le restan eficacia. Por su parte, los monetaristas le reconocen efectos reales a corto plazo, que desaparecen cuando el desempleo vuelve a su tasa natural. La NMC limita esos efectos a las sorpresas, mientras que la NEK los recupera, aunque transitorios, gracias a las rigideces de precios y salarios y a las fricciones financieras. La evidencia confirma la neutralidad a largo plazo y muestra efectos reales a corto, con retardos largos y variables y una eficacia asimétrica. En el límite inferior, las medidas no convencionales actúan sobre las expectativas, la prima por plazo y el crédito, con efectos financieros sustanciales y reales más moderados.

La transmisión completa el diseño de la política monetaria y sus instrumentos (tema A36). La curva de Phillips neokeynesiana y la divina coincidencia explican por qué, en el modelo básico, estabilizar la inflación estabiliza la brecha de producción (tema A07), y el modelo de Barro y Gordon, por qué una política discrecional genera un sesgo inflacionista (tema A06). Esos efectos tienen como contrapartida los costes de la inflación (tema A41). Por otro lado, la transmisión no es independiente de la política fiscal, ni cuando el banco central financia al Estado (tema A39) ni en el límite inferior, donde la política fiscal gana eficacia (tema A40). Por último, el BCE ha aplicado las medidas no convencionales en una unión monetaria con riesgo de fragmentación (tema B43).

No hay acuerdo sobre el peso del canal de la redistribución. En los modelos neokeynesianos con agentes heterogéneos (HANK), la política monetaria actúa sobre todo por sus efectos indirectos sobre la renta y el empleo de los hogares con poca liquidez, y apenas por la sustitución intertemporal del agente representativo (Kaplan, Moll y Violante, 2018). Por ello, la potencia de la política monetaria depende de la distribución de la riqueza y de la respuesta fiscal. Sigue abierta, además, la cuestión del alcance de la política no convencional, acotado por el *forward guidance puzzle* y el tipo de reversión (apartado 3.5.2).

# A38. Política fiscal, crecimiento y ahorro

La política fiscal: efectos sobre el crecimiento económico y el ahorro.

## Introducción

Estabilizar el ciclo económico, impulsar la producción y el empleo o redistribuir la renta son algunos de los fines a los que sirve la **política fiscal**, es decir, el manejo de los ingresos y los gastos del sector público. Siguiendo a Richard Musgrave (1959), cumple tres funciones básicas: *asignativa*, cuando corrige los fallos de mercado y provee bienes públicos; *redistributiva*, cuando modifica la distribución de la renta resultante del mercado; y *estabilizadora*, cuando atenúa las fluctuaciones de la actividad. Su importancia deriva de que actúa directamente sobre la demanda agregada y, a través de los impuestos y del gasto productivo, también sobre la oferta, con efectos sobre variables reales como el empleo, la producción y el crecimiento.

Históricamente, su papel ha enfrentado a las escuelas herederas de Keynes, que han defendido una política fiscal activa, capaz de sostener la demanda en las recesiones a través del **multiplicador**, con las herederas de la tradición neoclásica, para las que la economía dispone de mecanismos de ajuste automáticos y el gasto público tiende a desplazar al privado (**efecto expulsión**). La **equivalencia ricardiana** de Robert Barro (1974), según la cual, dada la senda del gasto público, sustituir impuestos por deuda no altera la demanda si los agentes anticipan los impuestos futuros, llevó ese escepticismo a su expresión más radical. La política fiscal se enfrenta, además, a limitaciones teóricas, como la magnitud incierta de los multiplicadores y los efectos sobre los incentivos. Entre las empíricas figuran los retardos en su aplicación y la restricción presupuestaria intertemporal, que obliga a preservar la sostenibilidad de la deuda.

Estas limitaciones explican por qué chocan dos maneras de entender la política fiscal, una que la usa para estabilizar la demanda a corto plazo y otra que la concibe como palanca de crecimiento a largo plazo por su efecto sobre la oferta. La Gran Recesión de 2008 y la crisis de deuda soberana (2010-2014) reavivaron ambas, al enfrentar las políticas de estímulo con las de consolidación, y dieron lugar a una revisión al alza del valor estimado de los multiplicadores en las recesiones. Después, la pandemia de 2020 impulsó la mayor expansión fiscal coordinada desde la posguerra. Hoy, el aumento de la deuda pública y de los tipos de interés sitúa la **sostenibilidad fiscal** y el diseño de las **reglas fiscales** en el centro del debate.

Los efectos de la política fiscal sobre el nivel de renta a corto plazo, centrales en el debate sobre las políticas de estabilización y medidos por los multiplicadores (tema A40), son la base del análisis del ahorro (sección 1). Este tema analiza, sobre todo, sus efectos sobre el crecimiento económico y el ahorro, que constituyen hoy el núcleo del debate sobre el tamaño y el diseño del sector público.

El tema se organiza en torno a los tres canales por los que la política fiscal incide en el ahorro y el crecimiento:

- *Gasto público no productivo y su gravamen*: sus efectos se analizan desde la perspectiva neoclásica y desde la perspectiva neokeynesiana.

- *Diseño del sistema de pensiones*: constituye la principal vía por la que el sector público altera las decisiones de ahorro de los hogares.

- *Gasto público productivo*: sus efectos sobre el crecimiento se estudian en los modelos de crecimiento exógeno y endógeno y se contrastan con la evidencia empírica.

# 1. Efectos vía gasto público no productivo y gravamen

El análisis del gasto público no productivo y su gravamen parte del modelo IS-LM y pasa después a las perspectivas de la nueva macroeconomía clásica (NMC) y de la nueva economía keynesiana (NEK).

## 1.1. Antecedentes: el modelo IS-LM

El punto de partida es el modelo IS-LM, del que se exponen los efectos de la política fiscal y el debate sobre su importancia relativa.

### 1.1.1. Efectos de la política fiscal

En el corto plazo, con precios fijos, el nivel de renta viene determinado por la demanda agregada. Una expansión fiscal, que traslada la curva IS a la derecha, tiene dos efectos:

1. *Efecto multiplicador*: en el mercado de bienes, más gasto público o menos impuestos elevan la renta en una cuantía que fija el multiplicador, cuyo valor depende de la teoría del consumo que se adopte[1]. Con la función de consumo keynesiana, $C = C_0 + c(Y - T_0)$, donde $c \in (0, 1)$ es la propensión marginal a consumir y $T_0$ los impuestos de suma fija, los multiplicadores son:

- *Multiplicador del gasto*: a tipo de interés constante, $\partial Y/\partial G_0 = 1/(1-c) > 1$, porque cada unidad de gasto genera renta, que en parte se consume y vuelve a generar renta.

- *Multiplicador de los impuestos*: $\partial Y/\partial T_0 = -c/(1-c)$, menor en valor absoluto que el del gasto, porque una rebaja de impuestos solo eleva la demanda en la parte de la renta disponible que se consume.

- *Multiplicador del presupuesto equilibrado*: la suma de los dos anteriores es igual a la unidad (Haavelmo, 1945), de modo que un aumento del gasto financiado con impuestos de igual cuantía eleva la renta en esa cuantía.

- *Multiplicador en el IS-LM*: cuando el tipo de interés se ajusta en el mercado de dinero, el multiplicador es $\partial Y/\partial G_0 = 1/(1-c+vk/h)$, menor que a tipo de interés constante por el efecto expulsión. En esta expresión, $v$ es la sensibilidad de la inversión al tipo de interés y $k$ y $h$ las de la demanda de dinero a la renta y al tipo de interés.

2. *Efecto expulsión* (*crowding out*): con más renta, los agentes necesitan más dinero para sus transacciones y lo obtienen vendiendo bonos, cuyo precio baja, con lo que sube el tipo de interés ($\downarrow B^d \rightarrow \downarrow P_B \rightarrow \uparrow i$), lo que expulsa gasto privado.

Hay, además, otros efectos que dependen del contexto y de la forma de financiación:

1. *Efecto desbordamiento*: en economía abierta (Mundell-Fleming), la apertura del canal de importaciones reduce el multiplicador, pues parte del estímulo se filtra al exterior, y la efectividad depende del régimen cambiario y de la movilidad de capital.

2. *Financiación vía impuestos*: reduce el multiplicador, pero no lo anula, porque el del presupuesto equilibrado es positivo.

[1] Las hipótesis de la renta permanente y del ciclo vital reducen la propensión a consumir la renta corriente y, con ella, el multiplicador. La reducción es mayor en la primera (el consumo no responde a la renta transitoria) que en la segunda (la renta transitoria se reparte a lo largo de la vida restante).

3. *Financiación vía deuda*: añade el *efecto cartera*, por el que la mayor oferta de bonos reduce su precio, eleva la rentabilidad exigida y desplaza la LM a la izquierda, y, para los monetaristas, los efectos riqueza.

4. *Financiación vía emisión de dinero*: añade el *efecto liquidez*, por el que el aumento de la oferta monetaria desplaza la LM hacia la derecha y limita el efecto expulsión, si bien para los monetaristas sus efectos se diluyen a largo plazo por la teoría cuantitativa.

Gráficamente, en el plano $(Y, i)$, la IS recoge el equilibrio del mercado de bienes y es decreciente, porque un tipo menor eleva la inversión. La LM recoge el del mercado de dinero y es creciente, porque una renta mayor eleva la demanda de dinero. Las tres vías de financiación parten del punto 0, con la renta $Y_0$ y el tipo de interés $i_0$, y la expansión desplaza la IS de $\text{IS}_0$ a $\text{IS}_1$:

1. *Vía impuestos* (figura A38.1): con un aumento de impuestos igual al del gasto ($\Delta G_0 = \Delta T_0 > 0$), la LM no varía y la economía pasa al punto 1, con más renta ($Y_1 > Y_0$) y un tipo de interés más alto ($i_1 > i_0$).

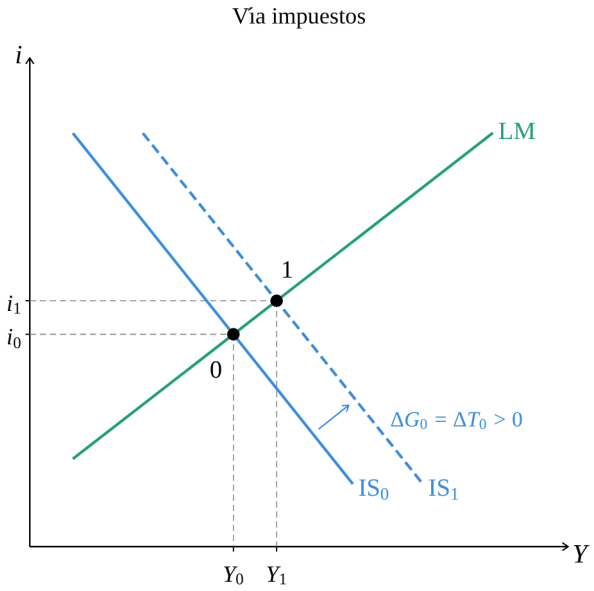



Figura A38.1: Financiación vía impuestos y multiplicador del presupuesto equilibrado en el IS-LM.

2. *Vía deuda* (figura A38.2): el aumento del gasto ($\Delta G_0 > 0$) lleva la economía al punto $0'$. Después, la mayor oferta de bonos ($\Delta B > 0$) desplaza la LM a la izquierda, de $\text{LM}_0$ a $\text{LM}_1$, hasta el punto 1, con un tipo de interés $i_1$ aún mayor y más expulsión.

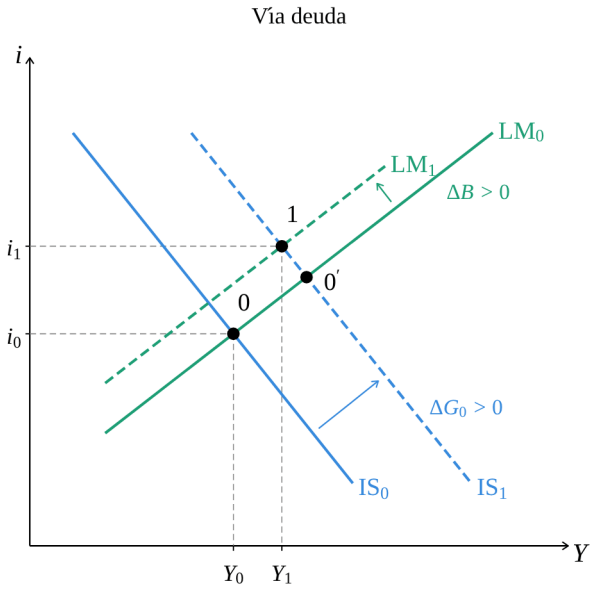



Figura A38.2: Financiación vía deuda. La expansión fiscal expulsa parcialmente la inversión.

3. *Vía emisión de dinero* (figura A38.3): la expansión monetaria ($\Delta M^s > 0$) desplaza la LM a la derecha, de $\text{LM}_0$ a $\text{LM}_1$, y la economía pasa al punto 1 sin subida del tipo de interés ($i_0 = i_1$), por lo que no hay expulsión.

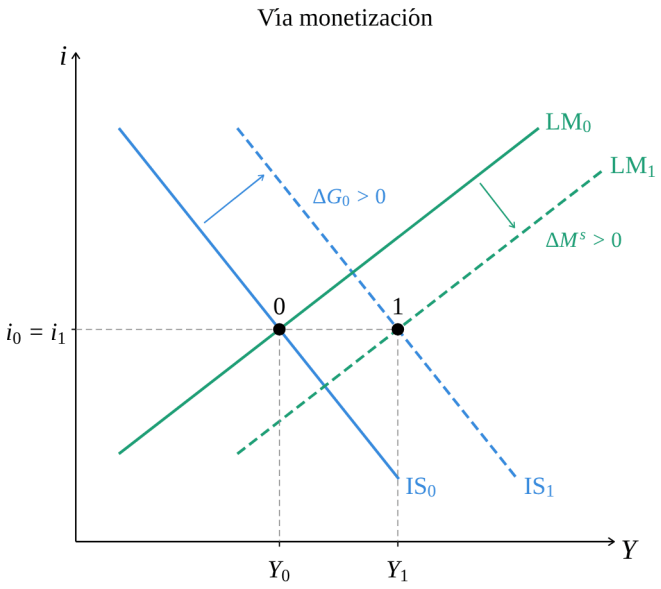



Figura A38.3: Financiación vía emisión de dinero. La acomodación monetaria neutraliza el efecto expulsión.

Los efectos de las distintas vías de financiación del déficit se desarrollan en el tema A39, apartado 1.2. Aquí se retienen en lo que afectan a la efectividad de la política fiscal.

### 1.1.2. Debate

Históricamente se ha discutido la importancia relativa de estos efectos, que depende de las pendientes de las curvas. Las pendientes reflejan cuánto responden al tipo de interés la inversión ($v$) y la demanda de dinero ($h$), como recoge el multiplicador del IS-LM.

**Perspectiva keynesiana**: la inversión, gobernada por los instintos animales, es muy volátil y apenas responde al tipo de interés ($v$ bajo), así que la IS es muy rígida a este. El consumo, por su parte, se apoya sobre todo en la renta disponible. Con el dinero sucede lo contrario, pues el motivo especulación vuelve su demanda, y con ella la LM, muy elástica al tipo de interés, y puede llegarse incluso a la trampa de liquidez[2].

La política fiscal es, así, muy efectiva, porque el efecto multiplicador es elevado y el efecto expulsión reducido, y el multiplicador del gasto es superior a la unidad.

**Perspectiva neoclásica**: en la visión monetarista, el gasto privado reacciona con fuerza al tipo de interés real, porque inversores y hogares atienden a la rentabilidad de los activos y recomponen su cartera de activos y de bienes duraderos cuando cambian los rendimientos relativos. La IS resulta, en consecuencia, muy elástica a este.

A su vez, la demanda de dinero, y con ella la LM, es muy rígida al tipo de interés ($h \approx 0$), dado que el dinero tiene muchos sustitutivos imperfectos.

Por ambas razones, la política fiscal es poco efectiva, porque el efecto multiplicador es reducido y el efecto expulsión es total o casi total, de modo que el multiplicador del gasto es inferior a la unidad.

Gráficamente, en el plano $(Y, i)$ (figura A38.4), el mismo aumento del gasto ($\Delta G_0 > 0$) desplaza la IS de $\text{IS}_0$ a $\text{IS}_1$ en ambas perspectivas. A tipo de interés constante, el multiplicador llevaría la economía del punto 0 al $0'$, pero la subida del tipo de interés la devuelve al punto 1. Esa vuelta, de $Y_{0'}$ a $Y_1$, mide la expulsión, que depende de las pendientes:

1. *Perspectiva keynesiana* (panel a): la subida del tipo de interés, de $i_0$ a $i_1$, apenas reduce la inversión, y $Y_1$ queda, por tanto, cerca de $Y_{0'}$.

2. *Perspectiva neoclásica* (panel b): la subida del tipo de interés expulsa casi tanta inversión como gasto añade, y $Y_1$ apenas supera a $Y_0$.

[2] Ambas sensibilidades son los "eslabones débiles" keynesianos del mecanismo de transmisión monetaria.

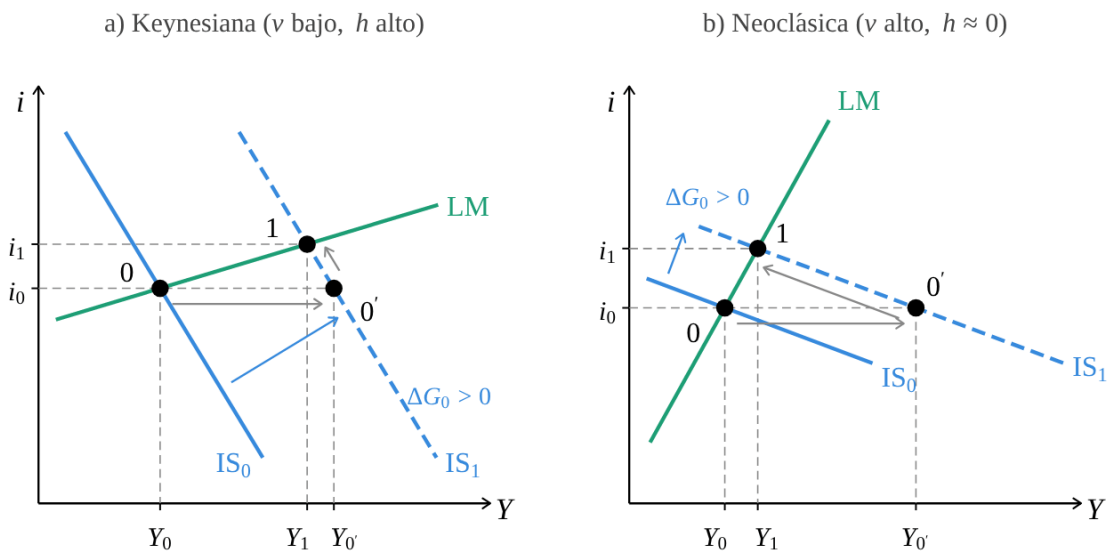



Figura A38.4: Expansión fiscal en el IS-LM según las pendientes de las curvas. La expulsión, de $Y_{0'}$ a $Y_1$, es pequeña en la perspectiva keynesiana (a) y casi total en la neoclásica (b).

## 1.2. Perspectiva NMC: el modelo del ciclo real con sector público

**Origen**: el modelo es el del ciclo real de la nueva macroeconomía clásica, ampliado con sector público en la línea de Baxter y King (1993).

**Idea**: la lectura neoclásica del IS-LM puede ilustrarse a través de un modelo del ciclo real con sector público.

### 1.2.1. Supuestos

**Entorno**: la economía es cerrada, con incertidumbre (*shocks* tecnológicos y fiscales), tiempo discreto y horizonte infinito. Existe también competencia perfecta, completitud en todos los mercados y plena flexibilidad de precios, salarios y tipo de interés.

**Agentes**: se cumple la hipótesis de expectativas racionales. Existe un número elevado de agentes idénticos, que son simultáneamente consumidores, productores y oferentes de trabajo, con vida infinita (existe altruismo intergeneracional perfecto).

El sector público realiza gasto público no productivo, recauda impuestos de suma fija y financia su déficit con deuda. Por consiguiente, el gasto público no tiene efecto distribución y las figuras impositivas son neutrales (no distorsionan las decisiones de consumo-ocio, aunque sí las afectan vía efecto riqueza).

**Tecnología**: la producción sigue una función Cobb-Douglas con tecnología neutral en el sentido de Hicks y rendimientos constantes a escala, y el capital se deprecia a una tasa $\delta$.

### 1.2.2. Modelo

El modelo se resume en las ecuaciones de comportamiento de hogares y empresas, las que describen la tecnología y el sector público, y una condición de equilibrio competitivo.

**Hogares**: deciden cuánto consumir y cuánto trabajar.

▪ *Planteamiento*: maximizan su utilidad intertemporal, aditivamente separable en el tiempo, en consumo y ocio, sujetos a su restricción presupuestaria:

$$\max_{\{c_t,l_t,k_{t+1},b_{t+1}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, 1-l_t), \qquad \beta \in (0,1)$$
$$\text{s.a.:} \begin{cases} c_t + [k_{t+1} - (1-\delta)k_t] + (b_{t+1} - b_t) = y_t - T_t + r_t b_t \\ k_0,\ b_0 \text{ dados} \\ \lim_{T\to\infty} \mathbb{E}_0 \left[b_{T+1}/R_T\right] \geq 0 \end{cases}$$

donde $c_t$ es el consumo, $l_t$ las horas trabajadas y $1-l_t$ el ocio (dotación de tiempo normalizada a uno), $\beta \in (0,1)$ el factor de descuento, $k_t$ el *stock* de capital y $\delta$ su tasa de depreciación. Por último, $b_t$ son los bonos públicos en manos del hogar, $r_t$ el tipo de interés real que rinden, $y_t$ la renta obtenida como productor (que remunera trabajo y capital) y $T_t$ el impuesto de suma fija. El capital y los bonos iniciales están dados, y la condición de no-Ponzi, con $R_T \equiv \prod_{s=0}^{T}(1+r_s)$, impide endeudarse indefinidamente.

▪ *CPO*: con solución interior y la restricción agotada, del lagrangiano se obtienen tres condiciones:

1. *Condición de ocio-consumo*: $\text{RMS}_{c,1-l} = (\partial u/\partial(1-l_t))/(\partial u/\partial c_t) = w_t$, de carácter intratemporal, donde $w_t$ es el salario real.

2. *Ecuación de Euler estocástica del consumo*: $\partial u/\partial c_t = \beta \mathbb{E}_t[(\partial u/\partial c_{t+1})(1+r_{t+1})]$, según la cual la utilidad marginal de una unidad consumida hoy equivale a la que reportaría, descontada, en el periodo siguiente. Se trata de la condición de compra de bonos, cuyo factor de rendimiento bruto es $1+r_{t+1}$.

3. *Ecuación de Euler del capital*: $\partial u/\partial c_t = \beta \mathbb{E}_t[(\partial u/\partial c_{t+1})(1+r^k_{t+1})]$, con $r^k_{t+1} = \partial f/\partial k_{t+1} - \delta$ la rentabilidad neta del capital, condición análoga a la de los bonos. En equilibrio, la ausencia de arbitraje entre bonos y capital exige $r^k_{t+1} = r_{t+1}$.

▪ *CSO*: concavidad estricta de $u$ en consumo y ocio, restricción lineal y condiciones de transversalidad sobre capital y bonos, por lo que las CPO son necesarias y suficientes.

**Empresas**: deciden cuánto producir y contratar.

▪ *Planteamiento*: maximizan su beneficio en cada periodo, tomando como dados el salario y el coste de uso del capital, sujetas a la tecnología:

$$\max_{l_t,k_t} \Pi_t = y_t - w_t l_t - (r_t + \delta)k_t \quad \text{s.a.:} \quad y_t = A_t k_t^{\alpha} l_t^{1-\alpha}$$

donde $\Pi_t$ es el beneficio, $w_t l_t$ la masa salarial y $(r_t+\delta)k_t$ el coste bruto de uso del capital.

▪ *CPO*: sustituyendo la tecnología en el beneficio se obtienen dos condiciones, análogas a las del contexto estático:

1. *Demanda de trabajo*: $\partial f/\partial l_t = (1-\alpha)\dfrac{y_t}{l_t} = w_t$, es decir, la empresa contrata trabajo hasta que su productividad marginal iguala al salario real.

2. *Demanda de capital*: $\partial f/\partial k_t = \alpha\dfrac{y_t}{k_t} = r_t + \delta$, es decir, alquila capital hasta que su productividad marginal iguala a su coste bruto de uso. En términos netos, $\partial f/\partial k_t - \delta = r_t$, que es la rentabilidad $r_t^k$ de la ecuación de Euler del capital de los hogares.

Con rendimientos constantes a escala, ambas condiciones agotan el producto (teorema de Euler), $y_t = w_t l_t + (r_t + \delta)k_t$, y el beneficio de equilibrio es nulo.

- *CSO*: la tecnología Cobb-Douglas es cóncava en $(k, l)$ y estrictamente cóncava en cada factor, luego las CPO caracterizan el máximo.

**Tecnología**: se resume en tres ecuaciones:

- *Función de producción*: la Cobb-Douglas per cápita del programa de la empresa, donde $A_t$ es el nivel tecnológico y $\alpha \in (0, 1)$ la participación del capital en la renta.

- *Progreso técnico*: la tecnología crece en torno a una tendencia a un ritmo exógeno y constante, afectada por un componente aleatorio,

$$\ln A_t = \ln \bar{A} + g_A t + \tilde{a}_t$$

donde $\bar{A}$ es el nivel inicial, $g_A$ la tasa de progreso técnico y $\tilde{a}_t$ la desviación respecto a la tendencia, que sigue el proceso de los *shocks* tecnológicos.

- *Shocks tecnológicos*: siguen un proceso AR(1),

$$\tilde{a}_t = \rho_A \tilde{a}_{t-1} + \phi_t, \quad \rho_A \in (-1, 1)$$

donde $\rho_A$ recoge la persistencia de los *shocks* y $\phi_t$ es una perturbación no correlacionada, es decir, un ruido blanco.

**Sector público**: su comportamiento se condensa en otras tres ecuaciones:

- *Restricción presupuestaria*:

$$(g_t - T_t) = b_{t+1} - (1 + r_t)b_t$$

donde $g_t$ es el gasto público no productivo y $T_t$ la recaudación del impuesto de suma fija. Según la restricción, el déficit primario (lado izquierdo) más los intereses de la deuda heredada, $r_t b_t$, se financian con la emisión neta de bonos, $b_{t+1} - b_t$.

- *Gasto público*: crece en torno a una tendencia a un ritmo constante que coincide con el del *output* en la senda de crecimiento equilibrado, $g_A/(1-\alpha)$ con tecnología neutral en el sentido de Hicks,

$$\ln g_t = \ln \bar{g} + \frac{g_A}{1-\alpha}\, t + \varepsilon_t^g$$

donde $\bar{g}$ es el nivel inicial de gasto, $g_A/(1-\alpha)$ su tasa de crecimiento tendencial y $\varepsilon_t^g$ la desviación respecto a la tendencia.

▪ *Shocks fiscales*: los *shocks* de gasto público siguen un proceso AR(1) no correlacionado con los tecnológicos,

$$\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \psi_t, \quad \rho_g \in (-1, 1)$$

donde $\rho_g$ recoge la persistencia de los *shocks* fiscales y $\psi_t$ es un ruido blanco independiente de $\phi_t$. La persistencia $\rho_g$ es el parámetro que distingue un *shock* transitorio ($\rho_g$ próximo a cero) de uno muy persistente, que en el límite $\rho_g \to 1$ equivale a uno permanente, en la dinámica que sigue.

**Equilibrio**: la condición de equilibrio competitivo exige que los mercados de trabajo y capital se vacíen en todos los periodos, es decir, que en cada momento la inversión iguale al ahorro.

**Estática comparativa**: si el gasto público no productivo aumenta por sorpresa, los agentes saben que en algún momento tendrán que pagarlo con más impuestos, por lo que su renta permanente cae. El *output* no crece por el canal directo de la demanda agregada, propio de los modelos keynesianos tradicionales, sino por el canal riqueza, porque los individuos, más pobres, ofrecen más trabajo. Las empresas, por su lado, no piden más horas a un salario dado, porque el gasto público deja intacta la productividad marginal del trabajo, guía de su demanda.

**Dinámica**: los efectos dependen de la permanencia del *shock*:

1. *Transitorio*: al ser pequeño el efecto riqueza negativo, producción y horas trabajadas suben poco, mientras que el consumo (efecto riqueza) y la inversión (efecto expulsión) retroceden. La renta crece, así, menos que el gasto público, y el multiplicador queda entre cero y uno.

2. *Permanente*: cuanto más persiste el *shock*, más fuerte es el efecto riqueza negativo, más se recorta el consumo en favor de las horas trabajadas y más sube la producción. A ello se suma el efecto ampliación que describen Baxter y King (1993). Con más empleo, el trabajo pierde productividad marginal y el capital la gana. El mayor rendimiento del capital eleva el tipo de interés real y estimula también la inversión privada, de modo que el multiplicador puede llegar a superar la unidad.

**Extensiones**: del modelo se derivan cuatro implicaciones de política económica:

1. *Equivalencia ricardiana* (Barro, 1974): en la restricción presupuestaria intertemporal de los hogares, una vez incorporada la del sector público, no aparecen los bonos. Por ello, no constituyen riqueza neta y, dada la senda de gasto público, la elección entre impuestos y deuda para financiarla es irrelevante. Como los individuos no padecen ilusión de deuda, el desahorro público se traduce en ahorro privado.

Con todo, la política fiscal sigue teniendo efectos reales sobre el ahorro, el consumo, el empleo y el crecimiento. Las limitaciones de la equivalencia (restricciones de liquidez, ilusión fiscal, altruismo imperfecto, impuestos distorsionantes y gasto productivo) motivan el resto del tema.

2. *Efectos no keynesianos* (Giavazzi y Pagano, 1990): si los agentes son ricardianos (tratan la deuda pública como impuestos diferidos), una consolidación fiscal creíble y permanente puede ser expansiva. Ocurre así porque la reducción del gasto anticipa menores impuestos distorsionantes futuros, eleva la riqueza percibida y reduce las primas de riesgo y el tipo de interés. El efecto atracción (*crowding in*) resultante sobre la inversión privada puede compensar la caída de la demanda pública.

3. *Impuestos distorsionantes y suavización impositiva* (Barro, 1979): con impuestos sobre el consumo, el trabajo y el capital, más realistas que los de suma fija, a los efectos riqueza se añaden efectos sustitución intra e intertemporales. En consecuencia, la senda temporal de impuestos y déficit sí es relevante.

Cada punto adicional del tipo impositivo distorsiona más que el anterior. Por eso la política óptima mantiene constante el tipo esperado y deja que operen los estabilizadores automáticos, con un déficit contracíclico y una deuda que amortigua los *shocks* (suavización impositiva). Como corolario, aumentos de los tipos impositivos pueden reducir la recaudación si la caída de la base imponible domina (curva de Laffer).

4. *Aritmética monetarista desagradable* (Sargent y Wallace, 1981): si la autoridad fiscal no ajusta sus déficits, la restricción presupuestaria intertemporal del sector público acaba imponiéndose a la autoridad monetaria.

> La equivalencia ricardiana y los efectos no keynesianos se desarrollan en el tema A06, apartado 3.2, y la suavización impositiva y la aritmética monetarista desagradable en el tema A39, apartados 1.2.2 y 3.2.

### 1.2.3. Conclusiones

**Aspectos positivos**: los modelos del ciclo real supusieron un avance metodológico, pues en ellos se basa la mayoría de los modelos posteriores. La hipótesis de equivalencia ricardiana es, además, un valioso punto de partida teórico. En cuanto al ahorro y el crecimiento, la lección de los modelos del ciclo real es que el gasto público no productivo solo eleva el *output* a costa del consumo privado, vía efecto riqueza negativo, y que su financiación con impuestos distorsionantes reduce la producción potencial.

**Limitaciones**: los resultados dependen, en parte, de la calibración, lo que puede suponer un reto econométrico. Con todo, la principal limitación es la naturaleza walrasiana del modelo, que supone que no existen rigideces de ningún tipo y que los ciclos son la respuesta óptima de agentes racionales a perturbaciones reales. Por ello, los modelos del ciclo real en su versión walrasiana pura apenas se aplican hoy en día.

## 1.3. Perspectiva NEK: el modelo IS-PC-MR con gasto público

**Origen**: el modelo es el canónico de la nueva economía keynesiana, ampliado con gasto público.

**Idea**: la lectura keynesiana del IS-LM, que enfatiza el efecto de las rigideces, se formaliza hoy en el modelo canónico de la nueva economía keynesiana.

> El modelo IS-PC-MR se desarrolla en el tema A07, apartado 2.1, y su relación con la política monetaria y el problema del límite inferior de los tipos de interés (*zero lower bound*, ZLB), en el tema A40, apartado 2.1. Aquí se retiene lo que interesa a los efectos del gasto público.

### 1.3.1. Supuestos

**Entorno**: competencia monopolística en el mercado de bienes finales, rigideces nominales de precios a la Calvo (cada periodo solo una fracción de las empresas puede revisar su precio) y

mercado de trabajo competitivo. Por las rigideces, la política monetaria no es neutral a corto plazo, aunque sí a largo plazo.

**Agentes**: los hogares y las empresas tienen expectativas racionales, como en el modelo del ciclo real. Una autoridad monetaria fija el tipo de interés nominal según una regla, y un sector público realiza gasto no productivo financiado con impuestos de suma fija. En el modelo canónico, los hogares son ricardianos (vida infinita y sin restricciones de liquidez).

**Tecnología**: $Y_t = A_t L_t^{1-\alpha}$, con el trabajo como único factor variable y sin acumulación de capital, lo que elimina la ecuación de Euler del capital del modelo anterior.

### 1.3.2. Modelo

**Planteamiento**: a diferencia del modelo del ciclo real, cuyo equilibrio se caracteriza por las condiciones de primer orden de hogares y empresas, el modelo neokeynesiano se resume en tres ecuaciones agregadas, que son las versiones log-linealizadas de aquellas. Así, la IS dinámica sustituye a la ecuación de Euler del consumo, la curva de Phillips a la condición de fijación de precios de las empresas, y la regla de Taylor incorpora a la autoridad monetaria.

**IS dinámica**: relaciona la brecha de producción (*output gap*) con la brecha esperada, el tipo de interés real y el gasto público:

$$\tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta}(r_t - \bar{r}_t) + \varepsilon_t^g$$
$$r_t = i_t - \mathbb{E}_t[\pi_{t+1}], \qquad \bar{r}_t = f(\rho^+, g^+, (\mathbb{E}_t[\Delta \ln A_{t+1}])^+)$$

donde $\tilde{y}_t = y_t - \bar{y}_t$ es la brecha de producción (*output* respecto de su nivel natural, el que se daría con precios flexibles, ambos en logaritmos), $\theta > 0$ la aversión relativa al riesgo, $i_t$ el tipo de interés nominal, $r_t$ el tipo real *ex ante* y $\varepsilon_t^g$ el *shock* fiscal[3]. Por su parte, $\bar{r}_t$ es el tipo de interés natural, que aumenta con la impaciencia ($\rho$, con $\beta = 1/(1+\rho)$), con el gasto público, $g$, y con el crecimiento esperado de la productividad.

**Curva de Phillips neokeynesiana**: $\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \kappa \tilde{y}_t$ determina la inflación dada la senda de la brecha, que depende a su vez de la senda de los tipos y del gasto público. En ella, $\pi_t$ es la inflación y $\kappa > 0$ la pendiente, que es tanto mayor cuanto menor es la rigidez de precios.

**Regla de Taylor**: $i_t = \bar{i}_t + \alpha_\pi \tilde{\pi}_t + \alpha_y \tilde{y}_t + \varepsilon_t^m$, donde $\bar{i}_t = \bar{r}_t + \pi^T$ es el tipo de interés nominal neutral (tipo real natural más objetivo de inflación) y $\tilde{\pi}_t = \pi_t - \pi^T$ la desviación de la inflación. Los coeficientes $\alpha_\pi > 1$ y $\alpha_y \geq 0$ son las respuestas a la inflación (principio de Taylor) y a la brecha, y $\varepsilon_t^m$ el *shock* de política monetaria.

**Equilibrio**: las tres ecuaciones determinan las sendas de $\tilde{y}_t$, $\pi_t$ e $i_t$ dadas las de los *shocks*. En ausencia de *shocks*, y si los precios no revisados se indexan al objetivo (o si $\pi^T = 0$), el equilibrio estacionario es $\tilde{y} = 0$, $\pi = \pi^T$ e $i = \bar{r} + \pi^T$, es decir, la economía produce su nivel natural y la inflación coincide con el objetivo. El principio de Taylor garantiza que ese equilibrio sea único.

**Desarrollo gráfico**: el modelo IS-PC-MR se representa en los planos brecha de producción-desviación del tipo de interés real respecto del natural, arriba, y brecha de producción-inflación, abajo (figura A38.5):

1. *Curva IS dinámica*, IS: es decreciente, porque un tipo real más alto respecto del natural, $r - \bar{r}$, contrae la demanda y la brecha de producción, $\tilde{y}$.

[3] El *shock* fiscal recoge solo la parte del impulso del gasto público no incorporada en $\bar{r}_t$, de modo que el mismo impulso no se computa dos veces.

2. *Curva de Phillips*, CP: crece con la brecha de producción, con pendiente $\kappa$.

3. *Curva MR*, $\text{MR}_0$: resulta de sustituir la regla de Taylor en la IS dinámica y es decreciente, porque el banco central responde a una mayor inflación con un tipo real más alto y una brecha menor (tema A07, apartado 2.1.2).

4. *Equilibrio*: en el punto 0, la brecha es nula, la inflación está en el objetivo, $\pi^T$, y el tipo real iguala al natural.

**Estática comparativa**: con rigideces nominales, un aumento del gasto público se transmite por dos canales:

1. *Canal de la riqueza*: los consumidores, igual que en el ciclo real, saben que el gasto se pagará con impuestos futuros, se sienten más pobres (efecto riqueza negativo), consumen menos y ofrecen más trabajo. Este canal eleva el *output* natural, $\bar{y}_t$.

2. *Canal del mark-up*: las empresas no pueden ajustar sus precios de inmediato y atienden el estímulo de demanda produciendo más, para lo que, a diferencia del modelo neoclásico, contratan más trabajo. Con ello, el margen precio-coste se estrecha y el salario real sube. Este canal eleva el *output* efectivo, $y_t$, aunque el consumo de los hogares ricardianos lo sigue fijando la ecuación de Euler. Por ello, si el aumento del gasto es transitorio, el consumo no varía cuando el tipo real se mantiene constante y cae cuando el banco central lo eleva (extensión 2).

Como un canal eleva el *output* natural y el otro el efectivo, el efecto neto sobre la brecha y la inflación es, *a priori*, ambiguo, y depende de la respuesta monetaria. Gráficamente (figura A38.5), cuando domina el canal del *mark-up*, el aumento del gasto ($\Delta g > 0$) eleva el tipo natural en $\Delta\bar{r}$ y desplaza la MR de $\text{MR}_0$ a $\text{MR}_1$. A tipo real constante, la economía pasaría del punto 0 al $0'$, con una brecha $\tilde{y}_{0'}$. Ante las presiones inflacionistas, sin embargo, el banco central, siguiendo la regla de Taylor, eleva el tipo real en $\Delta r$ y contrarresta parte del impulso. En consecuencia, la economía se sitúa en el punto 1, con una brecha $\tilde{y}_1$ menor y una inflación $\pi_1$ por encima del objetivo.

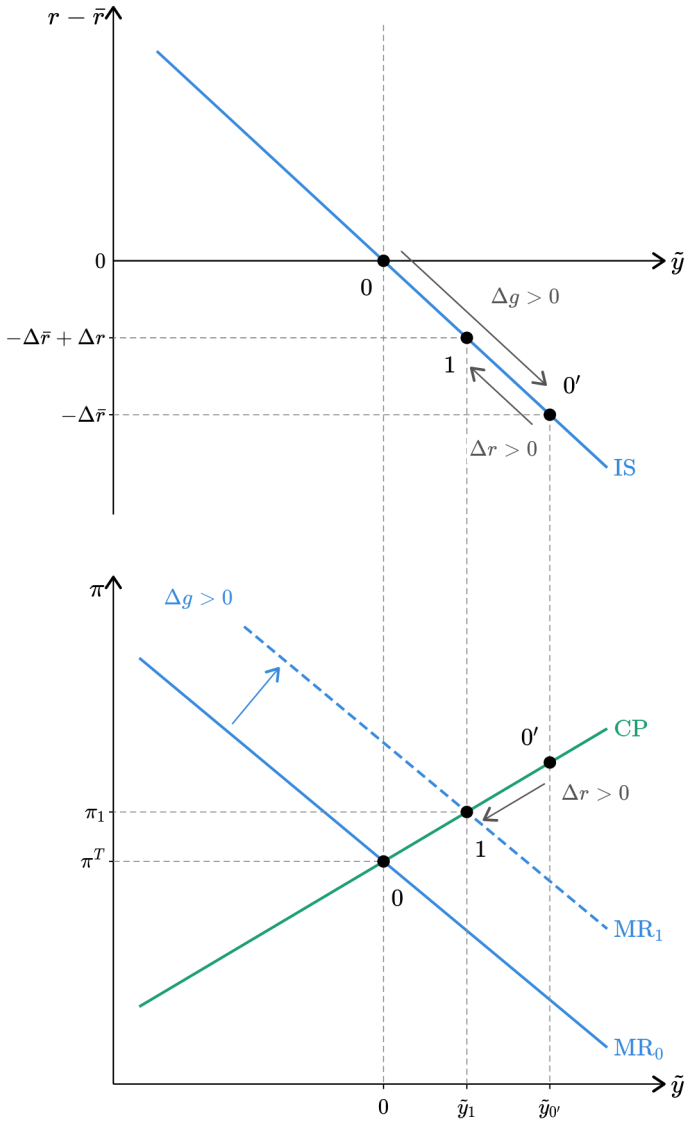



Figura A38.5: Expansión fiscal ($\Delta g > 0$) en el modelo IS-PC-MR. A tipo real constante, la economía pasaría de 0 a 0′, y con la regla de Taylor ($\Delta r > 0$) se sitúa en 1, con un multiplicador inferior a la unidad.

**Dinámica**: como en el modelo del ciclo real, la persistencia del *shock* importa:

1. *Transitorio*: genera un efecto riqueza pequeño y opera sobre todo por el canal del *mark-up*, con un multiplicador de impacto mayor que el del permanente.

2. *Permanente*: genera un efecto riqueza grande, eleva más el *output* natural y menos la brecha, y su multiplicador se aproxima al neoclásico.

**Extensiones**: las tres siguientes precisan las condiciones que determinan la efectividad de la política fiscal:

1. *Efectividad con restricciones de liquidez* (Galí, López-Salido y Vallés, 2007): parten del modelo canónico con dos tipos de agentes, ricardianos, que suavizan su consumo intertemporalmente, y

no ricardianos, que no pueden ahorrar ni endeudarse y consumen su renta corriente. Puesto que los no ricardianos no participan en los mercados financieros, no sufren el efecto riqueza negativo, y el canal del *mark-up* (mayor demanda de trabajo, mayor salario real) eleva directamente su consumo. Por tanto, si la proporción de agentes no ricardianos es suficiente y el gasto se financia en parte con deuda, el multiplicador del gasto público es superior a la unidad.

2. *Interrelación con la política monetaria* (Woodford, 2011): la respuesta de la autoridad monetaria determina el multiplicador en tres escenarios, a los que se añade una extensión del último:

- *Tipo de interés real constante* ($r_t = \bar{r}$): la política monetaria acomoda por completo el impulso, el consumo no varía por la ecuación de Euler y el multiplicador es exactamente igual a uno.

- *Objetivo estricto de inflación o regla de Taylor* ($\tilde{y}_t = 0$ como objetivo): el banco central eleva el tipo real ante las presiones inflacionistas, el consumo cae y el multiplicador es inferior a uno.

- *Límite inferior de los tipos de interés* ($i_t = 0$): el tipo nominal no puede responder al estímulo, la mayor inflación esperada reduce el tipo real, el consumo aumenta y el multiplicador es superior a uno. No obstante, el resultado exige que el gasto no se prolongue más allá del periodo en que el límite es vinculante.

- *Perturbaciones del factor de descuento*: una perturbación que eleva el ahorro precautorio de los hogares (p. ej. tras la crisis financiera de 2008) puede hacer negativo el tipo natural y vinculante la restricción de tipos nulos. Como el tipo real no puede bajar hasta él, la pérdida de *output* es grande y la política fiscal es especialmente efectiva.

3. *Uniones monetarias* (Galí y Monacelli, 2008): parten del modelo en economía abierta con países que conservan la política fiscal (consumo público) y comparten la política monetaria. El efecto del gasto público nacional no depende del grado de acomodación monetaria, en la medida en que la política monetaria común no responde a la inflación generada en el país. Por ello, el multiplicador es mayor que en una economía con política monetaria propia y objetivo estricto de inflación, aunque sigue siendo inferior a la unidad si el país financia el gasto con sus propios impuestos (Farhi y Werning, 2016). Ahora bien, la política fiscal óptima de cada país debe corregir las distorsiones que la política monetaria común no puede atender.

### 1.3.3. Conclusiones

**Aspectos positivos**: el modelo neokeynesiano recupera la efectividad de la política fiscal a corto plazo sin renunciar a los microfundamentos del ciclo real, y muestra que su magnitud depende de las restricciones de liquidez de los hogares y de los regímenes monetario y cambiario. En cuanto al ahorro, el gasto público sigue reduciendo el consumo privado vía efecto riqueza, pero el canal del *mark-up* eleva el empleo y el salario real a corto plazo y, si hay hogares no ricardianos, también el consumo de estos (extensión 1).

**Limitaciones**: el modelo canónico considera únicamente gasto público no productivo financiado con impuestos de suma fija, por lo que no permite analizar los efectos sobre el crecimiento a largo plazo, que estudian las secciones 2 y 3.

## 1.4. Evidencia empírica

La evidencia sobre el valor del multiplicador fiscal se reúne en el tema A40, apartado 3.2. De ella interesa aquí lo que permite contrastar las dos perspectivas anteriores.

**Respuesta del consumo privado**: es la predicción que separa ambos modelos. Con el enfoque de vectores autorregresivos estructurales (SVAR), el consumo privado aumenta tras un *shock* de gasto (Blanchard y Perotti, 2002), lo que motivó el modelo con agentes no ricardianos. Con *shocks* de gasto militar fechados cuando se anuncian, en cambio, el consumo privado cae, como predice el efecto riqueza negativo del ciclo real. Ramey (2011), que construye esos *shocks*, atribuye la diferencia a que los del SVAR llegan tarde, cuando los agentes ya conocían el aumento del gasto.

**Hogares no ricardianos**: entre el 20 % y el 40 % de las devoluciones del impuesto sobre la renta de 2001 en EE. UU. se gastó en bienes no duraderos en el trimestre en que se recibieron, aunque se conocían de antemano (Johnson, Parker y Souleles, 2006). Ese gasto contradice la hipótesis de la renta permanente, en la que se apoya la equivalencia ricardiana, y apoya la existencia de hogares restringidos por liquidez.

**Tamaño del multiplicador**: con datos de EE. UU. desde 1889, el multiplicador del gasto queda por debajo de la unidad tanto con desempleo alto como bajo (Ramey y Zubairy, 2018), como en la perspectiva neoclásica. El multiplicador relativo entre los estados de una unión monetaria, cuya política monetaria común no responde al impulso de uno solo, se sitúa, en cambio, en torno a 1,5 (Nakamura y Steinsson, 2014), un valor que respalda la perspectiva neokeynesiana.

## 1.5. Conclusiones

El análisis del gasto público no productivo y su gravamen arroja una conclusión común a ambas perspectivas. El gasto no productivo ejerce sobre el consumo privado un efecto riqueza negativo y, si es transitorio, reduce también la inversión y, con ella, el ahorro nacional.

Los efectos expansivos del gasto sobre el *output* a corto plazo dependen, en cambio, de la existencia de rigideces, de las restricciones de liquidez y del régimen de la política monetaria. En el marco neoclásico, los multiplicadores son inferiores a la unidad si el aumento del gasto es transitorio. En el neokeynesiano, el multiplicador alcanza la unidad cuando la política monetaria mantiene constante el tipo real. Si la política monetaria no responde o si una fracción suficiente de hogares está restringida por liquidez, el multiplicador puede superar la unidad. Los efectos sobre el crecimiento a largo plazo, sin embargo, discurren por otros dos canales (secciones 2 y 3).

# 2. Efectos vía diseño del sistema de pensiones

El diseño del sistema de pensiones condiciona el ahorro y, con él, la acumulación de capital. Una vez descritos los tipos de sistemas, el modelo de generaciones solapadas permite compararlos, y la evidencia empírica cierra la sección.

## 2.1. Sistemas de pensiones

Los sistemas públicos de pensiones, que responden a la función redistributiva del sector público y a la corrección de la miopía del ahorro individual, pesan mucho en las economías avanzadas, sobre todo en las europeas (en España, en torno al 12 % del PIB). Se financian principalmente con cotizaciones sociales, que actúan como un impuesto sobre el trabajo y, en los sistemas de reparto, reducen el ahorro.

Hay dos tipos puros, que difieren en quién financia la pensión de cada generación y, por tanto, en su rendimiento implícito, y la mayoría de países los combinan en sistemas mixtos.

### 2.1.1. Sistema de capitalización (*fully funded*)

**Antecedentes**: los primeros esquemas modernos de pensiones eran de capitalización, y la reforma chilena de 1981 la recuperó como sistema nacional obligatorio de gestión privada.

**Definición**: cada trabajador acumula sus cotizaciones en una cuenta individual invertida en activos financieros y cobra en la jubilación el capital y su rendimiento. No hay, por tanto, redistribución intergeneracional, y el factor de capitalización del ahorro es $(1+r)$, el rendimiento del mercado de capitales.

**Diseño**: el sistema es de aportación definida, con gestión privada o pública y conversión del saldo en una renta vitalicia según la esperanza de vida. Suele añadirse un pilar solidario, financiado con impuestos, que garantiza una pensión mínima.

**Ejemplos**: Chile, los países latinoamericanos que siguieron su modelo en los años noventa (Perú, Colombia, México) y Australia (*superannuation*, 1992).

**Conclusiones**:

- *Aspectos positivos*: es robusto al envejecimiento, porque no depende de la demografía. Asimismo, eleva el ahorro nacional en la fase de acumulación si obliga a ahorrar más de lo que se ahorraría voluntariamente o si sustituye al reparto, profundiza los mercados de capitales y, al ligar cotización y prestación, distorsiona menos la oferta de trabajo.

- *Limitaciones*: traslada al trabajador los riesgos financiero y de longevidad, y sus tasas de sustitución dependen de las comisiones y de la continuidad de las carreras de cotización, por lo que penalizan la informalidad y las lagunas. Además, la transición desde el reparto exige que una generación pague dos veces (la pensión de sus mayores y la propia), con un coste fiscal que puede ser muy elevado.

### 2.1.2. Sistema de reparto (*pay as you go*)

**Antecedentes**: el seguro social de Bismarck (1889) era contributivo y con reservas, pero el reparto puro se generalizó tras la II Guerra Mundial, cuando la destrucción de los fondos y la necesidad de pagar pensiones de inmediato lo impusieron.

**Definición**: lo que cotizan hoy quienes trabajan se destina a pagar las pensiones de quienes hoy están jubilados, de modo que sí hay redistribución intergeneracional. El factor de capitalización implícito es el crecimiento de la masa salarial, $(1+n)(1+g_y)$, que recoge el aumento del número de cotizantes $(n)$ y de la renta de cada uno $(g_y)$ (Samuelson, 1958; Aaron, 1966).

**Diseño**: el sistema es de prestación definida, calculada sobre las bases de cotización de un periodo de referencia, con la edad de jubilación, el tipo de cotización y la regla de revalorización fijados por ley. El Estado garantiza el pago, y los riesgos demográfico y político recaen sobre las generaciones futuras.

**Ejemplos**: lo aplican la mayoría de países europeos (Alemania, Francia, Italia, España) y la Seguridad Social de EE. UU. (1935).

**Conclusiones**:

- *Aspectos positivos*: paga pensiones desde el primer día, reparte los riesgos entre el Estado y los cotizantes, es inmune a las crisis financieras y es más sencillo y barato de administrar.

- *Limitaciones*: su rendimiento cae con el envejecimiento (menor $n$) y con la ralentización del crecimiento (menor $g_y$), lo que obliga a subir cotizaciones, bajar prestaciones o retrasar la jubilación. Además, reduce el ahorro nacional, y la desconexión entre cotización y prestación actúa como un impuesto sobre el trabajo.

### 2.1.3. Sistemas mixtos

**Antecedentes**: la convergencia entre ambos modelos desde los años noventa, impulsada por el informe del Banco Mundial *Averting the Old Age Crisis* (1994), que propuso una arquitectura de tres pilares.

**Definición**: sistemas que combinan un pilar de reparto con pilares de capitalización que lo complementan, de modo que el rendimiento del conjunto es una media ponderada de $(1+n)(1+g_y)$ y $(1+r)$.

**Diseño**: el sistema combina un pilar público de reparto (contributivo o asistencial), otro de capitalización obligatoria (de empresa o individual) y otro de ahorro voluntario con incentivos fiscales. En la variante de cuentas nocionales de aportación definida (NDC), el reparto calcula la pensión como si fuera de capitalización, con un rendimiento nocional ligado al crecimiento de la masa salarial.

**Ejemplos**: lo aplican Suecia, Países Bajos y Dinamarca.

**Conclusiones**:

- *Aspectos positivos*: diversifican los riesgos demográfico y financiero, y las cuentas nocionales restablecen el vínculo entre cotización y prestación sin el coste de transición.

- *Limitaciones*: la coordinación entre pilares es compleja, los pilares de capitalización reproducen la desigualdad de las carreras laborales y la transición hacia el segundo pilar conlleva el mismo doble pago, aunque atenuado.

## 2.2. Modelo base: generaciones solapadas (Allais, 1947; Samuelson, 1958; Diamond, 1965)

**Origen**: el modelo es el de generaciones solapadas de Allais (1947), Samuelson (1958) y Diamond (1965), en su versión de dos periodos.

**Idea**: las diferencias entre el sistema de capitalización y el de reparto pueden estudiarse a través de ese modelo.

> El modelo de generaciones solapadas de Diamond y su propiedad de ineficiencia dinámica se desarrollan en el tema A29, apartado 3.2. Aquí se aplica a los sistemas de pensiones.

### 2.2.1. Supuestos

**Entorno**: la economía es competitiva, no hay incertidumbre y los agentes viven dos periodos. En el primero los agentes trabajan, consumen y ahorran, y en el segundo consumen todo lo ahorrado y desaparecen del modelo. La población crece a una tasa constante y exógena $n$, y la renta laboral a una tasa $g_y$.

**Agentes**: dos cohortes conviven en cada periodo, jóvenes y ancianos, con las mismas preferencias logarítmicas, de modo que solo existe heterogeneidad en la edad. Las empresas contratan factores de acuerdo con su productividad marginal. En cuanto al sector público, se parte de un modelo sin sistema de pensiones y se introduce después cada sistema como una cotización $\tau$ sobre la renta laboral y una pensión en la vejez.

### 2.2.2. Modelo

**Planteamiento**: el joven elige su ahorro para maximizar la utilidad de su ciclo vital. Los tres casos comparten la función objetivo y difieren solo en la restricción presupuestaria, a través de la cotización, $\tau y_t$, que se paga en la juventud, y de la pensión, $p_{t+1}$, que se cobra en la vejez:

$$\max_{s_t} \; u^d = \ln c_t^j + \beta \ln c_{t+1}^a \quad \text{s.a.:} \begin{cases} c_t^j = (1-\tau)\, y_t - s_t \\ c_{t+1}^a = (1+r)\, s_t + p_{t+1} \end{cases}$$

donde $c_t^j$ y $c_{t+1}^a$ son el consumo en la juventud y en la vejez, $\beta$ el factor de descuento, $y_t$ la renta laboral de la juventud, $s_t$ el ahorro privado y $r$ el tipo de interés real. Cada sistema se traduce en una restricción distinta:

1. *Sin sistema de pensiones*: no hay cotización ni pensión, $\tau = 0$ y $p_{t+1} = 0$.

2. *Capitalización*: la pensión es la propia cotización capitalizada al tipo de interés de mercado, $p_{t+1} = (1+r)\tau y_t$.

3. *Reparto*: la pensión es la cotización de la generación siguiente, más numerosa y más rica, $p_{t+1} = (1+n)(1+g_y)\tau y_t$, porque hay $1+n$ jóvenes por cada anciano y cada uno cotiza sobre una renta $(1+g_y)y_t$.

**CPO**: con solución interior, sustituyendo las restricciones en la función objetivo, la condición $\partial u^d/\partial s_t = 0$ da la ecuación de Euler, común a los tres casos, $1/c_t^j = \beta(1+r)/c_{t+1}^a$. Dicho de otro modo, el joven ahorra hasta que la utilidad marginal del consumo presente iguala a la del consumo en la vejez, descontada y capitalizada.

**CSO**: $\partial^2 u^d/\partial s_t^2 = -1/(c_t^j)^2 - \beta(1+r)^2/(c_{t+1}^a)^2 < 0$, luego la CPO es suficiente.

**Solución**: despejando el ahorro de la ecuación de Euler se obtiene la función de ahorro privado,

$$s_t^* = \frac{\beta}{1+\beta}(1-\tau)\,y_t - \frac{p_{t+1}}{(1+\beta)(1+r)}$$

que cae con la cotización y con el valor actual de la pensión. Sustituyendo la pensión de cada sistema, se obtienen tres casos:

1. *Sin sistema de pensiones*: $s_t^* = \frac{\beta}{1+\beta}\,y_t$. El ahorro es exclusivamente privado, el único vehículo de transmisión de consumo entre periodos, y depende únicamente de las preferencias (impaciencia, $\beta$) y de la renta, por lo que podría ser insuficiente.

2. *Capitalización*: $s_t^* = \frac{\beta}{1+\beta}\,y_t - \tau y_t$. El ahorro privado cae exactamente en la cuantía del ahorro público forzoso, de modo que el ahorro nacional ($s_t^* + \tau y_t$) y la acumulación de capital no varían. El agente es indiferente respecto a quién ahorra, porque solo le interesa la tasa de retorno, que es la misma.

3. *Reparto*: el ahorro privado es

$$s_t^* = \frac{\beta}{1+\beta}\,y_t - \tau y_t\left[1 + \frac{(1+n)(1+g_y)-(1+r)}{(1+\beta)(1+r)}\right]$$

Como el sector público no ahorra la cotización, sino que la transfiere a los jubilados, el ahorro nacional y la acumulación de capital caen siempre. Además, si el rendimiento implícito del reparto, $(1+n)(1+g_y)$, supera al tipo de interés, el ahorro privado cae en más que la cotización.

**Bienestar**: la preferencia por un sistema u otro depende de la relación de los parámetros. Si $(1+n)(1+g_y) > (1+r)$, se prefiere el sistema de reparto, pues la contribución capitalizada es mayor que con capitalización, mientras que si $(1+n)(1+g_y) < (1+r)$ se prefiere el de capitalización.

Como corolario, la caída del ahorro nacional bajo reparto no tiene por qué producir pérdidas de bienestar, ya que, si la economía se encuentra en una zona dinámicamente ineficiente (sobreacumulación de capital, $r < n + g_y$), el sistema de reparto mejora el bienestar de todas las generaciones.

Ahora bien, en los países desarrollados las cohortes posteriores al *baby boom* son menos numerosas, lo que reduce $n$, y el ritmo de avance de la renta, $g_y$, se ha frenado desde los años setenta. El rendimiento del reparto se ha reducido por ello, y es más probable que acabe siendo inferior al del ahorro privado.

### 2.2.3. Conclusiones

**Aspectos positivos**: el modelo aísla con claridad el efecto de cada sistema sobre el ahorro nacional (nulo bajo capitalización y siempre negativo bajo reparto) y explica por qué el envejecimiento convierte en insostenibles sistemas que fueron óptimos.

**Limitaciones**: el modelo no incorpora incertidumbre sobre la longevidad ni sobre los rendimientos, que es la razón de ser de los seguros públicos. Supone, además, que la cotización no altera la oferta de trabajo, cuando en la práctica las cotizaciones actúan como un impuesto sobre el trabajo, y omite los costes de transición de un sistema a otro, que son la principal barrera a las reformas.

## 2.3. Evidencia empírica

Se estudia un caso por cada tipo de sistema (reparto en Alemania, capitalización en Chile y mixto en Suecia), más el de España, un sistema de reparto con pilares complementarios poco desarrollados.

**Sistemas de reparto: el caso de Alemania**: el sistema bismarckiano de reparto, con pensiones ligadas a las cotizaciones, ha respondido al envejecimiento con reformas paramétricas que reducen el rendimiento implícito del sistema:

- *Pensiones Riester* (2001): introducción de un pilar de capitalización voluntario subvencionado.
- *Factor de sostenibilidad* (2004): liga la revalorización de las pensiones a la ratio entre cotizantes y pensionistas.
- *Edad de jubilación* (2007): retraso gradual de 65 a 67 años, en aplicación entre 2012 y 2029.

La tasa de sustitución ha descendido, aunque el paquete de pensiones de 2025 prorrogó hasta 2031 un suelo del 48 % para el nivel de las pensiones (*Haltelinie*).

**Sistemas de capitalización: el caso de Chile**: la reforma de 1981 sustituyó el reparto por cuentas individuales obligatorias gestionadas por administradoras privadas de fondos de pensiones (AFP). El sistema profundizó el mercado de capitales, pero su efecto sobre el ahorro nacional es discutido, pues el coste fiscal de la transición redujo el ahorro público.

Además, las tasas de sustitución resultaron bajas por las lagunas de cotización, la informalidad y las comisiones. La insuficiencia de las pensiones obligó a introducir un pilar solidario financiado con impuestos (2008) y, tras un largo debate, una cotización adicional a cargo del empleador con un componente de reparto (2025).

**Sistemas mixtos: el caso de Suecia**: la reforma aprobada en 1998 y aplicada desde 1999 sustituyó el sistema de prestación definida por uno de aportación definida con una cotización del 18,5 % de la renta. De esa cotización, 16 puntos se destinan a cuentas nocionales de reparto (*inkomstpension*), cuyo rendimiento está ligado al crecimiento del salario medio, y 2,5 puntos a cuentas individuales de capitalización (*premiepension*).

Un mecanismo de equilibrio automático, activado por primera vez en 2010 tras la crisis financiera, reduce la revalorización de las pensiones cuando los pasivos del pilar de reparto superan a sus activos (incluido el valor de las cotizaciones futuras). El caso ilustra cómo un sistema mixto traslada a los cotizantes parte del riesgo demográfico y financiero a cambio de garantizar la solvencia del sistema.

**Un sistema de reparto con pilares complementarios poco desarrollados: el caso de España**: el sistema español es público y de prestación definida, con un fondo de reserva y pilares complementarios de capitalización (planes de empleo e individuales) voluntarios. El marco español presenta varios problemas:

- *Fecundidad*: según el Instituto Nacional de Estadística (INE), el indicador coyuntural se situó en 1,10 hijos por mujer en 2024, muy por debajo del nivel de reemplazo (en torno a 2,1). Por ello, cada generación de cotizantes es menos numerosa que la anterior. En términos del modelo de generaciones solapadas, cae $n$ y, con ella, el rendimiento implícito del reparto.

- *Efecto sustitución*: los nuevos jubilados, con carreras de cotización más largas y bases más altas, sustituyen a pensionistas con prestaciones menores, de modo que la pensión media crece por encima de la revalorización. Por tanto, el gasto aumenta aunque no aumente el número de pensionistas.

- *Envejecimiento*: la jubilación de la generación del *baby boom*, más tardío en España que en el resto de Europa, eleva la tasa de dependencia. A ello se suma el aumento de la esperanza de vida a los 65 años, que alarga el periodo de cobro. Según las proyecciones del INE (2026), la tasa de dependencia seguiría aumentando durante todo su horizonte, hasta un máximo del 73,2 % en 2076[4].

- *Tipos de interés*: próximos a cero en la década de 2010, recortaron la rentabilidad del fondo de reserva y de los planes de pensiones complementarios. En términos del modelo, un $r$ bajo hace relativamente más atractivo el reparto, pero obliga a ahorrar más para complementar la pensión pública.

Las reformas de 2011 (edad legal a 67 años, ampliación del periodo de cómputo) y 2013 (índice de revalorización y factor de sostenibilidad, después derogados) actuaron sobre el gasto. Las de 2021-2023, en cambio, actúan sobre el ingreso (mecanismo de equidad intergeneracional, destope gradual de las bases máximas, cuota de solidaridad) y recuperan la revalorización con el IPC, con una cláusula de evaluación periódica por la Autoridad Independiente de Responsabilidad Fiscal (AIReF).

## 2.4. Conclusiones

La capitalización rinde a quien cotiza el tipo de interés de mercado, $(1 + r)$, y el reparto, el crecimiento de la masa salarial, $(1 + n)(1 + g_y)$, mientras que los sistemas mixtos combinan ambos rendimientos. En el modelo de generaciones solapadas, la capitalización sustituye ahorro privado por ahorro público sin alterar el ahorro nacional. El reparto, por el contrario, lo reduce siempre, porque la cotización se transfiere a los jubilados en lugar de invertirse, y la caída crece con el diferencial entre su rendimiento implícito y el tipo de interés. Esa menor acumulación no implica por sí sola una pérdida de bienestar, y en una economía dinámicamente ineficiente el reparto mejora el de todas las generaciones.

El envejecimiento y la ralentización de la renta han reducido $n$ y $g_y$ y, con ellos, el rendimiento del reparto. De ahí la convergencia de las reformas hacia sistemas mixtos que ilustran los casos de Alemania, Chile y Suecia. La sección 3 pasa del ahorro a la productividad, con el gasto público que entra en la función de producción.

[4] El *Ageing Report* de la Comisión Europea, trienal, es la referencia para los efectos a largo plazo del envejecimiento sobre el gasto en pensiones, sanidad y dependencia de los Estados miembros.

# 3. Efectos vía gasto público productivo

Los efectos del gasto público productivo sobre el crecimiento se abordan primero con su caracterización y después con los modelos de crecimiento exógeno y endógeno, para terminar con la evidencia empírica.

## 3.1. Gasto público productivo

El **gasto público productivo** se distingue del no productivo (consumo público, transferencias) en que entra en la función de producción como *input* y eleva la productividad del sector privado. Destacan dos categorías: la educación y capacitación de la mano de obra, que eleva el capital humano, y las infraestructuras y la provisión de otros bienes públicos (transporte, energía, I+D, derechos de propiedad). El sector privado no proveería en cantidad suficiente las infraestructuras y los demás bienes públicos, porque presentan externalidades positivas.

Este gasto plantea una disyuntiva. Por un lado, aumenta la producción potencial de la economía y, por otro, hay que financiarlo, y en la práctica se hace con figuras impositivas distorsionantes, que reducen la renta disponible, la inversión y el consumo privados.

## 3.2. Crecimiento exógeno: el modelo de Ramsey-Cass-Koopmans con sector público

**Origen**: el modelo es el de Ramsey-Cass-Koopmans (RCK), ampliado con impuestos distorsionantes y gasto público productivo.

**Idea**: la disyuntiva que plantea el gasto público productivo puede analizarse a través de un modelo RCK ampliado.

> El modelo de Ramsey-Cass-Koopmans se desarrolla en el tema A43, apartado 3. Aquí se amplía con impuestos y gasto público.

### 3.2.1. Supuestos

**Entorno**: el modelo supone un mundo neoclásico, sin más fallo de mercado que la externalidad del gasto público productivo, y con certidumbre. La población crece a una tasa constante y exógena $n$.

**Agentes**: la economía la forma un agente representativo con tasa de ahorro endógena y utilidad isoelástica (CRRA), con aversión relativa al riesgo $\theta > 0$. El agente no toma decisiones de trabajo, de modo que ofrece una unidad de trabajo en cada periodo y $w_t$ es la renta laboral.

El sector público recauda impuestos distorsionantes sobre el consumo ($\tau_c$), la renta laboral ($\tau_w$) y la renta del capital ($\tau_r$), que solo afectan a los hogares, y los dedica a gasto público productivo con presupuesto equilibrado.

**Tecnología**: neoclásica, $y = A(g)f(k) = A(g)k^\alpha$, con depreciación $\delta > 0$, donde el gasto público productivo eleva la productividad total de los factores, $A'(g) > 0$, sin ser un factor acumulable.

### 3.2.2. Modelo

Como en el modelo del ciclo real, el equilibrio se caracteriza por las decisiones de hogares y empresas, la restricción del sector público y una condición de equilibrio de mercado. A diferencia de aquel, el tiempo es continuo, no hay incertidumbre ni decisión de trabajo y los impuestos son distorsionantes.

**Hogares**: deciden cuánto consumir y cuánto ahorrar.

▪ *Planteamiento*: maximizan su utilidad descontada sujeta a una restricción presupuestaria en la que los impuestos gravan cada uso y fuente de renta:

$$\max_{c} U = \int_0^{\infty} e^{-(\rho-n)t} \frac{c^{1-\theta}-1}{1-\theta}\, dt, \qquad \theta > 0$$

$$\text{s.a.:} \begin{cases} (1+\tau_c)c + \dot{b} = (1-\tau_w)w + [(1-\tau_r)r - n]\, b \\ b_0 > 0 \text{ dado} \\ \lim_{t\to\infty} \lambda(t)\, b(t) \geq 0 \end{cases}, \qquad \rho > n$$

donde $c$ es el consumo per cápita[5], $\rho$ la tasa de descuento (impaciencia), $n$ la tasa de crecimiento de la población y $\theta$ la aversión relativa al riesgo (inversa de la elasticidad de sustitución intertemporal). La variable $b$ es la riqueza financiera per cápita (que en equilibrio será capital). En cuanto a los precios y los impuestos, $w$ es el salario, $r$ el tipo de interés real antes de impuestos, y $\tau_c$, $\tau_w$ y $\tau_r$ los tipos impositivos sobre el consumo, la renta laboral y la renta del capital, respectivamente. La condición $\rho > n$ garantiza que la utilidad descontada sea finita. La riqueza inicial, $b_0$, está dada, y la condición de no-Ponzi impide acabar endeudado en el límite y se escribe con el multiplicador $\lambda$ del hamiltoniano que sigue[6]. El hamiltoniano en valor presente u origen es

$$H = e^{-(\rho-n)t} \frac{c^{1-\theta}-1}{1-\theta} + \lambda\left[(1-\tau_w)w + ((1-\tau_r)r - n)\, b - (1+\tau_c)c\right]$$

donde $\lambda$ es el precio sombra de la riqueza.

▪ *CPO*: de su resolución se obtienen tres condiciones:

1. *Condición de consumo*: $H_c = 0 \;\Rightarrow\; e^{-(\rho-n)t}c^{-\theta} = \lambda(1+\tau_c)$, es decir, la utilidad marginal descontada del consumo iguala a su precio sombra, incluido el impuesto.

2. *Condición de acumulación*: $H_b = -\dot{\lambda} \;\Rightarrow\; -\dot{\lambda}/\lambda = (1-\tau_r)r - n$, es decir, el precio sombra de la riqueza decrece a una tasa igual al rendimiento neto de impuestos del ahorro menos la tasa de crecimiento de la población.

3. *Condición de transversalidad*: $\lim_{t\to\infty} \lambda b = 0$, que excluye la acumulación de riqueza sin consumo[7].

▪ *CSO*: el hamiltoniano es estrictamente cóncavo en $c$ ($\theta > 0$) y lineal en $b$, por lo que las condiciones de Pontryagin, con la transversalidad, son suficientes (Mangasarian).

[5] En tiempo continuo se omite el argumento temporal: $c \equiv c(t)$, $b \equiv b(t)$, $w \equiv w(t)$, $r \equiv r(t)$ y $\lambda \equiv \lambda(t)$; se escribe solo en las condiciones inicial y terminal.

[6] Por la condición de acumulación, $\lambda(t) = \lambda(0)\, e^{-[(1-\tau_r)R(t)-nt]}$, con $R(t) \equiv \int_0^t r(v)\, dv$ y $\lambda(0) > 0$, de modo que la condición equivale a $\lim_{t\to\infty} b(t)\, e^{-[(1-\tau_r)R(t)-nt]} \geq 0$.

[7] La condición de transversalidad es la holgura complementaria de la no-Ponzi llevada al límite (tema A43, apartado 3.2).

- *Ecuación de Euler*: combinando las dos primeras CPO,

$$g_c \equiv \frac{\dot{c}}{c} = \frac{1}{\theta}\left[(1-\tau_r)\, r - \rho\right]$$

donde $g_c$ es la tasa de crecimiento del consumo. Sin sector público, $g_c = \frac{1}{\theta}(r-\rho)$, es decir, el hogar posterga consumo mientras el rendimiento del ahorro, $r$, supere a su impaciencia, $\rho$, con una intensidad que depende de $\theta$. Con sector público, el rendimiento relevante pasa a ser $(1-\tau_r)r$, el que el hogar efectivamente percibe:

1. *Impuesto sobre el capital*: introduce una *cuña* entre el rendimiento social del ahorro, que sigue siendo $r = A(g)\, df/dk - \delta$, y el privado. Por ello, para cualquier $k$, el consumo crece menos (o decrece más) que sin impuesto, es decir, el hogar ahorra menos.

2. *Impuestos sobre el consumo y el trabajo*: $\tau_c$ y $\tau_w$ no aparecen en la ecuación de Euler. El impuesto sobre el consumo grava por igual el consumo de hoy y el de mañana, sin alterar su precio relativo, y el impuesto sobre el trabajo grava una renta cuya oferta es inelástica por supuesto. Ambos reducen la renta disponible (efecto renta), pero dejan intacta la decisión intertemporal (sin efecto sustitución). Solo el impuesto sobre el capital distorsiona la acumulación, y de ahí que la teoría de la imposición óptima recomiende gravar el consumo antes que el ahorro.

3. *Gasto público*: no aparece en la ecuación de Euler del hogar, sino en el rendimiento del capital que este toma como dado. Al elevar $A(g)$, eleva $r$ para cada $k$ y empuja al hogar a ahorrar más, en sentido contrario al impuesto sobre el capital.

**Empresas**: deciden cuánto capital emplear por trabajador.

- *Planteamiento*: maximizan su beneficio por trabajador en cada instante, tomando como dados los precios de los factores y el gasto público:

$$\max_k \ \Pi = A(g)f(k) - (r+\delta)k - w$$

donde $\Pi$ es el beneficio por trabajador, $A(g)f(k)$ el *output* por trabajador, $(r+\delta)k$ el coste bruto de uso del capital y $w$ el salario de la unidad de trabajo que cada hogar ofrece.

- *CPO*: de la resolución se obtienen dos condiciones:

1. *Demanda de capital*: $A(g)df/dk = r+\delta$, es decir, la productividad marginal del capital iguala a su coste bruto de uso. En términos netos, $r = A(g)df/dk - \delta$.

2. *Salario*: con rendimientos constantes a escala el beneficio es nulo y el trabajo percibe el residuo, $w = A(g)\left[f(k) - kdf/dk\right]$.

- *CSO*: $f'' < 0$, por lo que el beneficio por trabajador es estrictamente cóncavo en $k$.

El gasto público productivo es una externalidad para la empresa, porque eleva la productividad de sus factores sin que la empresa lo elija ni pague por él. Por ello, la empresa no lo internaliza en su decisión de capital, y el equilibrio competitivo no es eficiente por sí solo.

**Sector público**: mantiene el presupuesto equilibrado en cada instante:

$$g = \tau_w w + \tau_r r b + \tau_c c$$

donde $g$ es el gasto público productivo per cápita y cada sumando la recaudación de cada figura impositiva. El sector público dedica a dicho gasto una fracción $g/y$ del *output*, que, dados los tipos impositivos, queda determinada por la recaudación.

Al no haber deuda ni impuestos de suma fija, cada unidad de gasto productivo tiene que financiarse con alguna de las tres figuras impositivas, y estas no son equivalentes:

- *Financiación con* $\tau_c$ *o* $\tau_w$: el hogar pierde renta, pero su decisión de ahorro no cambia, y el efecto sobre $k^*$ es el positivo de $A(g)$.

- *Financiación con* $\tau_r$: el mismo $g$ se obtiene a costa de la cuña sobre el ahorro, y el efecto neto sobre $k^*$ es ambiguo.

La composición de la recaudación importa tanto como su nivel. Además, como $g$ aparece a ambos lados (en el gasto y, vía $A(g)$, en las bases imponibles $w$, $r$ y $c$), el gasto productivo se autofinancia parcialmente, ya que, al elevar la productividad, eleva la recaudación para unos tipos dados.

**Equilibrio**: la condición de equilibrio en los mercados financieros es $b = k$, es decir, la riqueza de los hogares se materializa en capital. Sustituyendo el tipo de interés de las empresas en la ecuación de Euler, la dinámica del consumo es $g_c = \frac{1}{\theta}\left[(1-\tau_r)\left(A(g)\,df/dk - \delta\right) - \rho\right]$ y, en estado estacionario, $g_c = 0$:

$$(1-\tau_r)\left[A(g)\left.\frac{df}{dk}\right|_{k^*} - \delta\right] = \rho$$

El capital de estado estacionario es el que iguala el rendimiento neto de impuestos del capital a la impaciencia. Como en el modelo sin sector público, el equilibrio es un punto de silla. Dado $k_0$, existe una única senda de consumo que converge a $(k^*, c^*)$.

**Desarrollo gráfico**: la dinámica del modelo se representa en el plano $k$-$c$ (figura A38.6):

1. *Locus* $\dot{k} = 0$: recoge los pares de capital y consumo con los que el capital per cápita no varía. Sustituyendo $b = k$ y la restricción del sector público en la de los hogares, $c + g + \dot{k} + (n+\delta)k = A(g)f(k)$, de modo que el *locus*, $c = A(g)f(k) - (n+\delta)k - g$, tiene forma de U invertida, con su máximo en la regla de oro.

2. *Locus* $\dot{c} = 0$: es vertical en el capital $k^*$ de la condición de estado estacionario, porque la ecuación de Euler solo depende de $k$. A su izquierda el consumo crece y a su derecha decrece.

3. *Senda estable*: es la única trayectoria que converge al estado estacionario.

4. *Equilibrio*: en el punto 0, donde se cortan $\dot{k} = 0$ y $(\dot{c} = 0)_0$, la economía permanece en su estado estacionario, con capital $k_0^*$ y consumo $c_0^*$.

**Estática comparativa**: de la condición de estado estacionario se desprenden cuatro conclusiones sobre los efectos de la política fiscal:

1. *Impuestos sobre el consumo y el trabajo* ($\tau_c$, $\tau_w$): con tipos constantes, no alteran $k^*$ ni $c^*$.

2. *Impuesto sobre el capital* ($\tau_r$): reduce $k^*$ y $c^*$, porque eleva la productividad marginal del capital que los hogares exigen para ahorrar.

Gráficamente (figura A38.6), se supone una subida de $\tau_r$ no anticipada y permanente que no altera el gasto $g$, porque la mayor recaudación se compensa con una rebaja de $\tau_c$ o $\tau_w$. En ese caso, el *locus* $\dot{k} = 0$ no varía y el *locus* $\dot{c} = 0$ se desplaza a la izquierda, de $(\dot{c} = 0)_0$ a $(\dot{c} = 0)_1$.

En el impacto, con el capital dado, el consumo salta de $c_0^*$ a $c_{0'}$ (del punto 0 al $0'$) hasta la nueva senda estable, porque el ahorro es menos rentable. Después, el capital y el consumo decrecen a lo largo de ella hasta el punto 1, con $k_1^* < k_0^*$ y $c_1^* < c_0^*$.

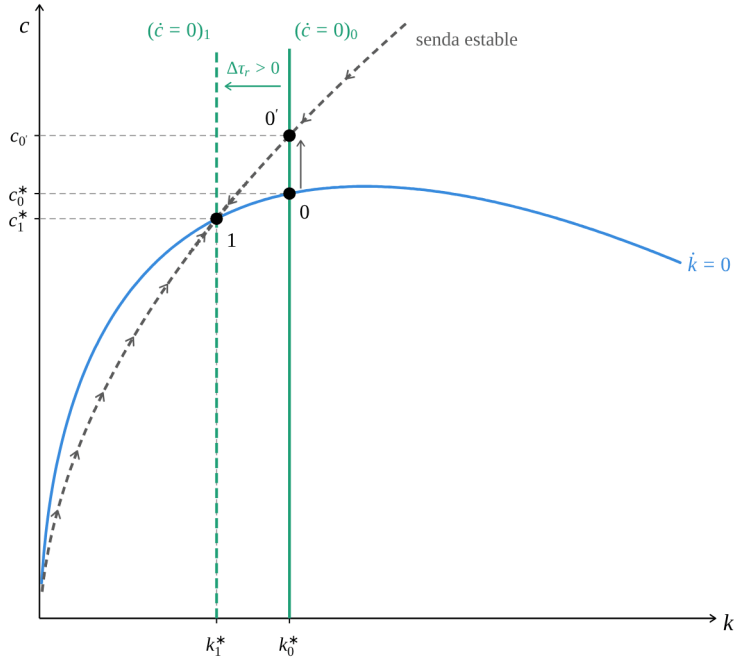



Figura A38.6: Subida del impuesto sobre el capital en el modelo de Ramsey-Cass-Koopmans. El consumo salta a la nueva senda estable ($0 \to 0'$) y la economía converge a un estado estacionario con menos capital y consumo (1).

3. *Gasto público productivo* ($g$): eleva $A(g)$ y, con ella, la productividad marginal del capital, $k^*$ y el nivel de renta.

4. *Disyuntiva*: se plantea entre el crecimiento impulsado por el sector privado (no distorsionado, pero con externalidades y bienes públicos infraprovistos) y el impulsado por el sector público (distorsionante, pero que internaliza externalidades y provee bienes públicos). El tamaño óptimo del sector público es aquel en que el efecto productividad marginal de $g$ iguala al coste marginal de la distorsión necesaria para financiarlo.

En un modelo de crecimiento exógeno, la política fiscal tiene efectos *nivel* (sobre $k^*$ y $c^*$) pero no efectos *tasa*, porque la tasa de crecimiento a largo plazo sigue siendo la del progreso técnico exógeno.

**Extensiones**: tres relajaciones de los supuestos matizan los resultados:

1. *Oferta de trabajo elástica*: si los hogares eligen entre consumo y ocio, $\tau_w$ y $\tau_c$ pasan a ser distorsionantes, pues reducen el salario real neto y, por el efecto sustitución, la oferta de trabajo. La jerarquía de daño entre figuras se convierte entonces en una cuestión empírica.

2. *Imposición óptima del capital* (Chamley, 1986; Judd, 1985): en un modelo de este tipo, el tipo impositivo óptimo sobre el capital a largo plazo es nulo, pues gravarlo equivale a gravar el consumo futuro a un tipo creciente con el horizonte. Sin embargo, el resultado se debilita con

horizontes finitos, restricciones de liquidez o rentas del capital que son en realidad rentas puras. Además, Straub y Werning (2020) muestran que, con una elasticidad de sustitución intertemporal inferior a la unidad ($\theta > 1$), el tipo óptimo puede ser positivo a largo plazo incluso en los modelos originales.

3. *Capital público como stock* (Arrow y Kurz, 1970): si el gasto público se acumula en un *stock* de capital público que se deprecia, el análisis se convierte en uno de dos capitales con una regla de oro para cada uno. En ese caso, la inversión pública solo afecta al crecimiento durante la transición.

#### 3.2.3. Conclusiones

**Aspectos positivos**: el modelo ordena los impuestos según su coste en términos de acumulación, y el impuesto sobre el capital resulta el más dañino, en línea con el resultado de Chamley-Judd de tipo óptimo nulo a largo plazo. Muestra, además, que el gasto productivo tiene efectos nivel sobre la renta.

**Limitaciones**: al tratarse de un modelo de crecimiento exógeno, la política fiscal influye solo en el nivel y no en la tasa de crecimiento a largo plazo. Además, la oferta de trabajo inelástica elimina la distorsión del impuesto sobre el trabajo.

### 3.3. Crecimiento endógeno: el modelo de Barro (1990)

**Origen**: el modelo es el de crecimiento endógeno AK con sector público de Barro (1990), presentado aquí en su versión con tasa de ahorro exógena, como extensión del modelo de Solow-Swan.

**Idea**: el gasto público productivo eleva la productividad del capital privado y su financiación detrae renta del sector privado, de modo que existe un tamaño del sector público que maximiza el crecimiento.

#### 3.3.1. Supuestos

**Entorno**: el modelo supone un mundo neoclásico, sin más fallo de mercado que la externalidad del gasto público productivo, en el largo plazo. El sector público mantiene en todo momento un presupuesto equilibrado.

**Agentes**: intervienen consumidores, empresas y sector público, los tres racionales. Lo distintivo del modelo es que el sector público realiza gasto productivo. Cada individuo ofrece una unidad de trabajo y la población crece a la tasa $n$, de modo que las variables se expresan en términos per cápita. La tasa de ahorro es constante y exógena, y el ahorro se invierte íntegramente.

**Tecnología**: la función de producción presenta las propiedades habituales (rendimientos constantes a escala en los *inputs* rivales, productividad marginal positiva y decreciente, condiciones de Inada), y el gasto público productivo entra como un *input*:

$$y = Ak^{\alpha}g^{1-\alpha}, \qquad A > 0,\ \alpha \in (0,1)$$

donde $y$, $k$ y $g$ son el *output*, el capital privado y el gasto público per cápita, $A$ el nivel tecnológico y $\alpha \in (0,1)$ la participación del capital privado[8]. La función presenta rendimientos constantes

[8] Barro (1990) escribe $y = Ak^{1-\alpha}g^{\alpha}$, con $\alpha$ la elasticidad del *output* respecto al gasto público, y obtiene por ello $\tau^* = \alpha$. Aquí y en el tema A44, $\alpha$ es la participación del capital, como en el resto del libro.

en el conjunto de $k$ y $g$. Por ello, si $g$ crece en la misma proporción que $k$, no hay rendimientos decrecientes del capital y se trata de un modelo AK.

### 3.3.2. Modelo

**Planteamiento**: la economía es cerrada y tiene sector público, de modo que la renta se reparte entre consumo, inversión y gasto público, $Y = C + I + G$. El gasto productivo se financia, con presupuesto equilibrado, mediante un impuesto proporcional sobre la renta de tipo $\tau$, que reduce la renta disponible y el ahorro. Con ahorro exógeno, un impuesto de suma fija de igual recaudación tendría el mismo efecto, porque la distorsión del rendimiento privado del capital solo aparece con hogares optimizadores. Como en el modelo de Solow-Swan (tema A43, apartado 2.2 y Anexo), cada ecuación de comportamiento (gasto público, renta, inversión y consumo, por este orden) se escribe en niveles y per cápita, y las de la renta y el consumo, también en tasas de crecimiento:

$$\begin{aligned}
&G = \tau Y && \Rightarrow\ g = \tau y \\
&Y = AK^{\alpha}L^{1-\alpha}(G/L)^{1-\alpha} && \Rightarrow\ y = A^{1/\alpha}\tau^{\frac{1-\alpha}{\alpha}}k\ \Rightarrow\ g_y = g_k, \quad A>0,\ \alpha\in(0,1) \\
&\dot{K} = s(1-\tau)Y - \delta K && \Rightarrow\ \dot{k} = s(1-\tau)y - (n+\delta)k \\
&C = (1-s)(1-\tau)Y && \Rightarrow\ c = (1-s)(1-\tau)y\ \Rightarrow\ g_c = g_y
\end{aligned}$$

donde $s$ es la tasa de ahorro, $(1-\tau)Y$ la renta disponible, $n$ la tasa de crecimiento de la población y $\delta$ la de depreciación. La renta per cápita se obtiene sustituyendo $g = \tau y$ en $y = Ak^{\alpha}g^{1-\alpha}$ y despejando $y$. Resulta proporcional al capital porque el gasto público crece con la renta y evita que el capital privado tenga rendimientos decrecientes. La tasa de crecimiento del capital, que comparten la renta y el consumo, se obtiene en el equilibrio.

**Equilibrio**: como el *output* es proporcional al capital, no hay dinámica de transición, y la renta, el capital y el consumo per cápita crecen desde el principio a la misma tasa constante, que depende del tipo impositivo:

$$g_k(\tau) = s(1-\tau)A^{1/\alpha}\tau^{\frac{1-\alpha}{\alpha}} - (n+\delta)$$

El tipo impositivo aparece dos veces:

- *Efecto distorsión*: como $(1-\tau)$, reduce la renta disponible y el ahorro.
- *Efecto productividad*: como $\tau^{(1-\alpha)/\alpha}$, eleva la productividad del capital vía gasto público.

El tamaño óptimo del sector público, el que maximiza el crecimiento, se alcanza cuando la participación del ingreso público en el PIB iguala a la participación del gasto productivo determinada por la tecnología:

$$\frac{\partial g_k}{\partial \tau} = 0\ \rightarrow\ \tau^* = 1-\alpha$$

**Desarrollo gráfico**: en el plano tipo impositivo-tasa de crecimiento (figura A38.7), la curva $g_k(\tau)$ tiene forma de U invertida:

1. $\tau = 0$: el Estado no recauda y no puede gastar, y sin ese *input* imprescindible nada se produce. El capital por trabajador varía entonces a la tasa $-(n+\delta) < 0$.

2. $\tau = 1$: el impuesto absorbe toda la renta, los hogares no tienen nada que ahorrar ni invertir y el capital vuelve a variar a la tasa $-(n+\delta)$.

3. *Tipo óptimo*: se corresponde con el máximo de la curva, $\tau^* = 1-\alpha$, con la tasa de crecimiento $g_k^*$. A la izquierda domina el efecto productividad del gasto, a la derecha el efecto distorsión del impuesto.

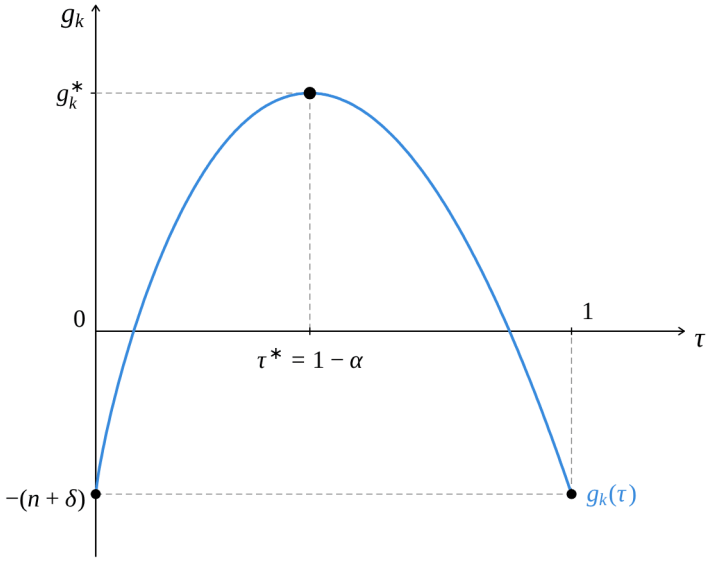



Figura A38.7: Modelo de Barro (1990). La tasa de crecimiento es máxima cuando el tipo impositivo iguala a la elasticidad del *output* respecto al gasto público productivo.

**Estática comparativa**: a diferencia del modelo de crecimiento exógeno, la política fiscal tiene aquí efectos tasa, porque el gasto productivo eleva permanentemente la tasa de crecimiento y su financiación distorsionante la reduce.

**Extensiones**: en la versión original con hogares optimizadores (Barro, 1990), la figura impositiva sí importa, dado que los hogares eligen su ahorro a la Ramsey (tema A44, apartado 1.4). Si el gasto productivo se financia con impuestos de suma fija, $T_0 = g$, estos no alteran el rendimiento privado del capital:

$$\frac{\partial y}{\partial k} = \alpha A^{1/\alpha} \left(\frac{g}{y}\right)^{\frac{1-\alpha}{\alpha}}$$

que es creciente con la ratio de gasto público sobre el PIB. La tasa de crecimiento del consumo, $g_c = \frac{1}{\theta}\left[\partial y/\partial k - \delta - \rho\right]$, es por ello creciente con $g/y$, si bien el tamaño que maximiza el bienestar sigue siendo $g/y = 1-\alpha$, donde la productividad marginal del gasto público iguala a su coste ($\partial y/\partial g = 1$). Con el impuesto proporcional, en cambio, el rendimiento privado del capital se reduce a $(1-\tau)\,\partial y/\partial k$ y el tipo que maximiza el crecimiento es el mismo que con ahorro exógeno, $\tau^* = 1-\alpha$.

Baxter y King (1993) endogenizan, además, la oferta de trabajo, con lo que los *shocks* fiscales alteran también el empleo. Los aumentos transitorios de la inversión pública tienen un efecto

expulsión sobre la inversión privada mayor que los permanentes y un efecto expansivo muy reducido.

En cambio, un aumento permanente puede llevar a un nuevo estado estacionario con mayor producción, inversión, consumo privado y empleo. Ocurre así porque, a medida que se acumula capital público, aumenta la productividad marginal del empleo y del capital privado, lo que estimula la inversión privada y la oferta de trabajo y compensa el efecto riqueza negativo de la financiación[9].

### 3.3.3. Conclusiones

**Aspectos positivos**: la aportación de Barro fue una de las primeras en estudiar el tamaño del sector público en un modelo de crecimiento endógeno, y permite ilustrar de una manera intuitiva el debate sobre ese tamaño[10].

**Limitaciones**: la principal es que el enfoque de Barro considera fija la oferta de trabajo. Además, la tasa de ahorro es exógena en la versión aquí presentada (no en la original de Barro) y el gasto público es un flujo y no un *stock*. El modelo tampoco distingue entre partidas de gasto productivo ni entre figuras impositivas distorsionantes, aunque estas dañan el crecimiento en grado muy distinto (Arnold et al., 2011).

## 3.4. Evidencia empírica

La validez de los modelos anteriores exige contrastar sus principales conclusiones sobre la inversión pública, el consumo público y los impuestos.

**Gasto productivo: efectos de la inversión pública** (Aschauer, 1989): estima una función de producción agregada para EE. UU. (1949-1985) con el *stock* de capital público como *input* y encuentra un fuerte efecto positivo de la inversión pública en infraestructuras básicas sobre la productividad y el crecimiento, con una elasticidad del *output* respecto al capital público en torno a 0,4.

**Gasto no productivo: peso del consumo público** (Barro, 1991): en su regresión de crecimiento para 98 países (1960-1985) encuentra una relación negativa entre la ratio de consumo público sobre el PIB y la tasa de crecimiento, y no significativa para la inversión pública. El resultado concuerda con la distinción entre gasto productivo e improductivo del modelo de Barro (1990).

**Efectos de la imposición** (Kneller, Bleaney y Gemmell, 1999): con un panel de 22 países de la OCDE (1970-1995) encuentran que los impuestos distorsionantes reducen el crecimiento y los no distorsionantes no, y que el gasto productivo lo eleva y el improductivo no.

**Magnitud del efecto del gasto productivo**: la literatura posterior a Aschauer encuentra un efecto positivo, pero mucho menor. En particular, el metaanálisis de Bom y Ligthart (2014), con 578 estimaciones de 68 estudios, sitúa la elasticidad media de la producción respecto al capital público en torno a 0,1, frente al 0,4 de Aschauer.

**Composición de los impuestos** (Arnold et al., 2011): con datos de 21 países de la OCDE (1971-2004) establecen una jerarquía de figuras impositivas según su daño al crecimiento. Los impuestos

[9] Por ello, el multiplicador a largo plazo de un aumento permanente de la inversión pública puede superar la unidad.

[10] El sector público, que apenas alcanzaba un 13 % del PIB en las economías desarrolladas antes de la I Guerra Mundial (Tanzi y Schuknecht, 2000), creció hasta mediados de los años ochenta y se estabilizó después en torno al 35-50 %.

recurrentes sobre la propiedad inmobiliaria son los menos dañinos, seguidos de los impuestos sobre el consumo, los impuestos sobre la renta personal y, los más dañinos, los impuestos sobre sociedades. En España corresponden, de menor a mayor daño, al impuesto sobre bienes inmuebles (IBI), al IVA, al impuesto sobre la renta de las personas físicas (IRPF) y al impuesto sobre sociedades (IS).

**Inversión pública en la OCDE**: la evidencia más reciente (Abiad, Furceri y Topalova, 2016) indica que un aumento de la inversión pública eleva el *output* a corto y largo plazo, con efecto atracción sobre la inversión privada. Los efectos son mayores en periodos de holgura económica y cuando la inversión es eficiente y se financia con deuda.

## 3.5. Conclusiones

El gasto público productivo eleva el nivel de renta en los modelos de crecimiento exógeno y la tasa de crecimiento en los de crecimiento endógeno, pero su financiación con impuestos distorsionantes impone un límite, que se traduce en un tamaño óptimo del sector público. La evidencia empírica respalda ambas caras de la disyuntiva, con efectos positivos de la inversión pública menores que los que estimó Aschauer (1989). Kneller, Bleaney y Gemmell (1999) confirman la distinción entre gasto productivo e improductivo, y Arnold et al. (2011) añaden que importa la composición de los impuestos distorsionantes, una distinción que el modelo de Barro no hace.

## Conclusiones

La política fiscal incide en el ahorro y el crecimiento por tres canales. El gasto no productivo y su gravamen presionan a la baja el consumo privado por el efecto riqueza y, si son transitorios, el ahorro nacional, y su efecto sobre el *output* a corto plazo depende de las rigideces, las restricciones de liquidez y la política monetaria. En pensiones, la capitalización deja inalterado el ahorro nacional y el reparto lo reduce, tanto más cuanto mayor es el diferencial entre su rendimiento implícito y el tipo de interés. El envejecimiento ha reducido ese rendimiento implícito. De ahí la convergencia de las reformas hacia sistemas mixtos que ilustran los casos de Alemania, Chile y Suecia. El gasto productivo eleva el nivel de renta o la tasa de crecimiento hasta un tamaño óptimo del sector público, y su composición, como la de los impuestos, importa tanto como su nivel. Todo ello supone que existe espacio fiscal, un margen para gastar más sin que la deuda entre en una senda explosiva. Sin él, la subida de los tipos puede expulsar la inversión privada y anular el impulso.

El análisis del gasto no productivo prolonga dos debates de las escuelas macroeconómicas, la equivalencia ricardiana, según la cual la deuda no altera el consumo (tema A06), y el multiplicador con rigideces y hogares sin acceso al crédito (tema A07). Las pensiones se analizan con el modelo de generaciones solapadas, en el que el reparto transfiere renta entre cohortes (tema A29), y el gasto productivo, con los modelos de Ramsey-Cass-Koopmans y de Barro, en los que el gasto público productivo entra en la función de producción (temas A43 y A44). Por último, la evidencia sobre el multiplicador fiscal (tema A40) y la sostenibilidad de la deuda, que delimita el espacio fiscal (tema A39), completan el análisis.

Straub y Werning (2020) han reabierto la discusión sobre la imposición óptima del capital al cuestionar el tipo nulo de Chamley (1986) y Judd (1985) que sustenta la jerarquía de impuestos (apartado 3.2.2). Sigue abierta también la eficacia de la inversión pública. Abiad, Furceri y Topalova (2016) la respaldan (apartado 3.4), pero Ramey (2020) advierte de que los retrasos de construcción debilitan su efecto a corto plazo y de que las estimaciones de la elasticidad del producto al capital público probablemente están sesgadas al alza.

# A39. Déficit público, su financiación y sostenibilidad

Déficit público. Conceptos. Financiación y sus consecuencias macroeconómicas. Dominación monetaria y dominación fiscal. La dinámica de la deuda pública y su sostenibilidad.

## Introducción

**Introducción común.** Los temas A38 y A39 comparten una introducción, que se desarrolla al comienzo del tema **A38**. Presenta las tres funciones de la política fiscal según Musgrave (asignativa, redistributiva y estabilizadora) y la evolución del debate sobre su eficacia, hasta la preocupación actual por la sostenibilidad de la deuda y las reglas fiscales.

Siguiendo a De Viti de Marco (1936), una vez el Estado ha recurrido a la deuda, es probable que lo vuelva a hacer.

El **déficit público** se puede definir como un desequilibrio en el cual los gastos del Estado superan a sus ingresos netos de transferencias. El concepto está unido al de política fiscal, que está en su origen.

Este desequilibrio debe financiarse[1][2], habitualmente mediante la emisión de deuda. Se distingue entre[3]:

1. *Financiación ortodoxa*: cuando el sector público recurre al sector privado (familias, empresas o resto del mundo). Incluye la emisión de deuda pública, el incremento de la presión fiscal (que, en rigor, reduce el déficit en lugar de financiarlo) y las privatizaciones.

2. *Financiación heterodoxa*: cuando el adquiriente es el banco central, el cual compra los títulos mediante la creación de base monetaria, en cuyo caso se habla de *monetización del déficit*.

El estudio del déficit público es relevante dado que el crecimiento del sector público en países desarrollados durante el s. XX ha sido muy intenso. Así, el peso del gasto público se ha multiplicado por más de tres, desde en torno al 13 % del PIB en 1913 hasta cerca del 45 % a finales del siglo (Tanzi y Schuknecht, 2000). En torno al 60 % de ese gasto se destina a gasto social.

La crisis financiera global y la pandemia llevaron la deuda pública de muchas economías avanzadas a niveles muy elevados, y la subida de los tipos de interés posterior ha devuelto al primer

[1] No confundir con la financiación del gasto. El gasto se financia mediante impuestos, mientras que el déficit (diferencia entre el gasto y los impuestos) se financia mediante la emisión de deuda o de dinero.

[2] Alternativamente, el Gobierno puede optar por no pagar la deuda (*default*), con costes para los tenedores de los títulos, para la economía en su conjunto y para la credibilidad del sector público.

[3] La diferencia entre ambos tipos, hoy en día, es difusa debido a la aplicación de políticas económicas no convencionales tanto en materia fiscal como monetaria.

plano su sostenibilidad. Esta depende, sobre todo, de la diferencia entre el tipo de interés de la deuda y el crecimiento de la economía, que fija el **superávit primario** necesario para estabilizar la ratio de deuda. En la UE, la reforma del Pacto de Estabilidad y Crecimiento (2024) traduce ese análisis en sendas de gasto neto con un ancla de sostenibilidad de la deuda, y las compras de deuda de los bancos centrales han reavivado el debate sobre la **dominancia fiscal** y la monetaria.

El tema estudia el déficit público, su financiación y sus consecuencias macroeconómicas, y se ordena en tres partes:

- *Fundamentos*: se presentan los conceptos esenciales y el debate sobre la efectividad de la política fiscal y su financiación.

- *Financiación y sostenibilidad*: se aborda la financiación vía deuda y vía monetización y la sostenibilidad del déficit en su formulación analítica, con sus extensiones a la economía abierta.

- *Dominancia monetaria y fiscal* ("dominación" en el título oficial del tema): se estudia la interrelación entre las autoridades fiscal y monetaria a través del modelo de Leeper y de las teorías de la dominancia monetaria y fiscal.

# 1. Fundamentos del análisis del déficit público y de la deuda

Antes de abordar la financiación del déficit y la dominancia, se fijan los conceptos esenciales del análisis del déficit y de la deuda y se repasa el debate sobre la efectividad de la política fiscal y su financiación.

> Esta sección fija los conceptos y el debate previo que el resto del tema da por sabidos. El desarrollo del tema está en las secciones 2 y 3, que son las que responden al título.

## 1.1. Conceptos esenciales

Los conceptos esenciales son los saldos presupuestarios, el señoraje y el impuesto inflacionario, y la dominancia monetaria y fiscal, con su formalización como políticas activas y pasivas.

**Saldo presupuestario**: analíticamente, el saldo presupuestario público observado es:

$$\mathrm{SP}_t = T_t - (G1_t + iB_{t-1}) + \varepsilon(y_t - \bar{y}_t) + \varphi_t$$

donde $T_t$ y $G1_t$ son la recaudación impositiva y el gasto público primario netos del componente cíclico y de las medidas *one-off*, $iB_{t-1}$ es el coste del servicio de la deuda y $\varepsilon(y_t - \bar{y}_t)$ es el saldo primario cíclico. Por último, $\varphi_t$ es el efecto de las medidas *one-off*, es decir, de carácter puntual[4] o contingente[5].

El parámetro $\varepsilon > 0$ recoge la semielasticidad del saldo presupuestario a la brecha de producción (*output gap*), es decir, cuántos puntos de PIB mejora el saldo por cada punto de brecha positiva a través de los estabilizadores automáticos (mayor recaudación, menor gasto en prestaciones). La Comisión Europea la estima en torno a 0,5-0,6 para España y la media de la UE.

A partir de esta expresión se pueden derivar los conceptos fundamentales:

1. *Saldo primario*: $\mathrm{SP}_t + iB_{t-1} = T_t - G1_t + \varepsilon(y_t - \bar{y}_t) + \varphi_t$, es el saldo observado corregido del coste del servicio de la deuda. Por tanto, excluye los intereses de la deuda y se aproxima más a las variables bajo control del Gobierno, aunque aún incluye el componente cíclico.

2. *Saldo primario cíclicamente ajustado*: $\mathrm{SP}_t + iB_{t-1} - \varepsilon(y_t - \bar{y}_t) = T_t - G1_t + \varphi_t$, es el saldo primario corregido del componente cíclico.

3. *Saldo primario estructural*: $\mathrm{SP}_t + iB_{t-1} - \varepsilon(y_t - \bar{y}_t) - \varphi_t = T_t - G1_t$, que corrige además las medidas *one-off* y es el relevante para evaluar la orientación de la política fiscal.

**Señoraje e impuesto inflacionario**: el ingreso real que el Estado obtiene al emitir dinero y el gravamen implícito que la inflación impone a los tenedores de saldos monetarios. No son idénticos y se desarrollan en el apartado 2.2.1.

**Dominancia monetaria y fiscal** (Sargent y Wallace, 1981):

[4] Son medidas puntuales las ayudas a la reestructuración del sector financiero, los ingresos extraordinarios de amnistías fiscales o las ventas de activos públicos no financieros.

[5] Ejemplos de pasivos contingentes son garantías públicas y pasivos de empresas públicas productoras de mercado (fuera del sector de las administraciones públicas, AAPP). Lo son también las obligaciones pendientes de pago relativas a proyectos de asociación público-privada (*public-private partnerships*, PPP) en que los activos no están registrados en los balances de las AAPP, y los préstamos de dudoso cobro.

- *Dominancia monetaria*: el banco central fija la senda de la cantidad de dinero (o de los tipos) y el Tesoro ajusta déficits y superávits para que la deuda sea sostenible dada esa senda, lo que constituye la conducta "habitual" de la política monetaria.

- *Dominancia fiscal*: el Tesoro fija la senda de déficits y superávits y el banco central tiene que generar el señoraje necesario para que la deuda no siga una dinámica explosiva, con lo que pierde el control del nivel de precios.

**Políticas activas y pasivas** (Leeper, 1991): formalizan la distinción anterior. Una autoridad sigue una *política activa* cuando fija su instrumento con independencia de la restricción presupuestaria intertemporal del sector público y una *política pasiva* cuando lo ajusta para garantizar su cumplimiento. Por tanto, la dominancia monetaria equivale a una política monetaria activa con una fiscal pasiva, y la dominancia fiscal, a la combinación inversa.

## 1.2. Debates sobre la efectividad de la política fiscal

La efectividad de la política fiscal y los efectos de su financiación han sido objeto de un intenso debate, primero en el paradigma IS-LM y, posteriormente, en los modelos con microfundamentos de la nueva macroeconomía clásica (NMC) y de la nueva economía keynesiana (NEK).

> Los efectos de la política fiscal en el IS-LM, el modelo del ciclo real con sector público y el modelo IS-PC-MR con gasto público se desarrollan en el tema A38, apartados 1.1 a 1.3. Aquí se retiene, de cada perspectiva, lo que interesa a la financiación del déficit.

### 1.2.1. Antecedentes: el modelo IS-LM

En el marco IS-LM de economía cerrada y con precios fijos, una política fiscal expansiva desplaza la IS hacia la derecha y genera un efecto multiplicador y un efecto expulsión (*crowding out*). La importancia relativa de ambos efectos depende de cuán sensibles sean al tipo de interés la inversión ($v$) y la demanda de dinero ($h$).

La forma en que se financie el déficit añade una serie de efectos adicionales, que difieren según la escuela:

1. *Financiación vía impuestos*: tiene un efecto contractivo que no anula el expansivo, pues el gasto actúa directamente sobre la demanda y los impuestos solo a través de la propensión marginal a consumir (*teorema del multiplicador del presupuesto equilibrado*, Haavelmo, 1945)[6]. Por tanto, la IS se desplaza hacia la derecha, pero menos que sin financiación impositiva. Para los monetaristas, esta financiación hace aún más ineficaz la política fiscal.

2. *Financiación vía emisión de deuda*: el keynesianismo contempla el *efecto reajuste de cartera*, según el cual el aumento de la oferta de bonos reduce el precio de los ya emitidos[7], eleva la rentabilidad exigida y desplaza la LM hacia la izquierda (figura A39.1). El efecto contractivo es tanto mayor cuanto menor sea la sustituibilidad entre bonos y dinero. Por su parte, el monetarismo añade dos efectos riqueza ligados a la hipótesis de la renta permanente:

- *Efecto riqueza directo*: más tenencias de bonos elevan la riqueza y el consumo, de modo que la IS se desplaza a la derecha.

[6] Analíticamente, $\partial Y/\partial G_0 + \partial Y/\partial T_0 = (1-c)/[1-c(1-\tau)+vk/h] > 0$, porque la caída de renta disponible se reparte entre menor consumo y menor ahorro según la propensión marginal a consumir.

[7] Siempre y cuando los nuevos bonos y los bonos antiguos tengan las mismas características.

▪ *Efecto riqueza indirecto*: la subida del tipo de interés reduce el valor de mercado de los bonos ya emitidos y, con él, la riqueza y el consumo, lo que hace la IS más elástica al tipo de interés y refuerza la expulsión. Con los parámetros monetaristas la expulsión es completa, y la política fiscal puede llegar incluso a contraer la actividad.

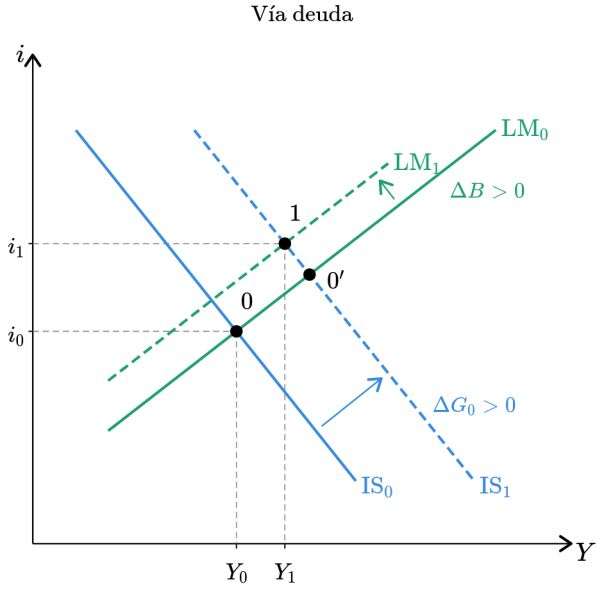



Figura A39.1: Financiación del déficit vía deuda en el IS-LM. El aumento del gasto ($\Delta G_0 > 0$) desplaza la IS de $IS_0$ a $IS_1$ y lleva la economía del punto 0 al $0'$. La mayor oferta de bonos ($\Delta B > 0$) desplaza la LM a la izquierda, de $LM_0$ a $LM_1$ (efecto reajuste de cartera), y la economía pasa al punto 1, con un tipo de interés $i_1$ más alto y más expulsión.

3. *Financiación heterodoxa vía monetización*: añade el *efecto liquidez* o *efecto Keynes*, por el cual el aumento de la oferta monetaria desplaza la LM hacia la derecha[8] y presiona a la baja el tipo de interés. Como hay más dinero para atender en parte la mayor demanda de saldos, el tipo de interés sube menos y el efecto expulsión es menor. En consecuencia, a corto plazo la renta aumenta más que con cualquier otra forma de financiación (los efectos a largo plazo se estudian en el apartado 2.2.3).

Para los primeros keynesianos, la política monetaria es eficaz para financiar el gasto público pero, debido a los dos eslabones débiles (la trampa de liquidez y la escasa sensibilidad de la inversión al tipo de interés), no lo es para estimular la demanda agregada. Para los monetaristas, en cambio, lo es únicamente en el corto plazo, pues sus efectos se diluyen a largo plazo por la teoría cuantitativa, y el efecto liquidez va acompañado de los efectos riqueza y del efecto cartera.

### 1.2.2. Perspectiva NMC: la financiación en el modelo del ciclo real

La principal limitación del marco IS-LM es que ofrece un enfoque estático. Por eso, tras la crítica de Lucas, el debate se traslada a modelos dinámicos con expectativas racionales, en los cuales lo relevante no es el efecto impacto de una medida sino la respuesta de los agentes a la senda

[8] El efecto final sobre el tipo de interés, $i = (1/h)(kY - M^s/P)$, es ambiguo, aunque, si sube, lo hace menos que con financiación vía deuda.

esperada de impuestos, déficits y deuda. Del modelo del ciclo real con sector público se derivan tres resultados sobre la financiación:

1. *Equivalencia ricardiana* (Barro, 1974): bajo mercados de capitales completos y perfectos, agentes con expectativas racionales y altruismo intergeneracional perfecto, e impuestos de suma fija, los bonos no aparecen en la restricción presupuestaria intertemporal de los hogares. Por ello, no constituyen riqueza neta y, dada la senda de gasto público, la elección entre impuestos y deuda para financiarla es irrelevante[9]. Con apertura al exterior, una rebaja de impuestos a gasto dado abre un déficit que el mayor ahorro de los hogares financia por completo, y el saldo por cuenta corriente queda igual (no hay déficits gemelos).

Las limitaciones de la equivalencia hacen que la financiación tenga efectos. Por ejemplo, con restricciones de liquidez el sector público obtiene financiación más barata que el privado, de ahí que la deuda sí constituya riqueza neta. Asimismo, la ilusión fiscal y el altruismo imperfecto trasladan la carga a generaciones posteriores. Otras limitaciones son los impuestos distorsionantes y el gasto productivo.

2. *Efectos no keynesianos* (Giavazzi y Pagano, 1990): con agentes ricardianos, un ajuste fiscal que recorte gasto de forma creíble y duradera puede estimular la actividad. Al anticipar una menor carga tributaria futura, los agentes se sienten más ricos y perciben menos incertidumbre, y las primas de riesgo bajan, con un efecto atracción (*crowding in*) sobre la inversión privada.

3. *Suavización impositiva* (Barro, 1979): si los impuestos son distorsionantes y su coste de eficiencia es convexo en el tipo impositivo, la política fiscal óptima mantiene constante el tipo impositivo esperado y utiliza el déficit como amortiguador de los *shocks* transitorios de gasto (guerras, recesiones). El déficit es, así, el resultado de una política óptima y no un síntoma de indisciplina. El Gobierno se endeuda en las recesiones y amortiza en las expansiones, y el tipo impositivo sigue un paseo aleatorio (martingala), al igual que la ratio de deuda, pues cada *shock* no anticipado altera permanentemente su nivel.

> La equivalencia ricardiana como resultado de la nueva macroeconomía clásica se desarrolla en el tema A06, apartado 3.2, y como frontera entre los modelos de horizonte infinito y los de generaciones solapadas en el tema A29, apartado 3.3.

### 1.2.3. Perspectiva NEK: la financiación en el modelo IS-PC-MR

Las rigideces nominales y las fricciones financieras devuelven a la política fiscal una función estabilizadora, y hacen que el efecto de una misma medida, y de su financiación, dependa del entorno:

1. *Restricciones de liquidez*: una fracción de hogares no ricardianos (*hand-to-mouth*) consume su renta corriente, de modo que las transferencias y las rebajas impositivas tienen efectos reales y la equivalencia ricardiana falla en proporción a esa fracción.

2. *Objetivo de inflación*: con política monetaria activa, el banco central contrarresta el impulso fiscal elevando el tipo de interés, por lo que el multiplicador es inferior a la unidad.

[9] Con emisión de deuda, aumentará el ahorro privado, que compensará el desahorro público, ante el previsible aumento futuro de impuestos. Con una reducción de impuestos, los agentes ahorrarán íntegramente la renta disponible liberada. Con monetización, las expectativas de emisión continua en el futuro generan un aumento del ahorro privado en previsión del impuesto inflacionario asociado.

3. *Tipo de cambio fijo*: en este régimen o en una unión monetaria, la política monetaria no reacciona al impulso fiscal de un país y el multiplicador es mayor.

4. *Límite inferior de los tipos de interés* (*zero lower bound*, ZLB): el tipo nominal no sube ante el impulso fiscal y la mayor inflación esperada reduce el tipo real, por lo que el multiplicador puede superar la unidad.

> La evidencia empírica sobre el valor de los multiplicadores fiscales, su dependencia del estado de la economía y la experiencia desde la Gran Recesión se desarrollan en el tema A40, apartado 3.

## 1.3. Conclusiones

El saldo observado se depura de los intereses, del componente cíclico y de las medidas *one-off* hasta llegar al saldo primario estructural, el relevante para juzgar la orientación de la política fiscal. La dominancia monetaria y la fiscal se formalizan, por su parte, como la combinación de una política activa con otra pasiva.

Del debate sobre la efectividad de la política fiscal se extraen dos lecciones. La primera es que la financiación del déficit no es neutral salvo bajo equivalencia ricardiana estricta. Así, en el IS-LM la monetización es la vía más expansiva a corto plazo, y con impuestos distorsionantes la suavización impositiva hace del déficit el resultado de una política óptima. La segunda es que el efecto de una misma medida depende del régimen monetario en el que se aplica, desde un multiplicador inferior a la unidad bajo objetivo de inflación hasta uno que puede superarla en el límite inferior de los tipos. Ambas lecciones conducen a las preguntas de las secciones 2 y 3, la de si puede sostenerse el déficit y la de qué relación guardan las dos autoridades.

# 2. Financiación del déficit y su sostenibilidad

Actualmente, el debate no gira en torno a cómo financiar el déficit público y qué financiación es más expansiva, sino en torno a la sostenibilidad del déficit y de la deuda pública y sus efectos sobre el crecimiento[10].

1. *Déficit público sostenible*: aquel que resulta de una política, fiscal y monetaria, capaz de mantenerse sin límite de tiempo, lo que equivale a estabilizar el peso de la deuda pública en el PIB a precios corrientes (PIB nominal). Por esa razón, la ratio deuda-PIB pesa mucho en la valoración del riesgo-país.

2. *Déficit público insostenible*: en cambio, si el tipo de interés real supera al crecimiento ($r > g_Y$), un déficit primario estructural y permanente desencadena un *efecto bola de nieve*, con una ratio deuda pública-PIB nominal que no deja de crecer.

## 2.1. Financiación vía deuda: la dinámica de la deuda pública y su sostenibilidad

**Origen**: la lógica de sostenibilidad de la restricción presupuestaria intertemporal del Gobierno (RPIG) se remonta a Domar (1944), y Blanchard (1990) la sistematizó en los indicadores fiscales que utilizan hoy los organismos internacionales.

**Idea**: los efectos macroeconómicos de la financiación vía deuda pueden entenderse a partir de la RPIG. El déficit es sostenible si la ratio deuda-PIB se estabiliza, lo que depende del saldo primario, del señoraje y de la relación entre el tipo de interés real y el crecimiento.

### 2.1.1. Supuestos

**Entorno**: economía cerrada, en la que el déficit solo puede financiarse con emisión de deuda o de base monetaria. Los agentes privados absorben la deuda emitida al tipo de interés de mercado y se cumple la ecuación de Fisher de tipos de interés. Los tipos de interés, la inflación y el crecimiento se consideran constantes en el análisis gráfico.

**Agentes**: el Gobierno se comporta de acuerdo con su restricción presupuestaria periodo a periodo.

### 2.1.2. Modelo

**Planteamiento**: el modelo parte de la restricción presupuestaria del Gobierno en términos nominales, la expresa en proporción del PIB para obtener la ecuación dinámica de la deuda y deriva de ella la condición de sostenibilidad:

- *Restricción presupuestaria del Gobierno*: en términos nominales es:

$$\underbrace{G1_t + i_t B_{t-1}}_{\text{Gasto total}} - T_t = (B_t - B_{t-1}) + (H_t - H_{t-1})$$

[10] La sostenibilidad del déficit remite a la solvencia, no a la liquidez (los problemas de liquidez pueden resolverse reestructurando los pagos). La solvencia se refiere al sujeto (que el valor actual de sus superávits primarios futuros cubra su deuda) y la sostenibilidad a la política (que pueda mantenerse sin una senda explosiva de la deuda). Se puede ser solvente sin que la política sea sostenible, al menos durante un tiempo, pero no al revés.

donde $G1_t$ es el gasto público primario (gasto en bienes y servicios), $i_t B_{t-1}$ el pago de intereses sobre la deuda heredada, $T_t$ la recaudación fiscal, $B_t$ el *stock* de deuda pública y $H_t$ la base monetaria. En cada periodo, el déficit público debe ser financiado vía emisión de deuda o creación de base monetaria. En una economía abierta, las reservas internacionales podrían considerarse una fuente de financiación adicional[11].

Alternativamente, la expresión se puede reescribir como:

$$\underbrace{G1_t - T_t}_{\text{Déficit primario}} = (B_t - B_{t-1}) + (H_t - H_{t-1}) - i_t B_{t-1}$$

- *Ecuación dinámica de la deuda*: para trabajar en términos de PIB[12] se procede en tres pasos:

1. *División por el PIB nominal*: se divide la restricción presupuestaria entre $P_t Y_t$ y se expresan las variables en minúsculas como ratios sobre el PIB. La deuda heredada queda valorada sobre el PIB corriente, es decir, $B_{t-1}/P_t Y_t = b_{t-1}\,(P_{t-1}Y_{t-1}/P_t Y_t)$.

2. *Ecuación de Fisher*: se utiliza $1 + i_t = (1 + r_t)(1 + \pi_t)$ y la descomposición del crecimiento nominal, $P_t Y_t / P_{t-1} Y_{t-1} = (1 + \pi_t)(1 + g_{Y,t})$, con lo que el coste de la deuda heredada es $(1 + i_t)/[(1 + \pi_t)(1 + g_{Y,t})] = (1 + r_t)/(1 + g_{Y,t})$.

3. *Señoraje*: se agrupa la financiación monetaria en un único término, el señoraje en porcentaje del PIB, aplicando a $H_{t-1}$ la misma valoración que a $B_{t-1}$.

El resultado es la *ecuación dinámica de la deuda* (EDD):

$$b_t = d1_t + \frac{1 + r_t}{1 + g_{Y,t}} b_{t-1} - \underbrace{\left[h_t - \frac{h_{t-1}}{(1 + \pi_t)(1 + g_{Y,t})}\right]}_{\sigma_t}$$

donde $b_t = B_t/P_t Y_t$ es la ratio deuda-PIB, $d1_t = g1_t - \tau_t$ el déficit primario en porcentaje del PIB (con $\tau_t$ la recaudación impositiva en porcentaje del PIB), $r_t$ el tipo de interés real *ex post*[13] y $g_{Y,t}$ la tasa de crecimiento del PIB real. En cuanto a la financiación monetaria, $h_t = H_t/P_t Y_t$ es la base monetaria en porcentaje del PIB y $\sigma_t$ el señoraje en porcentaje del PIB.

Restando a ambos lados el *stock* de deuda en el periodo anterior se obtiene la variación de la ratio:

$$\underbrace{b_t - b_{t-1}}_{\Delta b_t} = \underbrace{g1_t - \tau_t}_{d1_t} + \frac{r_t - g_{Y,t}}{1 + g_{Y,t}} b_{t-1} - \sigma_t$$

Por tanto, la dinámica de la deuda depende de tres factores:

1. *Déficit primario*: dado por la diferencia entre el gasto primario (excluido el pago de intereses) y la recaudación impositiva.

[11] Las reservas se incorporan a la restricción distinguiendo entre deuda bruta y deuda neta, que descuenta de la bruta los activos del Gobierno (reservas internacionales y depósitos en el sistema financiero y en el banco central).

[12] Exigir la amortización de la deuda puede ser un criterio demasiado estricto, por lo que no se pide al Gobierno generar superávits futuros para devolverla, sino que el peso de la deuda sobre el PIB no crezca indefinidamente. Como el nivel absoluto de deuda no indica por sí solo el riesgo de insolvencia (importan los precios y el crecimiento real), se usa la ratio deuda-PIB.

[13] El tipo *ex post* utiliza la inflación observada y no la esperada. Solo las sorpresas inflacionistas ($\pi_t > \pi_t^e$) reducen la carga real de la deuda, pues una mayor inflación esperada se traslada a los tipos nominales exigidos.

2. *Relación entre tipo de interés real y crecimiento*: mide cuánto cuesta en realidad la deuda viva una vez descontado el crecimiento del PIB. Si $g_{Y,t} > r_t$, el segundo término es negativo y la deuda va perdiendo peso en porcentaje del PIB. En cambio, si $r_t > g_{Y,t}$, el segundo término es positivo, de modo que los pagos por intereses pesan cada vez más en el PIB y es necesaria más deuda para financiar el déficit público. El proceso se realimenta así en un círculo vicioso, conocido como *efecto bola de nieve* o *financiación estilo Ponzi*[14].

3. *Señoraje*: supone un alivio, al menos transitorio, del recurso a la financiación vía emisión de deuda pública.

**Equilibrio**: cuando el Estado carece de la facultad de emitir dinero ($\sigma_t = 0$), p. ej. por haber cedido la política monetaria a una autoridad independiente, la sostenibilidad del déficit solo exige una ratio deuda-PIB estable ($b_t = b_{t-1}$). Se obtiene así la *condición de sostenibilidad de la deuda pública*:

$$-d1_t = \frac{r_t - g_{Y,t}}{1 + g_{Y,t}} b_{t-1}$$

El superávit primario en porcentaje del PIB ($-d1_t$) ha de cubrir el coste efectivo de la deuda viva, $(r_t - g_{Y,t})b_{t-1}$, dividido entre el factor de crecimiento del PIB, $(1 + g_{Y,t})$. Prescindiendo del denominador, cercano a la unidad, el superávit primario requerido es $(r - g_Y)\, b$.

Intuitivamente, el crecimiento del PIB agranda el denominador de la ratio y "paga" así parte de la deuda. La condición es útil porque el saldo primario recoge los pagos y cobros que el Gobierno controla.

**Estática comparativa**: la condición de sostenibilidad ordena las soluciones de política económica frente a una deuda insostenible según la variable de la EDD sobre la que inciden:

- *Política monetaria*: la monetización eleva el señoraje ($\sigma$) y reduce la deuda que es preciso emitir, a costa de más inflación.

- *Política fiscal*: la consolidación fiscal reduce el déficit primario ($d1$).

- *Otras políticas*: el *default*, la quita de deuda o la reprogramación reducen el *stock* de deuda ($b$), las políticas de fomento del crecimiento elevan $g_Y$ y la *represión financiera* contiene el tipo de interés real ($r$)[15].

**Dinámica**: fuera del equilibrio, la EDD es una ecuación en diferencias de primer orden en $b_t$ con pendiente $(1+r)/(1+g_Y)$. Si $r < g_Y$, la pendiente es menor que uno y la ratio converge a un valor estacionario, $b^* = d1(1 + g_Y)/(g_Y - r)$, cualquiera que sea el punto de partida. Si $r > g_Y$, la pendiente es mayor que uno y la ratio diverge, salvo que el superávit primario la compense exactamente. El análisis gráfico ordena los casos.

Gráficamente, la dinámica de la deuda se representa en el plano deuda inicial-deuda final sobre el PIB, $b_{t-1}$-$b_t$ (figuras A39.2 a A39.5):

[14] La situación preocupante, $r > g_Y$, se dio en las economías avanzadas en los años ochenta y noventa, aunque en el conjunto de la posguerra ha sido más frecuente $r < g_Y$ (Blanchard, 2019).

[15] La represión financiera canaliza el ahorro privado hacia la deuda pública para contener sus tipos de interés. Para ello recurre a topes a los tipos de los depósitos, a la nacionalización de fondos de pensiones privados para invertirlos en deuda pública o a límites a la inversión exterior que generan ahorro cautivo. La consolidación fiscal también reduce los tipos, pero de forma indirecta.

1. *Ecuación dinámica de la deuda*, $b_t = f(b_{t-1})$: recoge la ratio de deuda que resulta de cada ratio heredada. Esta ecuación es una recta con ordenada en el origen $d1_t - \sigma_t$, que, si no existe monetización ($\sigma_t = 0$), equivale al déficit primario ($d1$). Su pendiente, $(1+r)/(1+g_Y)$, compara el tipo de interés real con el ritmo de crecimiento del PIB real. Cuanto más alto sea el tipo de interés real respecto al crecimiento, más crecerá la deuda.

2. *Bisectriz*, $b_t = b_{t-1}$: recoge los puntos en los que la ratio de deuda es estable. Los puntos a la derecha de la bisectriz reflejan que la ratio de deuda del periodo anterior es superior a la actual, y los puntos a la izquierda, lo contrario.

3. *Equilibrio*: en el corte de la recta con la bisectriz, $b^*$, la ratio de deuda es estacionaria. Las líneas discontinuas recogen la senda de la ratio desde cada ratio inicial ($b_0$, o $b_0^0$ y $b_0^1$ cuando se parte de dos).

Caben cuatro situaciones, según exista déficit o superávit primario y según el tipo de interés real ($r$) sea superior o inferior a la tasa de crecimiento económico ($g_Y$), suponiendo constantes $r, g_Y, \pi$:

1. *Déficit primario y crecimiento robusto* ($d1_t > 0$, $r < g_Y$): cuando el PIB avanza a un ritmo superior al tipo de interés real, la ratio de deuda tiende a un valor estacionario, tanto si se parte de una ratio superior como inferior a él (figura A39.2). El déficit público es, así, sostenible.

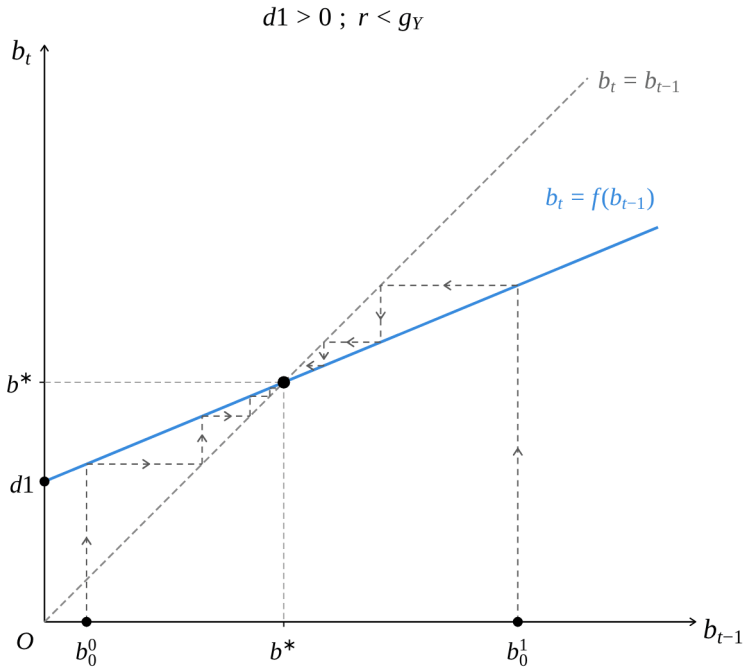



Figura A39.2: Dinámica de la deuda pública con déficit primario y crecimiento robusto ($d1 > 0$, $r < g_Y$). La ratio deuda/PIB converge a su valor estacionario, cualquiera que sea el punto de partida.

2. *Superávit primario y crecimiento robusto* ($d1_t < 0$, $r < g_Y$): la deuda converge hacia un valor estacionario negativo, es decir, el sector público acumula activos financieros y el superávit es sostenible. Basta con mantener la situación (figura A39.3).

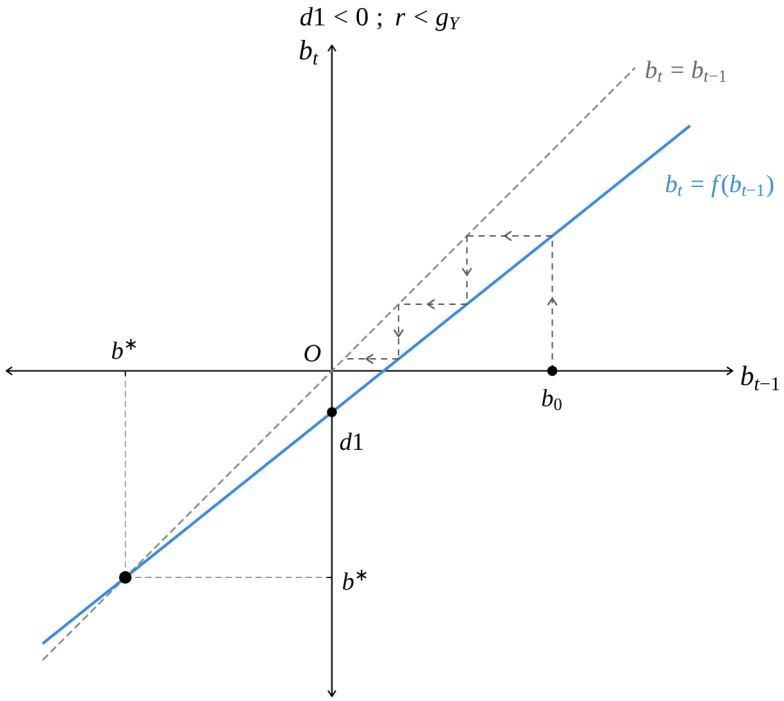



Figura A39.3: Dinámica de la deuda pública con superávit primario y crecimiento robusto ($d1 < 0$, $r < g_Y$). La ratio converge a un valor estacionario negativo y el sector público acumula activos.

3. *Déficit primario y crecimiento débil* ($d1_t > 0$, $r > g_Y$): si el tipo de interés real es superior a la tasa de crecimiento del PIB, la deuda sigue una senda explosiva, sea cual sea el punto de partida (figura A39.4). Se genera el efecto bola de nieve del segundo factor de la dinámica de la deuda. Por ello, el déficit no es sostenible y será necesaria una consolidación fiscal.

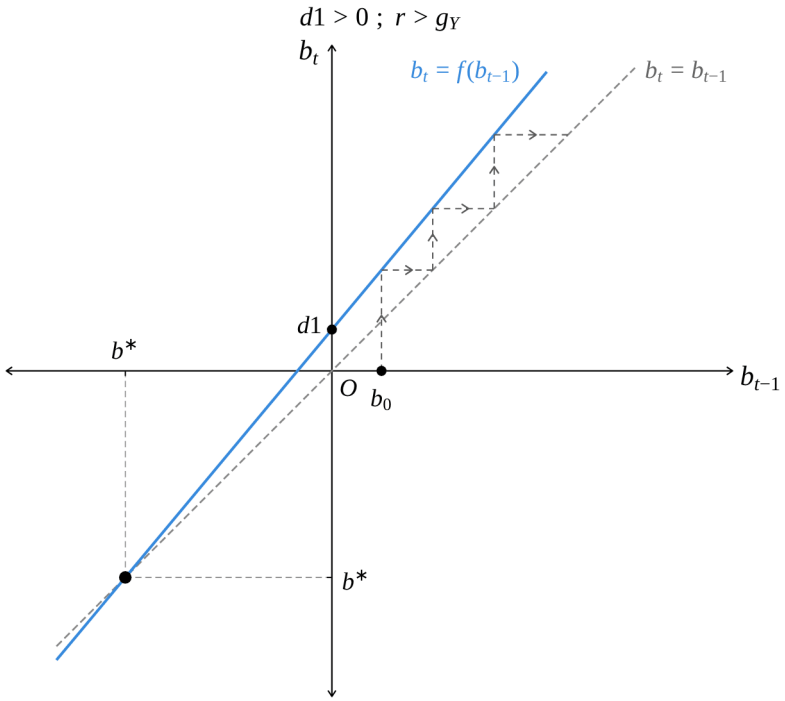



Figura A39.4: Dinámica de la deuda pública con déficit primario y crecimiento débil ($d1 > 0$, $r > g_Y$). La ratio deuda/PIB sigue una senda explosiva (efecto bola de nieve).

4. *Superávit primario y crecimiento débil* ($d1_t < 0$, $r > g_Y$): el resultado es ambiguo. Existe un nivel umbral de deuda, $b^* = d1(1 + g_Y)/(g_Y - r)$ (positivo, pues $d1 < 0$ y $g_Y < r$), por debajo del cual la deuda cae y por encima del cual crece. Es necesario ajustar el superávit en función de $b_0$ para compensar el diferencial $r - g_Y$ (figura A39.5).

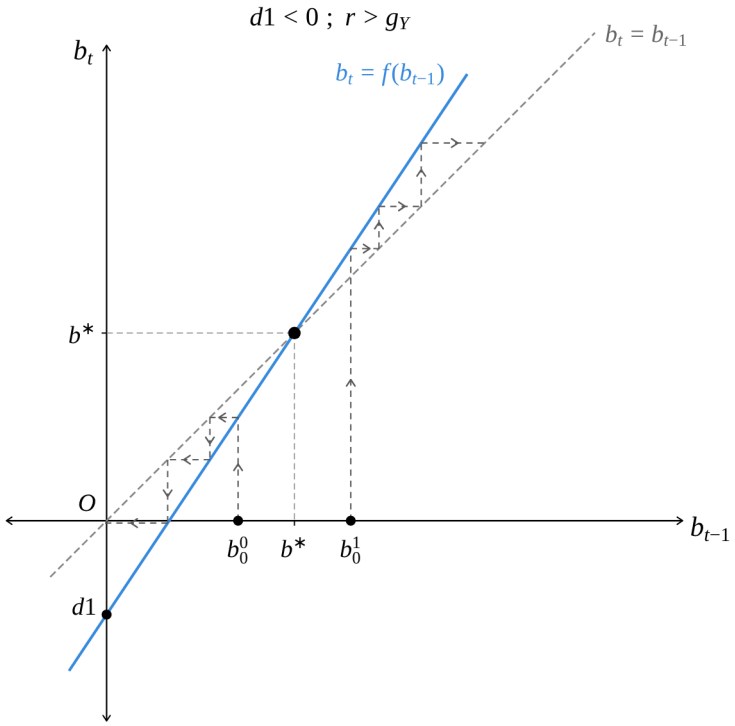



Figura A39.5: Dinámica de la deuda pública con superávit primario y crecimiento débil ($d1 < 0$, $r > g_Y$). El umbral $b^*$ separa las ratios iniciales que se reducen de las que crecen sin límite.

**Extensiones**: las dos siguientes relajan los supuestos del modelo:

▪ *Incidencia del sector privado*: la EDD supone que el sector privado absorbe toda la deuda emitida al tipo vigente. Sin embargo, en una economía cerrada la capacidad de absorción está limitada por el ahorro privado neto de inversión, de modo que la emisión neta de deuda no puede superar el excedente financiero del sector privado, $(B_t - B_{t-1})/P_tY_t \leq (S_t^p - I_t^p)/P_tY_t$. En la expresión, $S_t^p$ e $I_t^p$ son el ahorro y la inversión privados nominales. De este modo, la ecuación dinámica de la deuda con sector privado (EDDSP) hace depender la sostenibilidad de la capacidad de ahorro de la economía. Un déficit primario puede no ser sostenible, aun con $r < g_Y$, si el sector privado no puede o no quiere absorber la deuda al tipo vigente. El ajuste se produce entonces vía tipo de interés (prima de riesgo) o vía recurso a los mercados internacionales.

▪ *Función de reacción fiscal* (Bohn, 1998, 2007): como análisis empírico, la condición de sostenibilidad puede contrastarse estimando la respuesta del saldo primario a la deuda heredada:

$$-d1_t = \mu_t + \eta\, b_{t-1} + \varepsilon_t$$

donde $-d1_t$ es el saldo primario en porcentaje del PIB, $\mu_t$ recoge los determinantes exógenos (ciclo, gasto transitorio), $\eta$ la respuesta del saldo primario a la deuda y $\varepsilon_t$ un término de error. Bohn demuestra que $\eta > 0$ es condición suficiente para que se cumpla la restricción presupuestaria intertemporal, con independencia de la relación entre $r$ y $g_Y$. Basta con que el Gobierno reaccione sistemáticamente a los aumentos de deuda con mayores superávits primarios. Para

EE. UU., Bohn estima $\eta \approx 0{,}05$. En las economías avanzadas, Ghosh et al. (2013) encuentran además *fatiga fiscal*, es decir, una respuesta decreciente a partir de niveles elevados de deuda.

### 2.1.3. Conclusiones

**Aspectos positivos**: el análisis pone de manifiesto la importancia de destinar los fondos públicos a gasto corriente o a inversión pública, por su distinta capacidad de incidir en el crecimiento del PIB real.

También pone de manifiesto la importancia de obtener un tipo de interés competitivo, que reduce el diferencial $r - g_Y$. Generar confianza y credibilidad en los mercados de capitales permite obtener ese tipo más bajo.

**Limitaciones**: tal y como se ha definido (ratio deuda-PIB constante), el análisis de sostenibilidad es incompleto, porque ignora el valor que alcanza dicha ratio. El análisis indica la condición necesaria para que la ratio no crezca, pero una ratio del 20 % no equivale a una del 100 %, ya que en este último caso el superávit requerido supone un esfuerzo muy superior. Además, el análisis toma $r$ y $g_Y$ como dados cuando, en realidad, la propia deuda afecta a ambos (prima de riesgo, efecto expulsión sobre el capital privado). Cuanto mayor es la ratio deuda-PIB, mayor es, generalmente, el tipo que exigen los mercados, de modo que deuda y tipo se retroalimentan.

## 2.2. Financiación vía monetización en una economía cerrada

La monetización del déficit, es decir, la adquisición de deuda pública por la autoridad monetaria con dinero de nueva emisión, puede ser una práctica inflacionaria, lo que obliga a desarrollar los dos conceptos monetarios introducidos en el apartado 1.1, el señoraje y el impuesto inflacionario.

### 2.2.1. Señoraje e impuesto inflacionario

**Señoraje**: es el ingreso, en términos reales, que el Estado obtiene al crear dinero, medido por los bienes y servicios que su monopolio de emisión le permite adquirir[16]. Formalmente, se puede definir como el valor real del aumento de la cantidad nominal de dinero:

$$\sigma_t = \frac{\Delta M_t}{P_t} \;\rightarrow\; \sigma_t = g_{M,t}\frac{M_t}{P_t}$$

donde $\Delta M_t$ es la emisión de dinero en el periodo, $P_t$ el nivel de precios, $g_{M,t} = \Delta M_t / M_{t-1}$ la tasa de crecimiento monetario y $M_t/P_t$ los saldos reales (la segunda igualdad es exacta en tiempo continuo, $\sigma = \dot{M}/P = (\dot{M}/M)(M/P)$). El señoraje es, así, el producto del ritmo de emisión por los saldos reales sobre los que recae. Nótese que en este apartado el señoraje se mide en términos reales, mientras que en la EDD se medía en porcentaje del PIB.

**Impuesto inflacionario**: es el impuesto implícito que el señoraje representa, medido por la inflación. La inflación grava a los agentes privados, porque merma el valor real de sus saldos monetarios. Existen diferentes planteamientos a la hora de medir el impuesto implícito:

[16] Otras posibles definiciones del señoraje en la práctica podrían ser los ingresos del Tesoro Público procedentes del Banco Central o los ahorros que se obtienen por la emisión de deuda al Banco Central en comparación con la emisión al sector privado.

▪ *Criterio de Friedman* (1971): dada la producción total de la economía y la velocidad de circulación del dinero, la inflación vendrá dada por la tasa de crecimiento de la cantidad de dinero, de acuerdo con la teoría cuantitativa del dinero:

$$MV_0 = PY_0 \xrightarrow{\ln} \ln M + \ln V_0 = \ln P + \ln Y_0 \xrightarrow{d/dt} g_{M,t} = \pi_t \rightarrow T_t^{\pi} = \pi_t \frac{M}{P}$$

donde $V_0$ e $Y_0$ son la velocidad de circulación y la producción, constantes. El impuesto inflacionario, $T_t^{\pi}$, es el producto de la inflación (que representa el "tipo de gravamen") y los saldos reales de los agentes económicos, $M/P$ (que representan la "base imponible"). Con producción constante, coincide con el señoraje.

▪ *Criterio de Marty* (1967): si la producción crece a una tasa $g_{Y,t}$, la emisión de dinero genera impuesto inflacionario solo si la masa monetaria crece más que el PIB. Si $g_{M,t} = g_{Y,t}$, la oferta monetaria aumenta lo mismo que la demanda de dinero, que crece con la renta, y no hay inflación. En cambio, si $g_{M,t} > g_{Y,t}$, la oferta crece más que la demanda, se produce inflación y aparece el impuesto inflacionario. Formalmente, con elasticidad renta unitaria de la demanda de dinero, $g_{M,t} = \pi_t + g_{Y,t}$, de modo que:

$$\sigma_t = (\pi_t + g_{Y,t})\frac{M}{P} \rightarrow T_t^{\pi} = \sigma_t - g_{Y,t}\frac{M}{P}$$

donde $g_{Y,t}M/P$ es el *señoraje no inflacionario*, la parte de la emisión que acompaña al crecimiento real de la demanda de dinero, y $T_t^{\pi} = \pi_t M/P$ el impuesto inflacionario propiamente dicho.

### 2.2.2. Señoraje óptimo: el modelo de Cagan (1956)

> El modelo de Cagan se desarrolla en el tema A41, apartado 3.1.2. Aquí se retiene lo que interesa a la financiación del déficit, el señoraje óptimo.

**Origen**: P. Cagan (1956) estudió siete episodios de hiperinflación en seis países europeos durante el s. XX[17].

**Idea**: Cagan se preguntaba si el crecimiento de la masa monetaria y la inflación generada afectan, a su vez, a la demanda de dinero debido a las expectativas crecientes de inflación, de modo que la demanda de dinero se reduce y se perpetúa la inflación.

**Supuestos**:

▪ *Entorno*: el mercado monetario está en equilibrio en todo momento y las variables reales ($y_t$, $r_t$) presentan, durante las hiperinflaciones, variaciones insignificantes con respecto a las nominales[18].

[17] Cagan estudió hiperinflaciones en Alemania, Austria, Hungría, Polonia, Rusia y Grecia. Cuando estimó los coeficientes, encontró que en la mayoría de los episodios el producto $a\lambda$ era inferior a la unidad y la dinámica se explicaba, esencialmente, por los fundamentos (el crecimiento monetario). No obstante, para Rusia (1921-1924) y Alemania (1922-1923) el producto $a\lambda$ superaba ligeramente la unidad, lo que apuntaba a un componente autoalimentado de las expectativas.

[18] Se parte de la demanda canónica $m_t - p_t = \kappa_0 + \kappa_1 y_t - a i_t + \rho_t$, con $y_t$ el logaritmo de la renta real, $\rho_t$ una perturbación e $i_t = r_t + \pi_t^e$, de donde $c = \kappa_0 + \kappa_1 y_t - a r_t$.

▪ *Agentes*: forman sus expectativas de inflación según la hipótesis de expectativas adaptativas, $\dot{\pi}^e_t = \lambda(\pi_t - \pi^e_t)$, donde $\lambda > 0$ es la velocidad de ajuste[19]. Los agentes demandan saldos reales según la función de Cagan, $m_t - p_t = c - a\pi^e_t$, donde $m_t$ y $p_t$ son los logaritmos de la cantidad nominal de dinero y del nivel de precios.

En la función de Cagan, $c > 0$ recoge la demanda de dinero por factores reales distintos de la inflación esperada ($c = \kappa_0 + \kappa_1 y_t - a r_t$) y $a > 0$ mide la sensibilidad de la demanda de dinero a la inflación esperada e, indirectamente, al tipo de interés nominal[20]. Si la inflación esperada sube, la demanda de saldos reales baja, pues los agentes demandan sustitutivos del dinero. Además, la relación es exponencial ($M_t/P_t = e^{c - a\pi^e_t}$), por lo que la semielasticidad de los saldos reales respecto a la inflación esperada es constante ($-a$). Por tanto, cada punto adicional de inflación esperada los reduce en torno a un $a$ %, lo que, con las tasas propias de una hiperinflación, se traduce en caídas muy intensas.

**Planteamiento**: el Gobierno elige la tasa de crecimiento monetario que maximiza el señoraje de estado estacionario. Dado que la demanda de saldos reales depende inversamente de $\pi^e$, conforme aumenta la tasa de crecimiento monetario el señoraje primero aumenta y, a partir de cierto punto, disminuye. Se parte de la definición de señoraje, $\sigma_t = g_{M,t}(M_t/P_t)$. Como en estado estacionario la inflación esperada coincide con la tasa de crecimiento monetario, la demanda de dinero es $M_t/P_t = e^{c - a g_{M,t}}$, y el programa es:

$$\max_{g_M} \ \sigma = g_M \, e^{c - a g_M}$$

donde $g_M$ es la variable de decisión y $e^{c - a g_M}$ la demanda de saldos reales en estado estacionario ($c$ y $a$, como en la función de Cagan).

**CPO**: $\partial\sigma/\partial g_M = e^{c - a g_M}(1 - a g_M) = 0$. Elevar $g_M$ aumenta el "tipo" del impuesto, pero reduce su base (los saldos reales). El óptimo iguala en el margen ambos efectos, el de tipo y el de base.

**CSO**: $\partial^2\sigma/\partial(g_M)^2 = -a \, e^{c - a g_M}(2 - a g_M) < 0$ en el punto crítico, por lo que este es un máximo.

**Solución**: la tasa de crecimiento monetario que maximiza el señoraje y el señoraje máximo son:

$$g^*_M = \frac{1}{a} \ \rightarrow \ \sigma^* = \frac{e^{c-1}}{a} = f(c^+, a^-)$$

La tasa óptima es inversa a la sensibilidad de la demanda de dinero a la inflación esperada, de forma que el comportamiento de los agentes restringe la creación de oferta monetaria.

**Desarrollo gráfico**: en el plano crecimiento monetario-señoraje, $g_M$-$\sigma$ (figura A39.6), la *curva de Laffer del señoraje*, $\sigma(g_M)$, recoge el señoraje de estado estacionario para cada tasa de crecimiento monetario. Alcanza su máximo, $\sigma^*$, en $g^*_M = 1/a$, donde la elasticidad de la base impositiva respecto al tipo impositivo es igual a $-1$ (unitaria en valor absoluto), lo que coincide con la condición de recaudación máxima de cualquier impuesto.

[19] En tiempo discreto: $\pi^e_{t+1} = \pi^e_t + \lambda(\pi_t - \pi^e_t)$, o $\pi^e_{t+1} = \lambda\pi_t + (1-\lambda)\pi^e_t$. La principal limitación de esta forma de modelar las expectativas es que es *backward looking*. Por ello, en versiones posteriores la ecuación de precios se formuló con expectativas racionales, $\pi^e_{t+1} = \mathbb{E}_t[p_{t+1}|\Omega_t] - p_t$.

[20] Las estimaciones de Cagan, con la inflación medida en tasa mensual, sitúan $a$ aproximadamente entre 2 y 9 meses. Los sustitutivos del dinero pueden ser activos reales (p. ej. vivienda, obras de arte).

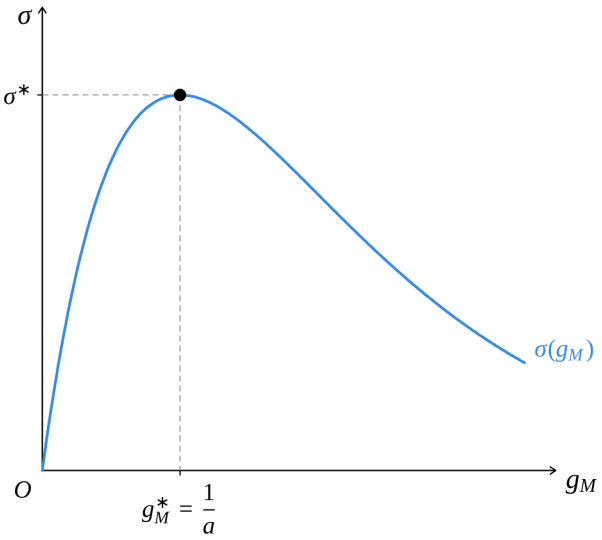



Figura A39.6: Modelo de Cagan. El señoraje de estado estacionario en función del crecimiento monetario tiene forma de curva de Laffer, con máximo en $g_M^* = 1/a$.

**Estática comparativa**: el señoraje máximo depende positivamente de la demanda de dinero por factores reales ($c$), pues una mayor demanda de saldos reales amplía la base sobre la que recae el impuesto (la tasa óptima, $g_M^* = 1/a$, no depende de $c$), y negativamente de $a$.

**Dinámica**: con expectativas adaptativas, la estabilidad del estado estacionario, $\pi^e = \pi = g_M$, depende de $a\lambda$. Si $a\lambda < 1$, el equilibrio es estable y la hiperinflación solo puede deberse a un crecimiento excesivo de la masa monetaria. En cambio, si $a\lambda > 1$, es inestable, porque un salto en la inflación revisa al alza la expectativa, reduce la demanda de saldos reales y genera más inflación, de modo que la hiperinflación es una profecía autocumplida.

> La condición de estabilidad, con su derivación y su análisis gráfico, se desarrolla en el tema A41, apartado 3.1.2.

**Conclusiones**:

- *Aspectos positivos*: el modelo muestra que la financiación monetaria del déficit está acotada, porque existe un señoraje máximo y, más allá de él, emitir más dinero recauda menos. Explica, además, las hiperinflaciones como el resultado de un Gobierno que intenta extraer más señoraje del que la demanda de dinero permite.

- *Limitaciones*: maximizar el señoraje no equivale a maximizar el bienestar social. La tasa $g_M^*$ genera una inflación con elevados costes de bienestar, que deben pesar al elegir la tasa recomendable. Ahora bien, sustituir la emisión de dinero por la de deuda no evita necesariamente la inflación, y puede incluso agravarla, como muestra la aritmética monetarista desagradable (apartado 3.2.1).

### 2.2.3. Efectos de la monetización a largo plazo (Blinder y Solow, 1973)

**Origen**: Blinder y Solow cuestionan la idea de que la monetización del déficit es más expansiva que su financiación a través de la emisión de bonos, y sistematizan la incorporación de la

restricción presupuestaria del Gobierno (RPG) al análisis IS-LM.

**Idea**: la RPG obliga a que el déficit se cierre a largo plazo, y la forma de financiación determina el *stock* de activos que acaba en manos del sector privado y, con él, la renta de equilibrio:

$$PG + iB = PT + \dot{B} + \dot{M}$$

donde $PG$ es el gasto corriente en términos nominales, $iB$ los pagos por intereses de la deuda, $PT$ la recaudación impositiva nominal, con $T = \tau Y$ creciente en la renta, $\dot{B}$ la emisión de deuda y $\dot{M}$ los ingresos de señoraje.

**Razonamiento**: el equilibrio a largo plazo se define por el presupuesto equilibrado, en el que los *stocks* de bonos y de dinero dejan de variar ($\dot{B} = \dot{M} = 0$). Imponiendo esta condición a la RPG, con financiación monetaria el multiplicador a largo plazo del gasto es $dY/dG = 1/\tau$, porque la renta debe crecer hasta que la recaudación adicional cubra el gasto. Con financiación vía bonos, el gasto a cubrir incluye los intereses de la deuda acumulada, $dY/dG = [1 + i\,(dB/dG)]/\tau > 1/\tau$.

Con financiación vía bonos, el público acumula más títulos y el tipo de interés sube. Si el *stock* de dinero no varía, la mayor riqueza financiera impulsa además el consumo y, con él, la renta. Con monetización, en cambio, el tipo de interés sube menos. Como el *stock* de bonos no cambia, el efecto riqueza solo procede de los nuevos saldos monetarios y no hay intereses adicionales que pagar. El efecto liquidez de la monetización hace que la renta se expanda antes. Con un horizonte largo, sin embargo, financiar el déficit con deuda resulta más expansivo, porque, hasta que el presupuesto se equilibra, obliga a crear un volumen total de activos mayor.

Los monetaristas sostenían que el efecto expulsión anula por completo un déficit que se cubre con deuda. Frente a esa tesis, los efectos riqueza y de cartera tienen en conjunto un sesgo expansivo, y la renta crece hasta que la recaudación adicional cubre el déficit. Ese resultado exige, no obstante, un sistema estable, algo que la financiación con bonos no asegura. Con un efecto riqueza débil, la carga de intereses aumenta más deprisa que los ingresos fiscales y la deuda se vuelve explosiva.

**Conclusiones**:

- *Aspectos positivos*: incorpora los efectos riqueza y la dinámica de la deuda al debate sobre la eficacia de la política fiscal.

- *Limitaciones*: hereda los supuestos del modelo IS-LM, con precios fijos y sin expectativas sobre la senda futura de impuestos, por lo que ignora la respuesta de los agentes que subraya la equivalencia ricardiana (apartado 1.2.2).

## 2.3. Financiación en una economía abierta

En una economía abierta, el Gobierno puede financiarse también en los mercados internacionales, y los efectos de la monetización dependen del régimen cambiario.

**Deuda externa**: si el sector privado nacional no absorbe la deuda, la economía acude a los mercados internacionales, con lo que la restricción presupuestaria del Gobierno incorpora la deuda externa, denominada en divisas y valorada al tipo de cambio nominal:

$$G1_t - T_t = [B_t - (1 + i_t)B_{t-1}] + (H_t - H_{t-1}) + E_t \left[B_t^f - (1 + i_t^f)B_{t-1}^f\right]$$

donde $B_t^f$ es la deuda externa denominada en divisas, $i_t^f$ el tipo de interés que paga y $E_t$ el tipo de cambio nominal (unidades de moneda nacional por divisa). Dividiendo por el PIB nominal se

obtiene una EDD para la deuda externa análoga a la interna, en la que el coste efectivo incorpora la depreciación observada.

El endeudamiento externo puede ser beneficioso a corto plazo, pero debe presentar una senda no explosiva que garantice su sostenibilidad. En ella influyen especialmente dos factores adicionales. El primero es la prima de riesgo, que eleva $i^f$ con el nivel de deuda y puede convertir en explosiva una senda que no lo era. El segundo es el tipo de cambio, pues una depreciación ($\uparrow E_t$) eleva el valor en moneda nacional de la deuda en divisas (*efecto hoja de balance*).

El FMI desarrolló el *enfoque de la hoja de balance* (*balance sheet approach*) tras las crisis emergentes de finales de los años noventa. Frente al análisis tradicional de la balanza de pagos, centrado en flujos, el enfoque estudia los balances (*stocks* de activos y pasivos) del sector público, el financiero y el privado no financiero, y su posición frente al resto del mundo. El objetivo es detectar descalces de plazos, de divisas y de estructura del capital.

**Monetización con tipo de cambio fijo**: la autoridad monetaria da prioridad a la financiación del déficit y puede elegir entre dos opciones (con CI el crédito interno y $R$ las reservas exteriores):

- *No esterilizar la financiación*: se amplía la base monetaria ($\uparrow H = \uparrow \text{CI} + R$) y, en consecuencia, la oferta monetaria.

- *Esterilizarla*: el banco central retira, a cambio de divisas, tanta moneda nacional como ha prestado al Gobierno, y la base monetaria se mantiene ($\overline{H} = \uparrow \text{CI} + \downarrow R$).

La defensa de la paridad exige esterilizar la monetización, y cada operación consume reservas, de modo que los fundamentos se deterioran poco a poco. Con unas reservas finitas, antes o después financiar el déficit y defender la paridad dejarían de ser compatibles para el banco central. Como los agentes lo prevén, un ataque especulativo se lleva de una vez las reservas que quedan sin esperar a su agotamiento gradual (crisis de balanza de pagos). Fiel a su prioridad, la financiación del déficit, el banco central renuncia entonces a la paridad (devaluación).

> Esta secuencia corresponde al modelo de crisis cambiarias de primera generación (Krugman, 1979a), desarrollado en el tema B17, apartado 1.

**Monetización con tipo de cambio flexible**: no se producen variaciones de reservas ni es necesario esterilizar. Sin embargo, la monetización amplía la oferta monetaria y baja el tipo de interés, con la consiguiente salida de capitales y depreciación de la moneda ($\downarrow i \ \rightarrow \uparrow E$, unidades de moneda nacional por divisa). La depreciación encarece las importaciones y puede elevar la inflación (inflación importada)[21].

## 2.4. Conclusiones

El análisis de la financiación arroja dos conclusiones. En primer lugar, la sostenibilidad de la deuda depende fundamentalmente de la relación $r - g_Y$ y de la respuesta del saldo primario a la deuda. Si el crecimiento supera al tipo de interés real, la deuda es sostenible incluso con déficit primario y, en caso contrario, se requiere un superávit primario permanente. En todo caso, la función de reacción fiscal (Bohn, 1998) muestra que lo decisivo es que el Gobierno reaccione a la deuda. Con deuda externa, la sostenibilidad depende además de la prima de riesgo y del tipo de cambio, por el efecto hoja de balance.

[21] La teoría monetaria moderna, heterodoxa (Wray, Mitchell), sostiene que un Estado que emite su propia moneda puede financiar el déficit con emisión de dinero, con la inflación como único límite.

En segundo lugar, la monetización puede aliviar transitoriamente la carga de la deuda, pero está acotada por el señoraje máximo (curva de Laffer) y genera inflación, con efectos que dependen del régimen cambiario. Queda por ver qué forma de financiación elegir y, sobre todo, quién decide, es decir, si existe dominancia de una autoridad sobre la otra.

# 3. Dominancia monetaria y dominancia fiscal

La restricción presupuestaria intertemporal del sector público vincula a la autoridad fiscal y a la monetaria, de modo que, si una de ellas sigue una política activa, la otra debe seguir una pasiva para que la restricción se cumpla (apartado 1.1). El concepto de dominancia fiscal y monetaria puede entenderse a través del modelo de Leeper (1991), que ofrece el marco en el que se ordenan las teorías posteriores.

## 3.1. El modelo de Leeper (1991)

**Origen**: Leeper (1991) formaliza la distinción de Sargent y Wallace (1981) entre dominancia monetaria y fiscal en un modelo dinámico con expectativas racionales en el que cada autoridad sigue una regla de política.

**Idea**: la combinación de las reglas de ambas autoridades, activas o pasivas (apartado 1.1), determina el régimen. Solo las combinaciones de una política activa y otra pasiva producen un equilibrio determinado.

### 3.1.1. Supuestos

**Entorno**: mundo neoclásico (plena flexibilidad de precios y tipos de interés, mercados de capitales perfectos y completos, racionalidad) en el que existe incertidumbre en cuanto a posibles *shocks* fiscales o monetarios.

**Agentes**: los consumidores tienen expectativas racionales, maximizan su utilidad intertemporalmente y sus preferencias se modelan con una función *money-in-utility* (MIU). El sector público engloba a la autoridad fiscal y a la autoridad monetaria, cada una de las cuales se guía por una regla de comportamiento.

**Tecnología**: economía de dotación, en la que la renta $y_t$ es exógena, de modo que el modelo aísla la determinación de la inflación y la deuda.

### 3.1.2. Modelo

**Planteamiento**: el modelo se resume en la decisión de los hogares, la restricción presupuestaria del sector público y las dos reglas de política, de cuya combinación resulta el régimen. Los hogares resuelven un programa de optimización intertemporal de la senda de consumo y saldos reales de dinero:

$$\max_{\mathbf{c},\mathbf{m}} \; u^d = \sum_{t=0}^{\infty} \beta^t u(c_t, m_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} c_t + (b_t - b_{t-1}) + \left(m_t - \frac{m_{t-1}}{1+\pi_t}\right) = y_t - T_t + r_t b_{t-1} \\ b_{-1},\ m_{-1} \text{ dados} \\ \lim_{T\to\infty} b_T / R_T \geq 0 \end{cases}$$

donde $c_t$ es el consumo real, $m_t$ los saldos reales, $\beta$ el factor de descuento, $u^d$ la utilidad descontada y $u$ la instantánea. En la restricción, $b_t$ es el *stock* real de bonos públicos en manos del agente, $\pi_t$ la inflación (que erosiona el valor real de los saldos heredados), $y_t$ su renta y $T_t$ los impuestos de suma fija. Por último, $r_t$ es el tipo de interés real que rinde en $t$ la deuda adquirida

en $t-1$ (los bonos adquiridos en $t$ rinden $r_{t+1}$). Los activos iniciales, $b_{-1}$ y $m_{-1}$, están dados, y la condición de no-Ponzi, con $R_T \equiv \prod_{s=1}^{T}(1+r_s)$, impide endeudarse indefinidamente.

**CPO**: con solución interior, sustituyendo la restricción en la función objetivo y derivando respecto a $m_t$ y $b_t$, se obtienen dos condiciones:

1. *Demanda de dinero*: $\partial u^d/\partial m_t = 0 \;\Rightarrow\; \dfrac{\partial u}{\partial c_t} = \dfrac{\partial u}{\partial m_t} + \beta\,\dfrac{\partial u/\partial c_{t+1}}{1+\pi_{t+1}}$, es decir, el coste de una unidad de dinero hoy (consumo sacrificado) iguala a su utilidad como liquidez más su valor real mañana.

2. *Ecuación de Euler*: $\partial u^d/\partial b_t = 0 \;\Rightarrow\; \text{RMSI} \equiv \dfrac{1}{\beta}\dfrac{\partial u/\partial c_t}{\partial u/\partial c_{t+1}} = 1+r_{t+1}$, es decir, la relación marginal de sustitución intertemporal iguala al rendimiento bruto de los bonos.

**CSO**: $u$ estrictamente cóncava en $(c, m)$, restricción lineal y condición de transversalidad sobre bonos y dinero, por lo que las CPO son necesarias y suficientes.

**Solución**: sustituyendo la segunda en la primera y reordenando:

$$\frac{\partial u/\partial m_t}{\partial u/\partial c_t} = \frac{i_t}{1+i_t}$$

donde $i_t$ es el tipo de interés nominal, con $1+i_t = (1+r_{t+1})(1+\pi_{t+1})$. Es decir, se demandan saldos reales hasta que la relación marginal de sustitución entre saldos reales y consumo se iguala al coste de oportunidad (en valor presente) de mantener una unidad de dinero en lugar de ahorro que rinde un interés $i_t$. Por tanto, la demanda de dinero depende negativamente del tipo nominal y, a través de él, de la inflación esperada. Por este canal, el del señoraje, la política fiscal puede acabar influyendo en los precios.

**Sector público**: se modela con su restricción presupuestaria y, a diferencia de otros modelos dinámicos, cada autoridad sigue una regla:

▪ *Restricción presupuestaria*: refleja las dos formas de financiación del déficit, vía emisión de deuda pública y vía emisión de masa monetaria:

$$(g_t - T_t) + r_t b_{t-1} = (b_t - b_{t-1}) + \left(m_t - \frac{m_{t-1}}{1+\pi_t}\right)$$

donde $g_t$ es el gasto público. El déficit primario más los intereses reales de la deuda (lado izquierdo) se financian con la emisión neta de bonos y con el señoraje (lado derecho).

▪ *Regla de la autoridad monetaria*: fija el tipo de interés como una función de la inflación corriente y de los *shocks* monetarios no anticipados:

$$i_t = \alpha_0 + \alpha\pi_t + \varepsilon_t^m, \qquad \varepsilon_t^m = \rho_1\varepsilon_{t-1}^m + \varphi_{1,t}, \quad |\rho_1| < 1, \quad \varphi_{1,t} \sim N(0, \sigma_1^2)$$

donde $\alpha_0$ es el tipo de interés objetivo a medio plazo, $\alpha$ la sensibilidad del tipo oficial a la inflación corriente y $\varepsilon_t^m$ el *shock* monetario, que sigue un proceso AR(1) con persistencia $\rho_1$ e innovación $\varphi_{1,t}$.

▪ *Regla de la autoridad fiscal*: ajusta los impuestos de suma fija en respuesta al nivel de deuda heredado y a los *shocks* fiscales no anticipados:

$$T_t = \gamma_0 + \gamma_1 b_{t-1} + \varepsilon_t^f, \qquad \varepsilon_t^f = \rho_2 \varepsilon_{t-1}^f + \varphi_{2,t}, \quad |\rho_2| < 1, \quad \varphi_{2,t} \sim N(0, \sigma_2^2)$$

donde $\gamma_0$ es el nivel de impuestos de estado estacionario, $\gamma_1$ la sensibilidad de los impuestos al *stock* de deuda heredado y $\varepsilon_t^f$ el *shock* fiscal, AR(1) con persistencia $\rho_2$ e innovación $\varphi_{2,t}$.

**Equilibrio**: linealizando el modelo en torno al estado estacionario se obtiene un sistema en $(\pi_t, b_t)$ cuya solución depende de dos parámetros, la respuesta de la política monetaria a la inflación ($\alpha\beta$) y la respuesta de la política fiscal a la deuda ($\beta^{-1} - \gamma_1$). Cada autoridad puede ser *activa* o *pasiva*:

1. *Autoridad monetaria*: es activa si $|\alpha\beta| > 1$, es decir, si responde a la inflación elevando el tipo de interés real (principio de Taylor), y pasiva en caso contrario.

2. *Autoridad fiscal*: es pasiva si $|\beta^{-1} - \gamma_1| < 1$, es decir, si los impuestos responden a la deuda en más de lo que cuesta su servicio ($\gamma_1 > \beta^{-1} - 1 \approx r$), de modo que la deuda se estabiliza, y activa en caso contrario.

**Desarrollo gráfico**: el modelo de Leeper se representa en el plano respuesta monetaria-respuesta fiscal, $|\alpha\beta|$-$|\beta^{-1} - \gamma_1|$ (figura A39.7):

1. *Fronteras*: las rectas $|\alpha\beta| = 1$ y $|\beta^{-1} - \gamma_1| = 1$ separan la postura activa de la pasiva de cada autoridad y delimitan cuatro regiones, según la postura de la política monetaria (PM) y la de la política fiscal (PF).

2. *Región I: dominancia monetaria* (PM activa, PF pasiva): la autoridad monetaria reacciona agresivamente a la inflación y la fiscal ajusta los impuestos para estabilizar la deuda. La inflación queda determinada por la política monetaria y la restricción presupuestaria se cumple vía impuestos. Corresponde al caso convencional estudiado por la nueva macroeconomía clásica y la nueva economía keynesiana. El equilibrio es único y determinado.

3. *Región II: dominancia fiscal* (PM pasiva, PF activa): la autoridad fiscal no ajusta los impuestos ante variaciones de la deuda y la monetaria acomoda la inflación. El nivel de precios queda determinado por la restricción presupuestaria del Gobierno, y en ello consiste el mecanismo central de la teoría fiscal del nivel de precios. El equilibrio es único y determinado.

4. *Región III: indeterminación* (ambas pasivas): las dos autoridades acomodan. El equilibrio está indeterminado, ya que múltiples trayectorias de inflación y deuda son compatibles con las reglas, lo que hace posibles equilibrios de manchas solares (*sunspots*).

5. *Región IV: sin equilibrio* (ambas activas): ni la autoridad monetaria ni la fiscal acomodan. De ahí que no exista equilibrio estacionario. La deuda sigue una senda explosiva, pues nadie garantiza la restricción presupuestaria.

6. *Equilibrio*: la determinación del equilibrio requiere que exactamente una de las dos autoridades sea activa y la otra pasiva (regiones I y II). La estimación de los parámetros de las reglas permite determinar si rige la dominancia monetaria o la fiscal, y la coordinación entre ambas políticas no es opcional sino un requisito de la determinación del equilibrio.

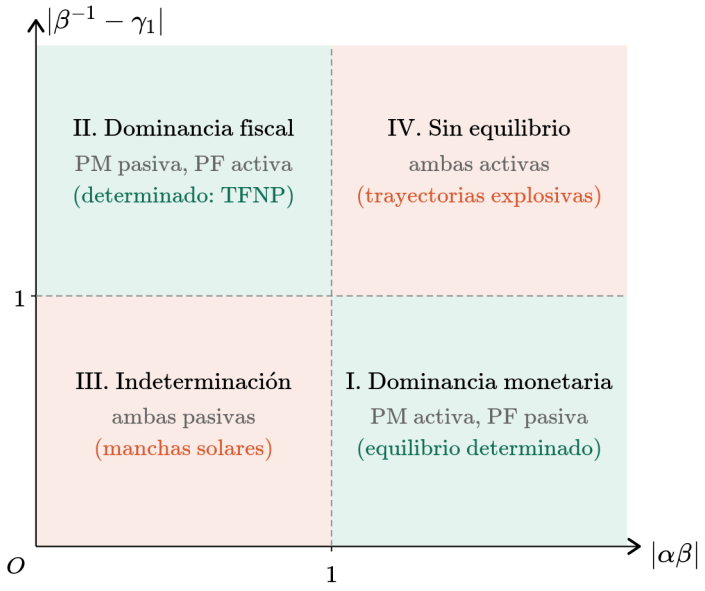



Figura A39.7: Modelo de Leeper (1991). El plano se divide en regiones de política monetaria activa/pasiva y fiscal activa/pasiva. Solo las combinaciones PM activa-PF pasiva y PM pasiva-PF activa generan un equilibrio determinado.

**Dinámica**: las funciones de respuesta a los *shocks* dependen del régimen, pues la combinación de posturas determina quién absorbe cada perturbación:

1. *Dominancia monetaria* (PM activa, PF pasiva): un *shock* fiscal (una rebaja de impuestos financiada con deuda) lo absorben los impuestos futuros. La deuda se estabiliza vía $\gamma_1$ y la inflación no varía (equivalencia ricardiana). Un *shock* monetario contractivo reduce la inflación, como en el análisis convencional.

2. *Dominancia fiscal* (PM pasiva, PF activa): el mismo *shock* fiscal no será compensado con impuestos futuros, de modo que la deuda emitida solo puede ser absorbida si su valor real cae, por lo que la inflación sube en el impacto. También un *shock* monetario contractivo eleva la inflación, porque la subida del tipo de interés aumenta la carga de intereses y, con ella, la deuda nominal que ningún impuesto respalda, que debe licuarse vía precios.

**Extensiones**: si las reglas pueden cambiar de régimen a lo largo del tiempo (*regime switching*), la expectativa de un régimen futuro de dominancia fiscal afecta a la inflación presente incluso bajo dominancia monetaria.

### 3.1.3. Conclusiones

**Aspectos positivos**: el modelo de Leeper ofrece un marco unificado que integra la interacción entre política monetaria y fiscal en un contexto dinámico con expectativas racionales. Su principal virtud es demostrar formalmente que la estabilidad de precios no depende exclusivamente de la política monetaria, sino de la combinación de reglas monetarias y fiscales. La idea tiene

implicaciones para el diseño institucional (p. ej. la arquitectura de la Unión Económica y Monetaria, UEM). Además, el modelo fundamenta la teoría fiscal del nivel de precios y proporciona una taxonomía clara de regímenes.

**Limitaciones**: el modelo supone reglas lineales y no contempla el sector exterior ni el tipo de cambio. Además, los agentes son homogéneos (agente representativo ricardiano), por lo que el modelo no captura las restricciones de liquidez que reforzarían la dominancia fiscal.

## 3.2. Dominancia fiscal

El modelo de Leeper permite entender diferentes teorías que han buscado explicar situaciones en cada una de las regiones. En la región II se sitúan dos aportaciones.

### 3.2.1. Aritmética monetarista desagradable (Sargent y Wallace, 1981)

**Origen**: Sargent y Wallace, en *Some Unpleasant Monetarist Arithmetic*, plantean el problema de un banco central monetarista que intenta controlar la inflación frente a una autoridad fiscal que mantiene déficits permanentes.

**Idea**: si la política fiscal es expansiva de manera persistente, el banco central pierde el dominio sobre la inflación, hasta el punto de que endeudarse para cubrir el déficit puede generar más inflación que monetizarlo.

**Razonamiento**: la secuencia es la siguiente:

1. *Financiación vía deuda*: cuando la expansión fiscal no puede monetizarse (p. ej. porque la autoridad monetaria mantiene una política restrictiva), al Gobierno no le queda otra vía que emitir deuda para cubrir sus déficits.

2. *Efecto bola de nieve*: como $r > g_Y$, los déficits crecen año tras año (apartado 2.1.2). La deuda acumulada ensancha además la diferencia entre $r$ y $g_Y$, porque el Estado disputa al sector privado el ahorro disponible.

3. *Monetización final*: cuando la deuda deja de ser sostenible, el mercado ya no la absorbe. Los intereses acumulados pueden ser entonces tan altos que solo la emisión de dinero permite atenderlos, y el Gobierno termina por recurrir a ella, lo que dispara la inflación.

4. *Expectativas racionales*: los agentes anticipan esa monetización, y la mayor inflación esperada reduce desde hoy la demanda de saldos reales y eleva los precios. Ni siquiera la inflación actual queda, por tanto, bajo el control del banco central.

Paradójicamente, una política monetaria contractiva hoy puede generar más inflación a largo plazo si la autoridad fiscal no coopera, pues la política monetaria contractiva exige mayor endeudamiento y, dado el límite a la absorción de deuda por el sector privado, un mayor crecimiento de $M^s$ mañana.

Dentro de este modelo, la inflación resulta así un fenómeno más fiscal que monetario, y sin disciplina fiscal no hay estabilidad de precios. Si la monetización futura no puede evitarse, retrasarla agrava sus efectos, porque aumentan los intereses que habrá que monetizar.

**Conclusiones**:

▪ *Aspectos positivos*: muestra que el control de la inflación no depende solo del banco central sino también de la política fiscal, y sienta la distinción entre dominancia monetaria y dominancia fiscal que Leeper (1991) formaliza después. Fundamenta así la exigencia de disciplina fiscal como complemento de la independencia del banco central.

▪ *Limitaciones*: el resultado descansa en supuestos fuertes, como un tipo de interés real superior al crecimiento, un límite exógeno a la deuda que el público está dispuesto a absorber y una autoridad fiscal que fija su política en primer lugar sin reaccionar a la monetaria. Si alguno de ellos falla, la paradoja deja de estar garantizada.

### 3.2.2. Teoría fiscal del nivel de precios (Leeper, 1991; Sims, 1994; Woodford, 1995; Cochrane, 2001)

**Origen**: la teoría fiscal del nivel de precios (TFNP) formaliza y generaliza la intuición de Sargent y Wallace en el marco del modelo de Leeper.

**Idea**: en el régimen de dominancia fiscal, el nivel de precios queda determinado no por la oferta monetaria sino por la restricción presupuestaria intertemporal del Gobierno, que deja de ser una restricción para convertirse en una condición de equilibrio. A diferencia de Sargent y Wallace, el mecanismo no requiere señoraje futuro, sino que opera a través del valor real de la deuda nominal, que el nivel de precios revalúa.

**Razonamiento**: la ecuación de valoración de la deuda pública establece que el valor real de la deuda nominal heredada debe igualar el valor actual de los superávits primarios esperados:

$$\frac{B_{t-1}}{P_t} = \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s \left(T_{t+s} - G_{t+s}\right)$$

donde $B_{t-1}$ es la deuda nominal heredada, $P_t$ el nivel de precios, $\beta$ el factor de descuento y $T_{t+s} - G_{t+s}$ los superávits primarios futuros. Si la autoridad fiscal fija la senda de superávits con independencia de la deuda (política activa) y la monetaria no reacciona (política pasiva), un aumento del gasto sin ajuste fiscal futuro reduce el lado derecho. El ajuste recae entonces en el nivel de precios, que sube hasta reducir el valor real de la deuda y restaurar la igualdad.

La inflación es, así, un fenómeno fiscal, porque la senda de $T$ y $G$ y la deuda heredada $B_{t-1}$ determinan $P_t$ y, con él, $\pi$. La implicación de política económica es que la independencia del banco central es condición necesaria pero no suficiente, porque sin coordinación fiscal la estabilidad de precios no está garantizada, tensión que se encuentra en el centro del debate sobre la arquitectura de la UEM.

**Conclusiones**:

▪ *Aspectos positivos*: integra la política fiscal en la determinación del nivel de precios y explica por qué la estabilidad de precios exige que la política fiscal sea coherente con la monetaria. Proporciona, además, el marco con el que se analizan la madurez de la deuda y la coordinación de ambas políticas.

▪ *Limitaciones*: su supuesto central, que la restricción presupuestaria intertemporal del Gobierno actúa como condición de equilibrio y no como una restricción que debe cumplirse para cualquier nivel de precios, es objeto de controversia (Buiter, 2002). Asimismo, su contraste empírico es difícil, porque los regímenes de dominancia fiscal y monetaria pueden generar datos observacionalmente equivalentes.

## 3.3. Dominancia monetaria

En la región I se sitúan el argumento simétrico al de Sargent y Wallace y el debate institucional sobre cómo garantizar una política fiscal pasiva.

### 3.3.1. Aritmética fiscal desagradable (King, 1995)

**Origen**: King plantea el argumento simétrico e inverso al de Sargent y Wallace, en el contexto de los bancos centrales independientes con objetivos de inflación de los años noventa.

**Idea**: cuando la autoridad monetaria es activa y se compromete de forma creíble con un objetivo de inflación, queda disciplinada la autoridad fiscal, ya que la restricción presupuestaria intertemporal del Gobierno se convierte en una restricción efectiva sobre la política fiscal.

**Razonamiento**: si el Gobierno eleva el gasto sin subir impuestos, no puede esperar que la inflación licúe la deuda (el banco central no lo permitirá), por lo que los mercados anticipan un ajuste fiscal futuro inevitable. En este contexto, la autoridad monetaria fija tipos altos, lo que compromete la capacidad de financiación vía deuda de la autoridad fiscal. Si el ajuste no se materializa, la prima de riesgo soberana sube con fuerza.

Una política monetaria creíble puede forzar la disciplina fiscal, pero al precio de que la carga del ajuste recaiga sobre el presupuesto, con sus costes políticos y sociales (p. ej. crisis de deuda soberana en la periferia de la zona del euro, 2010-2014).

**Conclusiones**:

- *Aspectos positivos*: completa la aritmética de Sargent y Wallace con el caso simétrico de dominancia monetaria y muestra que un banco central independiente y creíble puede disciplinar la política fiscal. Ofrece así un fundamento a las reglas fiscales que acompañan a los regímenes de objetivos de inflación.

- *Limitaciones*: el resultado presupone que la autoridad fiscal acaba cediendo. Si no ajusta, el conflicto entre ambas autoridades puede desembocar en una crisis de deuda o en una presión sobre el banco central para que acomode la política fiscal, es decir, en una dominancia fiscal *de facto* que el argumento no excluye.

### 3.3.2. Debate sobre las reglas fiscales

**Reglas fiscales**: Kopits y Symansky (1998) definen una *regla fiscal* como una restricción permanente sobre la política fiscal expresada mediante un indicador sintético de los resultados presupuestarios. Su generalización acompaña a la independencia de los bancos centrales en los años noventa.

La dominancia monetaria funciona mejor cuando se complementa con reglas fiscales que anclan las expectativas sobre la sostenibilidad de la deuda y construyen espacio fiscal.

Entre las reglas fiscales se incluyen cuatro tipos:

1. *Reglas de saldo*: límites al déficit público, es decir, reglas de equilibrio presupuestario (p. ej. la regla chilena de superávit estructural).

2. *Reglas de deuda*: límites a la deuda pública para garantizar su sostenibilidad.

3. *Reglas de gasto*: techos de gasto (p. ej. la regla suiza o la regla de gasto española, que limita el crecimiento del gasto computable a la tasa de referencia de crecimiento del PIB a medio plazo, $g_G \leq \bar{g}_Y$).

4. *Reglas de ingresos*: modulación de los ingresos públicos.

Otra solución es la creación de instituciones fiscales independientes (*watchdogs*) como la Autoridad Independiente de Responsabilidad Fiscal (AIReF) en España (2013). El Pacto de Estabilidad y Crecimiento (PEC, 1997) y el Pacto Presupuestario Europeo (2012) son ejemplos de marcos institucionales diseñados para garantizar la coherencia entre una política monetaria activa (BCE) y una política fiscal pasiva (Estados miembros).

**Conclusiones**:

▪ *Aspectos positivos*: limitan la tendencia de la política discrecional a acumular déficits y refuerzan la credibilidad del compromiso con la sostenibilidad de la deuda, lo que permite a la política fiscal comportarse como pasiva y al banco central centrarse en la estabilidad de precios.

▪ *Limitaciones*: la rigidez de estas reglas ha sido cuestionada por su carácter procíclico. La reforma del PEC (2024) introduce un esquema en dos niveles: un ancla de sostenibilidad (que la ratio de deuda quede en una senda descendente verosímil al final del periodo de ajuste) y una senda de gasto neto como instrumento principal para lograrlo gradualmente. El esquema se aplica mediante planes fiscales nacionales a medio plazo.

> El marco de gobernanza fiscal de la UEM y su reforma se desarrollan en el tema B45, apartados 1.2 y 1.3.5, y la relación entre el BCE y las políticas fiscales nacionales, en el tema B43.

## 3.4. Otros desarrollos y evidencia empírica

Dos resultados cualifican el marco anterior. El primero acota lo que puede lograr el balance del banco central, y el segundo subraya el papel de la estructura de los pasivos públicos. La evidencia empírica sobre el régimen de dominancia vigente cierra el apartado.

### 3.4.1. Teorema de irrelevancia (Wallace, 1981)

**Origen**: el teorema de Wallace (1981), también conocido como *teorema de Modigliani-Miller de la banca central*, traslada al balance del banco central la lógica de la irrelevancia de la estructura financiera.

**Idea**: bajo ciertas condiciones restrictivas (mercados completos, ausencia de fricciones, agentes ricardianos con expectativas racionales e información perfecta), la composición del balance del banco central es irrelevante para el equilibrio real.

**Razonamiento**: las operaciones de mercado abierto (intercambio de bonos por dinero) no afectan a la senda futura de déficits, a la deuda pública neta, a la demanda agregada ni a los precios. La razón es que los agentes neutralizan la operación ajustando su propia cartera, de modo que la asignación de equilibrio no varía.

**Extensiones**: Eggertsson y Woodford (2003) extienden el resultado al ZLB, donde las compras de activos son irrelevantes si no alteran las expectativas sobre la senda futura de tipos. Por ello, la potencia de la política monetaria, convencional y no convencional, descansa precisamente en

el incumplimiento de las condiciones del teorema (fricciones financieras, restricciones de liquidez, segmentación de mercados e información imperfecta).

**Conclusiones**:

- *Aspectos positivos*: trata las operaciones de mercado abierto como cambios en la composición de los pasivos del sector público consolidado. Ofrece, además, un resultado de referencia, análogo al de Modigliani-Miller, frente al que se evalúa la eficacia de la política de balance del banco central.

- *Limitaciones*: la evidencia sobre los programas de compras de activos apunta a efectos sobre los tipos a largo plazo que no proceden solo de la señal sobre la senda futura de tipos (tema A40, apartado 3.3). Esos efectos sugieren que, en la práctica, la composición del balance sí puede importar.

### 3.4.2. Composición y madurez de los pasivos públicos (Cochrane, 2001)

**Origen**: Cochrane (2001) extiende el análisis de la interacción monetario-fiscal al estudiar cómo la composición de los pasivos del sector público (dinero, deuda a corto y deuda a largo plazo), y no solo su nivel, condiciona el régimen y la transmisión de los *shocks*.

**Idea**: bajo dominancia fiscal, un *shock* fiscal debe reducir el valor real de los pasivos públicos, pero, según la madurez de la deuda, el ajuste puede recaer sobre el nivel de precios corriente o sobre el precio de los bonos, que refleja la inflación futura.

**Razonamiento**: con deuda a corto plazo, la revaluación solo puede producirse vía un salto inmediato del nivel de precios. Además, una subida de tipos encarece de inmediato el servicio de la deuda, lo que hace al Gobierno más vulnerable a la dominancia fiscal. En cambio, con deuda a largo plazo, la expectativa de inflación futura reduce hoy el precio de los bonos, de modo que el ajuste se reparte en el tiempo y el presupuesto queda parcialmente aislado de las decisiones del banco central.

La gestión de la madurez media de la deuda es, por tanto, un instrumento de política económica en sí mismo, con implicaciones para la coordinación monetario-fiscal. Por ejemplo, los programas de compras de activos, al sustituir deuda a largo plazo por reservas a corto, acortan la madurez efectiva de los pasivos consolidados del sector público.

**Conclusiones**:

- *Aspectos positivos*: incorpora la estructura de vencimientos de la deuda al análisis de la interacción monetario-fiscal y explica por qué un mismo *shock* fiscal puede traducirse en inflación inmediata o repartida en el tiempo. Conecta así la gestión de la deuda pública con la política monetaria.

- *Limitaciones*: se apoya en la teoría fiscal del nivel de precios y comparte, por tanto, sus críticas.

### 3.4.3. Evidencia empírica

**Déficit estructural como determinante**: la relación empírica entre el déficit observado y la inflación no es concluyente. Lo relevante es, en cambio, el déficit estructural, que, cuando es persistente, tiende a acompañarse de un mayor crecimiento monetario y de más inflación.

Esa asociación es coherente con la aritmética monetarista desagradable, pues, cuando el déficit es persistente, la autoridad monetaria termina acomodando, es decir, se produce dominancia fiscal *de facto*.

**Independencia del banco central y monetización**: en los países de la OCDE tras Bretton Woods, la monetización del déficit fue más frecuente donde el banco central era menos independiente que donde lo era más, como en Alemania con el Bundesbank. En Italia, por ejemplo, el "divorcio" entre el Tesoro y la Banca d'Italia (1981), que suprimió la obligación de esta de adquirir los títulos no colocados en subasta, precedió a la desinflación de los años ochenta. En la misma línea, el Tratado de Maastricht prohibió la financiación monetaria del sector público (hoy artículo 123 del Tratado de Funcionamiento de la Unión Europea, TFUE).

En consecuencia, el grado de independencia del banco central es un factor determinante del régimen de dominancia, y cuanto mayor es la independencia, mayor es la probabilidad de dominancia monetaria.

> La evidencia sobre la interrelación de ambas políticas desde la Gran Recesión (crisis de deuda soberana, respuesta a la pandemia y ciclo de endurecimiento posterior) se desarrolla en el tema A40, apartado 3.4.

## 3.5. Conclusiones

El modelo de Leeper ordena las teorías de la dominancia. La aritmética monetarista desagradable y la teoría fiscal del nivel de precios describen la dominancia fiscal, mientras que la aritmética fiscal desagradable y las reglas fiscales describen la dominancia monetaria. La evidencia asocia una mayor independencia del banco central con una inflación media más baja (Alesina y Summers, 1993), lo que inclina el régimen hacia la dominancia monetaria. Un déficit estructural persistente puede, sin embargo, imponer la dominancia fiscal *de facto*, como predice la aritmética monetarista desagradable.

## Conclusiones

El déficit público exige financiación, y la forma de financiarlo no es neutral salvo bajo equivalencia ricardiana estricta, porque impuestos, deuda y monetización tienen efectos distintos sobre la renta, el tipo de interés y la riqueza. La sostenibilidad de la deuda depende de la diferencia $r - g_Y$ y de la respuesta del saldo primario a la deuda heredada, y la monetización alivia transitoriamente su carga, pero está acotada por el señoraje máximo y genera inflación. En el modelo de Leeper solo hay un equilibrio determinado si una autoridad es activa y la otra pasiva, y la aritmética monetarista desagradable muestra que, sin disciplina fiscal, el banco central no puede controlar la inflación. Las reglas fiscales y las instituciones independientes construyen el espacio fiscal que hace compatible la dominancia monetaria con la sostenibilidad de la deuda. Esa compatibilidad es especialmente difícil en la UEM, con una política monetaria única y veintiuna políticas fiscales nacionales.

La financiación del déficit prolonga el análisis de los efectos del gasto y de los impuestos (tema A38) y pone a prueba la equivalencia ricardiana (tema A06). La monetización enlaza con la demanda de dinero, cuya sensibilidad a la inflación esperada acota el señoraje (tema A35), y con la hiperinflación, que aparece cuando el Estado intenta superar ese máximo (tema A41). Por su parte, el modelo de Leeper y las teorías de la dominancia ordenan la interacción entre las autoridades monetaria y fiscal, y la experiencia posterior a 2008 permite contrastarlas (tema A40). Por último, la compatibilidad entre una política monetaria única y políticas fiscales nacionales es el problema central del marco fiscal de la UEM (tema B45).

Los largos periodos con tipos de interés inferiores al crecimiento ($r < g_Y$) han llevado a reconsiderar el coste de la deuda. Blanchard (2019) sostiene que, si el tipo seguro se mantiene por debajo del crecimiento, la deuda puede no tener coste fiscal y su coste en bienestar es menor de lo que se suponía. La fatiga fiscal documentada por Ghosh et al. (2013) implica en cambio un límite al endeudamiento (apartado 2.1.2). Sigue abierta también la validez de la teoría fiscal del nivel de precios. Buiter (2002) la considera mal especificada por su supuesto central (apartado 3.2.2), mientras que Bianchi, Faccini y Melosi (2023) la emplean para explicar parte de la inflación posterior a la pandemia (tema A40, apartado 3.4.3).

# A40. Interrelación de políticas macroeconómicas

Efectividad e interrelación de las políticas monetaria y fiscal en las principales economías desarrolladas desde la Gran Recesión de 2008. El valor de los multiplicadores fiscales.

## Introducción

La macroeconomía toma como objeto la economía en su conjunto y la analiza con modelos rigurosos en lo formal y capaces de explicar los datos. El análisis de coyuntura y la política económica se interesan, por un lado, por el **nivel de actividad** (la renta, la producción y su desviación respecto del potencial, o brecha de producción) y, por otro, por los **precios** (su nivel y su tasa de variación, la inflación). Explicar cómo se determinan ambas variables y por qué fluctúan es el núcleo del análisis macroeconómico.

Esas fluctuaciones son el objeto de la **teoría del ciclo económico**. En su concepción tradicional, el ciclo es una oscilación de la actividad que atraviesa cuatro fases (depresión, recuperación, auge y recesión). La concepción moderna, iniciada en los años setenta, lo entiende, en cambio, como la sucesión de desviaciones del producto respecto de su tendencia y estudia por qué se producen esas desviaciones, cuál es su intensidad y cómo se relacionan con el resto de variables. Las teorías del ciclo pueden clasificarse según el tipo de perturbación inicial o *impulso* (de oferta o de demanda, real o nominal) y, sobre todo, según los mecanismos de *propagación* que convierten esa perturbación en una fluctuación persistente. Por este segundo criterio, las teorías de equilibrio, herederas del pensamiento neoclásico, explican el ciclo como una sucesión de equilibrios, y las de desequilibrio, herederas del keynesiano, lo atribuyen a imperfecciones que impiden que los mercados se vacíen.

Históricamente, la observación de ciclos regulares se remonta a Clément Juglar (1862), y la primera mitad del siglo XX añadió los ciclos de distinta duración de Kitchin y Kondratiev y la interpretación de Schumpeter. Keynes (1936) situó la demanda agregada en el centro del análisis, y sobre esa base Samuelson (1939) y Hicks combinaron el multiplicador y el acelerador. Más tarde, Mundell y Fleming extendieron el modelo IS-LM a la economía abierta, y el modelo de oferta y demanda agregadas incorporó los precios. Desde los años setenta, la nueva macroeconomía clásica explicó el ciclo por perturbaciones monetarias mal percibidas (Lucas) o por perturbaciones reales de productividad (Kydland y Prescott, 1982), mientras que la nueva economía keynesiana subrayó el papel de las rigideces. De la convergencia de ambas corrientes surge el modelo de tres ecuaciones (IS, curva de Phillips y regla monetaria), hoy la referencia para el análisis de la política macroeconómica.

El análisis de la actividad y los precios plantea cuatro cuestiones:

1. *Determinación*: la actividad y los precios se determinan conjuntamente, por la interacción de la demanda y la oferta agregadas, que en una economía abierta depende también del sector exterior y del régimen cambiario (tema A32).

2. *Políticas*: sobre ambas variables actúan las políticas monetaria y fiscal (temas A36 a A39), cuya eficacia, medida por sus multiplicadores, depende de cómo se combinan y de la perturbación a la que responden (tema A40).

3. *Inflación*: el aumento sostenido de los precios tiene causas, persistencia y costes propios (tema A41).

4. *Ciclo*: las fluctuaciones de la actividad y de los precios en torno a su tendencia son el objeto de las teorías del ciclo, de los modelos del acelerador a las modernas, que se juzgan por su capacidad para reproducir los hechos estilizados (tema A42).

Las políticas monetaria y fiscal son los dos grandes instrumentos de la política de estabilización macroeconómica. La **política monetaria** actúa sobre la cantidad y el precio del dinero y está en manos de bancos centrales generalmente independientes, mientras que la **política fiscal** actúa sobre los ingresos y gastos públicos y depende de los Gobiernos y de los Parlamentos. Ambas inciden sobre la demanda agregada, pero lo hacen a través de canales, con retardos y bajo restricciones distintas.

Además, sus efectos no son independientes, porque el tipo de interés que fija el banco central condiciona el coste de la deuda pública. El déficit y la deuda que genera la política fiscal condicionan, a su vez, el margen de maniobra de la política monetaria. Por ello, su efectividad, que se mide sobre todo con el multiplicador, y su interrelación (la forma en que se combinan en el *policy mix*) son dos cuestiones inseparables.

El debate arranca con la síntesis neoclásica, que discutió la eficacia relativa de ambas políticas en el modelo IS-LM, y con Mundell y Fleming, que mostraron que depende del régimen cambiario y de la movilidad del capital. Tinbergen (1952) estableció que la política económica necesita tantos instrumentos como objetivos, y Mundell (1962) propuso asignar cada instrumento al objetivo sobre el que es relativamente más eficaz.

Más tarde, Sargent y Wallace (1981) advirtieron, con su "aritmética monetarista desagradable", que una política fiscal insostenible puede acabar imponiendo la monetización de la deuda. Leeper (1991) formalizó las combinaciones de políticas activas y pasivas que dan lugar a la dominancia monetaria o a la dominancia fiscal. Durante la Gran Moderación, en consecuencia, el consenso asignó la estabilización a la política monetaria y la disciplina a la fiscal.

La Gran Recesión de 2008 reabrió ese consenso. Con los tipos de interés en su límite inferior, la política monetaria recurrió a instrumentos no convencionales y la política fiscal recuperó protagonismo, mientras la evidencia sobre el valor de los multiplicadores en las recesiones se revisaba al alza. Poco después, la crisis de deuda soberana europea mostró los riesgos de la interacción entre bancos, Estados y banco central. Por último, la respuesta a la pandemia de 2020, con una coordinación sin precedentes entre ambas políticas, y el posterior repunte de la inflación han devuelto al centro del debate la cuestión de la dominancia fiscal.

El objetivo del tema es analizar esta cuestión en tres bloques:

- *Efectividad*: la de ambas políticas, articulada en torno al concepto de multiplicador.
- *Interrelación*: la que se da entre ambas autoridades.
- *Evidencia empírica*: lo que muestran los datos desde la Gran Recesión.

# 1. Análisis de la efectividad de las políticas monetaria y fiscal

Antes de estudiar su interrelación, la efectividad de cada política se mide por separado a través de su multiplicador, primero el monetario y después el fiscal.

## 1.1. Multiplicador monetario

Del multiplicador monetario interesan su concepto, que tiene dos acepciones, y las perspectivas teóricas sobre su valor.

### 1.1.1. Concepto

El término **multiplicador monetario** admite dos acepciones, una operativa, que lo liga a la base monetaria, y otra teórica, que lo liga a la renta de equilibrio.

**Enfoque operativo**: en sentido estricto, designa la relación entre la base monetaria y la oferta monetaria, que depende de los coeficientes de efectivo ($c_D = E/D$) y de reservas ($q$) del sistema bancario:

$$M^s = \frac{1 + c_D}{c_D + q} H \quad \rightarrow \quad m = f(c_D^-, q^-)$$

donde $H$ es la base monetaria, $E$ el efectivo en manos del público, $D$ los depósitos y $m$ el multiplicador monetario, el cociente que multiplica a la base.

Su relevancia se debe a que la autoridad controla de forma imperfecta la oferta monetaria, ya que el multiplicador depende del comportamiento del público y de los bancos, es procíclico y cae bruscamente en las crisis financieras como la de 2008 (tema A36, apartado 1.4).

**Enfoque teórico**: en el contexto de este tema, sin embargo, la acepción relevante es la de *multiplicador de la política monetaria*, es decir, la variación del *output* de equilibrio ante una variación unitaria de la oferta real de dinero. En el marco IS-LM de economía cerrada, con $v, k, h > 0$:

$$M_M \equiv \frac{\partial Y}{\partial (M^s/P)} = \frac{v/h}{1 - c(1-\tau) + v\frac{k}{h}} > 0 \quad \rightarrow \quad M_M \equiv f(v^+, h^-, k^-, c^+, \tau^-)$$

donde $c$ es la propensión marginal al consumo, $\tau$ el tipo impositivo, $v$ el grado en que la inversión responde al tipo de interés, y $k$ y $h$ la sensibilidad a la renta y al tipo de interés, respectivamente, de la demanda de dinero.

La cadena de transmisión es indirecta ($\uparrow M^s \rightarrow \downarrow i \rightarrow \uparrow I \rightarrow \uparrow Y$). De ahí que la efectividad de la política monetaria dependa de la fortaleza de cada eslabón:

- *Fortaleza de los eslabones*: la política monetaria es tanto más efectiva cuanto más sensible es la inversión al tipo de interés ($\uparrow v$, IS más plana en el plano $(Y, i)$) y cuanto menos lo es la demanda de dinero ($\downarrow h$, LM más rígida).

- *Casos extremos*: en la *trampa de liquidez* ($h \rightarrow \infty$), el multiplicador monetario es nulo, porque la expansión monetaria no reduce el tipo de interés. En el caso clásico ($h = 0$), en cambio, el multiplicador alcanza su valor máximo ($M_M = 1/k$, la velocidad de circulación), coherente con la teoría cuantitativa.

Gráficamente, en el plano renta-tipo de interés (figura A40.1), la curva IS recoge las combinaciones que equilibran el mercado de bienes y es decreciente, porque un tipo de interés menor eleva la inversión. La curva LM recoge las que equilibran el mercado de dinero y es creciente, porque una renta mayor eleva la demanda de dinero y, con la oferta dada, el tipo de interés. Una expansión monetaria ($\Delta M^s > 0$) desplaza la curva LM de $\text{LM}_0$ a $\text{LM}_1$ a lo largo de la curva IS. El tipo de interés baja de $i_0$ a $i_1$, la inversión aumenta y el equilibrio pasa del punto 0 al 1, con una renta mayor ($Y_1 > Y_0$).

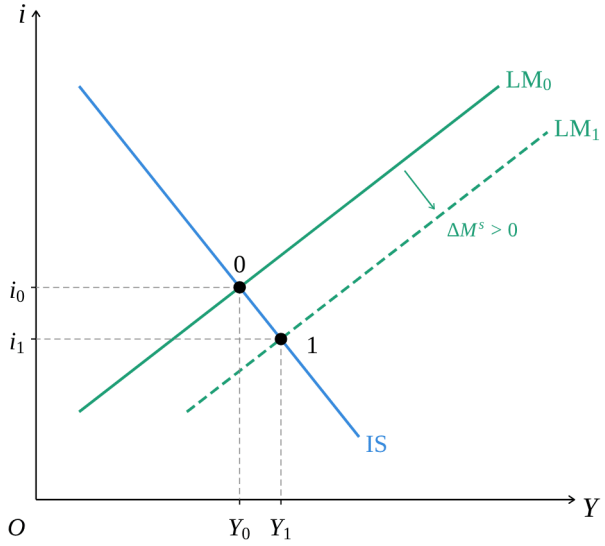



Figura A40.1: Multiplicador monetario en el IS-LM. La efectividad depende de las pendientes relativas de IS y LM.

### 1.1.2. Perspectivas teóricas sobre el multiplicador

**Multiplicador monetario en la visión keynesiana**: la política monetaria es poco efectiva. La cadena de transmisión es larga e incierta, porque la demanda de dinero es muy sensible al tipo de interés (motivo especulación, $h$ elevada, con la trampa de liquidez como caso límite). A ello se suma que la inversión responde poco al tipo de interés (dominada por los *animal spirits*, $v$ baja). La imagen clásica es que la política monetaria es como una cuerda, que sirve para tirar (frenar una expansión) pero no para empujar (estimular en una recesión). En consecuencia, los keynesianos priorizan la política fiscal.

**Multiplicador monetario en la visión de la teoría del ciclo real** (TCR): es nulo incluso a corto plazo. Con precios plenamente flexibles y agentes racionales, rige la dicotomía clásica, de modo que el dinero es neutral y los ciclos se explican por perturbaciones reales (tecnológicas). Por ello, la correlación observada entre dinero y *output* se reinterpreta como causalidad inversa (King y Plosser, 1984), en el sentido de que el dinero interno responde endógenamente a la actividad y no la causa.

**Multiplicador monetario en la visión de la nueva economía keynesiana** (NEK): esta escuela microfundamenta un multiplicador positivo a corto plazo mediante rigideces nominales (costes de menú, fijación escalonada de precios a la Calvo, salarios rígidos). Como los precios no se ajustan de inmediato, las variaciones del dinero y del tipo de interés nominal alteran el

tipo real y la demanda, aunque a largo plazo se recupera la neutralidad. Esta visión fundamenta el consenso de política previo a 2008, con la política monetaria como instrumento principal de estabilización, articulada mediante reglas de tipo de interés (principio de Taylor) en el marco IS-PC-MR, y la fiscal limitada a los estabilizadores automáticos.

**Multiplicador monetario y crecimiento**: a largo plazo, la cuestión es si la tasa de crecimiento monetario afecta a las variables reales (*superneutralidad*). Sidrauski (1967) demuestra la superneutralidad del dinero en un modelo de optimización con dinero en la función de utilidad, en el que el capital y el consumo de estado estacionario son independientes de la inflación. Tobin (1965), en cambio, obtiene un efecto real positivo (*efecto Tobin*), porque la inflación encarece mantener dinero y desplaza la cartera hacia capital, elevando el *stock* de capital por trabajador. La evidencia empírica matiza ambos resultados, pues la inflación elevada perjudica el crecimiento, con efectos no lineales a partir de ciertos umbrales.

## 1.2. Multiplicador fiscal

Como el monetario, el multiplicador fiscal se presenta a través de su concepto, por el lado del gasto y del ingreso, y de las perspectivas teóricas sobre su valor.

### 1.2.1. Concepto

El **multiplicador fiscal** mide la variación del *output* de equilibrio ante una variación unitaria del instrumento fiscal (gasto público, impuestos o transferencias). Conviene distinguir entre el multiplicador de impacto (efecto contemporáneo), el acumulado (cociente entre la variación acumulada del *output* y la del instrumento en un horizonte) y el de valor presente. Salvo indicación contraria, la evidencia empírica de la sección 3 se refiere al multiplicador acumulado del gasto.

**Perspectiva teórica del marco IS-LM, lado del gasto**:

$$M_G \equiv \frac{\partial Y}{\partial G_0} = \frac{1}{1 - c(1-\tau) + v\frac{k}{h}} > 0 \quad \rightarrow \quad M_G \equiv f(c^+, \tau^-, v^-, (k/h)^-)$$

El numerador recoge el impulso inicial del gasto, y la expresión en su conjunto refleja el mecanismo multiplicador keynesiano (el gasto genera renta, que genera consumo, que a su vez genera renta).

El denominador recoge, en cambio, las filtraciones que amortiguan ese mecanismo, que son el ahorro $(1-c)$, los impuestos $(c\tau)$ y el *efecto expulsión* (*crowding out*) monetario $(v\frac{k}{h})$. Por este último, el mayor nivel de renta incrementa la demanda de dinero y, con ella, el tipo de interés, lo que desplaza inversión privada. De ahí se siguen los determinantes del multiplicador y sus casos extremos:

- *Determinantes*: el multiplicador es mayor cuanto mayor es la propensión al consumo $(c)$ y menores el tipo impositivo $(\tau)$, la sensibilidad de la inversión $(v)$ y la pendiente de la LM $(k/h)$.

- *Casos extremos*: en la trampa de liquidez $(h \to \infty)$, el efecto expulsión desaparece y el multiplicador alcanza su máximo keynesiano $(M_G = 1/[1 - c(1-\tau)])$. En el caso clásico $(h = 0)$, por el contrario, el efecto expulsión es total y $M_G = 0$.

Gráficamente, en el plano renta-tipo de interés (figura A40.2), la IS recoge el equilibrio del mercado de bienes y la LM, el del mercado de dinero, con las pendientes de la figura A40.1. Un aumento del gasto público ($\Delta G_0 > 0$) desplaza la curva IS de $\text{IS}_0$ a $\text{IS}_1$ a lo largo de la curva LM. Con el tipo de interés constante en $i_0$, la renta llegaría a $Y_{0'}$ (punto $0'$), pero el alza del tipo hasta $i_1$ expulsa inversión privada y el equilibrio pasa del punto 0 al 1. La diferencia entre $Y_{0'}$ e $Y_1$ mide el efecto expulsión.

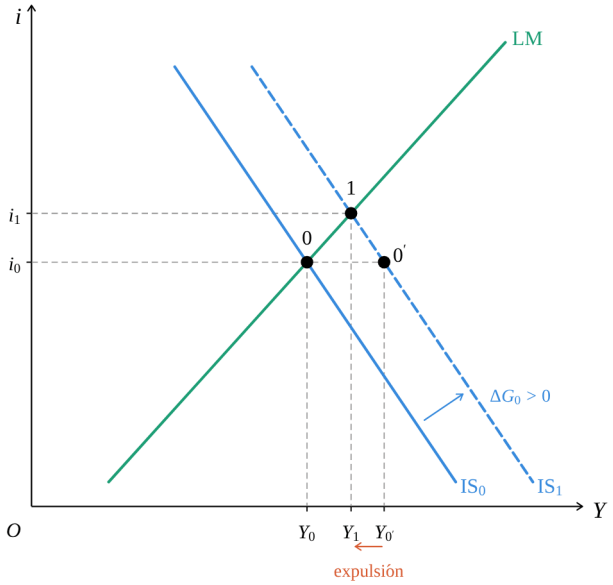



Figura A40.2: Multiplicador fiscal en el IS-LM. La efectividad depende de las pendientes relativas de IS y LM.

**Perspectiva teórica del marco IS-LM, lado del ingreso**:

$$M_\tau \equiv \frac{\partial Y}{\partial \tau} = \frac{-cY}{1 - c(1-\tau) + v\frac{k}{h}} < 0 \quad \rightarrow \quad M_\tau \equiv f(c^-, \tau^+)$$

Una subida del tipo impositivo reduce la renta disponible y, con ella, el consumo y la renta de equilibrio. Cabe hacer dos observaciones:

- *Menor valor absoluto*: para un impuesto de suma fija ($T_0$), el multiplicador impositivo es menor en valor absoluto que el del gasto:

$$M_T \equiv \frac{\partial Y}{\partial T_0} = \frac{-c}{1 - c(1-\tau) + v\frac{k}{h}}, \qquad |M_T| < M_G$$

La razón es que los impuestos operan indirectamente, a través del consumo (una parte $1-c$ de la variación de renta disponible se filtra al ahorro), mientras que el gasto impacta directamente sobre la demanda.

▪ *Presupuesto equilibrado*: de esta asimetría se deriva el *multiplicador del presupuesto equilibrado* (Haavelmo, 1945). En el modelo keynesiano simple (sin impuestos proporcionales ni efecto expulsión), un aumento del gasto financiado íntegramente con impuestos de suma fija tiene un multiplicador exactamente igual a la unidad, pues el efecto directo del gasto domina al efecto renta de los impuestos.

### 1.2.2. Perspectivas teóricas sobre el multiplicador

**Multiplicador fiscal en la visión keynesiana**: es elevado (superior a la unidad) y la política fiscal es el instrumento principal de estabilización. Los supuestos que lo sostienen son la capacidad ociosa (la oferta acomoda la demanda sin presiones de precios), el consumo dependiente de la renta corriente (propensiones estables, expectativas estáticas) y un efecto expulsión limitado (LM plana, especialmente en recesión).

**Multiplicador fiscal en la visión TCR**: es reducido (inferior a la unidad a corto plazo) o nulo. Dos mecanismos lo fundamentan:

▪ *Equivalencia ricardiana* (Barro, 1974): con agentes optimizadores de horizonte infinito (altruismo intergeneracional), expectativas racionales, mercados completos e impuestos de suma fija, los hogares anticipan los impuestos futuros que implica el déficit público y lo compensan con un mayor ahorro privado. Financiar el gasto con deuda o con impuestos es, por tanto, equivalente, y el consumo no reacciona a la forma de financiación.

▪ *Canal neoclásico del gasto*: un aumento permanente de $G$ genera un efecto riqueza negativo (mayores impuestos presentes o futuros) que eleva la oferta de trabajo. El *output* aumenta, pero menos que el gasto y a costa de una caída del consumo privado, de modo que el multiplicador es positivo pero, a corto plazo, inferior a la unidad. A largo plazo, en cambio, puede superarla por la acumulación de capital. Además, si los impuestos son distorsionantes, la financiación del gasto reduce la producción potencial.

**Multiplicador fiscal en la visión NEK**: es *estado-dependiente*. Con rigideces nominales depende de la respuesta de la política monetaria, y es reducido con una regla de Taylor activa y elevado en el límite inferior de los tipos de interés (*zero lower bound*, ZLB), según el modelo del apartado 2.1. Además, la existencia de hogares no ricardianos con restricciones de liquidez (consumidores *rule-of-thumb*) refuerza el canal keynesiano del consumo y eleva el multiplicador.

**Multiplicador fiscal y crecimiento**: a largo plazo, el efecto de la política fiscal depende de su composición y del nivel de deuda. Por ejemplo, la inversión pública productiva puede elevar la productividad del sector privado y la producción potencial (Aschauer, 1989), con multiplicadores de largo plazo superiores a los del gasto corriente.

La hipótesis de las *consolidaciones fiscales expansivas* (Giavazzi y Pagano, 1990) sostiene que un ajuste fiscal creíble puede ser expansivo vía expectativas (menores impuestos futuros, menor prima de riesgo). La evidencia posterior del FMI concluye, sin embargo, que las consolidaciones son, en general, contractivas a corto plazo.

Por último, el debate sobre deuda y crecimiento ilustra la dificultad de establecer relaciones robustas entre posición fiscal y crecimiento a largo plazo. Así, Reinhart y Rogoff (2010) asociaron una deuda superior al 90 % del PIB con un crecimiento sensiblemente menor, un resultado que cuestionaron Herndon et al. (2013).

## 1.3. Conclusiones

La principal conclusión es que la magnitud de los multiplicadores, y con ella la efectividad de las políticas monetaria y fiscal, no es absoluta, sino condicional. Depende del régimen cambiario, el grado de movilidad del capital, la posición cíclica, la formación de expectativas y el marco institucional.

- *Política monetaria*: en condiciones normales (economía cerrada o abierta con tipo flexible y libre movilidad de capital), es el instrumento principal de estabilización (Mundell-Fleming, principio de Taylor). Sin embargo, en el ZLB la política monetaria convencional pierde eficacia y la política fiscal recupera protagonismo, un resultado coherente tanto con el modelo IS-LM (trampa de liquidez) como con los modelos de equilibrio general dinámico y estocástico (DSGE) de la NEK.

- *Política fiscal*: su efectividad depende, a su vez, de las expectativas.

  - *Agentes ricardianos*: con equivalencia ricardiana, el multiplicador de una rebaja impositiva financiada con deuda tiende a cero.

  - *Agentes no ricardianos*: sin embargo, si existen restricciones de liquidez, agentes no ricardianos o rigideces nominales, el multiplicador es positivo y potencialmente superior a la unidad.

En consecuencia, la pregunta relevante no es cuál de las dos políticas es más efectiva "en abstracto", sino cuál es la combinación óptima (*policy mix*) dadas las condiciones del momento. Mundell (1962) formuló esta cuestión como el **principio de asignación**, y Leeper (1991) la formalizó como la necesidad de coordinación activa-pasiva entre ambas autoridades.

## 2. Análisis de la interrelación de las políticas monetaria y fiscal

Analizados sus efectos por separado, queda por ver cómo se interrelacionan ambas políticas. Lo hacen en dos planos:

- *Plano a corto plazo*: a través de la respuesta de la política monetaria al impulso fiscal, que el modelo IS-PC-MR permite analizar y que se vuelve crítica en el límite inferior de los tipos de interés.

- *Plano intertemporal*: a través de la restricción presupuestaria del sector público, que vincula a ambas autoridades y da lugar a los regímenes de dominancia.

### 2.1. Interrelación a corto plazo: el modelo IS-PC-MR y el límite inferior de los tipos de interés

> El modelo IS-PC-MR en economía cerrada se desarrolla en el tema A07, apartado 2.1, y su extensión a la economía abierta, en el tema A32, apartado 3. Aquí se retiene lo que interesa a la interrelación de políticas.

**Origen**: el modelo es el canónico de tres ecuaciones de la nueva economía keynesiana, en la versión con límite inferior de los tipos de interés de Eggertsson y Woodford (2003).

**Idea**: la efectividad de la política fiscal depende del régimen de la política monetaria. Es baja cuando esta es activa, pues el banco central contrarresta el impulso fiscal, y alta cuando está restringida por el límite inferior de los tipos de interés.

**Planteamiento**: el modelo de tres ecuaciones de la nueva economía keynesiana describe la economía con una IS dinámica, una curva de Phillips neokeynesiana y una regla de tipo de interés:

$$\tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta}(i_t - \mathbb{E}_t[\pi_{t+1}] - \bar{r}_t) + g_t$$

$$\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \kappa \tilde{y}_t, \qquad \kappa > 0,\ \beta \in (0,1)$$

$$i_t = \bar{i}_t + \alpha_\pi \tilde{\pi}_t + \alpha_y \tilde{y}_t, \qquad \alpha_\pi > 1,\ \alpha_y \geq 0$$

donde $\tilde{y}_t$ es la brecha de producción (*output gap*), $\theta$ el coeficiente de aversión relativa al riesgo, $\bar{r}_t$ el tipo de interés real natural, $g_t$ el impulso fiscal, $\beta$ el factor de descuento y $\kappa$ la pendiente de la curva de Phillips. En la regla, $\bar{i}_t = \bar{r}_t + \pi^T$ es el tipo nominal neutral y $\tilde{\pi}_t = \pi_t - \pi^T$ la desviación de la inflación respecto de su objetivo, $\pi^T$. Los coeficientes $\alpha_\pi$ y $\alpha_y$ miden la respuesta del tipo de interés a la inflación y a la brecha, con $\alpha_\pi > 1$ (principio de Taylor)[1].

La política fiscal entra como un desplazamiento de la IS ($g_t$) y, en su caso, como una variación del tipo de interés natural, $\bar{r}_t$.

**Estática comparativa**: el efecto de un impulso fiscal difiere en condiciones normales y en el límite inferior de los tipos de interés:

[1] La regla de Leeper del apartado 2.2, $i_t = \alpha_0 + \alpha\pi_t + \varepsilon_t^m$, es una regla de Taylor sin término de brecha ($\alpha_y = 0$), en la que $\alpha_0 = \bar{r} + (1-\alpha_\pi)\pi^T$ recoge la constante y $\alpha = \alpha_\pi$. Por ello, la condición de política monetaria activa de Leeper, $|\alpha\beta| > 1$, equivale aproximadamente al principio de Taylor, pues $\beta$ es próximo a la unidad.

▪ *Interrelación en condiciones normales*: ante un impulso fiscal ($\uparrow g_t$), la brecha de producción y la inflación aumentan y el banco central, siguiendo su regla, eleva el tipo de interés nominal más que proporcionalmente a la inflación, de modo que el tipo real sube y contrarresta parte del impulso. El multiplicador fiscal es, por tanto, reducido (inferior a la unidad), porque la política monetaria activa contrarresta en parte a la fiscal (véase la figura del tema A38, apartado 1.3.2).

▪ *El problema del límite inferior de los tipos de interés (ZLB)*: un *shock* negativo de demanda puede ser lo bastante grande como para que el tipo de interés natural caiga por debajo de $-\pi^T$ ($\bar{i}_t = \bar{r}_t + \pi^T < 0$). En ese caso, la regla de Taylor prescribe un tipo nominal negativo (véase la figura del tema A37, apartado 3.5.1). Sin embargo, el banco central no puede fijarlo, pues $i_t \geq 0$ (los agentes pueden mantener efectivo).

En consecuencia, el tipo real, $i_t - \mathbb{E}_t[\pi_{t+1}]$, queda por encima del natural, la brecha de producción es negativa y la inflación cae. La menor inflación esperada eleva, a su vez, el tipo real y agrava la brecha, en una espiral deflacionista. En este régimen, la interrelación se invierte:

1. *La política monetaria convencional pierde eficacia*: el banco central solo puede actuar sobre las expectativas, mediante el *forward guidance* (orientación prospectiva), o sobre los tipos a largo plazo y las primas de riesgo, mediante compras de activos, como se desarrolla en el apartado 3.3. El *forward guidance* es el compromiso de mantener el tipo en cero más allá de lo que la regla prescribiría.

2. *El multiplicador fiscal es elevado*: el tipo nominal no sube ante el impulso fiscal, y la mayor inflación esperada que este genera reduce el tipo real, con lo que el efecto expulsión desaparece y se convierte en un efecto atracción (*crowding in*).

Christiano, Eichenbaum y Rebelo (2011) obtienen en el ZLB multiplicadores muy superiores a la unidad (unos 3,7 en su modelo simple y 2,3 en su modelo de escala media), frente a valores en torno a la unidad o inferiores fuera de él. En general, el multiplicador supera la unidad siempre que el estímulo se limite al periodo en que el tipo permanece en el ZLB.

3. *La coordinación se vuelve necesaria*: la combinación de una expansión fiscal con un compromiso monetario de no reaccionar a la inflación resultante es la forma más efectiva de escapar del ZLB. Esa combinación difumina la frontera entre ambas políticas y anticipa la cuestión de la dominancia.

**Conclusiones**:

▪ *Aspectos positivos*: explica por qué la política fiscal volvió a ser un instrumento de estabilización cuando la Gran Recesión y la pandemia situaron a las principales economías desarrolladas en el límite inferior de los tipos durante más de una década.

▪ *Limitaciones*: descansa en dos simplificaciones:

- *Límite inferior*: supone $i_t \geq 0$, cuando el límite efectivo se sitúa algo por debajo de cero.
- *Política fiscal exógena*: entra como un desplazamiento de la IS, sin restricción presupuestaria intertemporal del sector público, por lo que el modelo no puede analizar la sostenibilidad del impulso fiscal ni los regímenes de dominancia, que se estudian a continuación.

## 2.2. Interrelación intertemporal: el modelo de Leeper (1991)

El modelo de Leeper (1991) cubre esa carencia al analizar la interrelación intertemporal de ambas políticas, a través de la restricción presupuestaria del sector público que vincula a las dos autoridades.

> El modelo de Leeper se desarrolla en el tema A39, apartado 3.1, y aquí se retiene su resultado central.

**Origen**: Leeper (1991) formaliza la distinción de Sargent y Wallace (1981) entre dominancia monetaria y fiscal en un modelo dinámico con expectativas racionales en el que cada autoridad sigue una regla de política.

**Idea**: la restricción presupuestaria intertemporal del sector público vincula a la autoridad fiscal y a la monetaria.

**Planteamiento**: Leeper modela consumidores con expectativas racionales y dinero en la función de utilidad y un sector público que financia el déficit vía deuda o vía base monetaria. La autoridad monetaria fija el tipo de interés en función de la inflación corriente, y la autoridad fiscal, los impuestos en función de la deuda heredada:

$$i_t = \alpha_0 + \alpha\pi_t + \varepsilon_t^m$$

$$T_t = \gamma_0 + \gamma_1 b_{t-1} + \varepsilon_t^f$$

donde $\alpha$ es la sensibilidad del tipo de interés a la inflación, $\gamma_1$ la de los impuestos al *stock* real de deuda heredado, $b_{t-1}$; $\alpha_0$ y $\gamma_0$, las constantes de cada regla; y $\varepsilon_t^m$ y $\varepsilon_t^f$ los *shocks* monetario y fiscal.

Cada autoridad es *activa* si fija su instrumento con independencia de la restricción presupuestaria y *pasiva* si lo ajusta para garantizar su cumplimiento.

**Equilibrio**: en el plano de los dos parámetros de respuesta, expresados como $|\alpha\beta|$ y $|\beta^{-1} - \gamma_1|$, con $\beta$ el factor de descuento (figura A40.3), el modelo delimita cuatro regiones, de las cuales solo dos generan un equilibrio determinado:

- *Dominancia monetaria*: la política monetaria es activa ($|\alpha\beta| > 1$) y la fiscal, pasiva ($|\beta^{-1} - \gamma_1| < 1$). La inflación la determina el banco central y los impuestos estabilizan la deuda.

- *Dominancia fiscal*: la política fiscal es activa y la monetaria, pasiva. El nivel de precios se ajusta para que la restricción presupuestaria se cumpla.

En las otras dos regiones no hay un equilibrio determinado. Si ambas autoridades son activas, la deuda es explosiva, y si ambas son pasivas, el equilibrio está indeterminado. La estabilidad macroeconómica requiere, por tanto, que exactamente una autoridad sea activa, de modo que la coordinación no es opcional sino un requisito de la determinación del equilibrio.

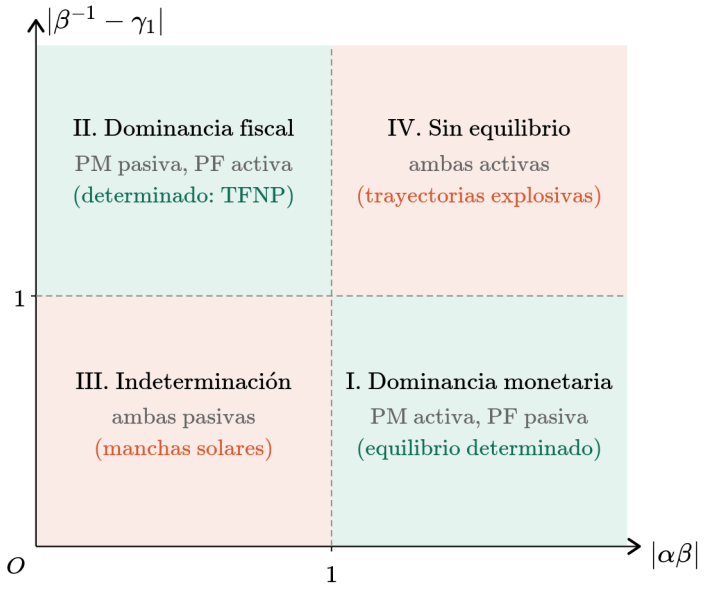



Figura A40.3: Modelo de Leeper (1991). En el plano de $|\alpha\beta|$ y $|\beta^{-1} - \gamma_1|$, las fronteras en 1 delimitan cuatro regiones. Solo la dominancia monetaria (I, política monetaria activa y fiscal pasiva) y la dominancia fiscal (II, política fiscal activa y monetaria pasiva) dan un equilibrio determinado. Si ambas son pasivas (III), el equilibrio está indeterminado, y si ambas son activas (IV), las trayectorias son explosivas.

**Conclusiones**:

▪ *Aspectos positivos*: demuestra que la estabilidad de precios depende de la combinación de reglas monetarias y fiscales, y no solo del banco central, y ofrece una taxonomía clara de regímenes en la que se apoya la teoría fiscal del nivel de precios.

▪ *Limitaciones*: supone reglas lineales, una economía cerrada y un agente representativo ricardiano, por lo que no recoge las restricciones de liquidez que reforzarían la dominancia fiscal.

## 2.3. Teorías de dominancia monetaria y fiscal

Al modelo de Leeper le siguen las teorías de la dominancia fiscal y de la dominancia monetaria, junto con otros desarrollos.

> Las teorías de la dominancia se desarrollan en el tema A39, apartados 3.2 a 3.4.

Cada una de ellas explica situaciones propias de una región del modelo de Leeper:

1. *Dominancia fiscal*: la *aritmética monetarista desagradable* (Sargent y Wallace, 1981) muestra que, si la autoridad fiscal no ajusta sus déficits, una política monetaria restrictiva hoy exige más deuda y más monetización mañana. Como los agentes lo anticipan, la inflación surge hoy como fenómeno fiscal y la autoridad monetaria es incapaz de controlarla.

La *teoría fiscal del nivel de precios* (Leeper, 1991; Sims, 1994; Woodford, 1995) generaliza el argumento, pues los precios se ajustan para que el valor real de la deuda nominal iguale al valor actual de los superávits primarios esperados.

2. *Dominancia monetaria*: la *aritmética fiscal desagradable* (King, 1995) es el caso opuesto, en el que un banco central comprometido de forma creíble con la estabilidad de precios fija tipos altos que comprometen la capacidad de financiación de la autoridad fiscal y fuerzan el ajuste presupuestario. El debate sobre las reglas fiscales (el Pacto de Estabilidad y Crecimiento, PEC, y su reforma de 2024) es su corolario institucional.

3. *Otros desarrollos*: el *teorema de irrelevancia* (Wallace, 1981) establece que, con mercados perfectos e información perfecta, las operaciones de mercado abierto no afectan a la senda de déficits, a la deuda neta ni a los precios, resultado que Eggertsson y Woodford (2003) extienden al ZLB. De ello se sigue que la potencia de la política monetaria descansa en el incumplimiento de esos supuestos.

Por su parte, el análisis de la madurez de los pasivos públicos (Cochrane, 2001) muestra que, bajo dominancia fiscal, el ajuste a un *shock* fiscal recae sobre el nivel de precios corriente si la deuda es a corto plazo. Si es a largo plazo, en cambio, el ajuste se reparte en la inflación futura a través del precio de los bonos.

El teorema de irrelevancia y el análisis de la madurez razonan con el balance consolidado del banco central y el Tesoro, una perspectiva que, desde 2022, ha tenido consecuencias prácticas sobre el presupuesto (apartado 3.4.4).

## 2.4. Conclusiones

La interacción entre política monetaria y fiscal constituye uno de los debates centrales de la macroeconomía moderna, y opera en dos planos. Los resultados de ambos planos tienen implicaciones directas para el diseño institucional de uniones monetarias como la Unión Económica y Monetaria (UEM). En ella, la independencia del BCE coexiste con veintiuna políticas fiscales nacionales y un marco de reglas fiscales en permanente evolución (véanse los temas B43 y B45).

# 3. Evidencia empírica desde la Gran Recesión: el valor de los multiplicadores y la interrelación de las políticas

Queda por contrastar con los datos lo que predice la teoría. Tras repasar las técnicas de medición, se recogen las estimaciones de los multiplicadores fiscal y monetario y la evidencia sobre la interrelación entre ambas políticas desde la Gran Recesión.

## 3.1. Técnicas de medición

**Técnicas basadas en modelos de vectores autorregresivos** (VAR): los VAR estructurales (SVAR) estiman el multiplicador fiscal y monetario a partir de las funciones de impulso-respuesta ante *shocks* identificados. El enfoque estándar (Blanchard y Perotti, 2002) explota los retardos institucionales de la política fiscal para identificar *shocks* exógenos de gasto e impuestos, bajo el supuesto de que el gasto público no reacciona contemporáneamente al *output*.

Esta identificación ha sido cuestionada porque los agentes anticipan los cambios fiscales, de modo que los *shocks* VAR están mal fechados y sesgan las respuestas estimadas (en particular, la del consumo privado). Ramey (2011) lo muestra al comprobar que tanto las previsiones profesionales como los *shocks* narrativos anticipan (causan en el sentido de Granger) los *shocks* VAR.

En política monetaria, la identificación canónica se debe a Christiano, Eichenbaum y Evans (1999). Su descomposición de Cholesky ordena el tipo de interés detrás del *output* y los precios y delante de los agregados de reservas y de dinero (la política monetaria reacciona contemporáneamente al *output* y los precios, pero no a la inversa). La principal limitación del enfoque es que los resultados son sensibles al esquema de identificación, al número de retardos y a la muestra temporal.

**Técnicas basadas en narrativas**: el enfoque narrativo identifica cambios exógenos de política a partir de registros históricos (actas de reuniones de política monetaria, discursos, documentos legislativos). Busca así aislar las decisiones motivadas por objetivos de largo plazo (p. ej. consolidación fiscal, lucha contra la inflación) de las respuestas endógenas al ciclo.

En el ámbito fiscal, Romer y Romer (2010) identifican cambios impositivos exógenos en EE. UU. (1945-2007) y estiman un multiplicador impositivo de alrededor de 3, es decir, cada dólar de subida de impuestos reduce el PIB en tres dólares a los dos años y medio. En el ámbito monetario, el mismo enfoque permite identificar *shocks* de política monetaria a partir de los registros y las previsiones internas de la Fed.

- *Ventaja*: evita los problemas de endogeneidad de los VAR.

- *Limitación*: subjetividad en la clasificación de los episodios y dificultad de replicación en países con menor tradición documental.

**Técnicas basadas en experimentos naturales**: explotan variaciones exógenas de política generadas por circunstancias institucionales o geográficas. Destacan tres ejemplos:

- *Gasto militar regional*: la distribución geográfica del gasto militar en EE. UU. permite estimar multiplicadores fiscales regionales (Nakamura y Steinsson, 2014).

- *Disolución de ayuntamientos*: la que se produce en Italia por infiltración mafiosa se ha utilizado como *shock* exógeno de gasto.

- *Trilema en un panel histórico*: en el ámbito monetario, Jordà, Schularick y Taylor (2015) explotan el trilema de Mundell-Fleming (trinidad imposible) con datos de 17 economías avanzadas desde 1870. Lo usan para estudiar los efectos de la política monetaria sobre el crédito hipotecario y los precios de la vivienda. Según el trilema, los países con tipo fijo y libre movilidad de capital importan la política monetaria del país ancla.

El enfoque presenta una ventaja y una limitación:

- *Ventaja*: identificación causal robusta.

- *Limitación*: validez externa limitada, pues los resultados locales no siempre son extrapolables al agregado nacional.

### 3.2. Evidencia empírica sobre el multiplicador fiscal

La Gran Recesión reavivó el debate sobre el tamaño del multiplicador fiscal, que había perdido centralidad durante la Gran Moderación. Los principales estudios, ordenados por técnica, son los siguientes:

**Enfoque narrativo** (Ramey y Shapiro, 1998; Ramey, 2011): identifica los episodios de aumento exógeno del gasto militar en EE. UU. (Corea, 1950-1953; Vietnam, 1965-1975; rearme de Carter-Reagan, 1980-1984) a partir de fuentes periodísticas y estima su efecto sobre el PIB. Ramey (2011) lo amplía con una serie de noticias que mide, desde 1939, el valor actual esperado del gasto militar futuro, de modo que el *shock* se fecha cuando se anuncia y no cuando se ejecuta.

Los multiplicadores del gasto así obtenidos se sitúan entre 0,6 y 1,2 según la muestra. El gasto militar eleva el *output*, pero el consumo privado y los salarios reales caen, en línea con el canal neoclásico del efecto riqueza negativo.

**Enfoque SVAR** (Blanchard y Perotti, 2002): con la identificación ya descrita y datos de EE. UU. (1960-1997)[2], en la especificación con tendencia determinista obtienen un multiplicador del gasto $M_G \in [0{,}84;\ 1{,}29]$ y un multiplicador impositivo $M_T \in [-0{,}78;\ -0{,}7]$. Con tendencia estocástica, el efecto máximo de los impuestos alcanza $M_T \approx -1{,}33$.

A diferencia del enfoque narrativo, el gasto público eleva el consumo privado (efecto multiplicador keynesiano) pero reduce la inversión privada (efecto expulsión). Lo difícil de conciliar con la teoría keynesiana estándar es que también las subidas de impuestos reduzcan la inversión.

**Enfoque de VAR de transición suave** (STVAR, Auerbach y Gorodnichenko, 2012): los modelos de cambio de régimen (*smooth transition VAR*) permiten que el multiplicador varíe con la posición cíclica y encuentran que el del gasto es mucho más expansivo en recesiones que en expansiones. Con datos trimestrales de EE. UU. de posguerra, alcanza 2,2-2,5 en recesión y no pasa de 0,6 en expansión. Este resultado cobró especial relevancia en el debate sobre los estímulos fiscales de la Gran Recesión. La explicación es que, en recesiones, la capacidad ociosa reduce el efecto expulsión y las restricciones de liquidez amplifican el canal keynesiano.

**Enfoque de proyecciones locales** (Ramey y Zubairy, 2018): cuestiona ese resultado con datos trimestrales de EE. UU. de 1889-2015, que incluyen varias guerras y recesiones profundas, y con regresiones separadas para cada horizonte. Con los *shocks* de noticias militares, el multiplicador se sitúa en 0,6-0,7 tanto con desempleo alto como bajo, y con los de Blanchard y Perotti queda también por debajo de la unidad.

[2] Los *shocks* impositivos se identifican con elasticidades externas de los impuestos al ciclo, que depuran el componente automático de la recaudación.

Los autores atribuyen la discrepancia sobre todo a que Auerbach y Gorodnichenko suponen que la economía sigue en recesión durante los 20 trimestres del horizonte, cuando en los datos de Auerbach y Gorodnichenko las recesiones duran 3,3 trimestres de media. Con esos mismos datos, el multiplicador a cinco años en recesión baja a 0,84. En el límite inferior de los tipos, en cambio, la evidencia es menos concluyente. Excluidos los años de racionamiento de la Segunda Guerra Mundial, algunas especificaciones dan hasta 1,5.

**Enfoque de metarregresiones** (Gechert y Will, 2012): a partir de una muestra de 89 estudios, cuyas estimaciones del multiplicador presentan una dispersión muy elevada, concluyen que el tamaño del multiplicador depende sobre todo del contexto y del método elegidos. La dispersión se explica por el tipo de instrumento, la técnica de estimación y la posición cíclica. Por instrumentos, la inversión pública presenta los multiplicadores más elevados, y las transferencias y los impuestos, los más bajos.

**Enfoque SVAR del FMI**: con VAR de cambio de régimen, el FMI ha estimado multiplicadores para EE. UU., Japón y Europa y ha encontrado también que el multiplicador es mayor en recesiones que en expansiones, especialmente en Japón (situación de ZLB prolongada). Además, ha propuesto un procedimiento en dos pasos para calibrar el multiplicador de un país según sus características estructurales (apertura, rigidez laboral, tamaño de los estabilizadores automáticos, régimen cambiario, nivel de deuda) y su posición cíclica.

**Enfoque de panel internacional** (Ilzetzki, Mendoza y Végh, 2013): con datos trimestrales de 44 países (1960-2007), muestran que el multiplicador del consumo público depende de las características del país:

- *Desarrollo*: es mayor en los países de renta alta que en los países en desarrollo.

- *Régimen cambiario*: es relativamente grande con tipo de cambio predeterminado y nulo con tipo flexible, como predice el modelo de Mundell-Fleming (tema A32, apartado 1.3).

- *Apertura*: es menor en las economías más abiertas, por la filtración del impulso hacia las importaciones.

- *Endeudamiento*: es negativo a largo plazo cuando la deuda supera el 60 % del PIB.

**Enfoque regional** (Nakamura y Steinsson, 2014): aprovechan que las movilizaciones militares elevan el gasto de defensa más en unos estados de EE. UU. que en otros y estiman un *multiplicador relativo de economía abierta* en torno a 1,5 (aumento del *output* de un estado respecto de los demás por dólar de gasto adicional).

El multiplicador relativo no es el multiplicador agregado, porque la respuesta de la política monetaria y de la fiscal federal, común a todos los estados, se cancela al comparar regiones, mientras que el multiplicador agregado depende de cuánto se oponga esa respuesta al impulso. Por eso el multiplicador relativo informa sobre el estímulo fiscal dentro de una unión monetaria y respalda los modelos en los que los *shocks* de demanda tienen efectos grandes (tema A38, apartado 1.3.2).

**Análisis del valor del multiplicador según el contexto**: de los estudios anteriores se desprende que el multiplicador depende críticamente del contexto:

1. *Régimen monetario*: Christiano, Eichenbaum y Rebelo (2011) demuestran teóricamente que el multiplicador puede superar ampliamente la unidad en el ZLB, pues la política monetaria no contrarresta la expansión fiscal (apartado 2.1).

2. *Tipo de instrumento*: Coenen et al. (2012) comparan siete modelos DSGE de bancos centrales y organismos internacionales. Encuentran que el multiplicador de gasto, en torno a 1,0-1,5 en el primer año, supera sistemáticamente al de transferencias y al de impuestos, que rondan, respectivamente, 0,3-0,8 y 0,2-0,5.

3. *Debate FMI*: Blanchard y Leigh (2013) estimaron que el FMI había subestimado sistemáticamente los multiplicadores en sus previsiones de crecimiento para las economías europeas al comienzo de la crisis (2010-2011). El FMI había utilizado valores de unos 0,5, cuando los reales se situaban entre 0,9 y 1,7. De hecho, los programas de austeridad aplicados en Grecia, Portugal, Irlanda y España (2010-2014) generaron contracciones del PIB superiores a las previstas, coherentes con multiplicadores elevados en contextos de recesión, ZLB y restricciones crediticias simultáneas.

**Persistencia del multiplicador**: la evidencia sobre cuánto duran los efectos indica que la persistencia del multiplicador no coincide con la del impulso fiscal. En particular, se observan tres resultados:

- *Perfil temporal*: el efecto sobre el *output* sigue una senda en forma de U invertida, con un máximo entre el primer y el segundo año y una duración máxima en torno a los cinco años.

- *Histéresis*: cuando existe, el efecto es mayor, pues la recesión evitada no destruye capital ni empleo de forma permanente (DeLong y Summers, 2012).

- *Acomodación monetaria*: el efecto es mayor cuando la política monetaria acompaña al impulso fiscal, en lugar de contrarrestarlo.

La tabla A40.1 resume la evidencia sobre el multiplicador del gasto por contexto (normalidad, recesión, límite inferior de los tipos, régimen cambiario y deuda elevada).

Tabla A40.1: Multiplicador del gasto público según el contexto.

| Contexto | Multiplicador del gasto | Estudios |
|---|---|---|
| Normalidad, con política monetaria activa | Inferior a la unidad (0,6-0,7 en EE. UU.) | Ramey y Zubairy (2018) |
| Recesión | En discusión: 2,2-2,5 o inferior a la unidad | Auerbach y Gorodnichenko (2012); Ramey y Zubairy (2018) |
| Límite inferior de los tipos | Superior a la unidad en los modelos (2,3-3,7); evidencia empírica mixta (hasta 1,5) | Christiano, Eichenbaum y Rebelo (2011); Ramey y Zubairy (2018) |
| Tipo de cambio fijo o unión monetaria | Relativamente elevado (relativo regional: 1,5) | Ilzetzki, Mendoza y Végh (2013); Nakamura y Steinsson (2014) |
| Tipo de cambio flexible o economía muy abierta | Nulo o reducido | Ilzetzki, Mendoza y Végh (2013) |
| Deuda pública superior al 60 % del PIB | Negativo a largo plazo | Ilzetzki, Mendoza y Végh (2013) |

### 3.3. Evidencia empírica sobre el multiplicador monetario

A diferencia del fiscal, el multiplicador monetario se ha medido sobre todo con modelos SVAR y, tras la Gran Recesión, a través de los efectos de las medidas no convencionales.

**Enfoque SVAR** (Bernanke y Mihov, 1998): antes de medir los efectos de la política monetaria es necesario medir la propia política. Para ello, Bernanke y Mihov proponen un VAR semiestructural del mercado de reservas bancarias que permite identificar las innovaciones de política monetaria sin imponer *a priori* cuál es el indicador de la Fed, y concluyen que deben utilizarse distintas medidas según el periodo[3].

Con el indicador resultante, un *shock* monetario expansivo eleva el *output* con relativa rapidez, con un máximo entre los doce y los dieciocho meses, y los precios responden más tarde (con poco efecto el primer año) y de forma persistente.

La Gran Recesión supuso un desafío sin precedentes para la política monetaria, ya que los tipos de interés alcanzaron el ZLB y los canales de transmisión convencionales se deterioraron. Por ello, la evidencia empírica posterior se centra en la efectividad de las medidas no convencionales, cuyos canales de transmisión se analizan en el tema A37, apartado 3.5.

**Expansión cuantitativa y tipos a largo plazo**: Gagnon et al. (2011) estiman que la primera ronda de expansión cuantitativa (*quantitative easing*, QE) de la Fed (QE1, 2008-2010, unos 1,7 BN$) redujo la prima por plazo a diez años entre 30 y 100 puntos básicos.

Krishnamurthy y Vissing-Jorgensen (2011) analizan varios canales (señalización, duración, liquidez, demanda de activos seguros, prepago, riesgo de impago e inflación) y concluyen que el efecto depende de los activos comprados. Encuentran evidencia de los canales de señalización, de demanda de activos seguros y de inflación, y de un canal hipotecario específico en el QE1, mientras que relativizan el de duración.

Para el BCE, Altavilla, Carboni y Motto (2015) estudian los 17 anuncios sobre el programa de compras de activos (APP) entre septiembre de 2014 y marzo de 2015. Estiman que, en conjunto, los anuncios redujeron los tipos soberanos a diez años en 30-50 pb en la zona del euro.

**Efectividad real**: la transmisión a la economía real fue más débil de lo esperado. Por ejemplo, según estimaciones del Banco de Inglaterra (Joyce, Tong y Woods, 2011), la primera ronda de QE (200.000 M£, 2009-2010) elevó el nivel del PIB del Reino Unido en torno a un 1,5-2 % en su efecto máximo. La asimetría entre el impacto financiero (elevado) y el real (moderado) se explica por las restricciones crediticias del sector bancario y el desapalancamiento del sector privado.

***Forward guidance***: complementó al QE. El compromiso de la Fed de mantener tipos bajos hasta 2014 buscaba reducir los tipos esperados y, con ellos, los tipos a largo plazo, más allá de lo que permitía el ZLB. No obstante, su efectividad depende de la credibilidad del compromiso, un problema de inconsistencia dinámica (*time inconsistency*) que formalizaron Kydland y Prescott (1977).

**Tipos negativos**: el BCE mantuvo el tipo de la facilidad de depósito en terreno negativo entre junio de 2014 y julio de 2022, con un mínimo de -0,50 % desde septiembre de 2019. Su experiencia y las del Riksbank, el Banco Nacional Suizo y el Banco de Japón (BoJ) mostraron que el límite inferior efectivo (*effective lower bound*) se sitúa por debajo de cero. Los tipos negativos tienen, sin embargo, rendimientos decrecientes y efectos adversos sobre la rentabilidad bancaria (*reversal rate*, Brunnermeier y Koby, 2018).

[3] El estudio abarca EE. UU. entre 1965 y 1996. El tipo de los fondos federales es un buen indicador para 1965-1979 y 1988-1996, y el volumen de reservas no prestadas lo es para 1979-1982 (en parte debido a la política de Volcker de control de agregados). Para el conjunto del periodo 1979-1996, que mezcla varios procedimientos operativos, el que mejor se ajusta es el de Strongin (1995), basado en el componente de las reservas no prestadas ortogonal a las reservas totales.

## 3.4. Evidencia empírica sobre la interrelación entre ambas políticas

**Evidencia previa a la Gran Recesión**: el análisis de las funciones de reacción de los bancos centrales de la OCDE tras Bretton Woods apunta a la monetización del déficit donde los bancos centrales eran menos independientes (como Italia), pero no donde lo eran más (EE. UU., Alemania).

> Esta evidencia se desarrolla en el tema A39, apartado 3.4.3.

**Evidencia pos-2008**: la Gran Recesión y la crisis de deuda soberana en la zona del euro proporcionaron un laboratorio natural para estudiar la interrelación, que se recorre por etapas.

### 3.4.1. La Gran Recesión en EE. UU. y el Reino Unido (2009-2015)

**EE. UU.**: la respuesta inicial combinó ambas políticas. Por un lado, la ley de recuperación de febrero de 2009 (American Recovery and Reinvestment Act, ARRA) ascendía a 787.000 M$ en la estimación inicial de la Oficina de Presupuesto del Congreso (CBO). Por otro lado, la Reserva Federal mantuvo el tipo de los fondos federales en el 0-0,25 % de diciembre de 2008 a diciembre de 2015, con tres programas de compras de activos (QE1 a QE3, 2008-2014) y *forward guidance*.

Desde 2011, sin embargo, la política fiscal giró hacia la restricción. Tras el pulso sobre el límite de deuda, la ley de control presupuestario (Budget Control Act, agosto de 2011) fijó techos al gasto y recortes automáticos (*sequester*), que entraron en vigor en marzo de 2013. El déficit federal bajó del 9,8 % del PIB en 2009 al 2,8 % en 2014 (CBO). Esa consolidación se produjo con el tipo oficial en su límite inferior, el caso en que el modelo del apartado 2.1 predice multiplicadores más altos.

**Reino Unido**: el presupuesto de emergencia de junio de 2010 abrió una consolidación plurianual. Mientras tanto, el Banco de Inglaterra, con el tipo oficial en el 0,5 % desde marzo de 2009, ampliaba sus compras de deuda pública de 200.000 M£ a 375.000 M£ entre octubre de 2011 y julio de 2012. Como en EE. UU., la restricción fiscal coincidió con una política monetaria que solo podía actuar mediante compras de activos.

### 3.4.2. La crisis de deuda soberana europea y Japón (2010-2024)

**Zona del euro (2010-2012)**: la combinación de consolidación fiscal agresiva (dominancia monetaria, pues el BCE mantuvo su mandato de estabilidad de precios) con multiplicadores fiscales elevados generó una espiral recesión-austeridad. El *whatever it takes* de Draghi (julio de 2012) y el posterior anuncio de las operaciones monetarias de compraventa (OMT) marcaron un punto de inflexión. A partir de entonces, la política monetaria asumió un papel cuasifiscal al reducir las primas de riesgo soberano y relajar así la restricción presupuestaria de los Gobiernos periféricos sin monetización directa.

**Japón (Abenomics, 2013-2020)**: la coordinación explícita de política monetaria expansiva (expansión monetaria cuantitativa y cualitativa del BoJ, QQE), política fiscal expansiva (estímulo fiscal) y reformas estructurales (las tres flechas de Abe) representa el caso más claro de interacción deliberada. Los resultados fueron mixtos, porque hubo un éxito parcial en romper las expectativas deflacionistas, pero la inflación objetivo del 2 % no se alcanzó de forma sostenida hasta 2022, y por un *shock* de oferta externo (energía), no por las políticas domésticas.

El Banco de Japón cerró esa etapa el 19 de marzo de 2024, al juzgar que el objetivo del 2 % estaba a la vista de forma sostenible. Abandonó entonces los tipos negativos (el tipo a un día pasó al 0-0,1 %), el control de la curva de rendimientos y las compras de fondos cotizados.

### 3.4.3. La respuesta a la pandemia de COVID-19 (2020-2021)

La coordinación monetario-fiscal alcanzó niveles sin precedentes en tiempos de paz. Los Gobiernos aumentaron con fuerza el gasto (estímulos fiscales equivalentes al 10-20 % del PIB en las principales economías). Al mismo tiempo, los bancos centrales compraban cantidades récord de deuda soberana en mercado secundario, más de 1,7 BN€ en el programa de compras de emergencia frente a la pandemia (PEPP) del BCE y unos 3,5 BN$ en el caso de la Fed. La línea entre política monetaria y fiscal se difuminó, y en la zona del euro las compras llegaron a calificarse de financiación monetaria, prohibida por los Tratados, una lectura que el BCE rechazó (Schnabel, 2020). Con ese precedente, la inflación pospandemia (2021-2023) reavivó el debate sobre si se había producido un episodio de dominancia fiscal.

**Inflación de origen fiscal**: el debate admite una lectura en clave de la teoría fiscal del nivel de precios (tema A39, apartado 3.2.2). Bianchi, Faccini y Melosi (2023) distinguen la deuda respaldada por superávits futuros de la que no lo está y estiman que, tras la pandemia, los *shocks* fiscales no financiados sostuvieron la recuperación, pero causaron un aumento persistente de la inflación. Su modelo predice, además, fuera de muestra, el efecto inflacionista del plan de rescate de marzo de 2021 (American Rescue Plan Act, ARPA, de 1,9 BN$ según la CBO).

La lectura rival atribuye la mayor parte del repunte a *shocks* de precios dados los salarios (energía, alimentos, escasez de bienes) y reserva a la tensión del mercado de trabajo un efecto menor, aunque más persistente (Bernanke y Blanchard, 2025).

### 3.4.4. El endurecimiento posterior a la pandemia (2022-2026)

La inflación pospandemia (impulsada por cuellos de botella, energía y estímulos fiscales) provocó el ciclo de subidas de tipos más rápido en décadas. La Fed elevó el tipo de referencia de 0-0,25 % a 5,25-5,50 % en 16 meses (de marzo de 2022 a julio de 2023). El BCE, por su parte, elevó el tipo de las operaciones principales de financiación de 0 % a 4,50 % (y el de la facilidad de depósito, de -0,50 % a 4 %) en 14 meses (de julio de 2022 a septiembre de 2023).

Sin embargo, los déficits fiscales permanecieron elevados, en torno al 6 % del PIB en EE. UU. en 2023 y al 3,5 % en la zona del euro. En términos del modelo de Leeper, la autoridad monetaria actuaba de forma activa sin que la fiscal se comportara, todavía, de forma pasiva.

Las consecuencias fueron las predichas por la aritmética fiscal desagradable de King (1995). El coste del servicio de la deuda aumentó con rapidez, y en EE. UU. pasó de unos 350.000 M$ en 2021 a unos 900.000 M$ en 2024. En Italia, la prima de riesgo repuntó hasta unos 250 pb en otoño de 2022. Todo ello intensificó la presión por un ajuste fiscal futuro.

**Reino Unido (2022)**: ofrece el ejemplo más nítido de conflicto entre la estabilidad financiera y el tono de la política monetaria causado por un fallo de credibilidad fiscal. El "minipresupuesto" del 23 de septiembre, con rebajas de impuestos financiadas con deuda, elevó unos 130 pb en tres sesiones el rendimiento de los bonos del Tesoro británico (*gilts*) a 30 años. La caída de su precio obligó a los fondos de inversión guiada por pasivos (*liability-driven investment*, LDI) de los planes de pensiones a vender *gilts* para atender peticiones de garantías, lo que realimentó la caída.

El Banco de Inglaterra compró 19.300 M£ de *gilts* entre el 28 de septiembre y el 14 de octubre, con indemnización del Tesoro, y aplazó las ventas de su cartera, pero el 3 de noviembre subió el tipo oficial 75 pb, hasta el 3 %. Separó así los instrumentos, con compras temporales para la estabilidad financiera y tipos de interés para la inflación.

En la zona del euro, el Instrumento para la Protección de la Transmisión (TPI, julio de 2022) buscó evitar la fragmentación financiera para preservar la dominancia monetaria sin desencadenar una crisis de deuda soberana. Este episodio ilustra la tensión central del modelo de Leeper, pues el endurecimiento monetario es sostenible solo si la autoridad fiscal se comporta de forma pasiva, es decir, si finalmente ajusta el déficit.

Tras un ciclo de bajadas en 2024-2025, la Fed y el BCE volvieron a subir los tipos en 2026. El BCE lo hizo en junio y en septiembre, hasta situar la facilidad de depósito en el 2,50 %, y la Fed, en septiembre, hasta un rango del 3,75-4,00 %.

**Pérdidas de los bancos centrales**: el endurecimiento abrió, además, un canal de interrelación a través del balance (apartado 2.3). Los bancos centrales habían financiado sus compras de activos con reservas remuneradas al tipo oficial. Cuando este superó el rendimiento de unas carteras compradas a tipo fijo y con tipos bajos, incurrieron en pérdidas, que llegan al Tesoro como menores transferencias de beneficios o como pagos directos:

- *BCE*: perdió 7944 M€ en 2024 y 1254 M€ en 2025. No reparte beneficios a los bancos centrales nacionales y compensará esas pérdidas con beneficios futuros.

- *Reserva Federal*: suspendió en 2022 sus transferencias al Tesoro y registra las pérdidas como un activo diferido, que alcanzó 243.481 M$ a finales de 2025, según sus estados financieros auditados, y bajaba ya a 232.751 M$ el 16 de septiembre de 2026, según su balance semanal (H.4.1). Ha de compensarlo con beneficios futuros antes de reanudar las transferencias.

- *Banco de Inglaterra*: el Tesoro indemniza al fondo de compra de activos del banco (Asset Purchase Facility, APF) y le transfiere fondos cada trimestre desde octubre de 2022. Las transferencias netas acumuladas a favor del Tesoro, 123.900 M£ en septiembre de 2022, se habían reducido a 21.300 M£ en marzo de 2026.

El episodio reabre el debate sobre si deben remunerarse todas las reservas al tipo oficial o solo un tramo. Asimismo, reabre otro sobre si el endurecimiento cuantitativo, que materializa pérdidas al vender bonos por debajo de su precio de compra, es también una decisión con efectos fiscales. El BCE dejó de remunerar las reservas mínimas en 2023 (tema A36, apartado 3.1.3), y el Banco de Inglaterra, que vende bonos además de dejarlos vencer, rebajó en septiembre de 2025 de 100.000 a 70.000 M£ la reducción anual de su cartera.

## Conclusiones

La efectividad de las políticas monetaria y fiscal es condicional. El multiplicador fiscal tiende a ser inferior a la unidad en condiciones normales y puede superarla cuando la política monetaria no contrarresta el impulso, en el límite inferior de los tipos o con tipo de cambio fijo. La inversión pública presenta, además, multiplicadores más altos que las transferencias y las rebajas de impuestos. Por su parte, la política monetaria no convencional relajó las condiciones financieras y evitó una espiral deflacionista, aunque con un efecto real más moderado. Ambas políticas están ligadas por la restricción presupuestaria del sector público. Según el modelo de Leeper, solo hay un equilibrio determinado si una autoridad es activa y la otra pasiva. La experiencia posterior a 2008 lo ilustra con la recesión de la austeridad sin acomodación monetaria en la zona del euro (2010-2012), los resultados mixtos de la coordinación explícita y la presión sobre la deuda del endurecimiento monetario sin consolidación fiscal (2022-2024). La independencia del banco central es, por ello, condición necesaria pero no suficiente de la estabilidad de precios.

Los multiplicadores fiscales proceden del análisis de la política fiscal (tema A38), y su dependencia del régimen cambiario, del modelo de Mundell-Fleming, en el que, con movilidad perfecta del capital, la política fiscal solo es eficaz con tipo fijo (tema A32). En el modelo de tres ecuaciones, la regla monetaria determina si el impulso fiscal se transmite o se neutraliza (tema A07), y en el límite inferior las medidas no convencionales desdibujan la frontera entre ambas políticas (tema A37). La restricción presupuestaria del sector público que las liga es el núcleo del modelo de Leeper y de las teorías de la dominancia (tema A39). Por último, la UEM es el caso extremo, con una política monetaria única y políticas fiscales nacionales (temas B43 y B45).

El tamaño del multiplicador fiscal en las recesiones divide a la literatura (apartado 3.2). Auerbach y Gorodnichenko (2012) sostienen que es mayor que en las expansiones, y Ramey y Zubairy (2018) lo niegan, pues atribuyen la discrepancia sobre todo al supuesto de que la economía sigue en recesión durante todo el horizonte de cálculo. Sigue abierto también el origen de la inflación posterior a la pandemia (apartado 3.4.3), en el que la lectura fiscal de Bianchi, Faccini y Melosi (2023) se opone a la de los *shocks* de precios de Bernanke y Blanchard (2025), que reservan a la tensión del mercado de trabajo un efecto menor, aunque más persistente.

# A41. Inflación

La inflación: causas y efectos sobre la eficiencia económica y el bienestar. Hiperinflación y deflación.

## Introducción

**Introducción común.** Los temas A40, A41 y A42 comparten una introducción, que se desarrolla al comienzo del tema **A40**. Presenta las dos variables que concentran el análisis macroeconómico, la actividad y los precios, y la teoría del ciclo que estudia sus fluctuaciones, y relaciona con ellas las políticas monetaria y fiscal, la inflación y las teorías del ciclo.

La **inflación** es el aumento sostenido del nivel general de precios a lo largo del tiempo, y es una de las variables macroeconómicas de mayor relevancia. En exceso, distorsiona la función señalizadora de los precios, envilece la moneda y redistribuye la renta de forma arbitraria. En el extremo opuesto, la deflación enfría el consumo y agrava las deudas. De ahí que exista un amplio consenso sobre la conveniencia de una inflación reducida y controlada, en torno al 2 %.

Históricamente, la inflación ha acompañado a los grandes cambios monetarios. La llegada de metales preciosos americanos provocó en Europa, desde mediados del siglo XVI hasta mediados del XVII, la llamada *revolución de los precios*, que inspiró las primeras formulaciones de la teoría cuantitativa del dinero en la Escuela de Salamanca y en Jean Bodin. En el siglo XX, las hiperinflaciones de entreguerras, la estanflación de los años setenta y la posterior desinflación de los ochenta hicieron de la inflación uno de los ejes del debate macroeconómico, desde la curva de Phillips hasta las expectativas racionales. Tras dos décadas de inflación baja y estable, e incluso de riesgo de deflación después de la Gran Recesión, el repunte de 2021-2023 devolvió la inflación al primer plano y puso a prueba la credibilidad de los bancos centrales.

La inflación se mide con índices de precios de consumo, como el **índice armonizado de precios de consumo** (IPCA) que usa el BCE, y su control es hoy el objetivo principal de la mayoría de los bancos centrales. Sobre sus causas, la teoría ha oscilado entre explicaciones por el exceso de demanda, por los costes y por las expectativas. Para la tradición cuantitativa y el monetarismo, a largo plazo precios y dinero se mueven a la par, de modo que la inflación es siempre un fenómeno monetario. Las escuelas posteriores han puesto el acento, en cambio, en las expectativas y en las rigideces, y el repunte reciente ha reabierto el debate sobre el peso de las perturbaciones de oferta y de la política fiscal.

El tema estudia la inflación en tres bloques:

- *Concepto y costes*: qué es la inflación y qué cuesta.
- *Causas*: las teorías que la explican, de la clásica a la nueva economía keynesiana.
- *Formas extremas*: la hiperinflación y la deflación.

Los modelos del mercado de trabajo en que se apoya la teoría moderna de la inflación se desarrollan en el tema A28. Aquí solo se recogen sus implicaciones para la inflación.

# 1. La inflación: concepto, costes y efectos sobre la eficiencia y el bienestar

El análisis empieza por definir y medir la inflación y sigue con sus costes, entre ellos los que impone su persistencia, y sus beneficios, de cuya comparación resulta la inflación óptima.

## 1.1. Concepto y medición

El carácter sostenido de la subida del precio medio de los bienes y servicios de la economía distingue la inflación de un aumento aislado o de una sola vez, que altera el nivel de precios y solo transitoriamente su tasa de variación. Como fenómeno dinámico, su análisis exige considerar el comportamiento intertemporal de los agentes y la formación de sus expectativas.

La inflación puede medirse de diferentes formas[1]:

- *Índice de precios de consumo* (IPC): es un índice de Laspeyres que mide la evolución de los precios de una cesta de bienes nacionales y extranjeros representativa del consumo de la sociedad. Sirve de medida de referencia a la mayoría de los objetivos de inflación y de la indiciación de salarios y pensiones. El BCE usa el índice armonizado de precios de consumo (IPCA), y la Reserva Federal es una excepción, pues usa el deflactor del consumo privado (PCE).

- *Deflactor del PIB*: es un índice de Paasche que mide la evolución de los precios de todos los bienes producidos en el país, y excluye por tanto los importados.

- *Inflación subyacente*: excluye del IPC los componentes más volátiles, energía y alimentos no elaborados, para aproximar la tendencia de los precios sobre la que actúa la política monetaria.

Para el análisis económico importan tres distinciones:

- *Intensidad*: la inflación es moderada (un dígito anual), galopante (dos o tres dígitos) o hiperinflación (más del 50 % mensual, según el umbral de Cagan que se usa en el apartado 3).

- *Previsibilidad*: es esperada, si los agentes la anticipan e incorporan a sus contratos, o no esperada, si los sorprende, y los costes de una y otra son distintos.

- *Causa*: es de demanda, de costes, estructural, inercial o monetaria, según la causa que la teoría del apartado 2 identifique.

La inflación relevante para la política económica es la sostenida y esperada, que los agentes incorporan a sus decisiones. La no esperada, en cambio, tiene efectos redistributivos y sobre la actividad, pero no puede mantenerse con una inflación estable, pues los agentes acaban anticipándola.

[1] Si el IPC crece más que el deflactor, el país importa inflación (los bienes extranjeros encarecen la cesta de consumo). Ambos índices pueden divergir de forma persistente si cambia la relación real de intercambio. El IPC tiene sesgos al alza conocidos, de sustitución (la cesta fija no recoge los cambios en el consumo) y de calidad y nuevos bienes. El informe Boskin (1996) estimó la sobreestimación total del IPC de EE. UU. en 1,1 pp anuales.

## 1.2. Los costes de la inflación

La inflación genera costes que dependen de su magnitud, de su volatilidad y de si es esperada o no. Conviene distinguir los costes de la inflación esperada, que subsisten aunque los agentes la anticipen perfectamente, de los costes adicionales de la inflación no esperada, que proceden de la sorpresa.

### 1.2.1. Costes microeconómicos

**Conducta de los agentes**: en el contexto de un ciclo expansivo, la inflación puede dar lugar, especialmente a corto plazo, a una aceleración del consumo y de la inversión que alimenta burbujas especulativas. La causa es que los agentes anticipan sus compras sabiendo que los precios futuros serán más altos.

Además, existen indicios de que los agentes sufren cierto grado de *ilusión monetaria*. Al confundir magnitudes nominales y reales, ignoran muchos efectos de la inflación y no optimizan adecuadamente.

**Mercado de bienes y servicios**: salvo que sea pura, la inflación hace que los precios informen peor de qué bienes son relativamente escasos. Como cuesta saber cuánto vale de verdad cada bien, los agentes incurren en costes de búsqueda y deciden con más incertidumbre, lo que sesga sus elecciones. Además, la necesidad de revisar los precios con frecuencia impone a las empresas *costes de menú*.

**Mercado de trabajo y de factores**: la inflación supone una pérdida del valor real de las rentas del trabajo, si los salarios no se ajustan, y también de la riqueza mantenida en activos nominales, lo que puede enfriar el consumo. Si se concentra en ciertos consumos intermedios, como las materias primas o la energía, produce disrupciones en las cadenas de valor por el encarecimiento de los insumos.

**Mercados financieros y monetarios**: la inflación afecta a las tres funciones del dinero:

- *Medio de pago*: al dudarse de su valor, el dinero se acepta peor en las transacciones.
- *Unidad de cuenta*: pierde fiabilidad como referencia de los precios.
- *Depósito de valor*: pierde poder de compra, se mantienen menos saldos reales y surgen los llamados *costes de suela de zapatos*. Los agentes reducen sus tenencias de efectivo, acuden a otros activos y multiplican las conversiones en saldos líquidos, con sus costes de transacción[2].

**Coste de bienestar de la inflación anticipada** (Bailey, 1956): mantener saldos reales cuesta el tipo de interés nominal, $i = r + \pi$, aunque crear dinero casi no cuesta nada. Con una demanda de saldos reales en proporción de la renta, $m(i)$, decreciente, la pérdida es el excedente al que los agentes renuncian al economizar efectivo, es decir, el área comprendida entre la curva de demanda y el eje del tipo de interés, desde $i = 0$ hasta el tipo vigente, menos el rectángulo $i \cdot m(i)$, que es una transferencia al emisor. Esa pérdida irrecuperable, $\mathrm{PB}(i)$, es:

$$\mathrm{PB}(i) = \int_0^i m(x)\,dx - i\,m(i)$$

[2]Bailey (1956) trata la inflación como un impuesto sobre las tenencias de efectivo, cuyo coste se mide a continuación.

Con la demanda semilogarítmica de Cagan (apartado 3.1.2), la pérdida crece aproximadamente con el cuadrado del tipo nominal, de modo que es pequeña con inflación moderada y crece deprisa con la alta. Así, Lucas (2000) estima que pasar de una inflación del 10 % a la nula equivale a algo menos del 1 % de la renta. La pérdida desaparece con $i = 0$, que es la regla de Friedman.

Sobre los *stocks* de deuda, la inflación no anticipada reduce el valor real de la deuda en moneda nacional, perjudica a los acreedores y favorece a los deudores, con un efecto redistributivo arbitrario. Además, una inflación alta y volátil lleva a los prestamistas a exigir primas de riesgo sobre los tipos nominales, lo que encarece la financiación en términos reales.

### 1.2.2. Costes macroeconómicos

**Tipo de interés real e inversión**: con inflación esperada, el tipo de interés nominal incorpora una prima que compensa la pérdida de valor del dinero, $i = r + \pi^e$. Por ello, a largo plazo la inflación se traslada íntegramente a los tipos nominales y el tipo real permanece determinado por factores reales (*efecto Fisher* completo), salvo el pequeño efecto de cartera de Mundell-Tobin (apartado 1.3.2). En cambio, una inflación no esperada reduce el tipo real *ex post*, lo que favorece a los deudores.

**Sector exterior**: con tipo de cambio fijo, o con una traslación incompleta al tipo de cambio, una inflación superior a la de los socios comerciales encarece los productos nacionales frente a los extranjeros. La pérdida de competitividad-precio deteriora la balanza comercial si se cumple la condición de Marshall-Lerner (tema B12).

**Crecimiento**: la incertidumbre que la inflación genera sobre los precios relativos y sobre la rentabilidad real de los proyectos acorta los horizontes de decisión y reduce la inversión productiva. En consonancia, la evidencia entre países muestra que las inflaciones elevadas y volátiles se asocian a un crecimiento menor, aunque el umbral a partir del cual el efecto es apreciable está por encima de las tasas moderadas.

### 1.2.3. Costes desde la política económica

**Política monetaria**: una inflación alta y volátil desancla las expectativas, con lo que la curva de Phillips de corto plazo se desplaza y la política pierde su capacidad de estabilizar. Además, volver a la estabilidad de precios tiene un coste en producción y empleo, la *ratio de sacrificio* (apartado 2.4.3).

Ese coste depende de la *persistencia de la inflación*, la tendencia de la tasa de inflación a permanecer próxima a su valor anterior, que puede asemejarse a la inercia física, la resistencia de un cuerpo a cambiar la dirección y la velocidad de su movimiento. Según la distinción de Fuhrer (2006), la persistencia tiene dos fuentes:

1. *Persistencia intrínseca*: la que proviene de la propia inflación y obliga a incluir los retardos de la inflación pasada entre los determinantes de la actual. Se manifiesta en dos ámbitos:

- *Desinflaciones*: cuanto mayor es la persistencia intrínseca, más intensa o prolongada ha de ser la restricción monetaria del banco central para desacelerar la inflación, y mayor es su coste en producción.

- *Respuesta a los shocks*: si los agentes revisan los precios de sus contratos en función de la inflación pasada, la inflación responde a los *shocks* con mayor persistencia.

2. *Persistencia heredada*: la que procede de los determinantes de la inflación como el desempleo, la brecha de producción (*output gap*) o el coste marginal, que tienen su propia dinámica temporal.

La distinción importa para la política monetaria, porque la persistencia heredada desaparece cuando lo hacen sus causas, mientras que la intrínseca subsiste y obliga a una desinflación gradual y costosa, cuya magnitud se mide con la ratio de sacrificio. Qué parte de la persistencia observada es intrínseca es el punto en que las teorías del apartado 2 discrepan, y el criterio con el que la evidencia las juzga.

**Política fiscal**: la inflación distorsiona el sistema tributario por dos vías:

- *Efecto Tanzi-Olivera*: como la recaudación tiene lugar con demora respecto al devengo, mientras los gastos corrientes crecen con la inflación, el valor real de lo recaudado resulta inferior al de lo gastado, y el déficit se amplía[3].

- *Rémora fiscal* o *progresividad en frío* (*fiscal drag*): con impuestos progresivos no indiciados, los aumentos nominales de salarios y rentas trasladan a los contribuyentes a tramos superiores y redistribuyen renta del sector privado al público sin decisión legislativa. Con todo, este coste no se debe a la inflación *per se* sino al diseño del sistema impositivo.

En suma, los costes de la inflación esperada son los de "suela de zapatos", los costes de menú y las distorsiones fiscales, pequeños a tasas moderadas pero crecientes con ella.

Los de la inflación no esperada son la redistribución arbitraria entre acreedores y deudores, la incertidumbre y la pérdida de contenido informativo de los precios, que distorsionan la asignación de recursos.

Ambos tipos justifican que la estabilidad de precios sea el objetivo central de la política monetaria, y la distinción explica por qué una inflación moderada y estable cuesta menos que una baja pero volátil.

## 1.3. Los beneficios de la inflación y la inflación óptima

La solución a muchos de los costes de la inflación pasa por la indiciación (de salarios, de tramos impositivos, de deuda). No obstante, la inflación no solo supone costes. Una tasa moderada tiene efectos beneficiosos, que se presentan desde las mismas tres perspectivas y que explican por qué el objetivo de los bancos centrales no es la inflación nula.

### 1.3.1. Beneficios microeconómicos

**Conducta de los agentes**: a tasas bajas, muchos agentes no incorporan del todo la inflación a sus decisiones de salarios y precios, porque ignorarla apenas les cuesta. Akerlof, Dickens y Perry (2000) muestran que esta conducta *casi racional* (*near-rational*) permite que, con una inflación baja y positiva, el desempleo sostenible sea menor que con inflación nula. El beneficio, sin embargo, se agota a tasas más altas, cuando ignorar la inflación sale caro y los agentes la incorporan a sus contratos.

**Mercado de trabajo**: siguiendo a James Tobin, en presencia de rigideces nominales de los salarios a la baja, la inflación permite reducir los salarios reales sin recortar los nominales. De este modo, "engrasa" el ajuste del mercado de trabajo y, si no es excesiva ni inestable, puede reducir el desempleo.

[3] Olivera (1967) y Tanzi (1977). El efecto es especialmente grave con inflación alta, cuando el desfase entre devengo y cobro pesa más.

**Economía sumergida**: comprende las actividades legales que se ocultan a la Administración para eludir impuestos, cotizaciones y regulaciones. Estrecha las bases imponibles, de modo que los contribuyentes que cumplen soportan tipos más altos, y da a quien evade una ventaja de costes frente a sus competidores. Como estas actividades se pagan sobre todo en efectivo, la inflación, al erosionar el valor del dinero, las grava de forma indirecta y alcanza una base que los impuestos ordinarios no tocan. Nicolini (1998) muestra que este argumento justifica una inflación óptima positiva, aunque la ganancia de bienestar es pequeña incluso en economías con un sector informal amplio.

### 1.3.2. Beneficios macroeconómicos

**Tipo de interés real e inversión**: a corto plazo, mientras la traslación de la inflación esperada al tipo nominal es incompleta, una inflación esperada superior al tipo de interés nominal ($\pi^e > i$) hace negativo el tipo de interés real ($r = i - \pi^e < 0$). Un tipo real negativo estimula la inversión y, con ella, el empleo.

**Acumulación de capital**: según el *efecto Mundell-Tobin*, la inflación esperada reduce la demanda de saldos reales y desplaza la cartera de los agentes hacia activos reales, lo que eleva la acumulación de capital y reduce el tipo de interés real. Por este canal, la inflación esperada puede tener efectos reales permanentes en un modelo con dinero, aunque su magnitud se considera pequeña.

**Desapalancamiento**: por el mismo efecto redistributivo que perjudica a los acreedores, la inflación reduce el valor real de las deudas de hogares, empresas y Estado, y facilita la salida de las crisis de sobreendeudamiento. Este efecto es la contrapartida de la deflación de deuda del apartado 3.2.2.

### 1.3.3. Beneficios desde la política económica y la inflación óptima

**Política monetaria**: una inflación positiva mantiene los tipos nominales alejados del límite inferior de los tipos de interés (*zero lower bound*), y da al banco central margen para bajarlos en las recesiones sin caer en la trampa de liquidez.

**Política fiscal**: la creación de dinero es una de las vías de financiación del Estado, junto con los impuestos y la deuda. El *señoraje* es el ingreso real que obtiene el Estado por la emisión de moneda, y le permite cubrir gasto público sin alterar los impuestos (tema A39).

Frente a estos beneficios, la teoría monetaria clásica ofrece un resultado opuesto, la *regla de Friedman* (1969), que establece que la cantidad óptima de dinero se alcanza con un tipo de interés nominal nulo, es decir, con una deflación igual al tipo de interés real. Solo así el coste de mantener dinero coincide con su coste de producción, que es nulo.

El consenso de la política monetaria moderna se aparta de ese óptimo teórico por tres razones: el sesgo al alza de los índices de precios, la rigidez a la baja de los salarios nominales (el argumento de Tobin) y la necesidad de un margen sobre el límite inferior de los tipos nominales para poder bajarlos en las recesiones. De ahí procede el objetivo del 2 %, una inflación reducida, controlada y anticipable.

## 1.4. Conclusiones

**Aspectos positivos**: el análisis de los costes y beneficios muestra que la inflación no es un fenómeno neutral, y ofrece el criterio con el que se fija el objetivo de la política monetaria.

**Limitaciones**: la magnitud de los costes es difícil de estimar, y los estudios empíricos encuentran costes reducidos para tasas moderadas y estables (Lucas, 2000), lo que hace que el debate sobre el nivel óptimo (2 %, 3 %, 4 %) siga abierto. Por su parte, la persistencia muestra que la inflación tiene una dinámica propia que las teorías estáticas no captan, lo que motiva las teorías dinámicas del apartado siguiente.

## 2. Las causas de la inflación: las escuelas y sus curvas de Phillips

Las causas de la inflación pueden ordenarse en una taxonomía de cinco tipos:

- *Inflación de demanda*: la demanda agregada crece más deprisa que la capacidad productiva de la economía.

- *Inflación de costes*: aumentan los costes de producción, salarios o insumos, con independencia de la demanda.

- *Inflación estructural*: el poder de mercado de trabajadores y empresas genera una puja distributiva sobre el reparto de la renta que se resuelve en precios.

- *Inflación inercial o de expectativas*: la inflación pasada se transmite a la presente a través de las expectativas y de la indiciación de contratos.

- *Inflación monetaria*: la causa última es el crecimiento de la cantidad de dinero por encima del de la producción.

Las escuelas de pensamiento difieren en cuál de esas causas destacan y en cómo la fundamentan. El recorrido sigue el mismo orden que el del mercado de trabajo en el tema A28, con los fundamentos clásico y keynesiano, la síntesis neoclásica y el monetarismo, la nueva macroeconomía clásica y la nueva economía keynesiana. La diferencia de énfasis es que allí la variable es el desempleo y aquí la producción, unidas por la ley de Okun.

### 2.1. Los fundamentos: teoría clásica y teoría keynesiana

Las dos teorías tradicionales de la inflación son la teoría cuantitativa del dinero, de raíz neoclásica, y la teoría keynesiana de la inflación de demanda y de costes.

#### 2.1.1. La teoría neoclásica: la teoría cuantitativa del dinero

**Origen**: la teoría cuantitativa es la más antigua de las teorías de la inflación y nace en la Escuela de Salamanca, donde Martín de Azpilcueta (1556) atribuyó la subida de los precios españoles a la llegada de la plata americana, antes que Jean Bodin (1568).

Su formulación moderna procede de Irving Fisher (1911) y de la escuela de Cambridge (Marshall, Pigou), que dieron forma algebraica a la relación entre dinero y precios que Hume (1752) había formulado en términos causales.

**Idea**: a largo plazo, precios y dinero se mueven a la par, en idéntica proporción, de ahí que la inflación sea siempre un fenómeno monetario.

**Razonamiento**: la teoría parte de una identidad contable y la reinterpreta como una relación causal, en dos versiones. La primera es la *ecuación de cambio* (Newcomb, 1885; Fisher, 1911). El valor de las transacciones de la economía puede escribirse de dos formas, como el dinero que circula por las veces que cambia de manos, o como las transacciones por su precio:

$$M \cdot V_T = P \cdot T$$

donde $M$ es la cantidad de dinero, $V_T$ su velocidad-transacción, $P$ el nivel general de precios y $T$ el volumen de transacciones.

Como identidad, es siempre cierta. Fisher, sin embargo, la convierte en una relación causal al leerla de izquierda a derecha y suponer que $V_T$ está determinada por la tecnología de los pagos y $T$ por factores reales, de modo que ambas son estables ($V_0$, $T_0$). Esto da lugar a una relación de proporcionalidad entre la cantidad de dinero en la economía y el índice de precios a razón de $V_0/T_0$: $P = (V_0/T_0)\, M$.

De ahí que los economistas clásicos considerasen que *el dinero es un velo*, es decir, que se cumpliese la *neutralidad monetaria* (pues los cambios en la cantidad de dinero solo afectan al nivel de precios) y la *dicotomía clásica*, pues las variables reales se determinan con independencia de las monetarias.

La segunda versión es la *ecuación de Cambridge* (Marshall, Pigou). La escuela de Cambridge escribe la misma relación en términos de renta y no de transacciones,

$$M = k \cdot P \cdot Y, \quad k = 1/V_Y$$

donde $Y$ es la renta real y $k$ la fracción de la renta nominal que los agentes desean mantener en forma de dinero, el inverso de la velocidad-renta ($k = 1/V_Y$, distinta de $V_T$).

La diferencia con Fisher no está en la conclusión, que es la misma si $k$ e $Y$ son estables, sino en la lectura. La ecuación de Cambridge es una demanda de dinero, en la que $k$ recoge una decisión de los agentes y no una tecnología, y por ello puede depender del tipo de interés y de las expectativas. De esa dependencia partirá Keynes (tema A35).

De cualquiera de las dos versiones se sigue, por tanto, que, con la cantidad de dinero exógena, sus cambios se trasladan proporcionalmente a los precios a largo plazo.

**Conclusiones**:

- *Aspectos positivos*: fija la conclusión de largo plazo que ninguna teoría posterior ha desmentido, la de que la inflación sostenida es un fenómeno monetario.

- *Limitaciones*: como se limita a comparar equilibrios estáticos, sirve para explicar el nivel de precios, o un salto aislado de este, más que la inflación, que es una subida sostenida. Para dar cuenta de esta última haría falta una política monetaria expansiva mantenida durante mucho tiempo, y la neutralidad del dinero que la propia teoría postula deja a esa política sin una justificación clara.

### 2.1.2. La teoría keynesiana: inflación de demanda y de costes

**Origen**: la Gran Depresión puso en duda el paradigma neoclásico, y la teoría keynesiana trajo consigo las rigideces de precios y salarios, con las que el equilibrio entre oferta y demanda deja de estar garantizado.

**Idea**: la inflación no es solo un fenómeno monetario. Un aumento de la cantidad de dinero puede tener efectos reales, porque con salarios nominales rígidos abarata el salario real y eleva la demanda de trabajo, y la inflación puede tener su origen en la demanda o en los costes con independencia del dinero.

**Razonamiento**: la teoría keynesiana distingue dos orígenes de la inflación. Keynes (1940) y Smithies (1942) formalizaron el primero con la *brecha inflacionista*, y el segundo procede del debate posterior sobre la inflación de costes (*cost-push*):

1. *Inflación de demanda*: aparece con pleno empleo de los recursos productivos o cerca de él. En esta situación, si la oferta no es capaz de adaptarse inmediatamente al aumento de la demanda agregada, los consumidores compiten por los productos y presionan al alza los precios de la economía.

2. *Inflación de costes*: aparece aun en épocas de elevado desempleo, es decir, cuando todavía existen recursos infrautilizados, por la fijación de *mark-ups* (poder de mercado). Así ocurre cuando los países productores de petróleo aumentan sus precios y, con ellos, suben los costes de producción de todas las empresas que utilizan el combustible en su proceso productivo.

**Desarrollo gráfico**: en el plano renta-precios (figura A41.1), la oferta agregada keynesiana es casi horizontal con desempleo elevado, creciente al acercarse al pleno empleo y vertical en la renta de pleno empleo, $Y_{\text{PE}}$. Un aumento de la demanda cerca del pleno empleo (panel a) eleva mucho los precios y poco la renta, y una subida de costes con desempleo elevado (panel b) desplaza la oferta hacia arriba y eleva los precios aunque la renta caiga.

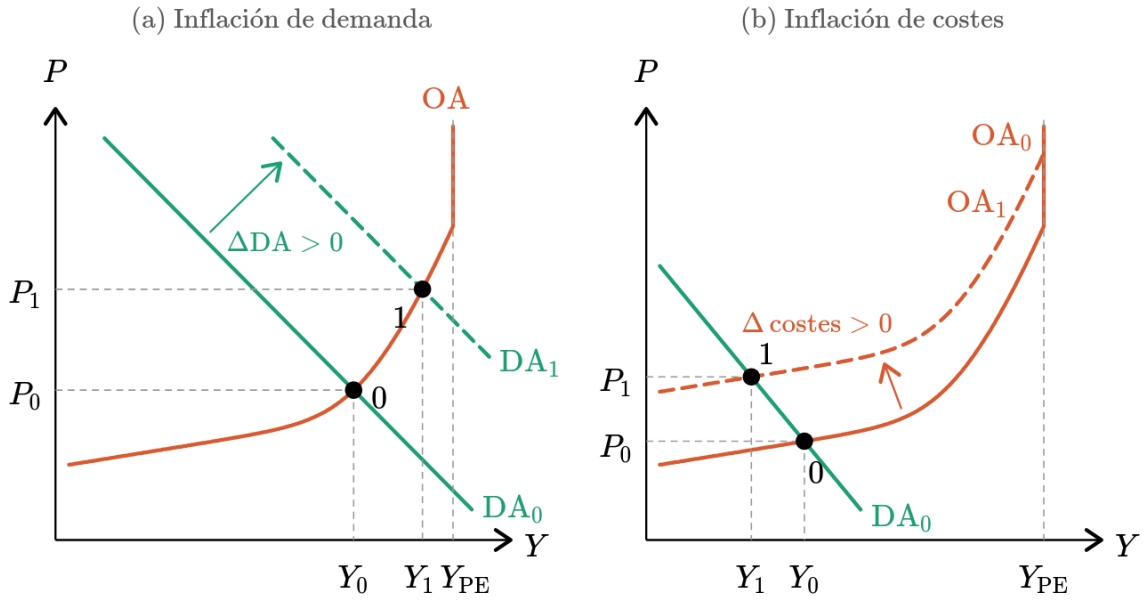



Figura A41.1: Inflación de demanda y de costes en la teoría keynesiana. La oferta agregada es casi horizontal con desempleo, creciente al acercarse al pleno empleo y vertical en $Y_{\text{PE}}$. Un aumento de la demanda cerca del pleno empleo (a) lleva de 0 a 1 con una fuerte subida de $P$ y poca de $Y$. Una subida de costes con desempleo elevado (b) desplaza la oferta de $\text{OA}_0$ a $\text{OA}_1$ y eleva $P$ aunque $Y$ caiga.

**Conclusiones**:

- *Aspectos positivos*: la distinción entre demanda y costes amplía el marco analítico e introduce el poder de mercado como factor explicativo. Además, sigue siendo relevante para el diagnóstico de política económica, porque un mismo nivel de inflación puede requerir respuestas de política muy distintas según su origen.

- *Limitaciones*: tampoco sale de la comparación de equilibrios estáticos, así que da cuenta de saltos puntuales de los precios mejor que de la inflación, cuya dinámica incluye retardos. Una

subida continuada de los precios requeriría que el exceso de demanda, o la escasez de oferta, durase también.

## 2.2. Síntesis neoclásica y monetarismo: la curva de Phillips y sus expectativas

> La curva de Phillips, su ampliación por las expectativas y la tasa natural de paro se desarrollan en el tema A28, apartados 1.2 y 1.3. Aquí se presentan como teoría de la inflación.

Las dos escuelas comparten el instrumento, la curva de Phillips, pero discrepan en su lectura. Para la síntesis es una relación estable y explotable, para el monetarismo una relación transitoria que solo las expectativas erróneas sostienen.

### 2.2.1. Síntesis neoclásica: la curva de Phillips como teoría de la inflación

**Origen**: Alban Phillips (1958) observó una relación inversa entre el crecimiento de los salarios nominales y el desempleo en el Reino Unido, Richard Lipsey (1960) la fundamentó y Paul Samuelson y Robert Solow (1960) la convirtieron en una relación entre inflación y desempleo[4].

**Idea**: la inflación nace de la presión que la demanda ejerce sobre el mercado de trabajo. Con desempleo bajo, los salarios crecen, las empresas trasladan los costes a los precios y la inflación sube, mientras que con desempleo alto ocurre lo contrario. La relación es estable, y ofrece a la política económica un menú entre inflación y desempleo.

**Razonamiento**: la curva de Phillips original relaciona la inflación con el desempleo cíclico, y por la ley de Okun con la brecha de producción:

$$\pi_t = -\kappa_u(u_t - \bar{u}) \qquad \Longleftrightarrow \qquad \pi_t = \kappa\,(y_t - \bar{y}), \qquad \kappa_u, \kappa > 0$$

Como teoría de la inflación es una teoría de demanda, en la medida en que la inflación es positiva cuando la producción supera a su nivel potencial, nula cuando lo iguala y negativa cuando se sitúa por debajo.

Su implicación de política, que la síntesis explotó, es que una expansión monetaria o fiscal sostenida permite mantener la producción por encima del potencial a cambio de una inflación más alta pero estable, y la autoridad elige el punto de la curva que prefiere.

**Desarrollo gráfico**: en el plano desempleo-inflación (figura A41.2), la curva de Phillips, CP, recoge la inflación asociada a cada tasa de desempleo. Su pendiente es negativa, porque un desempleo menor acelera los salarios y las empresas trasladan los costes a los precios, y cada uno de sus puntos es una combinación de inflación y desempleo entre las que la autoridad puede elegir.

**Estática comparativa**: una expansión monetaria o fiscal lleva la economía a lo largo de la curva del punto 0 al 1, con menos desempleo ($u_1 < u_0$) y más inflación ($\pi_1 > \pi_0$).

[4] Phillips (1958) utilizó datos del Reino Unido de 1861-1957. Irving Fisher (1926) había hallado ya, con datos de Estados Unidos, una relación inversa entre las variaciones de los precios y el desempleo.

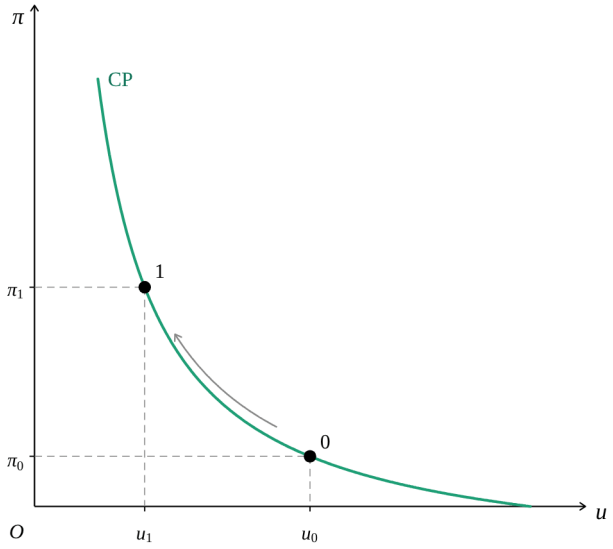



Figura A41.2: Curva de Phillips original (Samuelson y Solow, 1960). La relación estable entre inflación y desempleo ofrece un menú a la política económica, y una expansión de la demanda lleva la economía de 0 a 1.

**Conclusiones**:

- *Aspectos positivos*: la síntesis dio a la inflación una teoría operativa y una ecuación estimable.

- *Limitaciones*: la relación se apoyaba en el supuesto implícito de que los trabajadores no incorporan la inflación a sus demandas salariales. La estanflación de los años setenta, con inflación y desempleo crecientes a la vez, mostró que la curva se había desplazado.

### 2.2.2. Monetarismo: la reformulación de la teoría cuantitativa y la curva ampliada

**Origen**: el monetarismo de Milton Friedman cuestionó la síntesis en dos planos. Por un lado, en su ensayo de 1956 recuperó la teoría cuantitativa como teoría de la demanda de dinero. Por otro lado, en *The Role of Monetary Policy* (1968), a la vez que Edmund Phelps (1967) y de forma independiente, introdujo las expectativas en la curva de Phillips.

**Idea**: la inflación sostenida es siempre un fenómeno monetario, y la curva de Phillips solo existe a corto plazo, mientras los agentes se equivocan sobre la inflación, por lo que a largo plazo no hay *trade-off* que explotar.

**Razonamiento**: el argumento tiene dos pasos. El primero es la *reformulación cuantitativa*. Friedman lee la ecuación de Cambridge como una demanda de dinero, $M/P = k(\cdot)\,Y$, dependiente de la renta permanente y de los rendimientos de los activos alternativos, y sostiene que esa función es empíricamente estable, más que la función de consumo keynesiana. La ecuación cuantitativa, que como identidad se cumple siempre, se escribe en tasas de variación:

$$\pi = g_M + g_V - g_Y$$

de modo que, con velocidad estable y crecimiento real determinado por factores reales, la inflación es el exceso del crecimiento monetario sobre el crecimiento real. En esa ecuación se basa la

afirmación de que la inflación es "siempre y en todas partes un fenómeno monetario" (Friedman, 1963). Por tanto, la política monetaria determina la inflación a largo plazo, y las causas de demanda o de costes solo pueden sostenerla si el dinero las acomoda.

El segundo paso es la *curva de Phillips ampliada por las expectativas*. Si los trabajadores negocian en términos reales, incorporan la inflación que esperan a sus demandas salariales, y la curva de Phillips se desplaza con las expectativas:

$$\pi_t = \pi_t^e - \kappa_u(u_t - \bar{u}) \qquad \Longleftrightarrow \qquad \pi_t = \pi_t^e + \kappa\,(y_t - \bar{y}), \qquad \kappa_u > 0,\ \kappa > 0$$

donde $\bar{u}$ es la tasa natural de paro, e $\bar{y}$ la producción potencial que le corresponde. Como teoría de la inflación, la curva ampliada dice tres cosas:

1. *Componente inercial*: la inflación esperada, que con expectativas adaptativas es la inflación pasada[5], con lo que la inflación se autoalimenta.

2. *Componente de demanda*: la brecha de producción. Solo una producción por encima del potencial genera presión inflacionista, y la curva original es el caso $\pi_t^e = 0$.

3. *Aceleración*: en consecuencia, mantener la producción por encima del potencial exige una inflación creciente, pues cada periodo la esperada alcanza a la observada. La *teoría aceleracionista*[6] enlaza así con el primer paso, porque solo un crecimiento monetario cada vez mayor puede sostener las sorpresas.

**Desarrollo gráfico**: la curva de Phillips ampliada por las expectativas se representa en el plano desempleo-inflación (figura A41.3):

1. *Curva de Phillips de largo plazo*, $\text{CP}_{\text{LP}}$: es vertical en la tasa natural $\bar{u}$, donde la inflación observada coincide con la esperada.

2. *Curvas de Phillips de corto plazo*, $\text{CP}_{\text{CP},0}$, $\text{CP}_{\text{CP},1}$ y $\text{CP}_{\text{CP},2}$: una para cada inflación esperada (0 %, 5 % y 10 %). Son decrecientes, porque, dada la inflación esperada, solo una inflación superior a ella lleva el desempleo por debajo de la tasa natural, y cada una corta a la de largo plazo en su inflación esperada.

3. *Equilibrio*: en el punto 0, sobre $\text{CP}_{\text{LP}}$, el desempleo está en $\bar{u}$ y la inflación es nula, igual a la esperada.

**Dinámica**: si la autoridad quiere mantener el desempleo en $u_1 < \bar{u}$, la economía pasa del punto 0 al 1 a lo largo de $\text{CP}_{\text{CP},0}$, con una inflación del 5 %. En el periodo siguiente la inflación esperada alcanza ese 5 % ($\Delta\pi^e > 0$), la curva se desplaza a $\text{CP}_{\text{CP},1}$ y sostener $u_1$ exige una inflación del 10 % (punto 2). Después, con $\text{CP}_{\text{CP},2}$, la inflación sube al 15 % (punto 3). Dicho de otro modo, el desempleo solo se mantiene por debajo de la tasa natural con una inflación cada vez mayor, y cuando la autoridad deja de acelerarla la economía vuelve a $\bar{u}$ con la inflación ya alcanzada.

[5] Con expectativas adaptativas, $\pi_t^e = \pi_{t-1}^e + \lambda(\pi_{t-1} - \pi_{t-1}^e)$, con $0 < \lambda \leq 1$, lo que equivale a una media de la inflación pasada con pesos geométricamente decrecientes, $\lambda(1-\lambda)^{i-1}$ para $\pi_{t-i}$.

[6] La prescripción de política es que solo las políticas de oferta (formación, desregulación, reducción de costes institucionales) pueden reducir $\bar{u}$ de forma permanente.

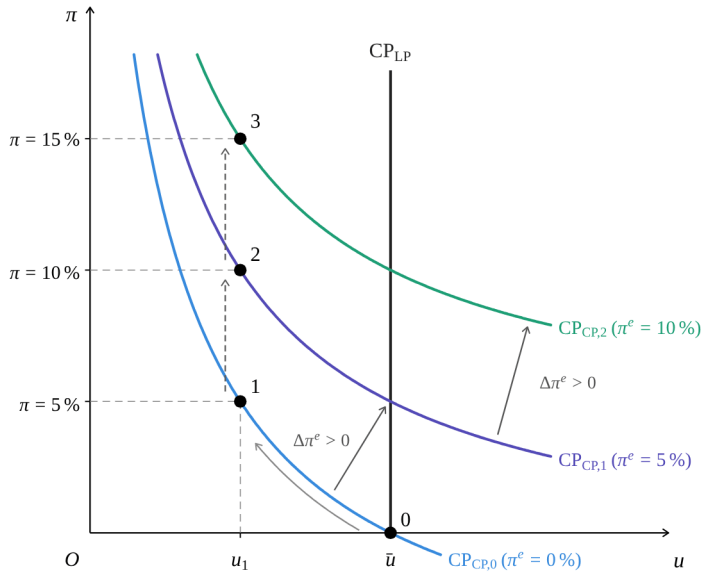



Figura A41.3: Hipótesis aceleracionista. Mantener el desempleo en $u_1$, por debajo de la tasa natural, exige una inflación creciente (puntos 1, 2 y 3), porque cada aumento de la inflación esperada desplaza hacia arriba la curva de Phillips de corto plazo.

**Conclusiones**:

▪ *Aspectos positivos*: la curva ampliada es el punto de partida de toda la teoría moderna de la inflación.

▪ *Limitaciones*: las de las expectativas adaptativas, que suponen errores sistemáticos, y la ausencia de una explicación de la tasa natural, que el modelo de Layard y Nickell sí aporta.

## 2.3. Nueva macroeconomía clásica: información imperfecta e inconsistencia dinámica

Las dos teorías de la inflación que aporta la nueva macroeconomía clásica son la oferta agregada de Lucas, basada en la información imperfecta, y la inconsistencia dinámica (*time inconsistency*) de la política monetaria.

### 2.3.1. Información imperfecta: la oferta agregada de Lucas (1972)

> El modelo de Lucas y la curva de Phillips con expectativas racionales se desarrollan en el tema A28, apartado 2.1.3.

**Origen**: Robert Lucas (1972) formuló el modelo de información imperfecta de la nueva macroeconomía clásica, del que se deriva la oferta agregada con sorpresas de precios, y Sargent y Wallace (1975) extrajeron de él la proposición de inefectividad de la política monetaria.

**Idea**: en la nueva macroeconomía clásica los agentes forman expectativas racionales en lugar de adaptativas, aprovechan cuanta información tienen y sus errores no siguen ningún patrón

sistemático. Si aun así existe un *trade-off* a corto plazo, ha de deberse a que su información sobre los precios agregados es incompleta.

**Razonamiento**: como les falta información sobre variables agregadas (la tasa de inflación y el crecimiento monetario), los productores no pueden distinguir con certeza las variaciones de su precio relativo de las del nivel general, es decir, afrontan un problema de extracción de señales. Por eso, ante una variación de su precio provocada por un *shock* de demanda, el productor no sabe a cuál de dos causas atribuirla:

- *Shock específico*: puede que los consumidores hayan cambiado sus preferencias y demanden más o menos de su producto, y en ese caso lo racional es ajustar la producción.

- *Shock general*: puede que haya variado la demanda agregada, y entonces lo conveniente es mantener la producción.

Los productores atribuirán al componente específico una fracción $\vartheta$ de la variación observada de su precio (coeficiente de extracción de señales):

$$\vartheta = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_p^2}$$

Con la experiencia acumulada, los productores estiman qué parte de la variabilidad total de los precios procede de los precios relativos ($\sigma_i^2$). Cuanto mayor sea esa parte, más verosímil les parecerá que el *shock* sea específico.

El valor de $\vartheta$ depende del grado general de variabilidad de los precios:

- *Inflación muy variable* ($\sigma_p^2$ grande): en una economía habituada a ella, como suele suceder cuando es elevada, el empresario interpreta casi cualquier variación de su precio como un *shock* global ($\vartheta$ cercana a cero) y no la traslada a la producción.

- *Precios estables* ($\sigma_p^2$ pequeña): $\vartheta$ es alta, y las sorpresas monetarias surten efectos reales transitorios.

La oferta agregada resultante (en logaritmos), $y_t = \bar{y}_t + \zeta\,(p_t - p_t^e) + \varepsilon_t^s$, con $\zeta > 0$ creciente en $\vartheta$, muestra que solo las sorpresas de precios tienen efectos reales, y que su efecto es menor cuanto más variable es la inflación. En consecuencia, la política monetaria sistemática es inefectiva y, como teoría de la inflación, el modelo implica que la inflación anticipada no compra actividad alguna.

El modelo de Lucas cierra así la vía de la inflación como instrumento, porque una autoridad que intente sorprender de forma sistemática deja de sorprender. El modelo siguiente plantea exactamente ese problema.

**Conclusiones**:

- *Aspectos positivos*: explica, con agentes racionales, por qué solo la inflación no anticipada tiene efectos reales y por qué el *trade-off* entre inflación y producción es más débil donde la inflación es más variable, implicación contrastable entre países. Con ello sitúa las expectativas y la credibilidad en el centro de la teoría de la inflación.

- *Limitaciones*: la confusión entre precios relativos y nivel general es poco plausible cuando los índices de precios se publican cada mes, y el modelo no explica por qué los efectos reales de las sorpresas persisten. Además, con datos de EE. UU., Mishkin (1982) encuentra que también la política monetaria anticipada tiene efectos reales.

### 2.3.2. Inconsistencia dinámica de la política monetaria (Kydland y Prescott, 1977; Barro y Gordon, 1983a)

**Origen**: Finn Kydland y Edward Prescott (*Rules Rather than Discretion*, 1977) mostraron que una política óptima anunciada puede dejar de serlo cuando llega el momento de aplicarla, y Robert Barro y David Gordon (1983a) aplicaron la idea a la inflación.

**Idea**: una política anunciada es *dinámicamente inconsistente* cuando, una vez que los agentes han actuado confiando en ella, a la autoridad le conviene incumplirla. Como los agentes lo anticipan, la política no es creíble y el resultado es peor que el prometido. Aplicada a la política monetaria, la falta de credibilidad genera por sí sola inflación, sin ganancia alguna de empleo, lo que se conoce como *sesgo inflacionista*.

**Supuestos**:

- *Agentes*:

- *Sociedad*: los agentes eligen sus expectativas de inflación según la hipótesis de expectativas racionales (HER), $\pi_t^e = \mathbb{E}_{t-1}[\pi_t]$, y sobre ellas fijan sus contratos de precios y salarios.

- *Gobierno*: intenta minimizar la pérdida social derivada de las fluctuaciones del nivel de desempleo y del nivel de inflación. Así, tiene un objetivo de inflación, $\pi^T$, y un objetivo de desempleo, $u^T$, inferior a la tasa natural ($u^T < \bar{u}$).

**Planteamiento**: el juego se analiza con las curvas de Phillips de corto plazo ($\text{CP}_{\text{CP}}$) del tema A28 y las curvas de indiferencia de la sociedad (CIS), elipses en torno al objetivo del Gobierno, $(u^T, \pi^T)$, tanto mejores cuanto más cerca de él. Las etapas del juego son cuatro:

1. *Punto 0*: al principio, la autoridad monetaria anuncia su compromiso de controlar la inflación. Los agentes toman ese anuncio como base de sus expectativas y pactan precios y salarios nominales.

2. *Punto 0′*: firmados los contratos, a la autoridad monetaria le conviene a corto plazo faltar a su palabra. Con los salarios nominales ya fijados, una inflación superior a la esperada abarata el trabajo en términos reales y reduce el desempleo. Si la autoridad monetaria faltara a su palabra, la economía se situaría en el punto $0'$.

3. *Punto 0″*: sin embargo, los agentes revisan al alza sus expectativas de inflación y reclaman salarios más altos. Esto provoca un desplazamiento de la $\text{CP}_{\text{CP}}$ hacia la derecha, de $\text{CP}_{\text{CP},0}$ a $\text{CP}_{\text{CP},0''}$, hasta el punto $0''$.

4. *Punto 1*: tras repetirse la secuencia durante varios periodos, la economía alcanza el punto 1. Allí la inflación iguala a la esperada ($u = \bar{u}$) y la $\text{CP}_{\text{CP}}$ toca la curva de indiferencia social (CIS) sin cortarla, de modo que ese punto es un equilibrio de Nash. Frente al punto 0, el empleo es el mismo, la inflación es más alta (*inflation bias*) y la CIS queda más alejada del objetivo $(u^T, \pi^T)$, lo que supone un empeoramiento.

**Equilibrio**: con expectativas racionales, los agentes anticipan la secuencia completa y piden desde el momento $t$ el salario propio del punto 1. Por eso el punto inicial queda descartado como equilibrio de Nash perfecto en subjuegos. Para estos autores, la inflación puede nacer de que el público no se fíe de los anuncios de política económica.

**Desarrollo gráfico**: la inconsistencia dinámica de la política monetaria se representa en el plano desempleo-inflación (figura A41.4):

1. *Curva de Phillips de largo plazo*, $\text{CP}_{\text{LP}}$: recoge los puntos en que la inflación coincide con la esperada y es vertical en la tasa natural de paro, $\bar{u}$, porque entonces el desempleo no depende de la inflación.

2. *Curvas de Phillips de corto plazo*, $\text{CP}_{\text{CP},0}$, $\text{CP}_{\text{CP},0''}$ y $\text{CP}_{\text{CP},1}$: una para cada inflación esperada, con pendiente negativa, porque solo una inflación superior a la esperada reduce el salario real y, con él, el desempleo.

3. *Curvas de indiferencia social*, $\text{CIS}_{0'}$ y $\text{CIS}_1$: elipses en torno al objetivo $(u^T, \pi^T)$, que representan una pérdida social mayor cuanto más se alejan de él.

4. *Equilibrio*: en el punto 1, la $\text{CP}_{\text{CP},1}$ es tangente a la $\text{CIS}_1$ sobre la curva de largo plazo, con el desempleo en $\bar{u}$ y una inflación $\pi_1$ superior al objetivo $\pi^T$ del punto 0.

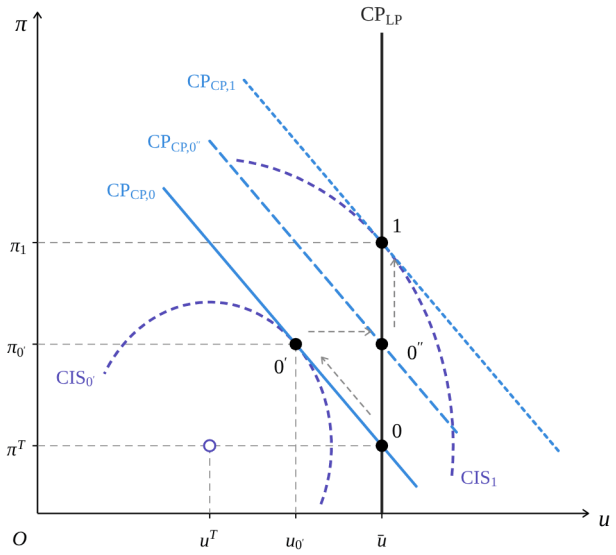



Figura A41.4: Modelo de Barro y Gordon (1983a). Desde el compromiso anunciado (0), el incentivo a incumplir (0′) y la revisión de expectativas (0″) llevan al equilibrio discrecional (1), con sesgo inflacionista y sin ganancia de empleo.

**Estática comparativa**: el sesgo inflacionista del equilibrio discrecional (punto 1) es tanto mayor:

- *Objetivo de desempleo*: cuanto más por debajo de la tasa natural fija el Gobierno su objetivo (mayor $\bar{u} - u^T$), porque mayor es la ganancia que promete una sorpresa inflacionista.

- *Preferencias del Gobierno*: cuanto más peso da el Gobierno al desempleo frente a la inflación en su función de pérdida, porque más inflación está dispuesto a aceptar a cambio de reducirlo. De ahí procede la solución del banquero conservador.

- *Pendiente de la curva de Phillips de corto plazo*: cuanto más reduce el desempleo una sorpresa de inflación dada, porque mayor es el incentivo a provocarla.

En todos los casos, el desempleo de equilibrio sigue siendo la tasa natural, y los parámetros solo alteran la inflación.

**Extensiones**: las soluciones al sesgo son cuatro:

- *Independencia del banco central*: separar el decisor de la presión política.

- *Reglas transparentes*: anclar las expectativas.

- *Banquero conservador* (Rogoff, 1985): nombrar un gobernador con mayor aversión a la inflación que la sociedad.

- *Contratos de desempeño* (Walsh, 1995): vincular la remuneración del gobernador al cumplimiento del objetivo de inflación.

**Conclusiones**:

- *Aspectos positivos*: el modelo explica el sesgo inflacionista de la política discrecional sin recurrir a errores de los agentes ni a perturbaciones, solo a la falta de compromiso, y fundamenta el diseño institucional de la política monetaria moderna.

- *Limitaciones*: supone que el objetivo de desempleo del Gobierno está por debajo de la tasa natural ($u^T < \bar{u}$), lo que no siempre ocurre, y que la sociedad conoce las preferencias del Gobierno. Con incertidumbre sobre ellas surgen la reputación y la señalización (Backus y Driffill, 1985).

## 2.4. Nueva economía keynesiana: rigideces reales y nominales

> El modelo de la tasa de paro que no acelera la inflación (NAIRU) de Layard y Nickell y la curva de Phillips neokeynesiana à la Calvo se desarrollan en el tema A28, apartados 2.2 y 3.2. Aquí se presentan como teorías de la inflación.

La nueva economía keynesiana explica la inflación por las rigideces reales, en su primera generación, y por las nominales, en la segunda. Sus desarrollos posteriores recuperan la inercia con las curvas de Phillips híbridas.

### 2.4.1. Primera generación: la puja distributiva (Layard y Nickell, 1985)

**Origen**: Richard Layard y Stephen Nickell (1985) formalizaron la NAIRU como resultado de la negociación salarial y de la fijación de precios con poder de mercado (tema A28, apartado 2.2).

**Idea**: la nueva economía keynesiana de primera generación explica la inflación como el resultado de una *puja distributiva*. En esa puja, los trabajadores negocian un salario real que crece con su poder negociador y las empresas fijan un margen sobre el coste. Cuando ambas aspiraciones son incompatibles, porque el desempleo es bajo, la puja se resuelve en inflación. La tasa de desempleo a la que son compatibles es la NAIRU[7].

**Razonamiento**: combinando la ecuación de salarios y la de precios se obtiene la misma curva de Phillips con expectativas, $\pi_t = \pi_t^e - \kappa_u(u_t - \bar{u})$, con la diferencia de que $\bar{u}$ ya no es un dato sino una función de los parámetros estructurales. En concreto, crece con el margen empresarial y con los factores institucionales que elevan las aspiraciones salariales (prestaciones, salario mínimo, cotizaciones), y decrece con la productividad. La inflación es aquí estructural, porque, para cada

[7] La NAIRU del modelo Layard-Nickell difiere de la tasa natural de Friedman en que es compatible con desempleo involuntario (por poder de mercado bilateral), no solo friccional y estructural. Esto la hace más realista para describir los mercados laborales europeos.

tasa de desempleo, depende del grado de competencia en los mercados de bienes y de trabajo, y solo las políticas que actúan sobre ellos reducen de forma permanente la NAIRU.

**Desarrollo gráfico**: la puja distributiva se representa en el plano desempleo-salario real (figura A41.5):

1. *Salario real negociado*, SRN: recoge el salario real que reclaman los trabajadores, $w/P^e = f(u, z)$, y es decreciente, porque un desempleo menor refuerza su poder negociador.

2. *Salario real pagado*, SRP: recoge el salario real que conceden las empresas al fijar su margen, $A_L/(1+\mu)$, y es horizontal, porque no depende del desempleo.

3. *Equilibrio*: en el punto 0 se cortan ambas curvas, y el desempleo es la NAIRU, $\bar{u}_0$.

- *Desempleo inferior a la NAIRU* ($u < \bar{u}_0$): el salario negociado supera al pagado, las aspiraciones son incompatibles y la puja se resuelve en una inflación creciente (zona sombreada).

- *Desempleo superior a la NAIRU* ($u > \bar{u}_0$): el salario negociado queda por debajo del pagado, y la inflación se reduce.

**Estática comparativa**: un aumento del margen empresarial ($\Delta\mu > 0$) reduce el salario real pagado y desplaza la curva de $\text{SRP}_0$ a $\text{SRP}_1$. Con ello, el equilibrio pasa del punto 0 al 1, con una NAIRU mayor ($\bar{u}_1 > \bar{u}_0$). Un aumento de los factores institucionales $z$, que desplaza el SRN hacia arriba, tiene el mismo efecto.

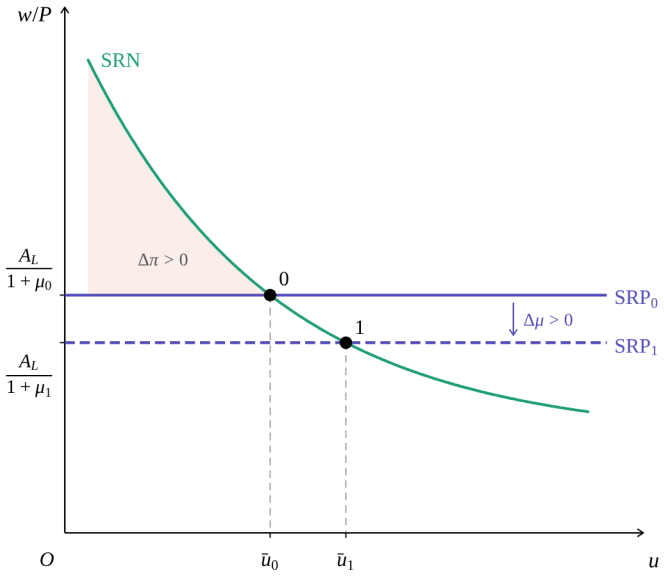



Figura A41.5: Puja distributiva (Layard y Nickell, 1985). Con un desempleo inferior a la NAIRU, el salario real negociado (SRN) supera al pagado (SRP) y la inflación crece, y un aumento del margen ($\Delta\mu > 0$) eleva la NAIRU de $\bar{u}_0$ a $\bar{u}_1$.

**Conclusiones**:

- *Aspectos positivos*: el modelo explica por qué economías con la misma política monetaria tienen inflaciones distintas, y fundamenta las políticas de rentas y las reformas estructurales como instrumentos antiinflacionistas.

- *Limitaciones*: la NAIRU de la que depende la inflación no es observable y sus estimaciones varían en el tiempo, lo que dificulta usarla como guía de la política. Además, el modelo explica la tasa de desempleo compatible con una inflación estable, pero no la dinámica del ajuste de los precios, que la segunda generación microfundamenta.

### 2.4.2. Segunda generación: la fijación escalonada de precios (Calvo, 1983)

**Origen**: Guillermo Calvo (1983) propuso, como alternativa a los contratos de Fischer (1977) y Taylor (1980), el modelo de fijación escalonada de precios (tema A28, apartado 3.2) que se ha convertido en el estándar de los modelos de equilibrio general dinámico y estocástico (DSGE).

**Idea**: la nueva economía keynesiana de segunda generación microfundamenta la curva de Phillips en la conducta de empresas con poder de mercado que no pueden ajustar sus precios en todo momento. Cada periodo solo una fracción $1-\xi$ de ellas los revisa[8], y al hacerlo fija un margen sobre los costes marginales presentes y esperados[9], porque el precio quedará vigente durante un tiempo.

**Razonamiento**: agregando los precios y loglinealizando se obtiene la *curva de Phillips neokeynesiana* (CP-NEK):

$$\pi_t = \beta \mathbb{E}_t[\pi_{t+1}] + \kappa\, \tilde{y}_t, \qquad \kappa > 0,\ \beta \in (0,1)$$

donde $\tilde{y}_t$ es la brecha de producción y $\kappa$ crece con la frecuencia de ajuste de precios[10]. Como teoría de la inflación, la curva dice que la inflación de hoy depende de la inflación esperada para mañana y del exceso de demanda actual. La curva es puramente prospectiva (*forward looking*)[11], y la inflación actual resume toda la senda esperada de brechas de producción futuras. De ahí que no haya inercia y que la inflación pueda reducirse sin coste si la política es creíble.

**Conclusiones**:

- *Aspectos positivos*: la curva neokeynesiana es la relación de oferta de los modelos con los que los bancos centrales analizan la política monetaria, y su carácter prospectivo hace de la credibilidad y la comunicación los instrumentos centrales de la desinflación.

- *Limitaciones*: no reproduce la persistencia observada de la inflación (limitación empírica).

[8] En Calvo la revisión de precios es estocástica (no depende del tiempo transcurrido desde la última), a diferencia de los contratos de duración fija de Fischer (1977) y Taylor (1980). Los tres generan dinámicas parecidas, aunque no idénticas (en el de Fischer la inflación depende de expectativas formadas en periodos anteriores), pero el de Calvo es analíticamente más manejable.

[9] El margen puede variar con el ciclo. Si los costes marginales son procíclicos y los precios se ajustan con retraso, en las expansiones los costes crecen más rápido que los precios y los márgenes se comprimen.

[10] Calibración típica (Galí, 2008): $\xi = 2/3$ (precios fijos durante unos tres trimestres), $\beta = 0{,}99$ y una elasticidad de sustitución entre variedades de seis (*mark-up* en torno al 20 %). La pendiente $\kappa$ es estrictamente decreciente con la rigidez ($\xi$) y, cuando el trabajo tiene rendimientos decrecientes (como en la calibración de Galí), con la sustituibilidad entre variedades, mientras que con rendimientos constantes no depende de esta.

[11] La formulación con la brecha ($\kappa\tilde{y}_t$) es la más utilizada en los modelos DSGE por su conexión directa con las variables de política, aunque la brecha de producción no es directamente observable.

### 2.4.3. Desarrollos: las curvas de Phillips híbridas (Galí y Gertler, 1999; Mankiw y Reis, 2002; Christiano, Eichenbaum y Evans, 2005)

**Origen**: los desarrollos de Galí y Gertler (1999), Mankiw y Reis (2002) y Christiano, Eichenbaum y Evans (2005) responden a las críticas empíricas a la curva de Phillips neokeynesiana pura, y recuperan la inercia que ya recogía el modelo del triángulo de Gordon (1982).

**Idea**: la CP-NEK pura (*forward looking*) se enfrenta a tres tipos de críticas, todas relacionadas con su carácter prospectivo:

▪ *Inmediatez del ajuste*: por su carácter prospectivo, predice un ajuste inmediato, mientras que la inflación cae con retraso y de forma gradual tras una contracción monetaria (Mankiw, 2001).

▪ *Ausencia de persistencia*: debido a su carácter prospectivo, no es capaz de explicar la persistencia de la inflación.

▪ *Evidencia empírica*: las encuestas a empresas muestran que la inflación apenas se percibe como importante para sus decisiones, por lo que las empresas dedican poco esfuerzo a informarse sobre ella y sus creencias son muy dispersas (Coibion, Gorodnichenko y Kumar, 2018).

**Razonamiento**: el modelo del triángulo, anterior a la CP-NEK, y varios desarrollos que la reformulan incorporan la inercia de la inflación:

▪ *Gordon (1982)* ("modelo del triángulo"): antes incluso de la CP-NEK, Robert Gordon formalizó la persistencia de la inflación (apartado 1.2.3) con una curva de Phillips con un retardo distribuido de la inflación, a la que la literatura ha vuelto para recoger la inercia. La persistencia se convirtió después en el principal contraste empírico de las teorías de la inflación. En la curva, $\pi_t = \sum_{i=1}^{n} a_i \pi_{t-i} - b(u_t - \bar{u}) + c \cdot s_t + \varepsilon_t$, la inflación $\pi_t$ depende de los tres vértices del triángulo, la inercia (base), la demanda y la oferta, más un componente estocástico:

- *Inflación pasada*: las tasas de inflación de los $n$ periodos anteriores, ponderadas respectivamente por los parámetros $a_i$. Recogen la persistencia intrínseca, que recuerda a la hipótesis aceleracionista de la escuela monetarista, basada en la hipótesis de expectativas adaptativas (HEA).
- *Diferencial de paro*: la diferencia entre la tasa de desempleo $u_t$ y la NAIRU $\bar{u}$, ponderada por un parámetro $b$.
- *Factores de oferta*: la incidencia de otros factores de oferta ($s_t$), ponderados por un parámetro $c$.
- *Término error*: un componente estocástico $\varepsilon_t$.

▪ *Galí y Gertler (1999)*: parten del modelo de Calvo e introducen reglas de ajuste sencillas. Así, llegan a una especificación de la CP-NEK con un solo retardo de la inflación (en lugar de un retardo distribuido) y con el coste marginal real, aproximado por la participación del trabajo en la renta, en lugar de la brecha de producción. Esta especificación es la más utilizada empíricamente.

▪ *Christiano, Eichenbaum y Evans (2005)*: consideran la indiciación generalizada, en la que las empresas que no ajustan su precio de modo óptimo en cada periodo sí lo ajustan con base en la inflación pasada, lo que genera persistencia endógena.

- *Mankiw y Reis (2002)*: sustituyen la rigidez de los precios por la de la información. Como informarse es costoso, en cada periodo solo una parte de las empresas actualiza la información con la que fija sus precios, y el resto los fija con información atrasada, lo que genera un ajuste gradual.

La curva híbrida, $\pi_t = \gamma_b \pi_{t-1} + \gamma_f \mathbb{E}_t[\pi_{t+1}] + \kappa \tilde{y}_t$, con un componente inercial y otro prospectivo, es la especificación que se usa en la práctica. Con ella, la desinflación tiene un coste, medido por la ratio de sacrificio (los puntos de producción perdidos por cada punto de inflación reducida), que es tanto mayor cuanto mayor es el peso de la inercia y menor la credibilidad de la política.

**Conclusiones**:

- *Aspectos positivos*: reconcilia la curva de Phillips neokeynesiana con la persistencia observada de la inflación y con el carácter gradual de las desinflaciones, y explica por qué la credibilidad reduce, pero no elimina, el coste de desinflar. El modelo del triángulo integra, además, en una sola ecuación la inercia, la demanda y los *shocks* de oferta, lo que le permite explicar episodios como la estanflación, que la curva de Phillips original no recogía. Su forma reducida se presta a la estimación y es una referencia empírica para medir la NAIRU y la ratio de sacrificio.

- *Limitaciones*: el componente inercial descansa en retardos de la inflación sin fundamento microeconómico, en el modelo del triángulo, o en la indiciación o en rigideces de información postuladas. Sus coeficientes y el grado de indiciación pueden cambiar con el régimen de política monetaria, lo que expone estas curvas a la crítica de Lucas. Además, el peso estimado de la inercia frente a las expectativas varía con la especificación y con la medida del coste marginal.

## 2.5. Conclusiones

Las teorías modernas de la inflación convergen en una estructura común, $\pi_t = \pi_t^e - \kappa_u(u_t - \bar{u})$, coinciden a largo plazo con la teoría cuantitativa y difieren en tres cosas:

- *Expectativas*: cómo se forman (adaptativas, racionales, con información imperfecta).

- *Tasa natural*: de qué depende (de factores reales, de la puja distributiva).

- *Inflación observada*: de dónde proceden sus efectos reales o su nivel (de las sorpresas, de la falta de credibilidad, de las rigideces).

Las dos teorías de la nueva macroeconomía clásica comparten la lección de que la inflación anticipada carece de efectos reales y difieren en lo que explican. Lucas explica los efectos reales, que proceden de las sorpresas, y Barro y Gordon, la inflación observada, que procede de la falta de credibilidad.

La segunda teoría ha organizado la política monetaria de las últimas décadas, ya que los bancos centrales independientes, los objetivos de inflación explícitos y la comunicación de la política son los mecanismos con los que se resuelve la inconsistencia dinámica. El éxito de esos mecanismos en anclar las expectativas contribuye a explicar la baja inflación de las economías avanzadas entre 1990 y 2020 (temas A36 y A37).

Las dos teorías de la nueva macroeconomía clásica descansan, sin embargo, en expectativas racionales y en precios flexibles, por lo que predicen que una desinflación creíble no cuesta desempleo y no explican la persistencia de la inflación, que la nueva economía keynesiana atribuye a las rigideces de precios y salarios. Con todo, ninguna de las teorías explica bien la persistencia sin recurrir a supuestos *ad hoc* sobre la indiciación o la información, y la inflación importada y de oferta que domina los episodios recientes solo entra en ellas como perturbación exógena.

Además, todas tratan la inflación como un fenómeno de economía cerrada, cuando en economías abiertas el tipo de cambio y los precios de las materias primas son canales de primer orden (tema B12).

# 3. Hiperinflación y deflación

Los dos extremos de la inflación tienen en común que en ellos el dinero deja de cumplir sus funciones. En la hiperinflación pierde la de depósito de valor y acaba perdiendo la de medio de pago. En la deflación, por el contrario, su valor crece y los agentes lo atesoran. Ambos extremos son fenómenos monetarios en su mecánica y fiscales o financieros en su origen.

## 3.1. La hiperinflación

Del primero de esos extremos, la hiperinflación, se ven su concepto y origen, el modelo de Cagan, que formaliza su dinámica, y su estabilización.

### 3.1.1. Concepto y origen

La **hiperinflación** es, según el criterio de Cagan (1956), una inflación superior al 50 % mensual, que equivale a multiplicar los precios por más de cien en un año. Se trata de un fenómeno raro, asociado a guerras, revoluciones y colapsos del Estado (Alemania en 1923, Hungría en 1946, Zimbabue en 2008, Venezuela entre 2017 y 2020), y en él las variables reales pierden relevancia frente a las nominales.

El origen de las hiperinflaciones es fiscal. Un Estado que no puede financiar su gasto con impuestos ni con deuda recurre al señoraje, la creación de dinero, y la inflación resultante reduce la demanda de dinero, lo que obliga a crear más dinero para obtener el mismo ingreso real. El proceso se retroalimenta y, cuando los agentes anticipan la inflación, puede acelerarse sin límite.

La hiperinflación es, por tanto, la forma extrema de la inflación monetaria, con la particularidad de que en ella las expectativas y la demanda de dinero, y no solo la oferta, son las que la aceleran. El modelo de Cagan formaliza ese mecanismo.

### 3.1.2. El modelo de Cagan (1956)

**Origen**: Phillip Cagan (1956) estudió siete episodios de hiperinflación del siglo XX en seis países europeos (Alemania, Austria, Hungría, Polonia, Rusia, Grecia).

**Idea**: Cagan se preguntaba si el crecimiento de la masa monetaria y la inflación consiguiente afectan, a su vez, a la demanda de dinero a través de las expectativas crecientes de inflación, de modo que esa demanda se reduce y la inflación se perpetúa.

**Supuestos**:

- *Entorno*: una economía en hiperinflación, en la que las variables reales (renta, tipo de interés real) varían de forma insignificante respecto a las nominales, de modo que pueden tomarse como constantes.

- *Agentes*: los individuos demandan saldos reales en función únicamente de la inflación esperada, y forman sus expectativas de forma adaptativa. El Gobierno fija exógenamente el crecimiento monetario $g_M$ para financiar su déficit, y la oferta de dinero se ajusta a la demanda a través del nivel de precios.

**Planteamiento**: el modelo consta de dos ecuaciones, la demanda de saldos reales y la formación de expectativas, y se cierra con el equilibrio del mercado monetario, en el que la cantidad de dinero crece a la tasa $g_M$:

▪ *Demanda de dinero*: con las variables reales constantes, la demanda de saldos reales depende solo de la inflación esperada,

$$m_t - p_t = c - a\pi_t^e$$

donde $m_t$ y $p_t$ son los logaritmos de la cantidad de dinero y del nivel de precios, $c > 0$ recoge la demanda de dinero por factores reales (componente autónomo, renta real y tipo de interés real) y $a > 0$ es la sensibilidad de la demanda de dinero a la inflación esperada. Por la relación de Fisher, una mayor inflación esperada eleva el tipo de interés nominal y, con él, el coste de oportunidad de mantener dinero, de modo que reduce la demanda de saldos reales.

▪ *Expectativas*: en la formulación original, Cagan se apoyó en la HEA, con la que la inflación esperada se revisa en proporción al error cometido,

$$\dot{\pi}^e = \lambda\,(\pi - \pi^e)$$

donde $\pi$ y $\pi^e$ son la inflación observada y la esperada y $\lambda > 0$ es la velocidad de aprendizaje, la fracción del error que se corrige por unidad de tiempo. Los desarrollos posteriores incluyen también la HER.

**Equilibrio**: despejando el nivel de precios en la demanda de dinero,

$$p_t = m_t - c + a\pi_t^e$$

el nivel de precios depende de la cantidad de dinero, de la demanda por factores reales y de la inflación esperada. En tasas de variación, la demanda de dinero se escribe $g_M - \pi = -a\,\dot{\pi}^e$, y al sustituir la HEA la inflación observada queda en función de la esperada:

$$\pi = \frac{g_M - a\lambda\,\pi^e}{1 - a\lambda}$$

En el estado estacionario la inflación esperada no cambia, y la observada, la esperada y el crecimiento monetario coinciden, $\pi^e = \pi = g_M$.

**Desarrollo gráfico**: el equilibrio del modelo se representa en el plano inflación esperada-inflación observada (figura A41.6):

1. *Inflación observada*, $\pi(\pi^e)$: recoge la inflación que corresponde a cada inflación esperada según la ecuación anterior. Su pendiente, $-a\lambda/(1 - a\lambda)$, es negativa si $a\lambda < 1$ (panel a) y positiva y mayor que uno si $a\lambda > 1$ (panel b).

2. *Bisectriz*, $\pi = \pi^e$: recoge los puntos en que la inflación observada y la esperada coinciden, de modo que la inflación esperada no cambia.

3. *Equilibrio*: en el punto 0 se cortan ambas rectas, con $\pi^e = \pi = g_M$.

**Estática comparativa**: de la ecuación de precios se deduce que:

▪ *Demanda por factores reales*: el nivel de precios es menor cuanto mayor es la demanda de dinero por factores reales ($c$).

▪ *Crecimiento monetario*: el nivel de precios aumenta con la cantidad de dinero ($m_t$), y la inflación de equilibrio, con su tasa de crecimiento ($g_M$). Por ello el modelo fue un argumento en favor del control de los agregados monetarios, ya que, si la cantidad de dinero se estabiliza y el equilibrio es estable, el nivel de precios también lo hace.

▪ *Sensibilidad y velocidad de aprendizaje*: cuando la inflación se acelera, el nivel de precios aumenta más cuanto mayores son la sensibilidad de la demanda de dinero a la inflación esperada ($a$) y la velocidad de aprendizaje ($\lambda$). La razón es que ambos parámetros intensifican la retroalimentación entre la inflación observada y la esperada.

**Dinámica**: con la HEA, la inflación esperada evoluciona según $\dot{\pi}^e = \lambda(g_M - \pi^e)/(1 - a\lambda)$, y la estabilidad del equilibrio depende de $a\lambda$ (flechas de la figura A41.6):

▪ *Equilibrio estable* ($a\lambda < 1$, panel a): la inflación esperada converge a $g_M$ tras cualquier perturbación. Una hiperinflación exige, en este caso, que la emisión de dinero sea excesiva y cada vez mayor.

▪ *Equilibrio inestable* ($a\lambda > 1$, panel b): un salto en la inflación revisa rápidamente al alza la expectativa y reduce la demanda de saldos reales. El exceso de oferta de dinero lleva a una inflación aún mayor que la esperada, que eleva de nuevo las expectativas. En consecuencia, la hiperinflación se explica sin que la cantidad de dinero crezca de forma continua, como una profecía autocumplida.

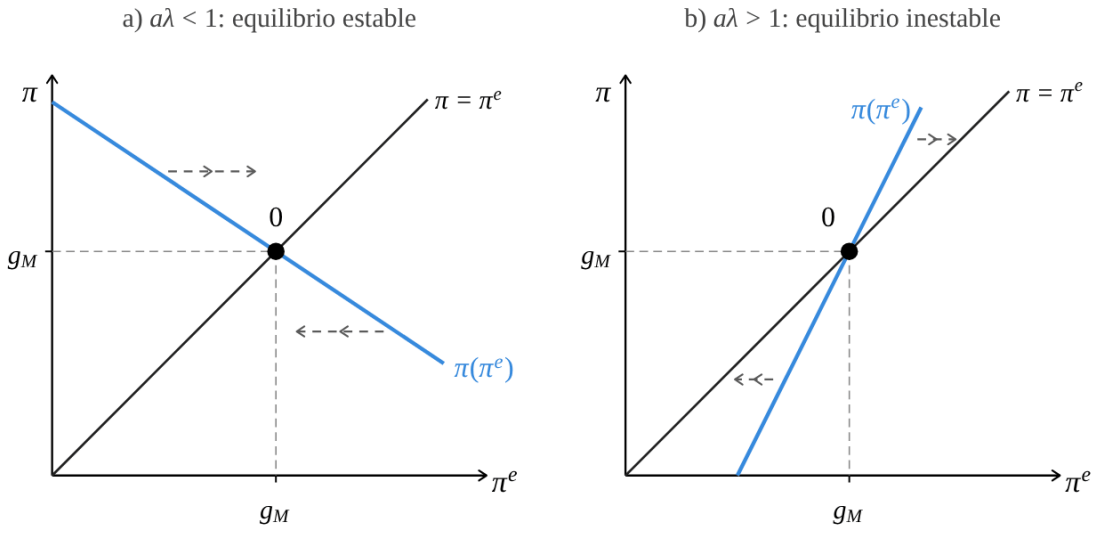



Figura A41.6: Estabilidad en el modelo de Cagan con expectativas adaptativas. Si $a\lambda < 1$ la inflación esperada converge a $g_M$ (a), mientras que si $a\lambda > 1$ se aleja de ella y la hiperinflación se autoalimenta (b).

**Extensiones**: el modelo admite dos desarrollos:

▪ *Expectativas racionales*: con la HER, las expectativas sobre el nivel de precios futuro, y no la inflación ya observada, determinan el nivel de precios en cada momento. Por ello, si el Gobierno carece de credibilidad, las expectativas de inflación pueden elevarla por sí solas, y la inflación se perpetúa no solo por el lado de la oferta de dinero, sino también por el de su demanda.

▪ *Señoraje*: el modelo permite calcular cuánto puede recaudar el Gobierno con la creación de dinero (tema A39, apartado 2.2.2). El señoraje, $\sigma = \dot{M}/P = g_M \cdot (M/P)$, es el producto de la tasa de crecimiento monetario, que actúa como tipo impositivo, por los saldos reales, que actúan como base imponible. En estado estacionario, $\pi^e = g_M$ y $M/P = e^{c-ag_M}$, con lo que $\sigma = g_M\, e^{c-ag_M}$. Aumentar $g_M$ eleva el tipo pero reduce la base, porque los agentes huyen del dinero.

En el señoraje, la condición de primer orden, $e^{c-ag_M}(1-ag_M) = 0$, da la tasa que maximiza la recaudación, $g_M^* = 1/a$, en la que la elasticidad de la base respecto al tipo es $-1$, como en la cima de la curva de Laffer. La condición de segundo orden se cumple, pues $\sigma''(g_M^*) = -a\, e^{c-1} < 0$. El señoraje máximo, $\sigma^* = e^{c-1}/a$, crece con la demanda por factores reales ($c$) y decrece con la sensibilidad a la inflación esperada ($a$). Las hiperinflaciones aparecen cuando el Gobierno intenta extraer más señoraje del que la demanda de dinero permite, y en los episodios que estudió Cagan el crecimiento monetario superó con mucho ese máximo.

**Conclusiones**:

▪ *Aspectos positivos*: es el primer modelo formal de la dinámica hiperinflacionaria, y conecta la política monetaria con la teoría de la imposición al tratar la inflación como un impuesto sobre los saldos reales.

▪ *Limitaciones*: el modelo trata la economía como cerrada. Si el tipo de cambio fuera fijo, monetizar el déficit haría perder reservas y acabaría forzando el abandono de la paridad. Además, ni la tasa $g_M^*$ que maximiza el señoraje es necesariamente la socialmente óptima ni endeudarse en lugar de emitir dinero evita la inflación, según la "aritmética monetarista desagradable" de Sargent y Wallace (1981). Las dos cuestiones se tratan en el tema A39, apartados 2.2.2 y 3.2.1.

### 3.1.3. La estabilización de las hiperinflaciones (Sargent, 1982)

**Origen**: Thomas Sargent (1982) estudió el final de las hiperinflaciones de Austria, Hungría, Polonia y Alemania en los años veinte.

**Idea**: las hiperinflaciones terminan de forma abrupta, no gradual, y con un coste en producción menor del que la ratio de sacrificio predeciría, cuando se produce un *cambio de régimen* creíble en la política fiscal y monetaria.

**Razonamiento**: la explicación es la del apartado 2.3. Si la inflación procede de las expectativas de financiación monetaria del déficit, lo que la detiene es cambiar esas expectativas, y solo un cambio institucional visible lo consigue. Por ello, el programa de estabilización combina tres elementos:

1. *Ajuste fiscal*: una reforma que elimina el déficit, la causa de la inflación.

2. *Independencia del banco central*: la prohibición de financiar al Tesoro, que hace creíble el ajuste.

3. *Ancla nominal*: a menudo una nueva moneda o un anclaje cambiario, que da a los agentes una referencia para fijar precios.

Los programas *heterodoxos* de los años ochenta (Argentina, Brasil, Israel) añadieron controles de precios y salarios para romper la inercia, con éxito solo cuando el ajuste fiscal los acompañó.

La estabilización de las hiperinflaciones es, por tanto, el caso extremo en que la credibilidad, y no la contracción monetaria gradual, es el instrumento.

**Conclusiones**:

▪ *Aspectos positivos*: aporta evidencia histórica sobre el papel de las expectativas en la dinámica de la inflación y muestra que el coste de una desinflación depende del régimen de política y no solo de la inercia. Ese resultado respalda la visión de la desinflación basada en la credibilidad y las expectativas racionales.

▪ *Limitaciones*: la evidencia procede de episodios extremos, en los que la inercia de la inflación es pequeña. Por eso, sus conclusiones se trasladan peor a las inflaciones moderadas, en las que los contratos y la indiciación hacen la desinflación más gradual y costosa.

## 3.2. La deflación

En el extremo opuesto, de la deflación interesan su concepto y causas, sus costes y la respuesta de la política económica.

### 3.2.1. Concepto y causas

La **deflación** es la caída sostenida del nivel general de precios. Tradicionalmente se ha analizado por analogía con la inflación, con efectos inversos y las mismas causas:

▪ *Shocks de oferta positivos* (mayor productividad, abaratamiento de las importaciones): generan una deflación benigna, acompañada de crecimiento.

▪ *Shocks de demanda negativos* (una crisis financiera, una contracción del crédito): generan una deflación recesiva, que es la que preocupa.

La analogía se rompe en un punto, porque la inflación y la deflación no son simétricas en sus efectos. Tres asimetrías, las deudas fijadas en términos nominales, la rigidez a la baja de los salarios nominales y el límite inferior de los tipos de interés nominales, hacen a la deflación más dañina y más difícil de combatir.

### 3.2.2. Los costes: la deflación de deuda y la trampa de liquidez

**Origen**: Irving Fisher (*The Debt-Deflation Theory of Great Depressions*, 1933) explicó la Gran Depresión por la interacción entre deflación y endeudamiento, y Keynes (1936) describió como caso límite, del que no conocía ejemplos, la trampa de liquidez, que acota la política monetaria.

**Idea**: la deflación eleva el valor real de las deudas y el tipo de interés real, y ambos efectos se refuerzan hasta convertir una recesión en una depresión.

**Razonamiento**: tres mecanismos actúan:

- *Deflación de deuda*: con deudas nominales fijas, la caída de precios eleva su valor real, los deudores venden activos para pagarlas, la venta deprime los precios de los activos y de los bienes, y la deflación se agrava. Así, cuanto más pagan los deudores, más deben.

- *Rigidez salarial*: si los salarios nominales no bajan, la deflación eleva el salario real y reduce el empleo, lo que reduce la demanda y alimenta la caída de precios.

- *Trampa de liquidez*: con tipos nominales en su límite inferior ($i = 0$), una deflación esperada eleva el tipo real, $r = i - \pi^e$, y contrae la inversión precisamente cuando la política monetaria convencional no puede bajar más los tipos. El banco central pierde así su instrumento, y la espiral deflacionista se autoalimenta.

La deflación recesiva es el principal escenario en el que la política monetaria convencional es inoperante, lo que explica que los bancos centrales la teman más que a una inflación moderada y que fijen su objetivo por encima de cero. Japón desde los años noventa es el caso de referencia, y la "japonización" el término con que se designa el riesgo de una deflación prolongada con tipos en cero y crecimiento débil.

**Conclusiones**:

- *Aspectos positivos*: explica por qué una deflación recesiva puede autoalimentarse y convertir una recesión en una depresión, y conecta la teoría de la inflación con las fricciones financieras y el límite inferior de los tipos. La deflación de deuda es, además, precursora del acelerador financiero (tema A42, apartado 2.3.2).

- *Limitaciones*: la deflación de deuda redistribuye riqueza de deudores a acreedores, de modo que su efecto agregado depende de que los deudores tengan una propensión al gasto mayor que la de los acreedores. Asimismo, la experiencia reciente muestra que el límite inferior efectivo se sitúa algo por debajo de cero y que las medidas no convencionales conservan cierta capacidad de actuación.

### 3.2.3. La política económica ante la deflación

> Los instrumentos no convencionales de la política monetaria se desarrollan en el tema A36, apartados 3.2 y 3.3.1, y los objetivos de PIB nominal y de nivel de precios, en los apartados 2.2.3 y 2.2.4.

Si la política monetaria convencional está agotada, la respuesta ha de venir de instrumentos no convencionales o de la política fiscal. La visión tradicional, keynesiana y monetarista, sostiene que es posible salir de la deflación con políticas expansivas, y el debate se ha centrado en cuáles:

1. *Política monetaria*: los instrumentos no convencionales actúan sobre las expectativas y sobre los tipos a largo plazo.

- *Forward guidance* (orientación prospectiva): compromete a mantener los tipos bajos.

- *Expansión cuantitativa* (*quantitative easing*, QE): compra activos a largo plazo para bajar sus tipos.

- *Tipos negativos*: rompen el límite inferior a costa de los bancos.

- *Objetivo de nivel de precios o de PIB nominal*: sustituye al objetivo de inflación y compromete al banco central a compensar la deflación pasada con inflación futura.

2. *Política fiscal*: recupera su eficacia, porque con tipos en cero el multiplicador es mayor, al no haber expulsión vía tipos de interés (tema A40).

La experiencia de Japón, de la zona del euro en 2014-2016 y de EE. UU. tras 2008 muestra que las políticas no convencionales evitan la espiral pero no siempre devuelven la inflación al objetivo, y que la combinación con la política fiscal es la que ha resultado eficaz. Por tanto, la deflación es el argumento más fuerte a favor de la coordinación de políticas.

### 3.3. Conclusiones

La hiperinflación y la deflación, los dos extremos de la dinámica de los precios, iluminan la teoría desde lados opuestos. La hiperinflación muestra que la inflación extrema tiene un origen fiscal y que las expectativas la aceleran, y la deflación, que las rigideces nominales y el límite inferior de los tipos hacen a la caída de precios más peligrosa que su subida. Ambos extremos justifican así un objetivo de inflación positivo y creíble.

Los episodios son, sin embargo, pocos y heterogéneos, lo que limita la generalización. Además, la frontera entre deflación benigna y recesiva no siempre es nítida en tiempo real, lo que dificulta la respuesta de política.

## Conclusiones

Los costes de la inflación recaen sobre la eficiencia, a través de las señales de precios y las distorsiones fiscales, y sobre el bienestar, a través de la redistribución y la incertidumbre. Aun así, sus beneficios, sobre todo el margen frente al límite inferior de los tipos, justifican un objetivo positivo y reducido. Cada escuela ha añadido un mecanismo a sus causas. La distinción keynesiana entre inflación de demanda y de costes orienta el diagnóstico. Friedman hace transitorio el *trade-off* de la curva de Phillips y Lucas lo limita a las sorpresas de precios. Barro y Gordon explican el sesgo inflacionista por la falta de credibilidad, y la nueva economía keynesiana atribuye la dinámica de corto plazo a las rigideces reales y nominales. A largo plazo, en cambio, la inflación sostenida es un fenómeno monetario y la curva de Phillips es vertical en la tasa natural. En la hiperinflación y la deflación, la credibilidad de las instituciones decide el desenlace más que la cantidad de dinero.

Cada causa de la inflación tiene su desarrollo propio. La tasa natural y la NAIRU proceden de los modelos del mercado de trabajo (tema A28), y la curva de Phillips neokeynesiana, del modelo de tres ecuaciones (tema A07). En el origen monetario y fiscal de las grandes inflaciones, la demanda de dinero de Cagan acota el señoraje (tema A35), y la teoría fiscal del nivel de precios explica la inflación cuando la deuda no está respaldada por superávits futuros (tema A39). Por último, la independencia del banco central y los objetivos de inflación, que resuelven la inconsistencia dinámica, organizan la política monetaria actual (tema A36).

El nivel del objetivo de inflación sigue en discusión. Blanchard, Dell'Ariccia y Mauro (2010) plantearon elevarlo hacia el 4 % para ganar margen sobre el límite inferior de los tipos, a costa de una inflación más costosa y peor anclada. También se discute el origen del repunte de 2021-2023. Bernanke y Blanchard (2025) lo atribuyen sobre todo a perturbaciones de precios, como las de la energía y los alimentos, con un efecto menor pero más persistente de la tensión del mercado de trabajo. Bianchi, Faccini y Melosi (2023), en cambio, lo explican con la teoría fiscal del nivel de precios, por el efecto de un gasto público no respaldado por superávits futuros.

# A42. Ciclos económicos

Teorías de los ciclos económicos: ciclos nominales y reales. Referencia al ciclo financiero y sus interrelaciones con el ciclo real.

## Introducción

**Introducción común.** Los temas A40, A41 y A42 comparten una introducción, que se desarrolla al comienzo del tema **A40**. Presenta las dos variables que concentran el análisis macroeconómico, la actividad y los precios, y la teoría del ciclo que estudia sus fluctuaciones, y relaciona con ellas las políticas monetaria y fiscal, la inflación y las teorías del ciclo.

Burns y Mitchell (1946) definieron los **ciclos económicos** como fluctuaciones recurrentes pero no periódicas de la actividad agregada, con fases de expansión y contracción que se suceden sin una duración fija. Su estudio interesa por sus costes, porque las recesiones destruyen empleo y renta y su huella puede durar años, y porque de su explicación depende la respuesta de la política económica. Medir el ciclo exige además separar la tendencia de las fluctuaciones, con métodos como el filtro de Hodrick-Prescott, que suponen una tendencia que varía suavemente, y la elección del método condiciona tanto el diagnóstico de la coyuntura como la contrastación de las teorías.

La teoría del ciclo se ha movido entre dos visiones. Para las teorías de equilibrio, el ciclo es la respuesta óptima de los agentes a perturbaciones, monetarias mal percibidas en Lucas o reales de productividad en Kydland y Prescott, y la política de estabilización tiene poco que corregir. Para las de desequilibrio, en cambio, las rigideces y los fallos de coordinación hacen que las fluctuaciones sean ineficientes y justifican la intervención. La nueva síntesis neoclásica reúne ambas tradiciones en el modelo de tres ecuaciones, la referencia actual de los bancos centrales.

Cada gran crisis, de la Gran Depresión a la crisis financiera global, ha renovado la teoría del ciclo, y la última añadió una tercera dimensión. Las fricciones financieras amplifican las perturbaciones, y el **ciclo financiero**, que Borio (2014) mide con el crédito al sector privado y los precios de los inmuebles, es más largo y más amplio que el ciclo económico y anticipa las crisis más graves. Las teorías se juzgan, en último término, por su capacidad para reproducir los **hechos estilizados** del ciclo, como la mayor volatilidad de la inversión que la del consumo, el carácter procíclico del empleo o la persistencia de las fluctuaciones.

El tema se organiza en tres partes:

- *Concepto y antecedentes*: el tema presenta el concepto de ciclo y sus antecedentes, desde las primeras aportaciones hasta los modelos del acelerador.

- *Teorías modernas*: se exponen los ciclos nominales del modelo de información imperfecta de Lucas, los ciclos reales de Kydland y Prescott y de Long y Plosser, los ciclos de desequilibrio de la nueva economía keynesiana, con las fricciones financieras y el ciclo financiero, y el modelo IS-PC-MR de la nueva síntesis neoclásica.

- *Análisis empírico*: el tema termina con la medición del ciclo, sus hechos estilizados y la contrastación de las teorías con esos hechos.

# 1. Concepto y antecedentes

El punto de partida es la definición de ciclo económico y los conceptos con que las teorías modernas lo analizan. Sobre esa base se repasan las primeras explicaciones del ciclo y los modelos del acelerador, que dieron su primera formalización.

## 1.1. Concepto de ciclo económico

Los **ciclos económicos** (*business cycles*) pueden definirse formalmente como fluctuaciones recurrentes pero no periódicas de la actividad económica agregada de un país, cuya duración es superior a un año (Burns y Mitchell, 1946). El análisis moderno los mide como desviaciones respecto de la tendencia secular.

Esta definición lleva a que las teorías modernas estudien tres conceptos:

1. *Impulso*: la perturbación que causa la desviación, sea exógena o endógena.

2. *Propagación*: los mecanismos de amplificación, difusión y persistencia que canalizan la perturbación a otras variables, entre ellos la suavización del consumo, los retardos en la comprensión de los *shocks*, la sustitución intertemporal de ocio y trabajo o la existencia de costes de ajuste.

3. *Comovimientos*: los movimientos conjuntos del PIB y de otras variables, medidos por el coeficiente de correlación $\rho$ entre el componente cíclico de cada variable y el del PIB. Según su signo, las variables son:

- *Procíclicas* ($\rho > 0$): evolucionan en la misma dirección que el ciclo (p. ej. el consumo, la inversión, el empleo, las importaciones o los beneficios empresariales).

- *Contracíclicas* ($\rho < 0$): evolucionan en sentido opuesto (p. ej. la tasa de desempleo, por la ley de Okun, las existencias no deseadas o el déficit público, por los estabilizadores automáticos).

- *Acíclicas* ($\rho \approx 0$): no muestran una correlación sistemática con el ciclo (p. ej. el gasto público en consumo, el salario real en muchas estimaciones o los tipos de interés reales).

Además, la correlación puede calcularse desfasando la variable $j$ periodos respecto del PIB, $\rho(x_t, y_{t+j})$, lo que permite clasificar las variables según su temporalidad:

- *Coincidentes* ($j = 0$): la correlación es máxima con el PIB contemporáneo (p. ej. la producción industrial, el empleo o las ventas al por menor).

- *Adelantadas* (*leads*, $j > 0$): la correlación es máxima con el PIB de periodos posteriores, es decir, la variable se mueve antes que el ciclo. Por ello son especialmente útiles para la predicción (p. ej. la inversión residencial, los índices de confianza como el de gestores de compras (PMI), la pendiente de la curva de tipos o los índices bursátiles).

- *Retrasadas* (*lags*, $j < 0$): la correlación es máxima con el PIB de periodos anteriores, es decir, la variable se mueve después del ciclo (p. ej. la inflación, los tipos de interés a corto plazo, los costes laborales unitarios o, ligeramente, el consumo).

## 1.2. Primeras aportaciones

Hasta la década de 1890 los economistas no abordaron de forma sistemática el análisis de la depresión o del ciclo económico. Una excepción es Clément Juglar (1862), que atribuyó el ciclo a fuerzas internas del propio sistema económico, no a factores externos, y propuso tres fases: prosperidad, crisis y liquidación.

Las aportaciones posteriores a Juglar pueden clasificarse, siguiendo a Gottfried von Haberler (1937), según la naturaleza del impulso que pone en marcha el ciclo.

### 1.2.1. Teorías no monetarias

**Teorías de fenómenos naturales** (Jevons, 1878; Moore, 1914): los ciclos responden a causas meteorológicas periódicas (las manchas solares, con un ciclo de unos once años, en Jevons; las lluvias, en Moore). Esas causas afectan a las cosechas y, a través de las rentas y los precios agrícolas, al conjunto de la actividad económica. Aunque hoy parecen inverosímiles, introdujeron la idea de un impulso exógeno regular y, por tanto, predecible.

**Teoría de la desproporcionalidad o de la sobreinversión no monetaria** (Tugan-Baranowsky, 1894; Spiethoff, 1902; Cassel, 1918b): el ciclo es un ciclo de crédito e inversión. El capital prestable acumulado durante la depresión se vuelca en la inversión fija y desata un auge concentrado en el sector de bienes de capital, que crece más deprisa que el de bienes de consumo (desproporción entre sectores). El auge dura hasta que el agotamiento de los fondos prestables interrumpe la inversión y provoca una crisis endógena al capitalismo.

**Teoría del infraconsumo** (Hobson, 1909): la desigualdad en la distribución de la renta lleva a que las rentas altas ahorren más de lo que puede invertirse rentablemente. Ese ahorro sobrante se convierte en una capacidad productiva excesiva, cuya producción el consumo no llega a absorber, y el desajuste termina en depresión. De ahí que el remedio sea redistributivo.

**Principio de aceleración** (Aftalion, 1909; Clark, 1917): la demanda de bienes de capital depende de la variación de la demanda de bienes de consumo, no de su nivel, por lo que pequeñas fluctuaciones del consumo se amplifican en grandes fluctuaciones de la inversión. Los retardos en la construcción del capital hacen que la respuesta llegue tarde y resulte excesiva. Sobre este mecanismo de propagación se construyen los modelos del acelerador y de ciclos endógenos.

**Teorías psicológicas** (Marshall, 1923; Pigou, 1927): las fluctuaciones se propagan por errores de optimismo y de pesimismo de los empresarios que se contagian y se generan mutuamente (el error de optimismo del auge engendra el de pesimismo de la crisis).

**Teoría keynesiana** (Keynes, 1936): el ciclo tiene su origen en las fluctuaciones de la eficiencia marginal del capital, que depende de expectativas formadas bajo incertidumbre y, por ello, volátiles (los *espíritus animales*). Por ejemplo, un colapso de la confianza hunde la inversión, y el multiplicador transmite la caída a la renta y al empleo. Sobre estas ideas se levantan después las teorías keynesianas que explican el ciclo endógenamente, entre ellas la de Kaldor.

**Teoría de la innovación** (Schumpeter, 1939): las innovaciones aparecen en racimos y se financian con crédito bancario, y su imitación por otros empresarios difunde la inversión y genera el auge. Al generalizarse la innovación, los beneficios extraordinarios se erosionan, el crédito se devuelve y los precios caen. La recesión es el proceso por el que la economía absorbe la innovación y liquida las estructuras obsoletas (la *destrucción creativa*). Schumpeter superpone tres ciclos de distinta duración (Kitchin, Juglar y Kondratiev).

### 1.2.2. Teorías monetarias

**Teoría puramente monetaria** (Hawtrey, 1919): el ciclo es un fenómeno del crédito. La expansión del crédito bancario reduce el tipo de interés y estimula la acumulación de existencias por parte de los comerciantes, cuya demanda es muy sensible al coste de financiación. La expansión continúa hasta que la pérdida de reservas obliga a los bancos a restringir el crédito, y la contracción sigue el camino inverso.

**Teoría de la sobreinversión monetaria** (Mises, 1912; Hayek, 1929): la expansión del crédito bancario sitúa el tipo de interés de mercado por debajo del natural (el que igualaría ahorro e inversión voluntarios). El tipo artificialmente bajo induce a alargar la estructura de producción (más bienes de capital, procesos más indirectos) sin que haya aumentado el ahorro real, de modo que el auge se financia con ahorro forzoso. Cuando la expansión crediticia se detiene, los proyectos resultan insostenibles (malinversión) y la crisis es la liquidación necesaria para readaptar la estructura productiva a las preferencias de los consumidores[1].

**Teoría monetarista** (Friedman y Schwartz, 1963): los ciclos tienen un origen eminentemente monetario, dada la estabilidad de la velocidad de circulación del dinero. Las variaciones de la oferta monetaria se transmiten al gasto nominal y, con retardos largos y variables, a la producción.

## 1.3. Modelos del acelerador y ciclos endógenos

Las teorías del apartado anterior identificaban impulsos y mecanismos plausibles, pero no los formalizaban. Con la *Teoría general* aparece el multiplicador, que, unido al principio de aceleración de Aftalion y Clark, permite hacerlo.

Combinados, ambos forman un sistema dinámico que se retroalimenta y cuya expresión matemática es una ecuación en diferencias, de modo que el ciclo se deriva de la estructura formal del modelo. Se presentan aquí el modelo de Samuelson, su revisión por Hicks y la alternativa de Kaldor.

### 1.3.1. Modelo del oscilador (Samuelson, 1939)

**Origen**: Paul Samuelson (1939) formaliza la combinación del multiplicador y el acelerador.

**Idea**: las fluctuaciones nacen de la interacción entre el multiplicador keynesiano y el principio de aceleración. Un impulso al gasto autónomo eleva la renta más que el propio impulso (multiplicador) y ese aumento induce nueva inversión (acelerador), y así sucesivamente.

**Planteamiento**: sobre el marco keynesiano, Samuelson especifica dos funciones de comportamiento:

- *Función de consumo keynesiana*: el consumo tiene una parte autónoma, $C_0$, y otra endógena, cuya respuesta a la renta mide la propensión marginal a consumir, $c$:

$$C_t = C_0 + cY_{t-1}, \quad c \in (0,1)$$

[1] De ahí la recomendación austriaca de no intervenir, porque sostener el auge con más crédito solo pospone y agrava la liquidación.

- *Inversión de Aftalion-Clark*: a una inversión autónoma constante se suma otra inducida, que, conforme al principio del acelerador, reacciona a las variaciones del consumo:

$$I_t = I_0 + v\,(C_t - C_{t-1}), \quad v > 0$$

donde $v$ es la sensibilidad de la inversión a las variaciones del consumo (el acelerador). A diferencia de la función de inversión keynesiana estándar, que responde al tipo de interés, la inversión inducida depende aquí de la variación del consumo, no de su nivel. Por ello, es positiva mientras el consumo crece, se anula cuando se estabiliza y se vuelve negativa cuando cae. Esta dependencia introduce la dinámica en el modelo.

**Equilibrio**: llevadas ambas funciones a la condición de equilibrio, $Y_t = C_t + I_t$, resulta una ecuación en diferencias de segundo orden, que vincula la producción corriente con sus dos valores retardados:

$$Y_t - c(1+v)Y_{t-1} + vcY_{t-2} = I_0 + C_0$$

**Estática comparativa**: el nivel estacionario de la renta es $Y^* = (C_0 + I_0)/(1-c)$ (Anexo A.1). Un aumento permanente del gasto autónomo lo eleva en el multiplicador keynesiano, $\partial Y^*/\partial I_0 = 1/(1-c)$, que no depende del acelerador. El acelerador solo decide cómo se recorre el camino hacia el nuevo nivel (Dinámica), de forma monótona o con ciclos amortiguados en las regiones A y D, y sin alcanzarlo nunca en las regiones B, C y E.

**Dinámica**: la ecuación de equilibrio es la *ecuación fundamental del ciclo de Samuelson*, cuya dinámica se lee en las raíces de su ecuación característica, $r^2 - c(1+v)\,r + vc = 0$ (véase el Anexo):

- *Raíces reales*: la trayectoria no presenta oscilaciones (no hay ciclos) y es convergente o divergente según su valor absoluto sea menor o mayor que la unidad.

- *Raíces complejas*: la trayectoria es cíclica, y el ciclo es amortiguado, constante o explosivo según su módulo, $\sqrt{vc}$, sea menor, igual o mayor que la unidad.

Como la dinámica depende solo de la propensión marginal a consumir ($c$), que determina el multiplicador, y del acelerador ($v$), las situaciones posibles se clasifican en el plano acelerador-propensión marginal a consumir, $v$-$c$ (figura A42.1):

1. *Frontera entre raíces reales y complejas*, $c = 4v/(1+v)^2$: recoge los pares con discriminante nulo, $c(1+v)^2 = 4v$. Por encima de ella las raíces son reales y, por debajo, complejas.

2. *Hipérbola*, $c = 1/v$: recoge las combinaciones con $vc = 1$, en las que el módulo de las raíces complejas es la unidad, y separa las trayectorias convergentes de las explosivas.

3. *Regiones*: las dos fronteras delimitan cinco regiones:

- *Región A*: convergencia sin oscilaciones (sin ciclos).

- *Región B*: divergencia sin oscilaciones (sin ciclos).

- *Región C*: ciclos de amplitud constante (sobre la hipérbola $c = 1/v$).

- *Región D*: ciclos amortiguados.

- *Región E*: ciclos explosivos.

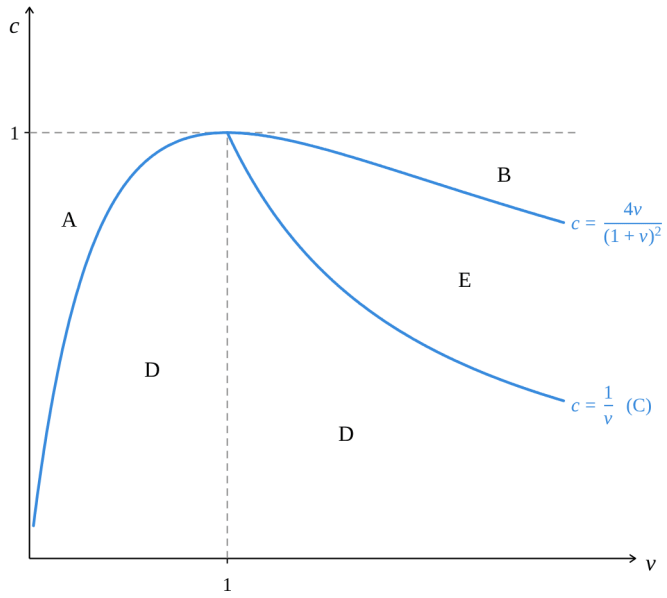



Figura A42.1: Modelo del oscilador (Samuelson, 1939). El plano de parámetros (acelerador, propensión marginal a consumir) se divide en regiones con trayectorias monótonas, cíclicas amortiguadas o explosivas.

**Conclusiones**:

- *Aspectos positivos*: es la primera aproximación formal basada en la concepción tradicional del ciclo e introduce la interacción multiplicador-acelerador como mecanismo de propagación.

- *Limitaciones*: los valores de los parámetros necesarios para que los ciclos sean estables (amortiguados) no casaban con la evidencia empírica. Con una propensión al consumo realista ($c \approx 0{,}8$) y un acelerador superior a $1/c = 1{,}25$, se tiene $vc > 1$ y la economía cae en la región E (o, si $v$ supera aproximadamente 2,62, en la B). El modelo predice, por tanto, ciclos explosivos (o una divergencia monótona) y no los ciclos persistentes pero acotados que se observan.

### 1.3.2. Modelo del oscilador revisado (Hicks, 1950)

**Origen**: John Hicks (1950) parte del modelo de Samuelson y, para conciliarlo con la evidencia empírica, lo hace explosivo (región E) pero lo contiene con límites no lineales.

**Idea**: el ciclo nace explosivo y rebota entre un techo y un suelo. Como ambos crecen con la tendencia (la senda de equilibrio $Y_t^*$, que crece a la tasa $\bar{g}_Y$), el ciclo se superpone a una senda de crecimiento en lugar de oscilar en torno a un nivel fijo.

**Planteamiento**: Hicks mantiene la estructura de Samuelson y añade dos restricciones no lineales:

- *Función de consumo*: se mantiene la expresión keynesiana, $C_t = C_0 + cY_{t-1}$.

▪ *Función de inversión y suelo*: la inversión tiene un componente inducido por el acelerador y un componente autónomo que crece a una tasa constante y exógena,

$$I_t = I_{0,t} + v\,(Y_{t-1} - Y_{t-2}), \quad I_{0,t} = I_0(1+\bar{g}_Y)^t$$

El *suelo* procede de que la desinversión inducida está acotada. La inversión bruta inducida no puede ser negativa, por lo que el *stock* de capital solo puede reducirse al ritmo de la depreciación. Mientras la inversión inducida toca ese mínimo, la inversión autónoma sigue creciendo, lo que define un nivel mínimo de renta, el suelo $\underline{Y}_t$, que crece con la tendencia.

▪ *Restricción de capacidad y techo*: cuando todos los factores están plenamente empleados, la producción topa con su *techo*, $Y_t \leq \bar{Y}_t = \bar{Y}_0(1+\bar{g}_Y)^t$, que avanza, como la inversión autónoma, al ritmo exógeno $\bar{g}_Y$ con que crece la dotación de factores.

**Desarrollo gráfico**: el modelo de Hicks se representa en el plano tiempo-logaritmo de la renta, $t$-$\ln Y$, en el que toda magnitud que crece a tasa constante es una recta (figura A42.2):

1. *Techo*, $\ln \bar{Y}_t$: la producción de pleno empleo, que crece a la tasa $\bar{g}_Y$.

2. *Tendencia*, $\ln Y_t^*$: la senda de equilibrio, que crece a la misma tasa.

3. *Suelo*, $\ln \underline{Y}_t$: el nivel mínimo de renta que sostiene la inversión autónoma.

4. *Senda del PIB*, $\ln Y_t$: oscila entre el techo y el suelo en torno a la tendencia.

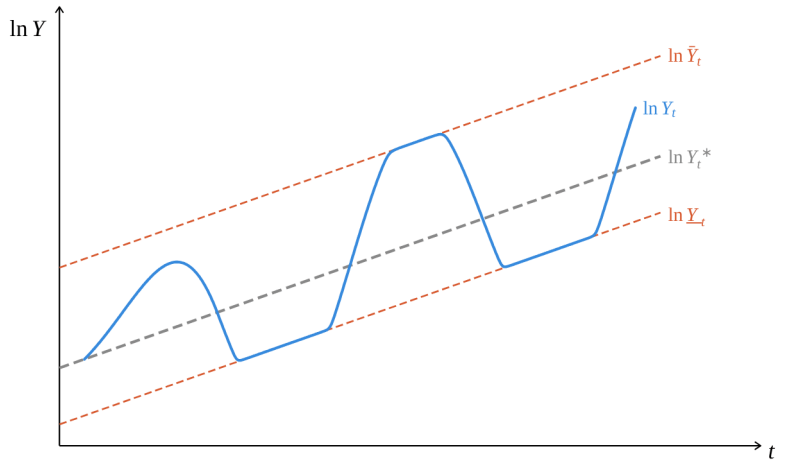



Figura A42.2: Modelo de Hicks (1950). El ciclo queda limitado por el techo de pleno empleo y el suelo de la inversión autónoma.

**Dinámica**: en la fase expansiva, el PIB crece por encima de la tendencia impulsado por la interacción del acelerador y el multiplicador hasta alcanzar el techo. A partir de ahí, el crecimiento se limita al ritmo de la oferta de factores y, como el acelerador solo responde a la variación de la renta, la inversión inducida cae y el PIB comienza a caer.

La caída se detiene en el suelo, donde la desinversión ya no puede acelerarse mientras la inversión autónoma sigue creciendo. En consecuencia, la renta vuelve a crecer, el acelerador se reactiva y se inicia un nuevo ciclo.

**Conclusiones**:

- *Aspectos positivos*: introduce límites realistas que impiden las oscilaciones explosivas del modelo de Samuelson e integra el ciclo en una senda de crecimiento.

- *Limitaciones*: se basa en el principio del acelerador rígido y en una interacción mecánica entre el multiplicador y la inversión que no ofrece una base satisfactoria para explicar los ciclos. Además, el techo y el suelo son restricciones impuestas, no derivadas del comportamiento de los agentes.

### 1.3.3. Ciclos endógenos (Kaldor, 1940)

**Origen**: Nicholas Kaldor (1940) ofrece la primera teoría keynesiana del ciclo plenamente endógeno, anterior a Hicks y sin recurrir al acelerador. Toma de Keynes las funciones de ahorro e inversión como determinantes de la renta y de Kalecki la idea de que la inversión depende del *stock* de capital acumulado.

**Idea**: el ahorro y la inversión planeados no guardan una relación lineal con el nivel de actividad, $Y$. Esa forma curva, sumada a los desplazamientos que la acumulación de capital provoca en ambas funciones, permite que el sistema oscile por sí solo, sin impulsos externos.

**Razonamiento**: el modelo muestra tres equilibrios, de los cuales solo los extremos son estables, y solo temporalmente:

- *Actividad baja*: el exceso de capacidad mantiene reducida la propensión marginal a invertir, mientras que la resistencia de los hogares a recortar su consumo, que les lleva a desahorrar cuando la renta cae, eleva la de ahorrar, y por eso este equilibrio es estable.

- *Actividad intermedia*: la propensión marginal a invertir supera a la de ahorrar, de modo que el equilibrio es inestable y cualquier desviación se amplifica hacia uno de los extremos.

- *Actividad alta*: ahora los límites de oferta deprimen la propensión marginal a invertir, y la saturación del consumo mantiene elevada la de ahorrar, con lo que también este equilibrio es estable.

La acumulación o desacumulación de capital desplaza las funciones de ahorro e inversión y vuelve inestables los equilibrios extremos, de modo que las fluctuaciones se generan de forma endógena (figura en el Anexo, apartado A.6).

**Conclusiones**:

- *Aspectos positivos*: muestra que la no linealidad de las funciones de ahorro e inversión basta para generar fluctuaciones endógenas y persistentes sin impulsos externos ni valores particulares de los parámetros, y anticipa el análisis del ciclo mediante la dinámica no lineal.

- *Limitaciones*: la forma de las funciones de ahorro e inversión se postula y no se deriva del comportamiento optimizador de los agentes, y el modelo genera un ciclo regular, mientras que los ciclos observados son recurrentes pero no periódicos.

> El principio del acelerador y sus versiones flexibles se estudian, como teorías de la inversión, en el tema A34.

## 1.4. Conclusiones

Las teorías tradicionales dejaron dos legados que el análisis moderno conserva, la distinción entre el impulso que desencadena el ciclo y el mecanismo que lo propaga y la primera formalización de este último en los modelos del acelerador. Las primeras aportaciones identifican, sin embargo, mecanismos plausibles que no formalizan ni contrastan, y los modelos del acelerador, que sí los formalizan, solo generan ciclos persistentes y acotados mediante supuestos *ad hoc*. Además, estos modelos descansan en reglas de comportamiento mecánicas, sin fundamento en la optimización de los agentes ni en la formación de sus expectativas.

La superación de ambas carencias, la microeconómica y la empírica, define, respectivamente, las teorías modernas del ciclo (apartado 2) y el análisis empírico con el que se contrastan (apartado 3).

## 2. Teorías modernas del ciclo

Las teorías tradicionales no son coherentes con el ciclo como fenómeno empírico. Por ello, y a raíz de la crítica de Lucas (1976), las teorías modernas del ciclo se basan en las aportaciones de la nueva macroeconomía clásica (NMC) y de la nueva economía keynesiana (NEK).

Las *teorías del ciclo de equilibrio* son la aportación de la NMC, con dos variantes según el impulso sea nominal (Lucas) o real (teoría del ciclo real). Las *teorías del ciclo de desequilibrio*, aportación de la NEK, conciben el ciclo como aumentos o disminuciones de los excesos de oferta y de demanda a lo largo del tiempo. La nueva síntesis neoclásica integra ambas tradiciones.

> Este apartado desarrolla los modelos como teorías del ciclo. Los fundamentos de las dos escuelas de las que proceden y sus implicaciones de política económica se estudian en los temas A06 (nueva macroeconomía clásica) y A07 (nueva economía keynesiana).

### 2.1. Ciclos nominales: el modelo de información imperfecta (Lucas, 1972; 1973; 1975)

**Origen**: el modelo de oferta agregada de Lucas (1972; 1973; 1975) precedió al desarrollo de los modelos del ciclo real. Se apoya en dos trabajos previos: el de Edmund Phelps (1969), del que nace la *parábola de las islas*, y el de Lucas con Leonard Rapping (1969), del que toma la importancia de la *elasticidad de sustitución intertemporal* en las decisiones de trabajo[2].

**Idea**: Lucas puso el énfasis en los determinantes nominales del ciclo. El impulso es una perturbación monetaria, el mecanismo de transmisión es el problema de extracción de señales y la persistencia no queda bien explicada.

#### 2.1.1. Supuestos

> Los supuestos y la derivación del modelo se desarrollan en el tema A06, apartados 2.1.1 y 2.1.2.

**Entorno**: el modelo supone un mundo neoclásico (competencia perfecta y plena flexibilidad de precios, salarios y tipo de interés), pero con información imperfecta. Los agentes, distribuidos en "islas", conocen el precio de su mercado, $p_{it}$, pero no el nivel general de precios, $p_t$. No pueden distinguir, por ello, un *shock específico* (un cambio en la demanda de su producto, ante el que lo racional es alterar la producción) de un *shock general* de demanda agregada (ante el que conviene no alterarla), dificultad denominada *problema de extracción de señales*. Todas las variables se expresan en logaritmos.

**Agentes**: cada agente produce y vende un único bien en su mercado y forma su expectativa del nivel general de precios según la hipótesis de expectativas racionales (HER), $p_t^e \equiv \mathbb{E}_{t-1}[p_t]$. En los periodos de mayor precio relativo sacrifica ocio para producir y consumir más (sustitución intertemporal de ocio y trabajo). En cuanto a la autoridad monetaria, la cantidad de dinero nominal tiene un componente previsible y otro no previsible, $\varepsilon_t$, que recoge las sorpresas monetarias (la política monetaria no anticipada por los agentes o los errores del banco central al controlar el dinero):

[2] En el artículo de Lucas y Rapping, no obstante, todavía se considera que la formación de expectativas se realiza de forma adaptativa.

$$m_t = \mathbb{E}_{t-1}[m_t \mid \Omega_{t-1}] + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_m^2)$$

### 2.1.2. Modelo

**Planteamiento**: el modelo se resume en una función de oferta agregada, derivada del problema de extracción de señales, y una función de demanda agregada, derivada de la teoría cuantitativa.

**Oferta agregada**: con lo aprendido en el pasado, los productores estiman la probabilidad $\vartheta \in [0, 1]$ de que un *shock* sea específico. Esa probabilidad coincide con el peso de la varianza de los precios relativos en la de todos los precios:

$$\vartheta = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_p^2}$$

El grado de variabilidad de los precios determina $\vartheta$:

- *Economía habituada a precios muy volátiles* ($\sigma_p^2$ grande): el productor atribuye casi todo cambio de precios a *shocks* globales ($\vartheta \to 0$) y apenas ajusta su oferta.
- *Economía con precios estables*: sucede lo opuesto ($\vartheta \to 1$).

La oferta de cada mercado suma un componente tendencial común ($\bar{y}_t$, el *output* asociado a la tasa natural de paro) y un componente cíclico. El componente cíclico depende del precio relativo percibido, de la probabilidad de que el *shock* sea específico ($\vartheta$) y de la capacidad de los productores para ajustar su producción ($\eta$). Formalmente, $y_{it} = \bar{y}_t + \zeta\,(p_{it} - p_t^e)$, con $\zeta \equiv \vartheta\eta > 0$. Agregando (el componente específico promedia a cero) y añadiendo los *shocks* exógenos de oferta, $\varepsilon_t^s$, se obtiene la oferta agregada de Lucas:

$$\text{OA: } y_t = \bar{y}_t + \zeta\,(p_t - p_t^e) + \varepsilon_t^s$$

Salvo por los *shocks* de oferta, no hay desviación respecto del nivel de largo plazo si $\vartheta = 0$ (p. ej. hiperinflación), si $\eta = 0$ o si los cambios en el nivel de precios están completamente anticipados ($p_t = p_t^e$).

**Demanda agregada**: de forma simplificada, se parte de la teoría cuantitativa del dinero, $m_t + v_t = p_t + y_t$, donde $v_t$ es una perturbación de la velocidad de circulación (de esperanza nula) que recoge los *shocks* exógenos de demanda[3]:

$$\text{DA: } y_t = m_t - p_t + v_t$$

**Equilibrio**: igualando oferta y demanda agregadas, imponiendo la HER y despejando el precio, se obtiene la producción como función de los *shocks* de la economía[4]:

[3] Aunque en el modelo original $m$ es la cantidad de dinero, puede interpretarse como cualquier variable de política que desplace la demanda agregada (en la práctica, el tipo de interés fijado por el banco central).

[4] Derivación. (i) En equilibrio, OA = DA: $\bar{y}_t + \zeta(p_t - \mathbb{E}_{t-1}p_t) + \varepsilon_t^s = m_t - p_t + v_t$. (ii) Desagregando el dinero ($m_t = \mathbb{E}_{t-1}m_t + \varepsilon_t$) y tomando expectativas con la HER, todo lo anticipable se cancela, de modo que la sorpresa de precios depende únicamente de los componentes no anticipados: $p_t - \mathbb{E}_{t-1}p_t = \frac{1}{1+\zeta}(\varepsilon_t + v_t - \varepsilon_t^s)$. (iii) Sustituyendo esta expresión en la OA se obtiene la producción de equilibrio como función de los *shocks* de la economía.

$$y_t = \bar{y}_t + \frac{\zeta}{1+\zeta}(\varepsilon_t + v_t - \varepsilon_t^s) + \varepsilon_t^s \;\rightarrow\; \tilde{y}_t = \frac{\zeta}{1+\zeta}(\varepsilon_t + v_t - \varepsilon_t^s) + \varepsilon_t^s$$

**Desarrollo gráfico**: el modelo de Lucas se representa en el plano renta-nivel general de precios (figura A42.3):

1. *Oferta agregada de largo plazo*, $\text{OA}_{\text{LP}}$: vertical en la producción tendencial, $\bar{Y}$, porque los cambios de precios anticipados ($p_t = p_t^e$) no alteran la producción.

2. *Oferta agregada de corto plazo*, $\text{OA}_{\text{CP},0}$: creciente, porque solo las sorpresas de precios llevan la producción por encima de su tendencia.

3. *Demanda agregada*, $\text{DA}_0$: decreciente en el nivel de precios, dada la cantidad de dinero, porque un precio mayor reduce los saldos reales ($y_t = m_t - p_t + v_t$).

4. *Equilibrio*: en el punto 0, sobre la tendencia, con precios $P_0$.

**Estática comparativa**:

$$\frac{\partial y_t}{\partial \varepsilon_t} = \frac{\partial y_t}{\partial v_t} = \frac{\zeta}{1+\zeta} > 0, \qquad \frac{\partial y_t}{\partial \varepsilon_t^s} = \frac{1}{1+\zeta} > 0$$

Los efectos dependen del tipo de perturbación:

- *Política monetaria y perturbaciones de demanda*: un cambio anticipado se traslada íntegramente a los precios y es neutral, y solo la parte no anticipada ($\varepsilon_t$, $v_t$) altera transitoriamente la producción.

- *Perturbación de oferta*: los *shocks* de oferta ($\varepsilon_t^s$) afectan a la propia capacidad productiva, por lo que sus efectos no son necesariamente transitorios.

En la figura A42.3, una expansión monetaria ($\Delta M^s > 0$) desplaza la demanda agregada de $\text{DA}_0$ a $\text{DA}_1$, con efectos distintos según los agentes la anticipen o no:

- *Expansión anticipada*: la oferta de corto plazo se desplaza en la misma medida, a $\text{OA}_{\text{CP},1}$, y la economía pasa del punto 0 al 1, sobre la tendencia y con precios más altos (neutralidad).

- *Expansión no anticipada*: los agentes forman sus expectativas con la información de $t-1$, la oferta de corto plazo no se desplaza (o solo parcialmente) y la producción supera transitoriamente su tendencia en el punto $0'$. Cuando conocen la verdadera cuantía de la perturbación, revisan sus expectativas ($\Delta P^e > 0$), la oferta de corto plazo se desplaza a $\text{OA}_{\text{CP},1}$ y la economía pasa de $0'$ a 1. Se produce así un ciclo, resultado de una sucesión de equilibrios.

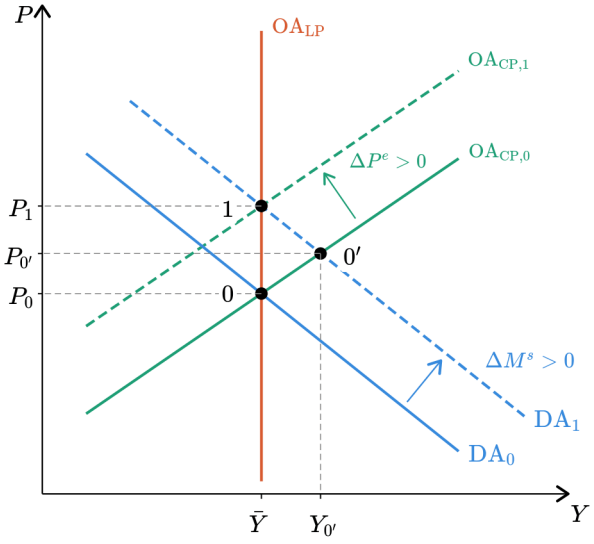



Figura A42.3: Modelo de Lucas. Solo una expansión monetaria no anticipada desplaza la producción por encima de su nivel natural, mientras que una expansión anticipada se traslada íntegramente a precios. El gráfico representa las variables en niveles ($P$, $Y$, $\bar{Y}$) y el texto las expresa en logaritmos.

### 2.1.3. Conclusiones

**Aspectos positivos**:

▪ *Primer modelo microfundamentado del ciclo*: con expectativas racionales, la oferta agregada se deriva del comportamiento optimizador de agentes racionales, y el modelo formaliza el papel de la información incompleta.

▪ *No neutralidad sin rigideces*: obtiene efectos reales de los *shocks* monetarios a corto plazo con precios plenamente flexibles y vaciado continuo, conciliando la evidencia de corto plazo con la hipótesis de la tasa natural y la neutralidad de largo plazo.

▪ *Predicción contrastable*: la dependencia de $\vartheta$ respecto de $\sigma_p^2$ implica que la disyuntiva producción-inflación es menos favorable (un mismo impulso nominal tiene menos efectos reales) en los países con demanda agregada más volátil, predicción que Lucas (1973) respaldó con datos de dieciocho países. Barro (1977) halló, además, que solo el crecimiento monetario no anticipado afecta al desempleo. Ball, Mankiw y Romer (1988) mostraron después que la pendiente depende sobre todo de la inflación media y no de la varianza de la demanda, como predicen los costes de menú (tema A07, apartado 1.3.2).

**Limitaciones**:

▪ *Persistencia*: no explica la duración de los ciclos, debido a la actuación de la HER y a la ausencia de un mecanismo de propagación. Por ello, para que el ajuste no sea inmediato, deben introducirse mecanismos *ad hoc* (costes de ajuste, retardos de información).

- *Plausibilidad de la fricción informativa*: dada la difusión rápida y barata de las estadísticas de precios, es difícil sostener que los agentes confundan de forma persistente las variaciones relativas y agregadas hasta el punto de generar ciclos.

- *Naturaleza real de muchos ciclos*: no explica las perturbaciones reales, como las crisis del petróleo.

- *Evidencia empírica*: las sorpresas monetarias solo dan cuenta de una parte de las fluctuaciones observadas, y el modelo necesita una oferta de trabajo más elástica de lo que indican los estudios microeconómicos.

Estas limitaciones, y en especial la incapacidad para generar persistencia, motivaron el segundo enfoque de la NMC sobre el ciclo.

## 2.2. Ciclos reales: la teoría del ciclo real (Kydland y Prescott, 1982; Long y Plosser, 1983)

**Origen**: Kydland y Prescott (1982) y Long y Plosser (1983) sientan las bases de la teoría del ciclo real (TCR) intentando replicar la dinámica observada en la economía estadounidense entre 1950 y 1979.

**Idea**: los ciclos se generan no por un problema de extracción de señales, sino por el comportamiento optimizador de los agentes ante perturbaciones reales, de carácter tecnológico, estocásticas y autocorrelacionadas. En el modelo se distinguen tres elementos:

- *Impulso*: las perturbaciones tecnológicas.

- *Mecanismo de transmisión*: el comportamiento optimizador (concretamente, la sustitución intertemporal del consumo y el ocio y la acumulación de capital).

- *Persistencia*: se explica por la autocorrelación de las perturbaciones y por los largos periodos de realización de los proyectos de inversión (*time to build*).

### 2.2.1. Supuestos

> Los supuestos y la solución analítica del modelo básico se desarrollan en el tema A06, apartado 2.2. Aquí se recogen para estudiar su dinámica.

**Entorno**: la economía es cerrada, con información perfecta, tiempo discreto y horizonte infinito, competencia perfecta, mercados completos y plena flexibilidad de precios, salarios y tipo de interés[5].

**Agentes**: idénticos, con expectativas racionales y vida infinita, que consumen, ofrecen trabajo y alquilan a las empresas el capital, su único activo. Su utilidad, logarítmica y separable en el tiempo, depende del consumo y del ocio[6]. No hay sector público.

**Tecnología**: la producción sigue una función Cobb-Douglas con progreso técnico aumentador del trabajo y rendimientos constantes a escala, afectada por perturbaciones tecnológicas autocorrelacionadas. La depreciación completa ($\delta = 1$), el caso de Long y Plosser, permite resolver el modelo analíticamente.

[5] Los agentes pueden, así, prestar o endeudarse sin límite entre periodos al tipo de mercado $r$.

[6] La separabilidad temporal facilita el tratamiento analítico, pero excluye la inercia y los hábitos de consumo que muestran los estudios empíricos.

### 2.2.2. Modelo

El modelo se resume en cuatro ecuaciones de comportamiento de hogares y empresas, tres ecuaciones que describen la tecnología y una condición de equilibrio competitivo.

**Hogares**: el hogar representativo consume, trabaja y ahorra.

▪ *Planteamiento*: maximiza el valor actual esperado de su utilidad, sujeto a su restricción presupuestaria intertemporal[7]:

$$\max_{c_t,\, l_t,\, k_{t+1}} U = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t, 1-l_t), \qquad \beta \in (0,1)$$
$$\text{s.a.:} \begin{cases} c_t + k_{t+1} - k_t = w_t l_t + r_t k_t \\ k_0 \text{ dado} \end{cases}$$

donde $c_t$ es el consumo, $l_t$ el trabajo y $k_{t+1}$ el capital del periodo siguiente (las variables de decisión), $u(\cdot)$ la utilidad instantánea en consumo y ocio, $\beta$ el factor de descuento, $w_t$ el salario, $r_t$ el rendimiento del capital y $k_0$ el capital inicial.

▪ *CPO*: con solución interior y la restricción agotada, del lagrangiano resultan la condición intratemporal entre ocio y consumo (la oferta de trabajo) y la ecuación de Euler estocástica. La condición de transversalidad, según la cual el valor actual esperado del capital que queda al final del horizonte es nulo, completa las condiciones de óptimo. La ecuación de Euler iguala la utilidad marginal del consumo presente con la esperada de invertir esa unidad al tipo $r_{t+1}$ y consumirla el periodo siguiente[8]:

$$\text{RMS}_{c_t,\, 1-l_t} = \frac{\partial u/\partial(1-l_t)}{\partial u/\partial c_t} = w_t, \qquad \frac{\partial u}{\partial c_t} = \beta\, \mathbb{E}_t\left[\frac{\partial u}{\partial c_{t+1}}\,(1+r_{t+1})\right]$$
$$\lim_{t\to\infty} \mathbb{E}_0\left[\beta^t \frac{\partial u}{\partial c_t}\, k_{t+1}\right] = 0$$

▪ *CSO*: la concavidad estricta de $u$ en consumo y ocio y la de la tecnología garantizan que las CPO, con la condición de transversalidad, son necesarias y suficientes.

**Empresas**: producen con trabajo y capital.

▪ *Planteamiento*: maximizan su beneficio[9], $\max_{l_t,k_t} \Pi_t = y_t - w_t l_t - (r_t + \delta)k_t$.

▪ *CPO*: cada factor se remunera según su productividad marginal, el trabajo al salario y el capital a su coste de uso, es decir, $\partial f/\partial l_t = w_t$ y $\partial f/\partial k_t = r_t + \delta$.

▪ *CSO*: la tecnología Cobb-Douglas es cóncava en $(k,l)$ y estrictamente cóncava en cada factor, luego las CPO caracterizan el máximo.

[7] Los autores de la NMC se apoyan en el modelo de Ramsey-Cass-Koopmans (tema A43) con dos diferencias fundamentales: la consideración del ocio en la función de utilidad y la aleatoriedad de la rentabilidad y del salario debida a la aleatoriedad de la tecnología. La segunda diferencia introduce la esperanza en la función objetivo.

[8] La expectativa de $r_{t+1}$ no es independiente de $c_{t+1}$ (a diferencia del modelo de Hall), ya que $r_t$ no es constante y depende de los *shocks* tecnológicos del modelo.

[9] Se considera que el *shock* se realiza antes de que las empresas tomen sus decisiones de empleo y capital, por lo que no hace falta el operador de expectativas.

**Tecnología**: la producción es Cobb-Douglas y la tecnología crece en torno a un nivel tendencial, $\bar{A}$, a un ritmo exógeno y constante, $g_A$. La tecnología está afectada también por un componente aleatorio, $a_t$, que sigue un proceso AR(1) con ruido blanco $\phi_t$:

$$y_t = k_t^{\alpha}(A_t l_t)^{1-\alpha}, \qquad \alpha \in (0,1)$$
$$\ln A_t = \ln \bar{A} + g_A\, t + a_t, \qquad a_t = \rho_A a_{t-1} + \phi_t, \quad \rho_A \in (-1,1)$$

**Equilibrio**: con vaciado de los mercados de trabajo y capital y depreciación completa, el gasto en consumo e inversión coincide con la renta, $c_t + k_{t+1} = y_t$. Al resolver el modelo se obtienen los valores solución de las variables en función de los parámetros:

$$k^*_{t+1} = s\, y_t, \quad s = \alpha\beta, \qquad c^*_t = (1-\alpha\beta)\, y_t, \qquad l^*_t = \bar{l}(\alpha, \beta, b)$$

donde $b$ es el peso del ocio en la utilidad, $u(c_t, 1-l_t) = \ln c_t + b \ln(1-l_t)$. El consumidor trabaja una cantidad constante de horas y ahorra una fracción constante de la renta, porque los cambios que la tecnología o el capital provocan en el tipo de interés y en el salario real se cancelan entre sí[10]. En cambio, el ahorro (el capital acumulado al final de cada periodo) y el consumo sí se ven afectados por los *shocks* tecnológicos.

**Estática comparativa**: en el plano renta-nivel general de precios (figura A42.4), un *shock* tecnológico positivo ($\Delta A > 0$) desplaza a la derecha la oferta agregada a largo plazo (vertical en la producción potencial, que fija la tecnología), de $\mathrm{OA}_{\mathrm{LP},0}$ a $\mathrm{OA}_{\mathrm{LP},1}$, y a corto plazo, de $\mathrm{OA}_{\mathrm{CP},0}$ a $\mathrm{OA}_{\mathrm{CP},1}$. En menor medida, desplaza también la demanda agregada, de $\mathrm{DA}_0$ a $\mathrm{DA}_1$, porque la mayor productividad eleva la renta esperada y la rentabilidad de la inversión. La economía pasa del punto 0 al 1, con más producción ($\bar{Y}_1 > \bar{Y}_0$) y un nivel de precios algo menor ($P_1 < P_0$).

Como el equilibrio se sitúa siempre sobre la oferta a largo plazo, el *shock* no abre una brecha respecto del potencial, y la fluctuación del *output* es la del propio potencial, que se desvía de su tendencia determinista con la persistencia que marca $\rho_A$.

[10] Un *shock* tecnológico positivo eleva el salario actual respecto al futuro esperado, lo que aumenta la oferta de trabajo, pero el *shock* eleva simultáneamente el ahorro y reduce el tipo de interés esperado, lo que la reduce. Con utilidad logarítmica y depreciación completa, ambos efectos se cancelan exactamente. Este resultado es propio de esta versión, ya que con una función de utilidad más general y depreciación incompleta, el empleo también fluctúa.

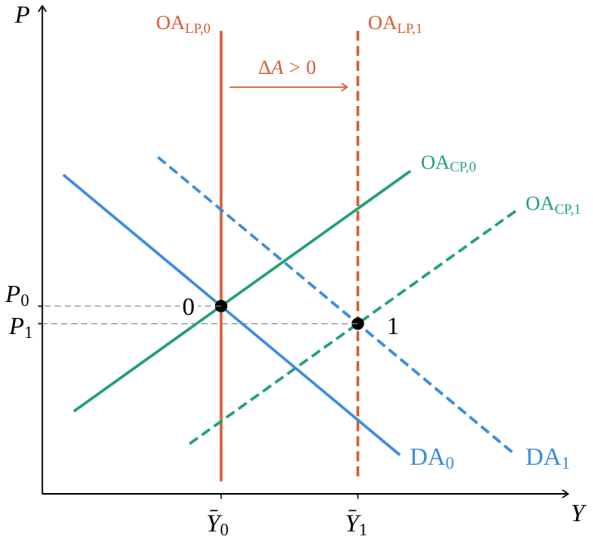



Figura A42.4: Ciclo real en el plano OA-DA. Un *shock* tecnológico positivo ($\Delta A > 0$) desplaza a la derecha la oferta agregada a largo y a corto plazo (la producción potencial pasa de $\bar{Y}_0$ a $\bar{Y}_1$) y, en menor medida, la demanda agregada. La economía pasa del equilibrio 0 al 1, con más producción y un nivel de precios algo menor. El gráfico representa las variables en niveles ($P$, $Y$, $\bar{Y}$) y el texto las expresa en logaritmos.

**Dinámica**: a partir de los valores solución se obtienen las ecuaciones dinámicas y estocásticas de cada variable endógena[11], expresadas en logaritmos y en desviaciones respecto de su trayectoria determinista[12]. En el caso del *output*, $\tilde{y}_t = \alpha\,\tilde{y}_{t-1} + (1-\alpha)\,a_t$ y, sustituyendo el AR(1) de la tecnología, la ecuación dinámica estocástica de la desviación del *output* respecto de su tendencia determinista es:

$$\tilde{y}_t = (\alpha + \rho_A)\,\tilde{y}_{t-1} - \alpha\rho_A\,\tilde{y}_{t-2} + (1-\alpha)\,\phi_t$$

El ciclo sigue un proceso autorregresivo de orden 2, de modo que, cuando se produce una perturbación tecnológica, el PIB se desvía de su tendencia en función de los ciclos de los dos periodos anteriores y de la persistencia de la perturbación. La duración del ciclo será mayor cuanto mayor sea la autocorrelación de la perturbación, $\rho_A$, que depende de la calibración del modelo[13].

Un aumento de la productividad total de los factores (un *shock* tecnológico positivo no anticipado) tiene efectos según su duración:

[11] Como se estudia el ciclo, el análisis se centra en el *output*, pero puede demostrarse que el ahorro y el consumo, en este modelo, también siguen un AR(2).

[12] Esquema de la derivación (variables en logaritmos, desviaciones respecto de la tendencia). (i) Log-linealizando la función de producción: $\ln y_t = \alpha \ln k_t + (1-\alpha)\ln A_t + (1-\alpha)\ln l_t$, con $\ln l_t = \ln \bar{l}$ constante. (ii) Con depreciación total y utilidad logarítmica, la tasa de ahorro es constante, $k_{t+1} = s\,y_t$, de modo que $\ln k_t = \ln s + \ln y_{t-1}$. (iii) Sustituyendo y restando la tendencia se llega a $\tilde{y}_t = \alpha\,\tilde{y}_{t-1} + (1-\alpha)\,a_t$ e, introduciendo el proceso AR(1) de la tecnología, $a_t = \rho_A a_{t-1} + \phi_t$, se obtiene el AR(2) del cuerpo del tema.

[13] Con $\alpha = 1/3$, si $\rho_A = 0$, bastan dos periodos para que desaparezca casi el 90 % del efecto inicial, y con $\rho_A = 0{,}5$, dos tercios desaparecen en tres periodos. La calibración típica usa $\rho_A \in [0{,}9;\ 0{,}95]$ para replicar la persistencia observada.

▪ *Transitorio*: produce efectos reales de carácter temporal y es expansivo sobre la producción, el consumo, la inversión y los salarios. Con tasa de ahorro constante ($s = \alpha\beta$), consumo e inversión varían en la misma proporción que la producción. Sobre el tipo de interés real el efecto es mixto, porque el tipo sube inicialmente, al aumentar la productividad marginal del capital, y cae en periodos posteriores, dado que el mayor *stock* de capital reduce su productividad marginal.

▪ *Permanente*: eleva duraderamente la producción potencial, con lo que se desplaza la trayectoria tendencial y no se genera un ciclo en sentido estricto. De hecho, Nelson y Plosser (1982) muestran que las series del PIB se caracterizan mejor como procesos con raíz unitaria, en los que las innovaciones afectan permanentemente a su valor futuro. Si los *shocks* de productividad son frecuentes y permanentes (caso límite $\rho_A = 1$), explican tanto el ciclo a corto plazo como el crecimiento a largo plazo, y las causas de ambos no son esencialmente distintas[14].

**Bienestar**: como la economía es walrasiana, las fluctuaciones no constituyen un fracaso de los mercados, sino la respuesta óptima de los agentes a las perturbaciones (el propio óptimo de Pareto fluctúa). En consecuencia, toda intervención pública para mitigarlas reduce el bienestar.

**Extensiones**: las versiones posteriores levantan los supuestos simplificadores del modelo básico para mejorar su ajuste al mercado de trabajo:

▪ *Indivisibilidades del trabajo* (Hansen, 1985; Rogerson, 1988): generan una mayor elasticidad de la oferta de trabajo agregada y, además, justifican la existencia de desempleo.

▪ *Depreciación incompleta* ($\delta < 1$): vuelve la inversión y el empleo más sensibles a las perturbaciones y reduce la prociclicidad del salario. Un *shock* positivo de productividad eleva la inversión y la tasa de ahorro, lo que modera la subida del consumo y eleva la oferta de trabajo. La mayor oferta de trabajo modera, a su vez, la subida del salario.

▪ *Sector público* (Christiano y Eichenbaum, 1992)[15]: rompe el vínculo directo entre producción y salario real, porque el efecto riqueza negativo de un aumento del gasto público eleva la oferta de trabajo y, sin cambio tecnológico, reduce el salario real. Producción y salario real se mueven así en direcciones opuestas (tema A38, apartado 1.2).

### 2.2.3. Conclusiones

**Aspectos positivos**: con modelos relativamente sencillos se reproducía gran parte de los hechos estilizados del ciclo de EE. UU. (1950-1979):

▪ *PIB*: los autores concluían que los *shocks* tecnológicos podían explicar en torno al 70 % de la volatilidad del PIB de posguerra en EE. UU. (Kydland y Prescott, 1991). Los modelos del ciclo real aportaron, además, un método que después adoptó también la NEK.

▪ *Consumo e inversión*: en sus versiones calibradas[16] (con depreciación incompleta), el modelo recoge la suavización del consumo y el exceso de volatilidad de la inversión.

[14] Sobre la separación entre tendencia estocástica y ciclo, véase el apartado 3.1.2. Sobre los efectos permanentes de las recesiones en el crecimiento endógeno, véase el tema A44, apartado 1.1. La magnitud de la respuesta del empleo depende de la persistencia del *shock*. Cuanto más persistente, mayor es el efecto riqueza (que reduce la oferta de trabajo) y menor el peso de la sustitución intertemporal, por lo que la respuesta del empleo se atenúa. Además, cuanto mayor sea la elasticidad de sustitución intertemporal de la oferta de trabajo, mayor será la fluctuación.

[15] La ampliación mejora el ajuste a los datos y reduce la correlación entre horas trabajadas y productividad, excesiva en el modelo básico.

[16] A diferencia del modelo de Lucas, en el que apenas tiene peso, la calibración es el método central de la TCR.

**Limitaciones**: el modelo falla en el impulso, en variables distintas del PIB y en el papel del dinero:

- *Mecanismo de impulso*: descansa en perturbaciones tecnológicas, pero a corto plazo el progreso técnico avanza con regularidad, y por eso resulta difícil atribuirle los ciclos y, sobre todo, las recesiones (¿una regresión tecnológica explica realmente una recesión?). Además, el modelo básico carece de un mecanismo de propagación que convierta una perturbación transitoria en una variación significativa y duradera de la producción.

- *Tasa de ahorro y oferta de trabajo constantes*: en la versión básica, inversión y consumo serían igualmente volátiles y el empleo no fluctuaría, en contra de la evidencia.

- *Mercado de trabajo*: la volatilidad de la oferta de trabajo es menor que la observada, y productividad y oferta de trabajo están mucho más correlacionadas que en los datos (donde parecen independientes). Además, el salario real y la productividad resultan excesivamente procíclicos y se requiere una elasticidad de la oferta de trabajo irrealistamente elevada.

- *Neutralidad del dinero*: se mantiene incluso en las versiones con dinero (p. ej. modelos *cash-in-advance*), en contra de la evidencia de efectos reales a corto plazo (p. ej. la ecuación de St. Louis o los modelos de vectores autorregresivos, VAR).

## 2.3. Ciclos de desequilibrio y fricciones financieras

Los modelos de la NEK parten de fallos de mercado (poder de mercado, información asimétrica) que se traducen en rigideces nominales y reales. En consecuencia, no hay vaciado continuo de los mercados, el ajuste es lento y se realiza vía cantidades, y pueden darse equilibrios múltiples subóptimos (fallos de coordinación).

En los mercados financieros, la información asimétrica raciona el crédito. Además, cuando las condiciones de financiación dependen del balance de los prestatarios, la información asimétrica amplifica las perturbaciones y enlaza el ciclo económico con el ciclo financiero. Cierran el apartado otras dos aportaciones de la NEK, los contratos escalonados y los fallos de coordinación.

> Los rasgos de la escuela y la fundamentación microeconómica de las rigideces reales y nominales se desarrollan en el tema A07, apartado 1. Aquí se presentan sus modelos como teorías del ciclo.

### 2.3.1. Racionamiento del crédito (Stiglitz y Weiss, 1981)

**Origen**: el modelo se debe a Joseph Stiglitz y Andrew Weiss (1981), en la línea de la NEK.

**Idea**: la información asimétrica lleva a los bancos a racionar el crédito en lugar de subir el tipo de interés. En el modelo se distinguen tres elementos:

- *Impulso*: perturbaciones de demanda agregada, tanto reales como monetarias (p. ej. un *shock* negativo en el flujo de caja, una política monetaria restrictiva o una innovación financiera).

- *Mecanismo de transmisión*: la rigidez del tipo de interés, que lleva a los bancos a ajustar la cantidad de crédito (racionamiento) en lugar de su precio.

- *Persistencia*: procede del ajuste escalonado de los precios en la economía real.

**Supuestos**:

▪ *Entorno*: información asimétrica. Los inversores conocen la probabilidad de éxito de su proyecto, pero los bancos no.

▪ *Agentes*: numerosos bancos en competencia perfecta, que conceden préstamos a un tipo de interés $i$, e inversores con proyectos arriesgados y sin financiación alternativa a la bancaria (p. ej. pymes), ambos neutrales al riesgo.

▪ *Tecnología*: los proyectos tienen el mismo vencimiento (un periodo) y la misma cuantía, $B$, y solo se diferencian por su probabilidad de éxito. El proyecto $j$ rinde $X_j$ con probabilidad $p_j$ y cero si fracasa (cuanto más seguro, menor el retorno), con un rendimiento esperado, $\bar{X} = p_j X_j$, idéntico para todos.

El banco presta la cuantía $B$ y el inversor se compromete a devolver $(1+i)B$, con responsabilidad limitada (si el proyecto fracasa, el inversor quiebra y el banco no recupera nada). Esta asimetría en los pagos genera el incentivo a asumir riesgos excesivos.

**Inversores**: demandan crédito para su proyecto.

▪ *Planteamiento*: cada inversor pide el préstamo si su beneficio esperado, $\mathbb{E}[\Pi_j] = p_j\,[X_j - (1+i)B] = \bar{X} - p_j(1+i)B$, no es negativo (no invertir rinde cero). Como $\bar{X}$ es común, el beneficio esperado es mayor cuanto más arriesgado es el proyecto (menor $p_j$).

▪ *CPO*: la elección es binaria, y la condición marginal se sustituye por la de participación, $p_j \leq \hat{p}(i) \equiv \bar{X}/[(1+i)B]$, con un umbral de probabilidad de éxito decreciente en el tipo de interés.

▪ *CSO*: el beneficio esperado es lineal y decreciente en $p_j$, así que el umbral es único y separa a quienes piden crédito de quienes no.

▪ *Solución*: al subir el tipo, el umbral baja y los proyectos seguros quedan expulsados (selección adversa), y la demanda de crédito, $L^d(i)$, tiene pendiente negativa.

**Bancos**: captan depósitos y ofrecen crédito.

▪ *Planteamiento*: eligen el tipo de interés que maximiza su rendimiento esperado por unidad prestada, $\bar{R}^e(i) = (1+i)\,\bar{p}(i)$, donde $\bar{p}(i)$ es la probabilidad media de éxito de los proyectos que demandan crédito a ese tipo, decreciente en $i$ por la selección adversa. Del rendimiento se obtiene la oferta de crédito, $L^s(i)$.

▪ *CPO*: en el tipo óptimo, $i^*$, se cumple $\bar{p}(i^*) + (1+i^*)\,(d\bar{p}/di)|_{i^*} = 0$. El primer sumando es el *efecto recaudación* (positivo), pues el banco cobra más por los proyectos que tienen éxito. El segundo es el *efecto deterioro de cartera* (negativo), pues la subida expulsa a los proyectos seguros.

▪ *CSO*: con un rendimiento esperado de un solo máximo, como suponen Stiglitz y Weiss (1981), domina el efecto recaudación por debajo de $i^*$ y el de deterioro de cartera por encima, así que $i^*$ es el máximo.

**Equilibrio**: el banco prefiere racionar a subir $i$ más allá de $i^*$. Si a ese tipo la demanda supera a la oferta, el resultado es un equilibrio con racionamiento del crédito, es decir, con exceso de demanda y ajuste vía cantidades, no vía precios.

**Desarrollo gráfico**: el modelo completo se representa en cuatro cuadrantes, con el tipo de interés ($i$) y los depósitos ($D$) en el eje horizontal y el rendimiento esperado ($\bar{R}^e$) y el crédito ($L$) en el vertical (figura A42.5):

1. *Rendimiento esperado del banco*, $\bar{R}^e(i)$: en el cuadrante I, creciente mientras domina el efecto recaudación y decreciente cuando domina el de deterioro de cartera, con su máximo en $i^*$.

2. *Oferta de depósitos*, $D^s(\bar{R}^e)$: en el cuadrante II, creciente en la rentabilidad.

3. *Conversión de depósitos en créditos*, $L = \chi D$, con $0 < \chi \leq 1$: en el cuadrante III, una recta, pues el banco presta una fracción constante de los depósitos.

4. *Oferta y demanda de crédito*, $L^s(i)$ y $L^d(i)$: en el cuadrante IV. La oferta se obtiene de los tres cuadrantes anteriores y es también máxima en $i^*$.

5. *Equilibrio*: en $i^*$, inferior al tipo walrasiano $i^W$, la diferencia $Z = L^d(i^*) - L^s(i^*)$ es el racionamiento del crédito.

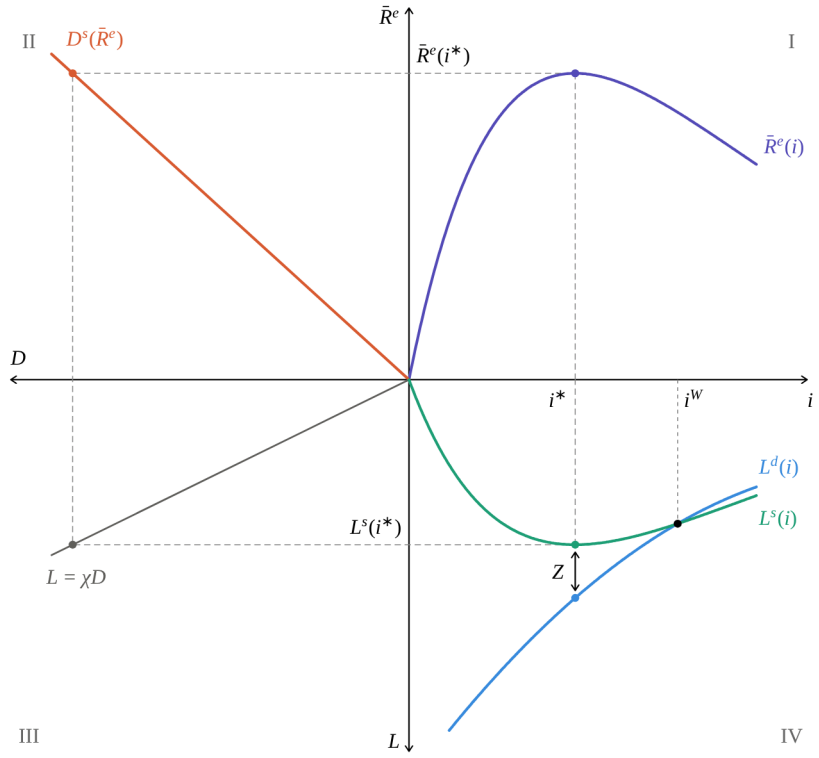



Figura A42.5: Racionamiento del crédito (Stiglitz y Weiss, 1981). El rendimiento esperado del banco, $\bar{R}^e(i)$, alcanza su máximo en $i^*$ (cuadrante I). A través de la oferta de depósitos, $D^s(\bar{R}^e)$ (cuadrante II), y del factor de conversión de depósitos en créditos, $L = \chi D$ (cuadrante III), se obtiene la oferta de crédito, $L^s(i)$, que también es máxima en $i^*$ y decrece por encima de ese tipo (cuadrante IV). En $i^*$, inferior al tipo walrasiano $i^W$ ($L^d = L^s$), la demanda de crédito, $L^d(i)$, excede a la oferta, y ese exceso de demanda, $Z$, es el racionamiento del crédito.

**Estática comparativa**:

▪ *Perturbación negativa en la oferta de depósitos* (p. ej. un aumento del desempleo, que reduce la renta de los hogares): desplaza hacia arriba la curva de oferta de depósitos en el segundo

cuadrante. Como el tipo de interés permanece en $i^*$, el ajuste se produce vía cantidad. Aumenta el racionamiento, se contraen el crédito, la inversión y el consumo y el PIB se desvía negativamente de su tendencia (ciclo recesivo).

- *Perturbación monetaria positiva* (p. ej. una política monetaria expansiva): desplaza hacia abajo la curva de oferta de depósitos, sin que varíe $i^*$. El racionamiento disminuye, se conceden más créditos y se realizan más proyectos, lo que aumenta la inversión y provoca un ciclo expansivo.

Los efectos son asimétricos, pues el coste del crédito y su racionamiento recaen sobre todo en las empresas más endeudadas y las percibidas como más arriesgadas. Así, en las restricciones monetarias, las pequeñas, dependientes del crédito bancario, se ven más afectadas que las grandes, con acceso al mercado de capitales. Además, la inversión agregada resulta sensible a los flujos de caja pasados.

**Conclusiones**:

- *Aspectos positivos*: ofrece un mecanismo de transmisión de las perturbaciones nominales y reales fundado en la información asimétrica, sin rigideces de precios en el mercado de bienes. Explica, además, el canal del crédito de la política monetaria y la asimetría de sus efectos entre empresas, y anticipa la literatura del acelerador financiero.

- *Limitaciones*: el racionamiento depende de que los bancos no puedan discriminar entre prestatarios, cuando en la práctica exigen garantías que revelan parcialmente el tipo de prestatario y atenúan la asimetría. Además, el modelo es de equilibrio parcial y la persistencia del ciclo se importa de fuera del mercado de crédito.

### 2.3.2. Acelerador financiero (Bernanke y Gertler, 1989; Kiyotaki y Moore, 1997)

> El acelerador financiero se presenta como determinante de la inversión en el tema A34, apartado 3.2.4, y como canal del crédito de la política monetaria en el tema A37, apartado 1.2. Aquí se desarrolla como mecanismo de propagación del ciclo.

**Origen**: el modelo de Stiglitz y Weiss aísla la fricción microeconómica, pero es estático y de equilibrio parcial. Bernanke y Gertler (1989) dan el paso siguiente al mostrar que el patrimonio neto de los prestatarios reduce los costes de agencia de la financiación y puede, por ello, propagar el ciclo. Por su parte, Kiyotaki y Moore (1997) llegan a un resultado análogo a través del colateral.

Bernanke, Gertler y Gilchrist, que denominan al mecanismo *acelerador financiero* (1996), lo integran después en un modelo de equilibrio general dinámico y estocástico (DSGE) neokeynesiano cuantitativo (1999). El mecanismo no es exclusivo de la NEK, pues también se obtiene en modelos de precios flexibles.

**Idea**: si el coste de la financiación depende del patrimonio de los prestatarios, y este cae en las recesiones, la fricción financiera amplifica y propaga el ciclo. Como el prestamista no puede observar sin coste el resultado de los proyectos, la financiación externa es más cara que la interna. La diferencia, la *prima de financiación externa*, disminuye con la parte de la compra de capital que el prestatario financia con recursos propios:

$$\rho_{\mathrm{fe},t} = \rho_{\mathrm{fe}}\left(\frac{\mathrm{PN}_t}{q_t K_t}\right), \qquad \rho'_{\mathrm{fe}}(\cdot) < 0$$

donde $\rho_{\text{fe},t}$ es la prima de financiación externa (la diferencia entre el coste de los fondos externos y el tipo sin riesgo), $\text{PN}_t$ el patrimonio neto de las empresas, $q_t$ el precio del capital (la $q$ de Tobin) y $K_t$ el capital que adquieren. Así, $q_t K_t - \text{PN}_t$ es la financiación externa que necesitan.

**Razonamiento**: una perturbación negativa (tecnológica, monetaria o de demanda) pone en marcha un bucle de amplificación que opera por dos canales:

- *Patrimonio neto* (Bernanke y Gertler, 1989): la perturbación reduce los beneficios y el precio de los activos y, con ellos, el patrimonio neto. La prima sube, la inversión cae más de lo que justificaría la perturbación y la menor actividad deprime de nuevo beneficios y precios.

- *Colateral* (Kiyotaki y Moore, 1997): un mismo activo duradero (la tierra o los inmuebles) sirve para producir y garantizar la deuda. Así, la caída de su precio endurece el límite de endeudamiento y fuerza ventas, y esas ventas vuelven a hundir el precio.

Como el patrimonio se recompone despacio, el efecto persiste. Por tanto, el mecanismo aporta lo que le faltaba al modelo de Stiglitz y Weiss, porque con el acelerador financiero perturbaciones pequeñas y transitorias generan fluctuaciones amplias y persistentes.

**Conclusiones**:

- *Aspectos positivos*: integra las fricciones financieras en los modelos de equilibrio general y explica por qué los precios de los activos y los balances de los prestatarios amplifican y prolongan el ciclo. La crisis financiera de 2008 confirmó su relevancia e impulsó su incorporación a los modelos de los bancos centrales.

- *Limitaciones*: la fricción se sitúa en el balance de los prestatarios no financieros, sin un papel propio para el capital de los intermediarios, que aquella crisis reveló decisivo. Además, estos modelos suelen resolverse en torno al estado estacionario, lo que capta mal la dinámica no lineal de las crisis.

### 2.3.3. Ciclo financiero

> La hipótesis de la inestabilidad financiera de Minsky se desarrolla en el tema B26, apartado 1.2.1. Aquí se recoge su idea central.

**Inestabilidad financiera** (Minsky, 1986): en el acelerador financiero las perturbaciones son exógenas. Para Minsky, en cambio, la fragilidad nace del propio auge, que desplaza la financiación de posiciones cubiertas a especulativas y Ponzi hasta que una perturbación menor desencadena ventas forzosas de activos y la crisis.

**Ciclo financiero**: ambas ideas confluyen en este concepto empírico, que Borio (2014) define como el conjunto de interacciones, mutuamente reforzadas, entre la percepción del valor y del riesgo, la actitud ante el riesgo y las restricciones de financiación, que generan auges seguidos de contracciones.

El ciclo financiero se mide por el componente de medio plazo de las fluctuaciones conjuntas del crédito al sector privado (en relación con el PIB) y de los precios de los inmuebles (Drehmann, Borio y Tsatsaronis, 2012). Sus hechos estilizados son tres:

- *Duración*: es mucho mayor que la del ciclo económico (de uno a ocho años). En siete economías avanzadas desde los años sesenta ha sido de unos 16 años en promedio, y de casi 20 en los ciclos con máximo posterior a 1998.

- *Crisis*: todas las crisis financieras de origen interno de la muestra se produjeron en un máximo del ciclo financiero o cerca de él.

- *Amplitud*: ha aumentado, como la duración, desde mediados de los años ochenta, con la liberalización financiera y los cambios en los marcos de política monetaria.

**Interacción con el ciclo económico**: las recesiones que coinciden con la fase contractiva del ciclo financiero son mucho más profundas, con caídas del PIB en torno a un 50 % mayores (Drehmann, Borio y Tsatsaronis, 2012). Tras las crisis bancarias graves, el producto cae en promedio más de un 9 %, el desempleo sube unos 7 puntos porcentuales en algo más de cuatro años y el precio real de la vivienda cae un 35 % en unos seis años (Reinhart y Rogoff, 2009).

> El fundamento de la política macroprudencial y el colchón de capital anticíclico de Basilea III se tratan en el tema B27, Anexo A.4 y apartado 3.1.3. Aquí se recoge su relación con el ciclo financiero.

**Política macroprudencial**: como el ciclo financiero es más largo que el económico, una economía puede crecer con inflación baja y estable mientras acumula desequilibrios que solo afloran en la crisis, y estabilizar el ciclo económico no basta para evitarlos.

El instrumento macroprudencial más ligado al ciclo financiero es el colchón de capital anticíclico, cuya referencia común es la desviación de la ratio de crédito sobre el PIB respecto de su tendencia[17].

### 2.3.4. Otras aportaciones: contratos escalonados y fallos de coordinación

Dos aportaciones más de la NEK explican por qué las perturbaciones nominales tienen efectos reales persistentes o se amplifican:

- *Contratos escalonados* (Fischer, 1977; Taylor, 1980): si los salarios nominales se fijan mediante contratos multiperiodo, la política monetaria posterior a la firma altera el salario real y genera desequilibrios que persisten hasta la renegociación. Además, si las negociaciones no están sincronizadas, cada empresa evita alejar sus salarios de los ya pactados por las demás, y las perturbaciones tienen efectos reales más allá de la duración de los contratos.

- *Fallos de coordinación* (Blanchard y Kiyotaki, 1987; Cooper y John, 1988): con competencia monopolística, las externalidades de demanda agregada hacen que el equilibrio sea subóptimo y, cuando hay complementariedades estratégicas, pueden generar equilibrios múltiples ordenados en el sentido de Pareto.

> Estas aportaciones se desarrollan en el tema A07, apartados 1.3.3 (Blanchard-Kiyotaki) y 1.4.1 (Fischer, Taylor y Calvo).

## 2.4. La nueva síntesis neoclásica: el modelo IS-PC-MR

**Origen**: el modelo canónico de la NEK (Clarida, Galí y Gertler, 1999) sintetiza tres ecuaciones, una curva IS dinámica, la curva de Phillips de la NEK y una regla de política monetaria, en un

[17] La guía orientativa de Basilea (2010) fija el colchón en cero mientras la desviación no supera 2 puntos porcentuales y lo eleva linealmente hasta el 2,5 % de los activos ponderados por riesgo cuando alcanza 10, aunque las autoridades no la aplican de forma mecánica.

modelo DSGE con rigideces nominales y competencia monopolística. Carlin y Soskice (2005) le dieron su exposición gráfica y el nombre de IS-PC-MR.

**Idea**: integra, así, ambas tradiciones, pues acepta la HER y la microfundamentación de la NMC, pero incorpora las rigideces de la NEK, de modo que el ciclo es a la vez resultado de *shocks* reales y de la no neutralidad del dinero a corto plazo.

> El modelo, su microfundamentación, su desarrollo gráfico, sus extensiones y sus implicaciones de política se desarrollan en el tema A07, apartados 2 y 3. Aquí se recogen sus tres ecuaciones, su lectura como teoría del ciclo y la respuesta a un *shock* tecnológico.

### 2.4.1. Supuestos

**Entorno**: competencia monopolística en el mercado de bienes y servicios (las empresas son precio-determinantes) y rigideces nominales en el mercado de bienes finales (precios *à la* Calvo), mientras que el mercado de trabajo es competitivo.

**Agentes**: se acepta la HER y hay cuatro: hogares (consumidores y oferentes de trabajo, aversos al riesgo), empresas en competencia monopolística, sector público y banco central.

**Tecnología**: por simplicidad, se prescinde del capital como factor productivo y, con ello, de la inversión privada.

### 2.4.2. Modelo

**Planteamiento**: las tres ecuaciones recogen la demanda de bienes, la fijación de precios y la conducta del banco central.

**Curva IS dinámica**: se obtiene de la ecuación de Euler de los hogares y de la condición de equilibrio en el mercado de bienes, log-linealizadas en torno al estado estacionario. En términos de la brecha de producción (*output gap*):

$$\tilde{y}_t = \mathbb{E}_t[\tilde{y}_{t+1}] - \frac{1}{\theta}\,(r_t - \bar{r}_t)$$

donde $\tilde{y}_t$ es la brecha de producción, $\mathbb{E}_t[\tilde{y}_{t+1}]$ su valor esperado (con $\mathbb{E}_t[\cdot]$ la esperanza condicionada a la información disponible en $t$), $\theta$ el coeficiente de aversión relativa al riesgo (inversa de la elasticidad de sustitución intertemporal), $r_t = i_t - \mathbb{E}_t[\pi_{t+1}]$ el tipo de interés real esperado ($i_t$ es el nominal a corto plazo y $\mathbb{E}_t[\pi_{t+1}]$ la inflación esperada) y $\bar{r}_t$ el tipo de interés real natural. La demanda agregada depende positivamente de la brecha esperada (los hogares alisan su consumo) y negativamente del tipo de interés real. El gasto público, que esta forma omite por simplicidad, desplaza la IS.

En línea con la concepción wickselliana, el tipo de interés real natural, determinado por factores reales, es el que, en ausencia de *shocks* de costes, garantiza la estabilidad de precios con el *output* de precios flexibles. En el modelo canónico, $\bar{r}_t = \rho + \theta\,\mathbb{E}_t[\Delta\bar{y}_{t+1}]$, que aumenta con la impaciencia ($\rho$) y con el crecimiento esperado del *output* natural, de modo que un *shock* tecnológico transitorio, que eleva el *output* natural hoy más que mañana, reduce el tipo natural. Cuando $r_t > \bar{r}_t$, la brecha de producción se vuelve negativa, y por ese canal el banco central afecta a la actividad real.

**Curva de Phillips de la NEK**: se obtiene de la fijación de precios *à la* Calvo y determina la inflación dada la senda de la brecha de producción:

$$\pi_t = \beta\, \mathbb{E}_t[\pi_{t+1}] + \kappa\, \tilde{y}_t, \qquad \kappa > 0,\ \beta \in (0,1)$$

donde $\beta$ es el factor de descuento de los hogares y $\kappa$ recoge la velocidad de ajuste de la inflación. La inflación depende positivamente de la inflación esperada, porque las empresas anticipan que el precio que fijan hoy permanecerá fijo y tienen en cuenta sus costes futuros, y de la brecha de producción, que presiona al alza los costes.

**Regla de política monetaria**: el comportamiento del banco central se resume en una regla de tipos de interés, la regla de Taylor:

$$i_t = \bar{i}_t + \alpha_\pi\, \tilde{\pi}_t + \alpha_y\, \tilde{y}_t, \qquad \alpha_\pi > 1$$

donde $\bar{i}_t = \bar{r}_t + \pi^T$ es el tipo de interés nominal neutral, $\pi^T$ el objetivo de inflación y $\tilde{\pi}_t = \pi_t - \pi^T$ la desviación de la inflación respecto del objetivo.

Si la inflación se desvía por encima de su objetivo, el tipo de interés nominal debe aumentar más que la inflación ($\alpha_\pi > 1$, el *principio de Taylor*) para que suba el tipo real y se estabilice la economía. Esta condición garantiza, además, la unicidad del equilibrio.

**Equilibrio**: al sustituir la regla de política en la IS dinámica se obtiene una relación negativa entre la brecha de producción y la desviación de la inflación respecto del objetivo, que hace las veces de curva MR (regla monetaria). En consecuencia, el compromiso del banco central con la estabilidad de precios se traduce, ante mayor inflación, en un menor *output*.

**Estática comparativa**:

- *Perturbación de costes*: la curva de Phillips se desplaza hacia arriba y el banco central acepta transitoriamente una brecha negativa, moviéndose a lo largo de la MR, para devolver la inflación al objetivo.

- *Política fiscal*: ante una expansión fiscal, la IS se desplaza y el banco central eleva el tipo, de modo que el multiplicador depende del grado de acomodación monetaria.

- *Perturbación tecnológica*: un *shock* tecnológico positivo transitorio ($\Delta a > 0$) reduce el tipo de interés natural. Si la regla traslada esa caída al tipo de interés, como la anterior ($\bar{i}_t = \bar{r}_t + \pi^T$), la brecha no se abre y no hay conflicto entre objetivos. En cambio, si no la traslada (p. ej. con un tipo neutral constante), el tipo real queda por encima del natural.

La figura A42.6 muestra arriba, en el plano brecha de producción-desviación del tipo de interés real respecto del natural, la IS, decreciente porque un tipo real más alto reduce la demanda. Abajo, en el plano brecha de producción-inflación, la curva de Phillips (CP) crece porque la brecha presiona los costes, y la MR, decreciente, da la brecha que la regla monetaria fija para cada inflación. Sin esa traslación, la MR se desplaza a la izquierda, de $\text{MR}_0$ a $\text{MR}_1$, y la economía pasa del punto 0 al 1, con brechas negativas de producción y de inflación.

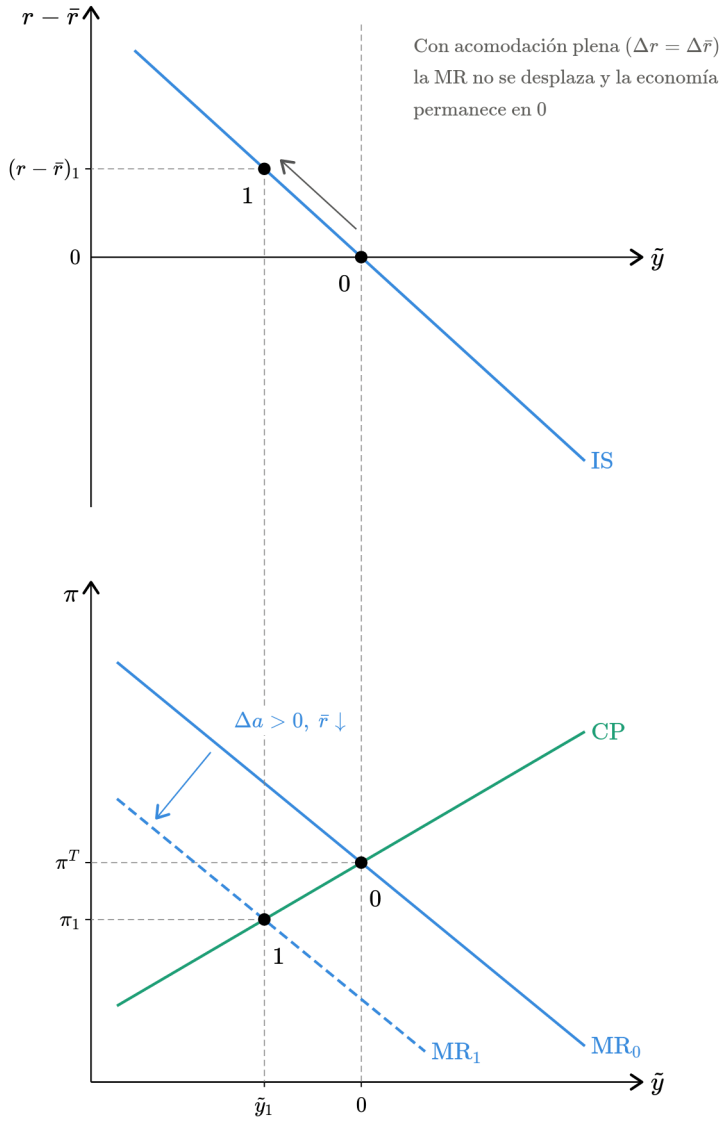



Figura A42.6: Modelo IS-PC-MR ante un *shock* tecnológico positivo ($\Delta a > 0$). Si la regla de Taylor no traslada la caída del tipo de interés natural, el tipo real queda por encima del natural y la MR se desplaza a la izquierda, de $MR_0$ a $MR_1$. La economía pasa de 0 a 1, con brechas de producción e inflación negativas. Con acomodación plena, la economía permanece en 0.

El mismo *shock* que en la teoría del ciclo real eleva el potencial sin abrir brecha (figura A42.4) genera aquí, con precios rígidos y una política monetaria que no lo acomoda, una brecha negativa, porque la demanda crece menos que el *output* natural. Por tanto, a diferencia de lo que sostiene esa teoría, la fluctuación depende de la respuesta del banco central y no solo del impulso real.

### 2.4.3. Conclusiones

**Aspectos positivos**: el modelo integra las dos tradiciones anteriores en un modelo DSGE con expectativas racionales (en principio, robusto a la crítica de Lucas). Permite estudiar la interre-

lación entre *output*, inflación y tipo de interés en un marco coherente, lo que lo ha convertido en la referencia de los bancos centrales.

**Limitaciones**: como teoría del ciclo, hereda las de sus componentes. La curva de Phillips puramente *forward-looking* no explica la persistencia de la inflación, y el modelo básico prescinde del capital, la inversión y el sector financiero. Sus supuestos (agente representativo, log-linealización) limitan, además, su capacidad ante grandes perturbaciones o el límite inferior de los tipos de interés (*zero lower bound*).

## 2.5. Conclusiones

Las cuatro teorías comparten, en mayor o menor medida, la exigencia de microfundamentación asociada a las expectativas racionales y a la crítica de Lucas. Sin embargo, se distinguen por el impulso y el mecanismo de propagación que privilegian:

- *Lucas*: una sorpresa nominal filtrada por la extracción de señales.
- *TCR*: una perturbación tecnológica propagada por la sustitución intertemporal y la acumulación de capital.
- *NEK*: perturbaciones de demanda transmitidas por la rigidez del tipo de interés y el racionamiento del crédito, y amplificadas por el acelerador financiero.
- *Nueva síntesis*: *shocks* reales y nominales que actúan sobre una economía con rigideces de precios y un banco central que sigue una regla.

A todas ellas se superpone el ciclo financiero, más largo que el económico, cuya fase contractiva hace más profundas las recesiones con las que coincide. Cuál de las cuatro teorías describe mejor el ciclo es una cuestión empírica, que exige medir el ciclo y confrontar cada modelo con los hechos estilizados.

# 3. Análisis empírico del ciclo

Medir el ciclo exige separar la tendencia del componente cíclico y estimar la brecha de producción. Con esa medida se obtienen los hechos estilizados que toda teoría debe ser capaz de reproducir, y frente a ellos se confrontan finalmente las teorías modernas.

## 3.1. Medición de los ciclos

La medición de los ciclos parte de las series temporales del PIB observado, cuyo componente de especial interés es el tendencial-cíclico. La **tendencia** ilustra la evolución a largo plazo del PIB (la producción potencial, cuyo estudio lleva al análisis del crecimiento) y el **componente cíclico**, su evolución a corto y medio plazo (la **brecha de producción** u *output gap*, cuyo estudio lleva al análisis del ciclo). Dado que interesan las fluctuaciones recurrentes pero no periódicas, el primer paso es la desestacionalización de la serie, seguida de la separación tendencia-ciclo.

### 3.1.1. Descomposición y desestacionalización de series

**Descomposición de la serie**: la serie observada, $Y_t$, se descompone en una tendencia, $T_t$, un componente cíclico, $C_t$, un componente estacional, $S_t$, y un componente irregular o residual, $\varepsilon_t$, que se combinan de dos formas:

1. *Aditiva* ($Y_t = T_t + S_t + C_t + \varepsilon_t$): apropiada cuando la magnitud de las fluctuaciones no varía en el tiempo.

2. *Multiplicativa* ($Y_t = T_t \cdot S_t \cdot C_t \cdot \varepsilon_t$): mejor cuando las fluctuaciones son proporcionales al nivel de la serie, como en la mayoría de los agregados macroeconómicos, y tomando logaritmos se convierte en aditiva.

**Desestacionalización**: la estacionalidad es la variación periódica y predecible de la serie con un periodo de hasta un año, debida a causas como el clima, las vacaciones o los festivos. El componente estacional se estima como el patrón que se repite cada año (habitualmente mediante medias móviles centradas o modelos autorregresivos integrados de media móvil, ARIMA) y se elimina de la serie, restándolo o dividiendo por él según el modelo sea aditivo o multiplicativo.

### 3.1.2. Estimación de la tendencia y el ciclo

Una vez desestacionalizada la serie del PIB real, y expresada en logaritmos para que su descomposición sea aditiva, quedan dos componentes no observables, la tendencia y el ciclo.

Formalmente, con $y_t = \ln Y_t$ y $\tau_t = \ln \bar{Y}_t$, la brecha de producción es:

$$\tilde{y}_t = y_t - \tau_t \approx \frac{Y_t - \bar{Y}_t}{\bar{Y}_t}$$

es decir, la desviación porcentual del PIB observado respecto del potencial. Toda la dificultad está en $\tau_t$, que debe estimarse.

La producción potencial (PIB potencial, en las estimaciones oficiales) admite una doble lectura. En la lectura estadística es el componente tendencial de la serie, de oscilaciones más suaves, y en la económica, el nivel de *output* compatible con una inflación estable, con los factores de oferta o con el uso pleno de los recursos. A cada lectura corresponde una familia de métodos.

**Métodos univariantes**: extraen la tendencia de la propia serie del PIB, sin información económica adicional:

▪ *Tendencia lineal* ($y_t = \tau_0 + \bar{g}_Y\, t + \varepsilon_t$): supone que la tendencia es determinista, es decir, que la producción potencial crece a una tasa constante, supuesto que la evidencia de raíz unitaria pone en duda (apartado 2.2.2).

▪ *Filtro de Hodrick-Prescott* (1997): supone que la tendencia es estocástica y varía suavemente. Se elige la senda $\tau_t$ que equilibra dos objetivos contrapuestos, ajustarse a los datos y ser suave:

$$\min_{\{\tau_t\}} \sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2$$

donde $y_t$ es la serie observada del PIB, $\tau_t$ la tendencia que se elige, $T$ el número de observaciones y $\lambda$ el parámetro que penaliza los cambios en la pendiente de la tendencia.

- *CPO*: derivando respecto a cada $\tau_t$ se obtiene un sistema lineal cuya solución es $\tau = (I + \lambda K'K)^{-1}\mathbf{y}$, donde $K$ es la matriz de segundas diferencias, de modo que la tendencia es un promedio ponderado de la serie observada, tanto más suave cuanto mayor es $\lambda$.

- *CSO*: el objetivo es una suma de cuadrados, estrictamente convexa en $\tau$, por lo que la CPO caracteriza el mínimo.

Si $\lambda \to \infty$, la tendencia es lineal y, si $\lambda = 0$, coincide con la serie y no hay ciclo. Las convenciones habituales son $\lambda = 1600$ para datos trimestrales, 6,25 para anuales (Ravn y Uhlig, 2002) y 129 600 para mensuales.

**Métodos multivariantes**: incorporan la lectura económica de la producción potencial. Los métodos basados en la función de producción (FMI, Comisión Europea, OCDE) estiman el nivel de *output* compatible con la utilización plena de los factores. Para ello especifican una función de producción agregada con el capital, el trabajo ajustado por la tasa de paro que no acelera la inflación (NAIRU) y la productividad tendencial. Con todo, no existe un único método correcto para separar la tendencia del ciclo.

**Usos de la brecha de producción**: la importancia de su análisis es triple:

▪ *Posición cíclica*: informa de si la economía se encuentra en una situación de infrautilización de recursos (brecha negativa) o de sobreutilización (brecha positiva).

▪ *Política monetaria*: sirve de guía, pues una brecha positiva anticipa presiones inflacionistas, y viceversa.

▪ *Política fiscal*: es el paso previo para medir el tono de la política fiscal, pues el saldo estructural se obtiene descontando del saldo observado el componente cíclico (véase el tema A39).

## 3.2. Hechos estilizados

Estimados los componentes cíclicos de los principales agregados, pueden calcularse los hechos estilizados de las fluctuaciones, que se resumen en la volatilidad, los comovimientos y la persistencia de cada serie[18]:

[18] La inversión es unas tres veces más volátil que el PIB y el volumen de comercio, entre dos y tres veces. Las horas por trabajador, en cambio, son menos volátiles que el PIB.

▪ *Volatilidad*: se mide como la cuasidesviación típica del componente cíclico de cada agregado, en relación con la del PIB. En muchas economías de la OCDE, la inversión, el consumo de bienes duraderos y el volumen de comercio (exportaciones más importaciones) son más volátiles que el PIB, y el empleo es de volatilidad similar. Por el contrario, el *stock* de capital, la productividad del trabajo, el salario real, el consumo de bienes no duraderos y el gasto público son menos volátiles que el PIB.

▪ *Comovimientos*: además de las regularidades del apartado 1.1, la productividad del trabajo es débilmente procíclica y el salario real es débilmente procíclico o acíclico, según la estimación. La inflación es débilmente procíclica cuando dominan los *shocks* de demanda y contracíclica cuando dominan los de oferta.

▪ *Persistencia*: se calcula como el coeficiente de autocorrelación de primer orden del componente cíclico de cada serie, y en el PIB es elevada.

### 3.3. Las teorías frente a los hechos

Medidas frente a los hechos estilizados, las teorías modernas se complementan más que se excluyen:

▪ *Ciclos nominales (Lucas)*: el modelo explica los efectos reales de las perturbaciones monetarias no anticipadas, pero no la persistencia, ya que, sin un mecanismo de propagación, la brecha de producción que genera es ruido blanco, cuando la autocorrelación de primer orden observada es elevada.

▪ *Ciclos reales (TCR)*: el modelo calibrado reproduce la ordenación de volatilidades, la prociclicidad del consumo y la inversión y la persistencia del *output* (que hereda de $\rho_A$), pero falla en el mercado de trabajo y en los efectos reales del dinero (apartado 2.2.3).

▪ *Ciclos de desequilibrio (NEK)*: explican la no neutralidad del dinero y sus efectos asimétricos (apartado 2.3.1), pero no generan por sí mismos el conjunto de comovimientos agregados.

▪ *Nueva síntesis (IS-PC-MR)*: recoge a la vez los efectos reales de los *shocks* tecnológicos y de la política monetaria y la prociclicidad de la inflación ante perturbaciones de demanda, pero no la persistencia de la inflación sin indexación (apartado 2.4.3).

Ninguna teoría reproduce por sí sola todos los hechos. De ahí la evolución hacia modelos DSGE de tamaño medio (Smets y Wouters, 2007) que combinan *shocks* reales y nominales, rigideces de precios y salarios, hábitos de consumo y costes de ajuste de la inversión, y que se estiman con los datos observados.

### 3.4. Conclusiones

La medición del ciclo es inseparable de su teoría, porque el filtro con que se separa la tendencia y la definición de producción potencial que se adopte condicionan los hechos estilizados que después sirven de criterio. Así, una tendencia lineal supone un potencial que crece a tasa constante, el filtro de Hodrick-Prescott deja la suavidad de la tendencia al valor de $\lambda$ y los métodos de función de producción dependen de una NAIRU que también se estima. Por eso las estimaciones de la brecha se revisan con frecuencia.

Con esa cautela, los hechos estilizados ofrecen un banco de pruebas común a todas las teorías, y cada una acierta en unos hechos y falla en otros. La respuesta han sido los modelos DSGE de tamaño medio, que reúnen elementos de todas ellas.

## Conclusiones

Las teorías del ciclo distinguen el impulso que lo desencadena del mecanismo que lo propaga. Los modelos del multiplicador y el acelerador formalizan por primera vez la propagación, pero solo generan ciclos persistentes y acotados con supuestos *ad hoc* y reglas de conducta mecánicas. Las teorías modernas responden a la crítica de Lucas con microfundamentos. Para la nueva macroeconomía clásica el ciclo es una sucesión de equilibrios, provocada por sorpresas nominales (Lucas) o por perturbaciones tecnológicas (ciclo real), y la política estabilizadora sobra. Para la nueva economía keynesiana, en cambio, es un fenómeno de desequilibrio, en el que las rigideces justifican la intervención y en el que las fricciones financieras, del racionamiento del crédito al ciclo financiero, amplifican las perturbaciones. La nueva síntesis integra ambas tradiciones en el modelo IS-PC-MR. Sin embargo, ninguna teoría reproduce todos los hechos estilizados, lo que explica el paso a los modelos DSGE de tamaño medio.

Las teorías del ciclo proceden de las escuelas macroeconómicas, el equilibrio con sorpresas de Lucas y el ciclo real de la nueva macroeconomía clásica (tema A06) y las rigideces dinámicas y el modelo de tres ecuaciones de la nueva economía keynesiana (tema A07). El acelerador es la versión agregada de la teoría de la inversión (tema A34), y el ciclo real es el modelo de Ramsey-Cass-Koopmans sometido a perturbaciones tecnológicas (tema A43). Por su parte, las fricciones financieras unen el ciclo con las crisis. La inestabilidad de Minsky y el racionamiento del crédito explican cómo el endeudamiento acumulado en la expansión desemboca en crisis bancarias (tema B26), y el colchón de capital anticíclico intenta suavizar esa dinámica (tema B27).

Está en discusión si la política monetaria debe "luchar contra el viento" del ciclo financiero. Borio (2014) sostiene que estabilizar la inflación no basta para evitar los desequilibrios que ese ciclo acumula. Svensson (2017), en cambio, estima que los costes de subir los tipos para frenar el crédito superan con holgura sus beneficios en menor probabilidad de crisis, y reserva esa tarea a la política macroprudencial. Sigue abierta también la separación entre tendencia y ciclo. Blanchard, Cerutti y Summers (2015) documentan que muchas recesiones van seguidas de una producción duraderamente inferior a la tendencia previa, lo que cuestiona tratar el ciclo como una desviación transitoria y aproxima su teoría a la del crecimiento.

# Anexo. Ecuación característica, raíces y ciclo

El comportamiento cíclico de un modelo dinámico lineal se lee en las raíces de su ecuación característica. El resultado se aplica tanto al modelo de Samuelson (apartado 1.3.1) como a la ecuación de la desviación del *output* respecto de su tendencia en la teoría del ciclo real (apartado 2.2.2), ya que ambos son ecuaciones en diferencias lineales de segundo orden.

## A.1. La ecuación en diferencias y su solución

Una ecuación en diferencias lineal de segundo orden con coeficientes constantes tiene la forma general:

$$Y_t = a_1\, Y_{t-1} + a_2\, Y_{t-2} + b$$

En el modelo de Samuelson, $a_1 = c(1+v)$, $a_2 = -vc$ y $b = I_0 + C_0$. En la ecuación del ciclo real, $a_1 = \alpha + \rho_A$, $a_2 = -\alpha\rho_A$ y el término independiente es la perturbación $(1-\alpha)\phi_t$. La solución se descompone en dos partes: una *solución particular*, que es el valor de equilibrio o estado estacionario ($Y^* = b/(1-a_1-a_2)$, constante), y la *solución de la ecuación homogénea* ($Y_t = a_1 Y_{t-1} + a_2 Y_{t-2}$), que describe cómo se comporta la desviación respecto de ese equilibrio a lo largo del tiempo. El ciclo, si existe, está en la segunda.

## A.2. La ecuación característica

Para resolver la ecuación homogénea se ensaya una solución de la forma $Y_t = r^t$, es decir, una trayectoria que en cada periodo se multiplica por un factor constante $r$. Sustituyendo, $r^t = a_1 r^{t-1} + a_2 r^{t-2}$, y dividiendo por $r^{t-2}$ se obtiene la **ecuación característica**:

$$r^2 - a_1\, r - a_2 = 0$$

que en el caso de Samuelson es $r^2 - c(1+v)\, r + vc = 0$. La ecuación es de segundo grado en $r$, y sus dos raíces, $r_1$ y $r_2$, se obtienen con la fórmula cuadrática:

$$r_{1,2} = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2}$$

La solución general de la ecuación homogénea es una combinación lineal de las dos trayectorias, $Y_t = A_1 r_1^t + A_2 r_2^t$, donde las constantes $A_1$ y $A_2$ quedan fijadas por las dos condiciones iniciales $(Y_0, Y_1)$. Todo lo que hay que saber sobre la dinámica del modelo está, por tanto, en $r_1$ y $r_2$, pues cada raíz es el factor por el que se multiplica su componente de la desviación en cada periodo.

## A.3. Raíces reales: basta el valor absoluto

Si el discriminante es positivo o nulo ($a_1^2 + 4a_2 \geq 0$, que en Samuelson equivale a $c(1+v)^2 \geq 4v$), las raíces son reales. Cada componente $A_i r_i^t$ crece o decrece geométricamente sin cambiar de signo (si $r_i > 0$) o alternando de signo cada periodo (si $r_i < 0$, un caso sin interés económico aquí). La trayectoria no oscila y no hay ciclos. Con raíces positivas no es necesariamente monótona, pues la suma de las dos componentes puede presentar un único punto de giro (la "joroba" del ciclo real, apartado A.4) antes de converger o divergir. Que la desviación se extinga o se amplifique depende solo de la magnitud de las raíces, de modo que el análisis se reduce a comparar el valor absoluto de cada raíz con la unidad:

- $|r_1| < 1$ *y* $|r_2| < 1$: ambas componentes tienden a cero y la economía converge sin oscilaciones al equilibrio (región A del plano $(v, c)$).

- *Alguna raíz con* $|r_i| > 1$: esa componente domina y la desviación crece sin límite, con divergencia sin oscilaciones (región B).

## A.4. Raíces complejas: hace falta el módulo

Si el discriminante es negativo ($a_1^2 + 4a_2 < 0$, que en Samuelson equivale a $c(1+v)^2 < 4v$), la raíz cuadrada no existe en los reales y las raíces son un par de números complejos conjugados, $r_{1,2} = h \pm \omega i$, con parte real $h = a_1/2$ y parte imaginaria $\omega = \sqrt{-(a_1^2 + 4a_2)}/2$. Para interpretarlos, se escriben en forma polar, $r_{1,2} = R\,(\cos\varphi \pm i\,\sin\varphi)$, donde:

$$R = \sqrt{h^2 + \omega^2} = \sqrt{-a_2} \qquad \text{y} \qquad \cos\varphi = \frac{h}{R}$$

$R$ es el **módulo** de las raíces (su distancia al origen en el plano complejo) y $\varphi$ su argumento. Por la fórmula de De Moivre, $r^t = R^t(\cos\varphi t \pm i\,\sin\varphi t)$, de modo que, combinando ambas raíces para que la solución sea real, la desviación sigue la trayectoria:

$$Y_t = A\,R^t \cos(\varphi t + \psi)$$

Aquí se ve por qué las raíces complejas generan ciclos y por qué el criterio cambia. El coseno hace que la desviación oscile con un periodo de $2\pi/\varphi$ periodos, de modo que la trayectoria es cíclica por construcción, cualquiera que sea el valor de las raíces. Además, la amplitud de esas oscilaciones viene multiplicada en cada periodo por $R$, por lo que ya no tiene sentido hablar del valor absoluto de cada raíz por separado (un número complejo no tiene signo ni orden), sino de su módulo común:

- $R < 1$: la amplitud decrece y el ciclo es amortiguado. La economía oscila en torno al equilibrio y converge a él (región D). En Samuelson, $R = \sqrt{vc} < 1$, es decir, $vc < 1$.

- $R = 1$: la amplitud es constante y el ciclo se mantiene indefinidamente (región C, un caso límite sin relevancia empírica, pues exige $vc = 1$ exactamente).

- $R > 1$: la amplitud crece y el ciclo es explosivo (región E).

En la ecuación del ciclo real, con $a_1 = \alpha+\rho_A$ y $a_2 = -\alpha\rho_A$, el discriminante es $(\alpha+\rho_A)^2-4\alpha\rho_A = (\alpha-\rho_A)^2 \geq 0$, por lo que las raíces son siempre reales ($r_1 = \alpha$, $r_2 = \rho_A$) y menores que la unidad. El modelo del ciclo real no genera, por tanto, oscilaciones propias, sino, tras cada perturbación, una respuesta sin oscilaciones (que, cuando $\alpha + \rho_A > 1$, primero crece y luego decrece, en forma de "joroba"), cuya duración depende de la mayor de las dos raíces, en la práctica $\rho_A$. Que el ciclo observado sea recurrente se debe a la llegada continua de perturbaciones $\phi_t$, no a la dinámica interna del modelo, lo que ilustra la distinción moderna entre impulso y mecanismo de propagación (apartado 1.1).

## A.5. Resumen

Tabla A42.1: Raíces de la ecuación característica, trayectoria y criterio de estabilidad.

| Discriminante | Raíces | Trayectoria | Criterio de estabilidad |
|---|---|---|---|
| $a_1^2 + 4a_2 \geq 0$ | Reales | Sin oscilaciones (sin ciclos) | $\lvert r_i\rvert < 1$ para ambas raíces |
| $a_1^2 + 4a_2 < 0$ | Complejas conjugadas | Cíclica, de periodo $2\pi/\varphi$ | Módulo $R = \sqrt{-a_2} < 1$ |

La misma lógica se extiende a ecuaciones de orden superior (tantas raíces como orden) y a los procesos autorregresivos de la econometría de series temporales. Un proceso AR(p) es estacionario cuando todas las raíces de su ecuación característica tienen módulo menor que la unidad, y presenta oscilaciones cuando alguna de ellas es compleja.

## A.6. Representación gráfica del modelo de Kaldor (1940)

El modelo del apartado 1.3.3 se representa en el plano renta-ahorro e inversión, $Y$-($S$, $I$), con las funciones de ahorro e inversión planeados (figura A42.7):

1. *Función de inversión*, $I(Y)$: tiene forma de S. Es plana en los extremos, porque el exceso de capacidad y los límites de oferta deprimen la propensión marginal a invertir, y empinada en el tramo intermedio.

2. *Función de ahorro*, $S(Y)$: tiene la forma contraria. Es empinada en los extremos, por el desahorro cuando la renta cae y la saturación del consumo cuando es alta, y más plana en el tramo intermedio.

3. *Equilibrios* (panel a): las dos funciones se cortan tres veces. En A (actividad baja) y en 0 (actividad alta), la pendiente de $S(Y)$ supera a la de $I(Y)$ y el equilibrio es estable, mientras que en B (actividad intermedia) ocurre lo contrario y el equilibrio es inestable. Las flechas indican el sentido del ajuste, pues la renta sube cuando $I > S$ y baja cuando $I < S$.

4. *Acumulación de capital* (panel b): en el equilibrio alto la inversión es elevada y el *stock* de capital crece, lo que desplaza $I(Y)$ hacia abajo, de $I_0(Y)$ a $I_1(Y)$, y $S(Y)$ hacia arriba, de $S_0(Y)$ a $S_1(Y)$. Los equilibrios alto e intermedio se acercan hasta fundirse en $0'$, donde ambas funciones son tangentes, y la economía cae al único equilibrio que queda, el bajo (punto 1). Allí el capital se desacumula, los desplazamientos se invierten y el proceso se repite en sentido contrario hasta que la economía salta de nuevo al equilibrio alto, lo que cierra el ciclo.

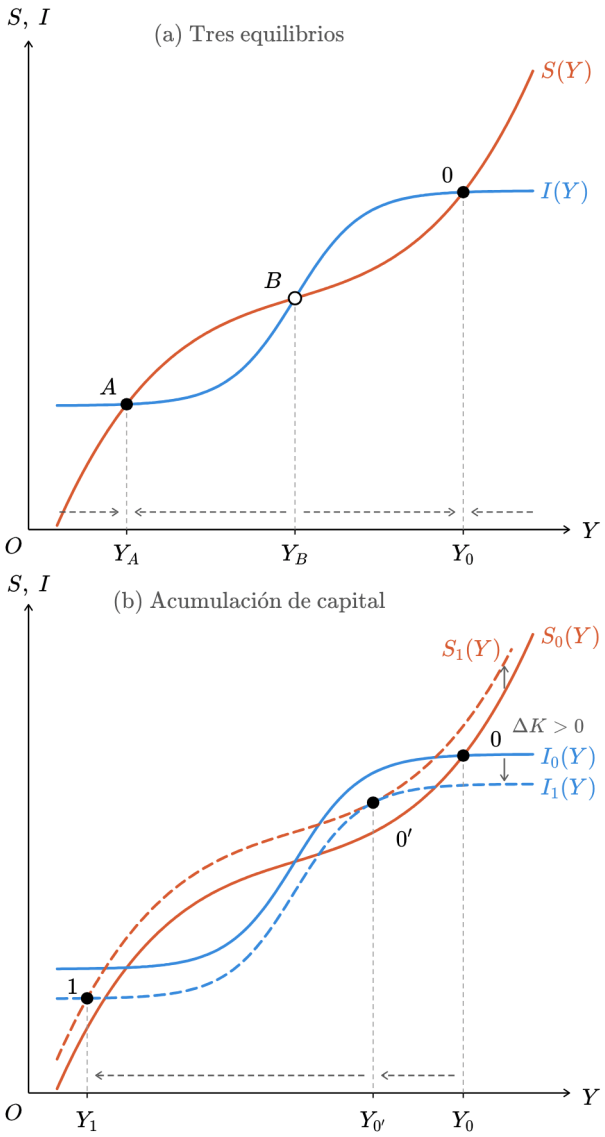



Figura A42.7: Modelo del ciclo de Kaldor (1940). En el panel (a), $S(Y)$ e $I(Y)$ se cortan en dos equilibrios estables, A y 0, y en uno inestable, B. En el panel (b), la acumulación de capital desplaza $I(Y)$ hacia abajo y $S(Y)$ hacia arriba hasta que 0 y B se funden en $0'$ y la economía cae al equilibrio bajo (de 0 a 1).

# A43. Crecimiento económico exógeno

Crecimiento económico (I). Acumulación de capital y progreso técnico exógeno. El modelo de Solow. El modelo de Solow aumentado con acumulación de capital humano. El modelo de Ramsey-Cass-Koopmans.

## Introducción

De todos los fenómenos que estudia la macroeconomía, el crecimiento económico es probablemente el de mayores consecuencias para el bienestar. Como escribió Robert Lucas (1988), cuando uno empieza a pensar en el crecimiento económico, le resulta difícil pensar en otra cosa. Por **crecimiento económico** se entiende el aumento sostenido de la producción agregada, y en particular de la producción por habitante, a lo largo del tiempo.

El crecimiento sostenido es un fenómeno reciente en la historia. Según las estimaciones de Angus Maddison, entre 1820 y 1998 la renta real per cápita se multiplicó por más de veinte en Estados Unidos y por casi quince en Europa occidental, mientras que durante los milenios anteriores apenas había variado[1]. Ese avance, sin embargo, ha sido desigual según el país y la época. Como las tasas se acumulan igual que el interés compuesto, crecer algo más deprisa durante varias décadas basta para que la distancia entre rentas per cápita se haga enorme. De ahí surge la pregunta central de la teoría del crecimiento, por qué crecen las economías y por qué unas crecen más que otras, a la que la literatura ofrece tres grandes respuestas, no excluyentes: la acumulación de capital físico (inversión), la formación de capital humano (educación) y el progreso tecnológico (innovación).

La teoría moderna del crecimiento tiene sus antecedentes en los clásicos, que vincularon la riqueza de las naciones a la división del trabajo y a la acumulación (Smith) y anticiparon el estado estacionario por los rendimientos decrecientes de la tierra (Ricardo, Malthus). El modelo keynesiano de Harrod (1939) y Domar (1946) mostró la inestabilidad de la senda de crecimiento. En cambio, el modelo neoclásico de Solow y Swan (1956), completado por Cass y Koopmans (1965) con el enfoque de Ramsey (1928), explicó la **convergencia** hacia un **estado estacionario**, pero dejó el progreso técnico fuera del modelo. Los modelos de **crecimiento endógeno**, iniciados por Paul Romer (1986) y Robert Lucas (1988), trataron de explicarlo desde dentro, y la contrastación empírica, desde Barro (1991) y Mankiw, Romer y Weil (1992), ha confrontado todas estas teorías con los datos. El objetivo último de la teoría del crecimiento es, así, doble, pues busca determinar cómo elevar el crecimiento global y cómo aproximar los niveles de vida de los países más pobres a los de los más ricos.

El tema estudia los modelos de crecimiento con progreso tecnológico exógeno en el orden en que la teoría los produjo:

[1] Maddison, A. (2001), *The World Economy: A Millennial Perspective*, París, OCDE.

- *Modelo de Harrod-Domar*: el modelo keynesiano relaciona el crecimiento con el ahorro y con la relación entre capital y producto.

- *Modelo de Solow-Swan*: el modelo neoclásico se estudia junto con su extensión posterior, el modelo aumentado con capital humano (Mankiw, Romer y Weil, 1992).

- *Modelo de Ramsey-Cass-Koopmans*: este modelo microfundamenta la decisión sobre la tasa de ahorro.

# 1. Modelo de Harrod-Domar (1939, 1946)

> **Sobre esta sección**
>
> El modelo de Harrod-Domar se expone por su valor conceptual, pues plantea el problema de la estabilidad del crecimiento al que responde el modelo de Solow-Swan. Desde la revisión del programa de 2022 no figura en el título del tema, de modo que en la exposición conviene tratarlo con brevedad.

**Origen**: Roy Harrod (1939) y Evsey Domar (1946) propusieron, con pocos años de diferencia, dos modelos de raíz keynesiana, que varían en detalles pero cuyas implicaciones coinciden.

**Idea**: la inversión tiene una naturaleza dual, crea demanda vía multiplicador y capacidad vía acelerador, y solo una tasa de crecimiento concreta, la garantizada, hace que ambas crezcan al mismo ritmo.

Nada asegura, sin embargo, que la economía se sitúe en ella, y si se desvía se aleja. En consecuencia, el crecimiento basado en la acumulación del capital es inestable, y la economía transita al filo de la navaja.

## 1.1. Supuestos

**Entorno**: mundo keynesiano en tiempo discreto, con variables fechadas por periodos, a diferencia de los modelos de Solow-Swan y de Ramsey-Cass-Koopmans, que se formulan en tiempo continuo. Los precios relativos no ajustan las proporciones de factores ni el ahorro a la inversión, y las decisiones de inversión dependen de las expectativas de los empresarios. La economía es cerrada, desarrollada, con rendimientos constantes a escala y sin sector público, de modo que la producción se destina a consumo e inversión, $Y = C + I$, y el ahorro iguala *ex post* a la inversión, $S = I$. *Ex ante*, la igualdad entre ahorro e inversión deseados es la condición de equilibrio. Familias y empresas no se modelizan por separado.

**Agentes**: las familias ahorran una fracción constante y exógena de la renta, $S = sY$. Las empresas ajustan su *stock* de capital al deseado periodo a periodo según el principio del acelerador, $K^* = vY$, donde $v$ es la ratio capital-output deseada, constante y determinada por la tecnología. Se cumple el multiplicador keynesiano, $\Delta Y = \Delta I/s$.

**Tecnología**: la producción sigue una función de Leontief con dos factores complementarios, $Y = \min(K/v, L/u)$, sin sustituibilidad entre ellos. El capital no se deprecia ($I_{\text{neta}} = I_{\text{bruta}}$), la fuerza de trabajo crece a la tasa constante y exógena $n$, $L_{t+1} = (1+n)L_t$, y no hay progreso tecnológico.

## 1.2. Modelo

**Planteamiento**: no hay programa de optimización. Los rendimientos constantes permiten trabajar con la función de producción agregada en forma intensiva, $y = \min(k/v, 1/u)$. El modelo se reduce a comparar tres tasas de crecimiento que se derivan de esta expresión: la que exige el equilibrio entre ahorro e inversión, la que permite la población y la que efectivamente se produce:

▪ *Tasa garantizada* (Harrod): de la igualdad entre ahorro e inversión, $S_t = sY_t = I_t = K_{t+1} - K_t = v(Y_{t+1} - Y_t)$, se obtiene la tasa de crecimiento que mantiene el equilibrio del mercado de bienes en cada periodo,

$$g_w = \frac{Y_{t+1} - Y_t}{Y_t} = \frac{s}{v}$$

donde $s \in (0, 1)$ es la tasa de ahorro y $v$ la ratio capital-output, por lo que la tasa garantizada aumenta con la propensión al ahorro y disminuye con la intensidad de capital de la tecnología.

▪ *Tasa natural*: es la máxima compatible con el pleno empleo, dada por el crecimiento de la fuerza de trabajo, $g_n = n$ (con progreso técnico, $n + g_A$).

▪ *Tasa efectiva*: $g_Y = (Y_{t+1} - Y_t)/Y_t$, la que resulta de las decisiones de inversión de los empresarios, guiadas por los *animal spirits*.

▪ *Condición de Domar*: la tasa garantizada se obtiene también partiendo del doble efecto de la inversión. Su aumento amplía la demanda agregada a través del multiplicador, $\Delta Y^d = \Delta I/s$, mientras que su nivel añade una capacidad productiva igual a $I/v$, donde $1/v$ es la productividad del capital. Igualando ambos incrementos, la inversión ha de crecer a la tasa $\Delta I/I = s/v$, es decir, a la tasa garantizada.

**Dinámica**: Harrod argumenta que el equilibrio $g_Y = g_w$ es inestable, y la economía discurre por el *filo de la navaja* (*knife-edge*). Concretamente, identifica dos desviaciones:

▪ *Exceso de ahorro* ($g_Y < g_w$, $S > I$): la producción no encuentra comprador suficiente. Ante ese excedente, los empresarios rebajan la producción que esperan vender y, como trabajan con proporciones fijas de capital, el acelerador reduce el capital que desean tener. Invierten entonces menos, la demanda se contrae aún más y el desequilibrio inicial se amplía, de modo que el crecimiento se frena periodo tras periodo en una espiral implosiva, con la que Harrod y Domar explicaban la década de 1930.

▪ *Exceso de inversión* ($g_Y > g_w$, $S < I$): el proceso opera al revés y el crecimiento se vuelve explosivo.

A largo plazo se añade un segundo problema. Nada garantiza que $g_w = g_n$, pues $s$, $v$ y $n$ son parámetros independientes. Caben dos casos:

▪ $g_w > g_n$: el capital crece más deprisa que el trabajo y, sin sustituibilidad, aparece capacidad ociosa permanente.

▪ $g_w < g_n$: aparece desempleo creciente y permanente.

**Extensiones**: dada la inestabilidad intrínseca de la economía, la intervención pública se vuelve imprescindible. Además, si el ahorro no basta para financiar el crecimiento objetivo, $g_Y^T$ (es decir, si $g_w = s/v < g_Y^T$), existe un *savings gap*, una brecha entre la tasa de ahorro corriente, $s$, y la necesaria para financiar la inversión que exige ese crecimiento, $v\, g_Y^T$.

▪ *Política fiscal anticíclica*: cuando sobra ahorro, la inversión pública debe aumentar, y debe reducirse cuando escasea.

▪ *Políticas de desarrollo*: la brecha de ahorro, en cambio, solo se cierra elevando el ahorro disponible, ya sea interno (ahorro público) o externo (ayuda exterior).

Ante la inestabilidad del filo de la navaja, surgen dos soluciones. O bien se adopta una función de producción continua, que permite que $v$ se ajuste (solución adoptada por Solow-Swan), o bien se endogeniza la tasa de ahorro y se permite que $s$ se ajuste.

La segunda vía es la de la escuela de Cambridge (Kaldor, 1956; Pasinetti, 1962), en la que $s$ depende de la distribución de la renta entre salarios y beneficios. Si los perceptores de beneficios ahorran una fracción $s_\pi$ mayor que la de los asalariados, $s_w$, la tasa de ahorro agregada es $s = s_w + (s_\pi - s_w)\,\Pi/Y$, con $\Pi$ los beneficios. La condición $s/v = n$ determina entonces la participación de los beneficios que hace compatible la tasa garantizada con la natural, $\Pi/Y = (nv - s_w)/(s_\pi - s_w)$, lo que exige $s_w < nv < s_\pi$. Con pleno empleo, el ajuste lo realizan los precios, ya que un exceso de inversión eleva los precios respecto de los salarios nominales, aumenta la participación de los beneficios y, con ella, el ahorro.

Si los asalariados no ahorran, la tasa de beneficio es $\Pi/K = n/s_\pi$. Pasinetti demostró que, con propensiones al ahorro por clase social y no por tipo de renta, el resultado se mantiene aunque los asalariados ahorren, siempre que su ahorro obtenga la misma tasa de beneficio (*teorema de Pasinetti*). La tasa de beneficio de equilibrio, $\Pi/K = n/s_c$, depende entonces solo de la tasa de crecimiento natural y de la propensión al ahorro de los capitalistas, $s_c$. Ramsey-Cass-Koopmans endogeniza también el ahorro, pero sobre una función de producción neoclásica en la que $v$ ya se ajusta.

## 1.3. Conclusiones

**Aspectos positivos**: con Harrod-Domar arranca la teoría moderna del crecimiento. El modelo lleva además la interacción entre multiplicador y acelerador a un marco dinámico y formula la cuestión que vertebra el tema, la de si el crecimiento equilibrado es estable. Su noción de brecha de ahorro sirvió de fundamento a las políticas de desarrollo apoyadas en la ayuda exterior.

**Limitaciones**: la función de producción de coeficientes fijos genera la inestabilidad del filo de la navaja, que la evidencia no respalda, pues las economías no siguen sendas de crecimiento explosivas ni implosivas. Asimismo, como $s$, $v$ y $n$ son parámetros independientes, la igualdad entre la tasa garantizada y la natural sería fortuita, y fuera de ella el modelo predice capacidad ociosa o desempleo permanentes.

De las dos salidas del filo de la navaja, la escuela de Cambridge ajusta la tasa de ahorro mediante la distribución de la renta. La sección 2 sigue la otra y deja que la ratio capital-output se ajuste con la función de producción neoclásica de Solow-Swan, y la sección 3 endogeniza el ahorro sobre esa misma función.

# 2. Modelo de Solow-Swan (1956)

**Origen**: lo propusieron Robert Solow (1956) y, de forma independiente, Trevor Swan (1956).

**Idea**: el modelo sustituye la función de Leontief por una función de producción neoclásica con sustituibilidad entre capital y trabajo. Con ello, la ratio capital-output se ajusta y la inestabilidad del filo de la navaja desaparece.

El modelo estudia el papel de la inversión como motor del crecimiento y concluye que, por los rendimientos decrecientes del capital, solo el progreso técnico sostiene el crecimiento per cápita a largo plazo.

## 2.1. Supuestos

**Entorno**: el modelo se formula en tiempo continuo, con información y competencia perfectas. La economía es cerrada, desarrollada y sin sector público, como en Harrod-Domar.

**Agentes**: las familias son agentes del tipo Robinson Crusoe, a la vez productores, consumidores y propietarios de los factores y de la tecnología. Ahorran e invierten una fracción constante de la renta, $S = I = sY$. Esa fracción es un parámetro exógeno y no se deriva de ninguna decisión.

**Tecnología**: la función de producción es una Cobb-Douglas neoclásica:

$$Y = AK^{\alpha}L^{1-\alpha}, \quad \alpha \in (0,1)$$

donde $A$ es la tecnología (Hicks-neutral, inicialmente constante, $\dot{A} = 0$), y $\alpha$ y $(1-\alpha)$ son las participaciones de capital y trabajo en la renta (constantes, dadas por la tecnología). El capital se deprecia a una tasa constante $\delta$ y la población crece a una tasa constante $n$.

La función cumple tres propiedades neoclásicas:

- *Rendimientos constantes a escala*: $f(\lambda K, \lambda L) = \lambda f(K, L)$.

- *Productividades marginales positivas y decrecientes*: $f'_K > 0$, $f''_K < 0$, con productividades marginales cruzadas positivas, $f''_{KL} = f''_{LK} > 0$ (la simetría se sigue del teorema de Young).

- *Condiciones de Inada*:

  - *Condición de Inada en el origen*: la productividad marginal del capital tiende a ser infinita cuando el *stock* de capital tiende a ser nulo, $\lim_{K\to 0} f'_K = \infty$.

  - *Condición de Inada en el infinito*: la productividad marginal del capital tiende a ser nula cuando el *stock* de capital tiende a ser infinito, $\lim_{K\to\infty} f'_K = 0$.

## 2.2. Modelo sin progreso técnico

**Planteamiento**: no hay programa de optimización, porque la tasa de ahorro es exógena. La economía no tiene sector exterior ni sector público, por lo que la renta se destina a consumo o a inversión:

$$Y = C + I$$

El modelo determina estas tres variables con una ecuación de comportamiento para cada una. Cada ecuación se escribe en niveles, en términos per cápita ($x = X/L$) y en tasas de crecimiento ($g_x \equiv \dot{x}/x$), que se obtienen tomando logaritmos y derivando respecto al tiempo con las reglas que recoge el Anexo.

- *Renta*: la genera la función de producción, que, con rendimientos constantes a escala, se escribe en forma intensiva dividiendo por $L$:

$$Y = AK^{\alpha}L^{1-\alpha} \quad \Rightarrow \quad y = Ak^{\alpha} \quad \Rightarrow \quad g_y = \alpha g_k$$

La renta per cápita depende solo del capital per cápita, y su tasa de crecimiento es una fracción $\alpha$ de la de este. Por eso, entender el comportamiento del capital equivale a entender el crecimiento.

- *Inversión*: las familias invierten la fracción $s$ de la renta, $I = sY$, y el capital aumenta con la inversión bruta y disminuye con la depreciación:

$$\dot{K} = sY - \delta K \quad \Rightarrow \quad \dot{k} = sAk^{\alpha} - (n+\delta)k \quad \Rightarrow \quad g_k = sAk^{\alpha-1} - (n+\delta)$$

La forma per cápita es la *ecuación fundamental de Solow-Swan* o ley de acumulación del capital per cápita. A diferencia de la renta, no basta con dividir por $L$, porque el capital per cápita cae también cuando crece la población (Anexo). La variación del capital per cápita es la diferencia entre el ahorro bruto per cápita, $sAk^{\alpha} = sf(k)$, y la inversión de reposición, $(n+\delta)k$, que mantiene constante el *stock* de capital per cápita ante el crecimiento de la población y la depreciación. Así, mientras el ahorro bruto supere la reposición ($\dot{k} > 0$), la economía acumula capital y crece.

En tasas, la economía crece siempre que la tasa de ahorro bruto por unidad de capital, $sAk^{\alpha-1}$, supere la tasa de reposición, $n + \delta$. La tasa de crecimiento depende positivamente de $s$ y $A$, y negativamente de $n$ y $\delta$. El efecto de $\alpha$, en cambio, es ambiguo, porque $\partial g_k/\partial\alpha = sAk^{\alpha-1}\ln k$ es positivo solo si $k > 1$.

- *Consumo*: es la parte de la renta que no se ahorra:

$$C = (1-s)Y \quad \Rightarrow \quad c = (1-s)Ak^{\alpha} \quad \Rightarrow \quad g_c = g_y$$

Con una tasa de ahorro constante, el consumo per cápita crece al mismo ritmo que la renta per cápita, y ambos siguen la evolución del capital per cápita.

**Equilibrio**: la renta per cápita crece hasta que $\dot{k} = 0$. En ese punto el ahorro bruto iguala la inversión de reposición y no quedan recursos para incrementar el capital per cápita una vez reemplazado el depreciado y equipados los nuevos trabajadores. Esa situación, en la que las tasas de crecimiento de todas las variables son constantes, es el *estado estacionario*. Sus valores per cápita son

$$k^* = \left(\frac{sA}{n+\delta}\right)^{1/(1-\alpha)}, \quad y^* = A^{1/(1-\alpha)}\left(\frac{s}{n+\delta}\right)^{\alpha/(1-\alpha)}, \quad c^* = (1-s)y^*$$

y los precios de los factores, que retribuyen a cada uno según su productividad marginal, son constantes, $w^* = (1-\alpha)Ak^{*\alpha}$ y $r^* = \alpha Ak^{*(\alpha-1)} - \delta$. Sin progreso tecnológico, $g_y^* = g_k^* = g_c^* = 0$ y, en agregado, $g_Y^* = g_K^* = n$. El estado estacionario tiene cuatro propiedades.

- *Existencia*: gracias a las dos condiciones de Inada. La del origen garantiza que el ahorro supere a la inversión de reposición para $k$ pequeño, y la del infinito, que quede por debajo para $k$ grande, de modo que ambas curvas se cortan en un $k^* > 0$.

- *Unicidad*: porque la función de ahorro por trabajador es cóncava y la de reposición lineal.

- *Estabilidad*: la economía gravita automáticamente hacia $k^*$ desde cualquier punto de partida, lo que constituye la diferencia fundamental con Harrod-Domar.

- *Ausencia de crecimiento endógeno*: gracias a la condición de Inada en el infinito (la productividad marginal del capital tiende a cero), que impide que la acumulación sostenga por sí sola el crecimiento per cápita. Por los rendimientos decrecientes del capital, solo el progreso técnico exógeno lo mantiene.

**Desarrollo gráfico**: el estado estacionario se representa en el plano capital por trabajador-producto por trabajador (figura A43.1):

1. *Función de producción*, $f(k)$: recoge la producción por trabajador y es cóncava, por los rendimientos decrecientes del capital.

2. *Curva de ahorro*, $sf(k)$: recoge el ahorro y la inversión bruta por trabajador, una fracción $s$ de la producción, y es cóncava como ella.

3. *Recta de reposición*, $(n+\delta)k$: recoge la inversión que mantiene constante el capital por trabajador ante el crecimiento de la población y la depreciación, y es lineal porque es proporcional a $k$.

4. *Tasa de crecimiento del capital*: en el panel inferior, en el plano $k$-$g_k$, es la distancia vertical entre la curva $sAk^{\alpha-1}$ y la recta $n+\delta$.

5. *Equilibrio*: en el punto 0, el cruce de la curva de ahorro y la recta de reposición determina $k^*$ y la inversión por trabajador del estado estacionario, $i^* = sf(k^*) = (n+\delta)k^*$, y el consumo es $c^* = y^* - i^*$. A la izquierda el ahorro supera a la reposición y $k$ crece ($g_k > 0$), mientras que a la derecha $k$ cae.

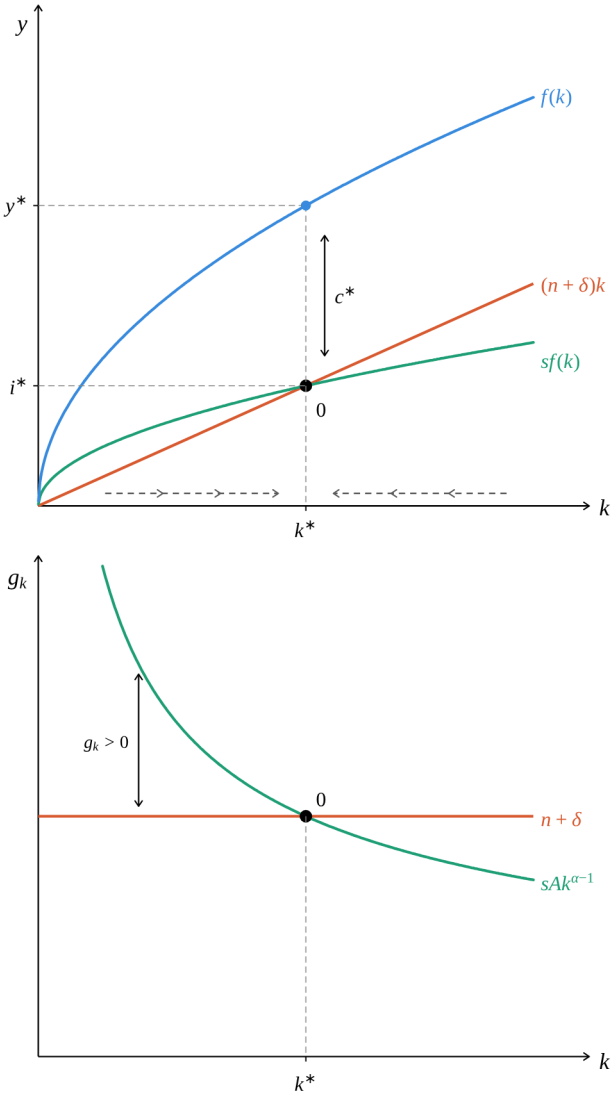



Figura A43.1: Modelo de Solow-Swan. El estado estacionario se alcanza donde el ahorro por trabajador iguala a la inversión de reposición.

**Estática comparativa**: la alteración permanente de un parámetro produce un efecto *nivel* permanente y un efecto *tasa* solo transitorio sobre las variables per cápita.

▪ *Aumento del ahorro* ($\Delta s > 0$): gráficamente, en el plano $k$-$y$ (figura A43.2, panel superior), la curva de ahorro pivota hacia arriba, de $s_0 f(k)$ a $s_1 f(k)$, y corta a la recta de reposición en un capital estacionario mayor, del punto 0 al 1 ($k_0^* < k_1^*$). En el plano $k$-$g_k$ (panel inferior), la curva de ahorro por unidad de capital se desplaza también hacia arriba, de $s_0 A k^{\alpha-1}$ a $s_1 A k^{\alpha-1}$. En $k_0^*$ la tasa de crecimiento del capital salta de cero a un valor positivo (del punto 0 al $0'$) y después decrece a lo largo de la nueva curva hasta anularse en el punto 1.

Por tanto, durante la transición la economía crece más deprisa, y una vez alcanzado el nuevo estado estacionario la tasa per cápita vuelve a ser la inicial. El consumo cae a corto plazo, porque la inversión absorbe una fracción mayor de la producción, y a largo plazo el consumo

solo aumenta si $k_0^*$ era inferior al capital de la regla de oro, que se define en este mismo apartado. Según el modelo, la recomendación habitual de los organismos internacionales a los países pobres, ahorrar más, eleva su nivel de renta, pero no su tasa de crecimiento a largo plazo.

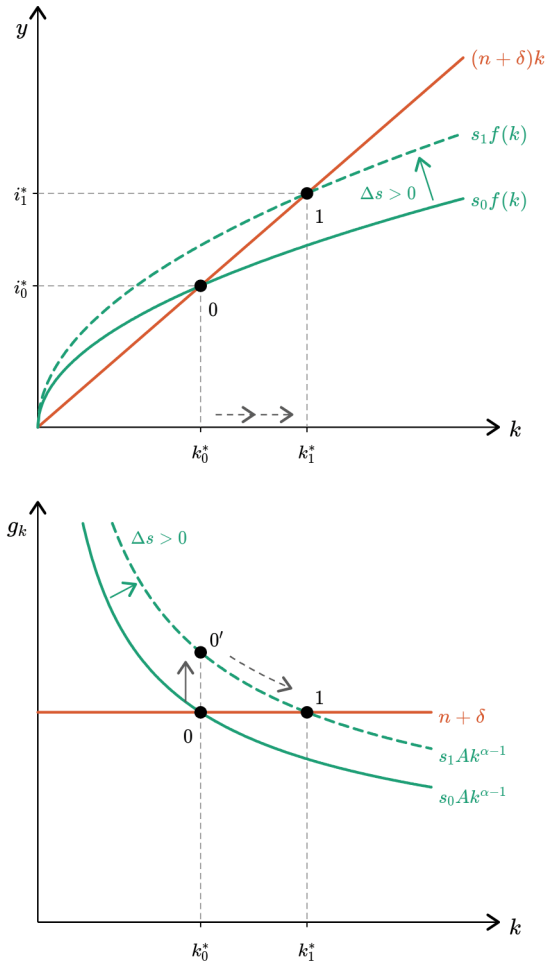



Figura A43.2: Aumento de la tasa de ahorro ($\Delta s > 0$) en el modelo de Solow-Swan. El capital estacionario pasa de $k_0^*$ a $k_1^*$ ($0 \to 1$), y la tasa de crecimiento del capital salta en $k_0^*$ ($0 \to 0'$) y se anula de nuevo en 1.

▪ *Caída de la tasa de crecimiento de la población* ($\downarrow n$): hace pivotar hacia abajo la recta de reposición, pues hay que equipar a menos trabajadores nuevos, y el capital estacionario por trabajador aumenta, con el mismo patrón.

▪ *Mejora tecnológica*: un aumento de $A$ de una vez desplaza la curva de ahorro y tiene el mismo patrón. En cambio, solo un crecimiento continuo de $A$ genera un efecto tasa permanente.

▪ *Destrucción de capital*: una pérdida de una vez del *stock* de capital (una guerra, una catástrofe) no altera ninguna curva. La economía se sitúa a la izquierda de $k^*$ y vuelve a él creciendo transitoriamente por encima de su tasa estacionaria. Así interpreta el modelo la reconstrucción de Alemania y Japón tras 1945.

**Bienestar**: la *regla de oro* (Phelps, 1961) es la tasa de ahorro que maximiza el consumo de estado estacionario. Partiendo de $c^* = f(k^*) - (n+\delta)k^*$ y maximizando respecto a $k$,

$$\left.\frac{df}{dk}\right|_{k_{\text{oro}}} = n + \delta \quad \Rightarrow \quad s_{\text{oro}} = \alpha$$

La tasa de ahorro oro coincide con la participación del capital en la renta (ahorrar toda la renta del capital y consumir toda la renta laboral), y el *stock* de capital oro es $k_{\text{oro}} = (\alpha A/(n + \delta))^{1/(1-\alpha)}$.

Gráficamente, la regla de oro se representa en el plano capital por trabajador-producto por trabajador (figura A43.3, panel superior):

1. *Función de producción*, $f(k)$: la producción por trabajador, cóncava por los rendimientos decrecientes del capital.

2. *Recta de reposición*, $(n + \delta)k$: la inversión necesaria para mantener constante el capital por trabajador.

3. *Curva de ahorro de la regla de oro*, $s_{\text{oro}}\, f(k)$: la inversión bruta por trabajador con la tasa $s_{\text{oro}} = \alpha$, que corta a la recta de reposición en $k_{\text{oro}}$.

4. *Consumo de estado estacionario*, $c^*(k) = f(k) - (n + \delta)k$: en el plano $k$-$c$ (panel inferior), recoge para cada capital estacionario la distancia vertical entre $f(k)$ y $(n+\delta)k$ del panel superior. Crece mientras $df/dk > n + \delta$ y decrece a partir de $k_{\text{oro}}$.

5. *Óptimo*: en $k_{\text{oro}}$ las pendientes de $f(k)$ y de $(n+\delta)k$ coinciden y el consumo alcanza su máximo, $c_{\text{oro}}$, en la cima de $c^*(k)$.

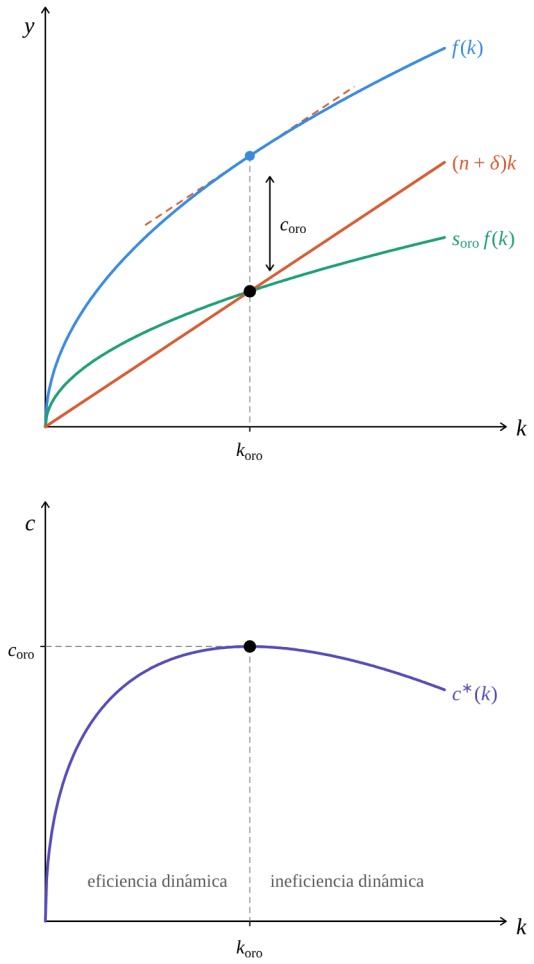



Figura A43.3: Regla de oro (Phelps, 1961). El consumo de estado estacionario, $c^*(k)$, alcanza su máximo, $c_{\text{oro}}$, en $k_{\text{oro}}$, donde la pendiente de la función de producción iguala a $n + \delta$.

Caben dos situaciones, a cada lado de $k_{\text{oro}}$ en el panel inferior:

- *Exceso de ahorro* ($k^* > k_{\text{oro}}$, $r^* < n$): reducir $s$ hasta $s_{\text{oro}}$ produce un salto inmediato del consumo hacia arriba, seguido de una convergencia al nuevo $c_{\text{oro}}$, superior al consumo inicial en todos los periodos. La economía era *dinámicamente ineficiente*, pues acumulaba capital que no le servía para consumir más. En la práctica, esto podría corregirse penalizando fiscalmente el ahorro financiero privado.

- *Insuficiencia de ahorro* ($k^* < k_{\text{oro}}$): aumentar $s$ hasta $s_{\text{oro}}$ requiere un periodo de sacrificio a corto plazo, pues el consumo cae al absorber la inversión más producción, aunque a largo plazo aumenta. A diferencia del caso anterior, ninguna trayectoria factible eleva el consumo en todos los periodos, de modo que la economía es *dinámicamente eficiente*. Ahorrar más es entonces una elección que depende de la impaciencia de los consumidores, porque una sociedad que valore mucho el presente puede preferir no sacrificar consumo corriente.

En el modelo de Solow nada garantiza que se ahorre exactamente $s_{\text{oro}}$, pues la tasa de ahorro es exógena. En el de Ramsey-Cass-Koopmans, en cambio, el ahorro óptimo excluye el exceso de

ahorro (apartado 3.2).

**Extensiones**: tres corolarios del estado estacionario completan el modelo:

▪ *Tipo de interés e ineficiencia dinámica*: como $r^* = df/dk|_{k^*} - \delta$ y en la regla de oro $df/dk|_{k_{oro}} = n + \delta$, la zona de exceso de ahorro ($k^* > k_{oro}$) es aquella en la que $r^* < n = g_Y^*$. Gráficamente, en el plano capital por trabajador-tipo de interés y crecimiento, $k$-$(r, n)$ (figura A43.4), la curva $df/dk - \delta$ es el tipo de interés real, decreciente por los rendimientos decrecientes del capital. Corta en $k_{oro}$ a la recta horizontal $n = g_Y^*$, la tasa de crecimiento de la economía, que no depende de $k$. A la izquierda de $k_{oro}$ la economía es dinámicamente eficiente ($r > n$). A la derecha es dinámicamente ineficiente ($r < n$), como en el $k^*$ de la figura, con un tipo $r^*$.

En esa zona el capital rinde menos de lo que crece la economía. Para contrastarlo, Abel, Mankiw, Summers y Zeckhauser (1989) propusieron comparar la renta bruta del capital con la inversión bruta, y en las economías avanzadas la primera supera sistemáticamente a la segunda, lo que indica eficiencia dinámica. Aun así, el debate se ha reabierto con los tipos de interés inferiores al crecimiento posteriores a 2008 (tema A39).

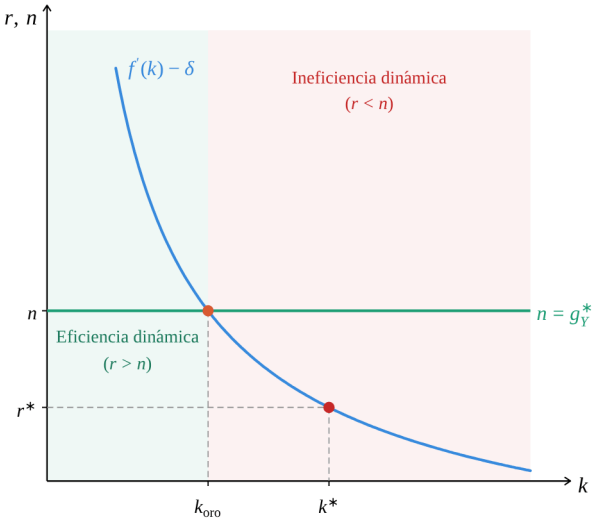



Figura A43.4: Eficiencia e ineficiencia dinámica en el modelo de Solow. A la derecha de la regla de oro ($k^* > k_{oro}$) el tipo de interés real queda por debajo del crecimiento de la economía ($r^* < n$) y reducir el ahorro elevaría el consumo en todos los periodos.

▪ *Rentabilidad del capital y flujos internacionales*: con la Cobb-Douglas, $df/dk = \alpha A k^{\alpha-1}$ y $y = Ak^\alpha$, de donde $df/dk \propto y^{(\alpha-1)/\alpha}$. Por ejemplo, si la renta per cápita de un país es diez veces la de otro y $\alpha = 1/3$, su capital per cápita es mil veces mayor y el rendimiento del capital cien veces menor. Con esas diferencias el capital debería fluir masivamente hacia los países pobres, y no lo hace (*paradoja de Lucas*, 1990). Sin riesgos ni fricciones en los mercados de capitales, la explicación exige que las diferencias de renta se deban a la tecnología o al capital humano y no al capital físico, o que la elasticidad de sustitución sea mayor que la unidad.

▪ *Elasticidad de sustitución*: la Cobb-Douglas impone una elasticidad de sustitución entre capital y trabajo $\sigma = 1$. Con una función de elasticidad de sustitución constante (CES), $Y = [\alpha K^{(\sigma-1)/\sigma} + (1-\alpha)(AL)^{(\sigma-1)/\sigma}]^{\sigma/(\sigma-1)}$, las conclusiones cambian en los extremos:

- $\sigma > 1$: la productividad marginal del capital está acotada inferiormente por una constante positiva, se incumple la condición de Inada en el infinito y el crecimiento per cápita puede sostenerse sin progreso técnico (tema A44).

- $\sigma < 1$: el estado estacionario puede no existir para tasas de ahorro bajas.

Además, con $\sigma \neq 1$ las participaciones factoriales dejan de ser constantes, lo que la evidencia reciente sobre la participación del trabajo hace relevante (tema A45).

## 2.3. Modelo con progreso técnico

El modelo anterior supone $\dot{A} = 0$ y concluye que el crecimiento per cápita se agota. Para que exista crecimiento sostenido, el modelo incorpora un **progreso técnico exógeno** que crece a la tasa constante $g_A = \dot{A}/A$ y que, como el maná, no consume recursos.

**Planteamiento**: la forma en que el progreso técnico entra en la función de producción importa. Puede ser:

▪ *Neutral en el sentido de Hicks*, $Y = AF(K, L)$: eleva la productividad de ambos factores por igual.

▪ *Aumentador del trabajo* o *neutral en el sentido de Harrod*, $Y = F(K, AL)$: equivale a disponer de más trabajadores.

▪ *Aumentador del capital* o *neutral en el sentido de Solow*, $Y = F(AK, L)$.

Con la Cobb-Douglas las tres formas son indistinguibles, pero en general no lo son. El *teorema de Uzawa* (1961) establece que solo el progreso técnico aumentador del trabajo es compatible con un estado estacionario de crecimiento equilibrado, es decir, con tasas de crecimiento constantes y participaciones factoriales constantes. Cualquier otra forma haría variar la relación capital-producto sin límite. Por ello el modelo se escribe $Y = K^\alpha (AL)^{1-\alpha}$, donde $AL$ es el *trabajo efectivo* o trabajo en unidades de eficiencia.

Las tres ecuaciones de comportamiento son las del modelo sin progreso técnico (renta, inversión y consumo, por este orden), pero se expresan por unidad de trabajo efectivo, $\tilde{x} = X/(AL)$, en lugar de por trabajador:

$$\begin{aligned} Y = K^\alpha (AL)^{1-\alpha} &\Rightarrow \tilde{y} = \tilde{k}^\alpha &&\Rightarrow g_{\tilde{y}} = \alpha g_{\tilde{k}}, \qquad \alpha \in (0,1) \\ \dot{K} = sY - \delta K &\Rightarrow \dot{\tilde{k}} = s\tilde{k}^\alpha - (g_A + n + \delta)\tilde{k} &&\Rightarrow g_{\tilde{k}} = s\tilde{k}^{\alpha-1} - (g_A + n + \delta) \\ C = (1-s)Y &\Rightarrow \tilde{c} = (1-s)\tilde{k}^\alpha &&\Rightarrow g_{\tilde{c}} = g_{\tilde{y}} \end{aligned}$$

La ley de acumulación es formalmente idéntica a la anterior, con la tasa de progreso técnico añadida a la de reposición, porque ahora hay que equipar no solo a los nuevos trabajadores sino también a las nuevas unidades de eficiencia. Las variables per cápita se recuperan multiplicando por $A$, $y = A\tilde{y}$, de modo que $g_y = g_A + g_{\tilde{y}}$ (Anexo).

**Equilibrio**: el estado estacionario, $\tilde{k}^* = [s/(g_A+n+\delta)]^{1/(1-\alpha)}$, es constante en unidades efectivas, lo que implica que las variables per cápita crecen a la tasa del progreso técnico y las agregadas a la del progreso técnico más la población:

$$g^*_{\tilde{y}} = g^*_{\tilde{k}} = 0, \qquad g^*_y = g^*_k = g^*_c = g_A, \qquad g^*_Y = g^*_K = g_A + n$$

El *crecimiento equilibrado* resultante reproduce los hechos estilizados de Kaldor, pues la relación capital-producto, el tipo de interés y las participaciones factoriales son constantes, mientras que el salario real crece a la tasa $g_A$ (tema A45).

**Estática comparativa**: un aumento de $g_A$ eleva permanentemente la tasa de crecimiento per cápita, no solo el nivel, y $g_A$ es el único parámetro del modelo con ese efecto, mientras que la acumulación de capital explica la transición, no el estado estacionario. Esa conclusión es a la vez el mayor logro del modelo, pues coincide con el residuo que la contabilidad del crecimiento atribuye a la tecnología, y su mayor limitación, pues deja sin explicar la variable que lo explica todo. Esta laguna motivará los modelos de crecimiento endógeno (tema A44).

## 2.4. Dinámica de transición y convergencia

**Velocidad de convergencia**: la economía converge gradualmente al estado estacionario, y la velocidad mide la tasa a la que la renta por unidad de trabajo efectivo, $\tilde{y}$, se aproxima a su nivel estacionario (la per cápita, a su senda de crecimiento equilibrado).

La velocidad se obtiene log-linealizando la ecuación fundamental en torno al estado estacionario. La tasa de crecimiento del capital, $g_{\tilde{k}} = s\tilde{k}^{\alpha-1} - (g_A + n + \delta)$, es decreciente en $\tilde{k}$, y su pendiente respecto a $\ln \tilde{k}$ en $\tilde{k}^*$ es $-(1-\alpha)(g_A + n + \delta)$, de modo que, para desviaciones pequeñas, $\dot{\ln \tilde{k}} \approx -\lambda(\ln \tilde{k} - \ln \tilde{k}^*)$ con

$$\lambda = (1-\alpha)(g_A + n + \delta)$$

La solución, $\ln \tilde{k}(t) - \ln \tilde{k}^* = e^{-\lambda t}(\ln \tilde{k}(0) - \ln \tilde{k}^*)$, muestra que la distancia al estado estacionario se cierra a un ritmo exponencial constante, de modo que cada año se recorre una fracción $\lambda$ de la distancia que queda, y el tiempo para recorrer la mitad es $t_{1/2} = \ln 2/\lambda$.

La velocidad es mayor cuanto menor es la participación del capital, porque con $\alpha$ pequeña los rendimientos decrecientes son más intensos. Con parametrización estándar ($\alpha \approx 1/3$, $g_A + n + \delta \approx 0{,}06$), el modelo predice $\lambda \approx 4$ % ($t_{1/2} \approx 17$ años), el doble de la observada empíricamente (en torno al 2 %). Con capital ampliado, en cambio, $\lambda$ baja a la velocidad observada.

**Convergencia entre países**: la predicción admite tres formulaciones:

- *$\beta$-convergencia absoluta*: los países pobres crecen más que los ricos. Se estima con $g_y = \beta_0 - \beta \ln y_0 + \varepsilon$ y existe si $\beta > 0$. Requiere parámetros estructurales idénticos, es decir, el mismo estado estacionario (véase la figura del tema A45, apartado 2.1).

La evidencia de 1960-1990 no la apoya para muestras grandes de países (sí la de las últimas décadas, tema A45), lo que en un principio se leyó como evidencia contraria al modelo neoclásico. En cambio, entre regiones de un mismo país sí se observa, a tasas de alrededor del 2 %, como ocurre entre los estados de EE. UU. (Barro y Sala-i-Martin, 1991), con resultados similares para las regiones europeas y las prefecturas japonesas.

- *$\beta$-convergencia condicional*: controlando por las diferencias en parámetros, $g_y = \beta_0 - \beta \ln y_0 + \gamma z + \varepsilon$, sí se obtiene una correlación parcial negativa. Basta con que un parámetro no coincida para que cada país se dirija a su propio estado estacionario (véase la figura del tema A45, apartado 2.1), y deje de estar asegurado que los pobres crezcan más. Con esos controles, Barro y Sala-i-Martin (1992a) obtienen en 98 países una tasa de alrededor del 2 % ($t_{1/2} \approx 35$ años).

- *$\sigma$-convergencia*: la dispersión de la renta per cápita entre países disminuye con el tiempo. La $\beta$-convergencia absoluta es condición necesaria pero no suficiente para la $\sigma$-convergencia.

## 2.5. Extensiones

La primera extensión añade el capital humano, y las demás llevan el modelo a la economía abierta e incorporan la tierra y los recursos naturales.

### 2.5.1. Modelo de Solow aumentado con capital humano (Mankiw, Romer y Weil, 1992)

**Origen**: Gregory Mankiw, David Romer y David Weil (1992) propusieron, ante los fallos cuantitativos del modelo canónico, una versión aumentada con capital humano que conserva la estructura neoclásica.

**Idea**: la participación del capital en la renta, en torno a un tercio, es lo que hace que el modelo de Solow prediga una convergencia demasiado rápida y unos diferenciales de renta entre países demasiado pequeños. Si el capital relevante incluye el capital humano, la participación del capital en sentido amplio se acerca a dos tercios, los rendimientos decrecientes de la acumulación son más suaves y ambas predicciones se corrigen sin renunciar al progreso técnico exógeno.

**Supuestos**:

▪ *Agentes*: las familias destinan fracciones constantes y exógenas de la renta a acumular cada tipo de capital, $s_k$ y $s_h$, y ambos capitales se deprecian a la misma tasa $\delta$.

▪ *Tecnología*: la función de producción incorpora el capital humano $H$ como tercer factor, $Y = K^\alpha H^\beta (AL)^{1-\alpha-\beta}$, con $\alpha + \beta < 1$, de modo que se mantienen los rendimientos decrecientes del capital en sentido amplio y la condición de Inada, lo que excluye el crecimiento endógeno. El resto de supuestos son los del modelo con progreso técnico.

**Planteamiento**: la renta se reparte ahora entre el consumo y la inversión en cada tipo de capital, $Y = C + I_K + I_H$, con $I_K = s_k Y$ e $I_H = s_h Y$. En unidades de trabajo efectivo ($\tilde{k} = K/(AL)$, $\tilde{h} = H/(AL)$), las ecuaciones de comportamiento de la renta, de la inversión en cada capital y del consumo son, por este orden:

$$\begin{aligned}
&Y = K^\alpha H^\beta (AL)^{1-\alpha-\beta} \;\Rightarrow\; \tilde{y} = \tilde{k}^\alpha \tilde{h}^\beta \;\Rightarrow\; g_{\tilde{y}} = \alpha g_{\tilde{k}} + \beta g_{\tilde{h}}, \quad \alpha, \beta > 0,\ \alpha + \beta < 1 \\
&\dot{K} = s_k Y - \delta K \qquad\quad \Rightarrow\; \dot{\tilde{k}} = s_k \tilde{y} - (g_A + n + \delta)\tilde{k} \\
&\dot{H} = s_h Y - \delta H \qquad\quad \Rightarrow\; \dot{\tilde{h}} = s_h \tilde{y} - (g_A + n + \delta)\tilde{h} \\
&C = (1 - s_k - s_h)Y \qquad \Rightarrow\; \tilde{c} = (1 - s_k - s_h)\tilde{y} \;\Rightarrow\; g_{\tilde{c}} = g_{\tilde{y}}
\end{aligned}$$

Cada capital tiene su propia ley de acumulación, con la forma de la del modelo canónico, y ambas se enlazan a través de la producción, de la que se financian. Dividiendo cada una por su capital se obtienen sus tasas de crecimiento, $g_{\tilde{k}} = s_k \tilde{y}/\tilde{k} - (g_A + n + \delta)$ y $g_{\tilde{h}} = s_h \tilde{y}/\tilde{h} - (g_A + n + \delta)$.

**Equilibrio**: resolviendo simultáneamente $\dot{\tilde{k}} = 0$ y $\dot{\tilde{h}} = 0$,

$$\tilde{k}^* = \left( \frac{s_k^{1-\beta} s_h^\beta}{g_A + n + \delta} \right)^{1/(1-\alpha-\beta)}, \quad \tilde{h}^* = \left( \frac{s_k^\alpha s_h^{1-\alpha}}{g_A + n + \delta} \right)^{1/(1-\alpha-\beta)}$$

y, sustituyendo ambos valores en $\tilde{y} = \tilde{k}^\alpha \tilde{h}^\beta$ y deshaciendo el cambio a unidades de trabajo efectivo, $y = A\tilde{y}$ con $A = A_0 e^{g_A t}$, la renta per cápita estacionaria en logaritmos es

$$\ln y^* = \ln A_0 + g_A t + \frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\beta}{1-\alpha-\beta}\ln s_h - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(g_A + n + \delta)$$

Las tasas de crecimiento estacionarias son las del modelo canónico, $g^*_{\tilde{y}} = 0$ en unidades efectivas y $g^*_y = g_A$ per cápita.

**Desarrollo gráfico**: el estado estacionario con dos tipos de capital se representa en el plano capital físico-capital humano por unidad de trabajo efectivo, $\tilde{k}$-$\tilde{h}$ (figura A43.5):

1. *Locus* $\dot{\tilde{k}} = 0$: combinaciones en las que el capital físico por unidad de trabajo efectivo no varía, porque su inversión, $s_k\tilde{y}$, iguala su reposición. Es creciente, porque un capital físico mayor exige más reposición, que solo una producción mayor cubre, obtenida con más capital humano. Es convexo porque, con $\alpha + \beta < 1$, cada aumento de $\tilde{k}$ exige un aumento mayor de $\tilde{h}$.

2. *Locus* $\dot{\tilde{h}} = 0$: combinaciones en las que el capital humano por unidad de trabajo efectivo no varía, porque su inversión, $s_h\tilde{y}$, iguala su reposición. Es creciente, porque un capital físico mayor eleva la producción y la inversión en capital humano, que solo se compensan con un capital humano mayor. Es cóncavo porque, con $\alpha + \beta < 1$, ese aumento de $\tilde{h}$ es cada vez menor.

3. *Equilibrio*: en el punto 0, $(\tilde{k}^*, \tilde{h}^*)$, ambos *loci* se cortan en el estado estacionario, hacia el que la economía converge desde cualquier punto.

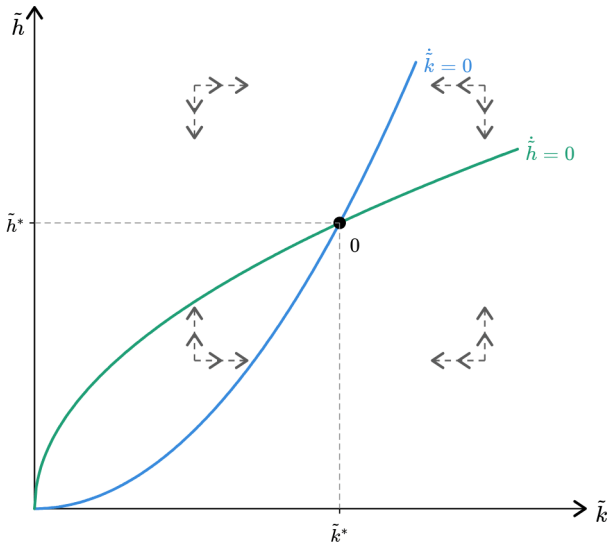



Figura A43.5: Estado estacionario con dos tipos de capital en el modelo de Solow-Swan con capital humano (Mankiw, Romer y Weil, 1992).

**Estática comparativa**: como en el modelo canónico, la alteración permanente de un parámetro tiene efecto nivel y no tasa, es decir, cambia la renta per cápita estacionaria, pero su tasa de crecimiento a largo plazo sigue siendo $g_A$.

▪ *Tasas de inversión*: un aumento de $s_k$ o de $s_h$ eleva $y^*$, con elasticidades $\alpha/(1-\alpha-\beta)$ y $\beta/(1-\alpha-\beta)$. La elasticidad respecto a $s_k$ es mayor que en el modelo canónico, $\alpha/(1-\alpha)$, porque el mayor ahorro eleva la renta y, con ella, el capital humano estacionario, aunque $s_h$ no varíe.

▪ *Crecimiento de la población*: un aumento de $n$ reduce $y^*$, con una elasticidad respecto a $g_A+n+\delta$ de $-(\alpha+\beta)/(1-\alpha-\beta)$, mayor en valor absoluto que en el modelo canónico.

**Dinámica**: log-linealizando en torno al estado estacionario, la tasa de crecimiento entre dos fechas depende de la renta inicial y de los determinantes del estado estacionario,

$$\ln \tilde{y}_t - \ln \tilde{y}_0 = (1-e^{-\lambda t})\left[\ln \tilde{y}^* - \ln \tilde{y}_0\right], \qquad \lambda = (1-\alpha-\beta)(g_A+n+\delta)$$

de modo que la velocidad de convergencia se reduce con la participación conjunta del capital físico y humano. La literatura empírica estima esta ecuación como regresión de convergencia condicional (tema A45).

**Conclusiones**:

▪ *Aspectos positivos*: Mankiw, Romer y Weil (MRW) obtienen $\alpha \approx \beta \approx 1/3$, de modo que la participación del capital en sentido amplio es de unos dos tercios. Ello supone tres mejoras. La velocidad de convergencia cae, aproximadamente, del 4 % al 2 %, coherente con la evidencia, y el $R^2$ de las regresiones del nivel de renta por persona en edad de trabajar sube de unos 0,6 (modelo canónico, 98 países) a unos 0,8. Además, los diferenciales de renta predichos se aproximan más a los observados.

▪ *Limitaciones*: MRW aproximan la tasa de inversión en capital humano, $s_h$ (un flujo, no el *stock* de capital humano), mediante el producto de la tasa de escolarización secundaria por la proporción de la población en edad de trabajar que está en edad escolar. Esta medida ha sido ampliamente criticada por imprecisa, ya que no captura la calidad de la educación, la experiencia laboral ni el aprendizaje en el puesto de trabajo.

### 2.5.2. Otras extensiones

**Economía abierta** (Barro, Mankiw y Sala-i-Martin, 1995): en una economía pequeña y abierta con movilidad perfecta del capital, el tipo de interés mundial fija de inmediato el capital por trabajador, $df/dk = r^w + \delta$, y la convergencia sería instantánea, lo que la evidencia desmiente. Si, por el contrario, el capital humano no puede usarse como colateral y solo el capital físico se financia en el exterior, la economía salta de inmediato al capital físico deseado, pero acumula capital humano con ahorro interno. La velocidad de convergencia es entonces mayor que en la economía cerrada, pero finita. Con la participación del capital físico en torno a un tercio, el efecto cuantitativo es pequeño y la velocidad predicha sigue próxima al 2 %.

**Tierra y recursos naturales**: con tierra ($T$, de oferta fija) y recursos no renovables ($R$, que se agotan a la tasa $b$), $Y = AK^{\alpha}T^{\gamma}R^{\eta}L^{1-\alpha-\gamma-\eta}$, la tasa de crecimiento per cápita estacionaria es $g_y^* = [g_A-(\gamma+\eta)n-\eta b]/(1-\alpha)$. La expresión se obtiene tomando tasas de crecimiento en la función de producción a lo largo de una senda de crecimiento equilibrado, con $g_T = 0$, $g_R = -b$ y $g_L = n$, y restando $n$. El progreso técnico es aquí neutral en el sentido de Hicks, de modo que $g_A$ mide el crecimiento de la productividad total de los factores (PTF) y no el del progreso aumentador del trabajo.

La tierra fija y el agotamiento de recursos actúan, pues, como un freno proporcional al crecimiento de la población y a la tasa de extracción, y el crecimiento per cápita solo es positivo

si el progreso técnico los compensa, $g_A > (\gamma + \eta)n + \eta b$. Así responde el enfoque neoclásico a la predicción maltusiana y al Informe Meadows (1972). Los límites físicos son reales, pero la experiencia de los dos últimos siglos muestra que el progreso técnico los ha superado con holgura. La economía de los recursos agotables, de la regla de Hotelling al modelo de Solow ampliado con recursos naturales, se examina en el tema B31 §2.1.

## 2.6. Conclusiones

**Aspectos positivos**: toda la literatura posterior toma el modelo como referencia. Al admitir que capital y trabajo se sustituyen, el modelo prueba que el estado estacionario es estable y deja atrás el filo de la navaja de Harrod-Domar. Muestra, además, que solo el avance tecnológico hace que la renta por habitante siga aumentando a largo plazo, por mucho capital que se acumule.

**Limitaciones**: el modelo tiene cinco:

- *Progreso tecnológico exógeno*: el teorema de Euler implica que, con rendimientos constantes y competencia perfecta, la innovación no tiene con qué financiarse. El propio Solow (1957) atribuyó al "residuo de Solow", es decir, al progreso tecnológico no explicado, el 87,5 % del crecimiento de la producción por hora trabajada en EE. UU. entre 1909 y 1949.

- *Calibración*: con $\alpha \approx 1/3$, la velocidad de convergencia y los diferenciales de renta son irrealistas.

- *Modelo unisectorial*: no captura la transformación estructural (ley de Engel, industrialización en U invertida, ley de Wagner).

- *Ahorro exógeno*: la tasa de ahorro es un parámetro y no el resultado de una decisión de las familias, carencia que corrige el modelo de Ramsey-Cass-Koopmans (sección 3).

- *Rentabilidad del capital*: los diferenciales entre países predichos son muy superiores a los observados (paradoja de Lucas).

# 3. Modelo de Ramsey-Cass-Koopmans (1928, 1965)

**Origen**: Frank Ramsey (1928) planteó el problema del ahorro óptimo. David Cass (1965) y Tjalling Koopmans (1965) lo formalizaron como modelo de crecimiento con optimización intertemporal, el modelo de Ramsey-Cass-Koopmans (RCK).

**Idea**: el modelo endogeniza la tasa de ahorro a partir de la decisión intertemporal de las familias. Su principal resultado es una tasa de ahorro óptima que puede subir, bajar o mantenerse a lo largo de la transición, sin que cambien las conclusiones cualitativas de Solow.

## 3.1. Supuestos

**Entorno**: el modelo comparte los supuestos neoclásicos de Solow (competencia perfecta, tiempo continuo, horizonte infinito).

**Agentes**:

▪ *Familias*: en ellas está el principal elemento diferenciador respecto al modelo de Solow-Swan, pues la tasa de ahorro es una variable endógena. En cada momento del tiempo los consumidores-productores deciden la fracción de su renta que destinan a ahorrar. Los hogares son idénticos y su tamaño crece a la tasa de la población, $n$, con $L(t) = e^{nt}$. Operan bajo altruismo intergeneracional puro (dinastías), lo que justifica el horizonte infinito. Alternativamente, este puede interpretarse como el de agentes con una probabilidad de muerte constante, de tipo Poisson. Los hogares poseen activos $a$ (positivos si prestan, negativos si se endeudan) que generan un tipo de interés real $r$, y fuerza de trabajo que alquilan a un salario real $w$.

▪ *Preferencias*: utilidad instantánea isoelástica (CRRA), $u(c) = (c^{1-\theta} - 1)/(1 - \theta)$, con $\theta > 0$ el coeficiente de aversión relativa al riesgo, constante, y $1/\theta$ la elasticidad de sustitución intertemporal (ESI). Si $\theta = 1$, la utilidad es logarítmica. Si $\theta$ es bajo, los hogares toleran más variabilidad en el consumo, mientras que si $\theta$ es alto desean alisar fuertemente. La tasa de descuento $\rho > 0$ cumple $\rho > n > 0$, lo que garantiza que la utilidad no diverge.

▪ *Empresas*: alquilan $K$ a su coste de uso, $r + \delta$, y contratan $L$ al salario $w$.

**Tecnología**: la función de producción es la Cobb-Douglas del modelo de Solow-Swan (apartado 2.1).

## 3.2. Modelo

**Hogares**: eligen en cada momento cuánto consumen y, con ello, qué fracción de su renta ahorran.

▪ *Planteamiento*: maximizan la utilidad descontada en un horizonte infinito, sujetos a la restricción presupuestaria dinámica, a sus activos iniciales y a la condición de no-Ponzi, que impide acabar endeudados en el límite:

$$\max_{\{c\}} \int_0^\infty e^{-(\rho-n)t} \frac{c^{1-\theta} - 1}{1-\theta} dt, \qquad \theta > 0$$

$$\text{s.a.:} \begin{cases} c + \dot{a} = w + (r - n)a \\ a_0 \text{ dado} \\ \lim_{t\to\infty} \lambda(t)\, a(t) \geq 0 \end{cases}$$

donde $c$ y $a$ son el consumo y los activos per cápita[2], $a_0$ los activos iniciales, dados, y el factor $e^{-(\rho-n)t}$ descuenta la utilidad a la tasa $\rho$ y la pondera por el tamaño del hogar, $e^{nt}$. El hamiltoniano en valor presente u origen es $\mathcal{H} = e^{-(\rho-n)t}u(c) + \lambda[(r-n)a + w - c]$, donde $\lambda$, el multiplicador dinámico, actúa como precio sombra de los activos y expresa cuánto vale en $t = 0$ para el consumidor disponer de una unidad más de ellos. La condición de no-Ponzi se escribe con este multiplicador[3].

- *CPO*: por el principio del máximo,

$$\frac{\partial \mathcal{H}}{\partial c} = 0 \quad \Rightarrow \quad e^{-(\rho-n)t}c^{-\theta} = \lambda \tag{1}$$

$$\frac{\partial \mathcal{H}}{\partial a} = -\dot{\lambda} \quad \Rightarrow \quad (r-n)\lambda = -\dot{\lambda} \tag{2}$$

$$\lim_{t\to\infty} \lambda(t)\, a(t) = 0 \tag{3}$$

La condición (3), de transversalidad[4], excluye conservar activos sin consumirlos nunca, de manera que en el óptimo la condición de no-Ponzi se cumple con igualdad.

Tomando logaritmos en (1), derivando respecto al tiempo y sustituyendo (2), se obtiene la *ecuación de Keynes-Ramsey*:

$$g_c = \frac{\dot{c}}{c} = \frac{1}{\theta}(r - \rho) \tag{KR}$$

Según (KR), el consumo per cápita crece cuando $r > \rho$, es decir, cuando la rentabilidad del ahorro supera a la impaciencia y los agentes ahorran hoy para consumir más mañana, y el consumo decrece en el caso contrario, $r < \rho$.

En ambos casos, la variación del consumo es más pronunciada cuanto menor sea $\theta$, es decir, cuanto mayor sea la elasticidad de sustitución intertemporal, $1/\theta$, lo que implica también mayor volatilidad de la senda de consumo ante *shocks*.

- *CSO*: el hamiltoniano es estrictamente cóncavo en $c$ y lineal en $a$, lo que hace que (1), (2) y (3) sean condiciones necesarias y suficientes.

[2] En tiempo continuo se omite el argumento temporal: $c \equiv c(t)$, $a \equiv a(t)$, $w \equiv w(t)$, $r \equiv r(t)$ y $\lambda \equiv \lambda(t)$; se escribe solo en las condiciones inicial y terminal.

[3] Por la CPO (2), $\lambda(t) = \lambda(0)\, e^{-[R(t)-nt]}$, donde $R(t) \equiv \int_0^t r(v)\, dv$ es el tipo de interés acumulado hasta $t$, y como $\lambda(0) > 0$, la condición equivale a $\lim_{t\to\infty} a(t)\, e^{-[R(t)-nt]} \geq 0$, es decir, que el valor presente de los activos no sea negativo en el límite.

[4] La condición de transversalidad es la holgura complementaria de la no-Ponzi (tema A08, apartado 1.4.1, condiciones (4) a (6)) llevada al límite. Con horizonte finito $T$, la restricción terminal sería $a(T) \geq 0$, con multiplicador $\lambda(T) \geq 0$, y la holgura complementaria exigiría $\lambda(T)\, a(T) = 0$, es decir, que no se dejen activos o que su valor sombra sea nulo. Con horizonte infinito, la misma condición se escribe en el límite y, como $\lambda(t) > 0$, obliga a que la no-Ponzi se cumpla con igualdad (Barro y Sala-i-Martin, 2004, apéndice matemático, A.3).

**Empresas**: maximizan el beneficio en mercados competitivos.

▪ *Planteamiento*: alquilan el capital a su coste de uso y contratan el trabajo al salario, que toman como dados, $\max_{K,\,L}\ AK^{\alpha}L^{1-\alpha} - (r+\delta)K - wL$, con la tecnología Cobb-Douglas del modelo de Solow-Swan (apartado 2.1).

▪ *CPO*: contratan cada factor hasta que su productividad marginal iguala a su coste real, $r = df/dk - \delta = \alpha A k^{\alpha-1} - \delta$ y $w = f(k) - k\,df/dk = (1-\alpha)Ak^{\alpha}$. Con rendimientos constantes a escala, los pagos a los factores agotan el producto y el beneficio es nulo.

▪ *CSO*: la tecnología Cobb-Douglas es cóncava en $(K, L)$ y estrictamente cóncava en cada factor, luego las CPO caracterizan el máximo.

**Equilibrio**: en el mercado de capitales la riqueza de las familias iguala al capital, $a = k$, porque en una economía cerrada sin sector público el único activo es el capital físico y la deuda neta entre familias es cero. La ley de acumulación del capital es, con ahorro endógeno, la restricción de recursos de la economía, $c + \dot{k} + (n+\delta)k = f(k)$. A diferencia de Solow, $c$ no es una fracción fija de $f(k)$ sino una variable de control. El sistema en $(k, c)$ formado por (KR) y esta ecuación describe el modelo.

En el estado estacionario, imponiendo $\dot{c} = 0$ en (KR), el tipo de interés iguala la tasa de descuento, $r = \rho$, y la productividad marginal del capital cubre la impaciencia y la depreciación, $df/dk|_{k^*} = \rho + \delta$, de donde

$$k^* = \left(\frac{\alpha A}{\rho + \delta}\right)^{1/(1-\alpha)}$$

El *stock* de capital estacionario es inferior al de la regla de oro, $k^* < k_{\text{oro}}$, siempre que $\rho > n$, porque la regla de oro impone $df/dk = n + \delta$ mientras que RCK impone $df/dk = \rho + \delta$, de modo que la productividad marginal exigida es mayor y, por rendimientos decrecientes, el capital menor. Intuitivamente, los agentes impacientes acumulan menos capital del que prescribe la regla de oro porque los sacrificios a corto plazo para alcanzarla son demasiado grandes. Sin progreso técnico, $g_c^* = g_k^* = 0$.

En cuanto a la tasa de ahorro, en el estado estacionario el ahorro repone exactamente el capital, $s^* = (n+\delta)k^*/f(k^*)$, y sustituyendo $df/dk|_{k^*} = \alpha A k^{*\alpha-1} = \rho + \delta$:

$$s^* = \alpha\,\frac{n+\delta}{\rho+\delta}$$

La tasa de ahorro estacionaria es inferior a la de la regla de oro, $s^* < s_{\text{oro}} = \alpha$, precisamente porque $\rho > n$, y decrece con la impaciencia.

Fuera del estado estacionario, en cambio, la tasa de ahorro no es constante. Con la Cobb-Douglas se demuestra (Barro y Sala-i-Martin, 2004) que $s(t)$ es monótona a lo largo de la transición y que su sentido depende de la comparación entre $s^*$ y $1/\theta$.

La intuición descansa en los efectos sustitución y renta. En una economía pobre el capital es escaso y el tipo de interés alto, lo que por efecto sustitución incentiva el ahorro, mientras que por efecto renta los agentes, que se saben más ricos en el futuro, desean adelantar consumo y ahorrar menos. Cuanto mayor es la elasticidad de sustitución intertemporal, $1/\theta$, más pesa el primero. Caben tres casos:

- $s^* < 1/\theta$ (elasticidad de sustitución intertemporal alta): los agentes de una economía pobre prefieren consumir poco y acumular deprisa, de modo que la tasa de ahorro decrece hacia $s^*$.

- $s^* > 1/\theta$ (fuerte deseo de alisar): consumen relativamente más al principio y la tasa de ahorro crece.

- $s^* = 1/\theta$: la tasa de ahorro es constante y el modelo reproduce exactamente a Solow.

Endogenizar el ahorro no altera las conclusiones cualitativas del modelo neoclásico, y en un caso particular ni siquiera las cuantitativas.

**Desarrollo gráfico**: la dinámica del modelo se representa en el plano capital por trabajador-consumo por trabajador, $k$-$c$ (figura A43.6):

1. *Locus* $\dot{c} = 0$: línea vertical en el $k^*$ que cumple $df/dk|_{k^*} = \rho + \delta$. A su izquierda el capital es escaso, su productividad marginal supera a $\rho + \delta$, el tipo de interés supera a la impaciencia y el consumo crece, mientras que a su derecha ocurre lo contrario.

2. *Locus* $\dot{k} = 0$: curva cóncava $c = f(k) - (n + \delta)k$, que alcanza su máximo en $k_{\text{oro}}$ y corta el eje horizontal, a la derecha del área representada, en el capital máximo sostenible, $\bar{k}$, donde $f(\bar{k}) = (n + \delta)\bar{k}$. Por encima de la curva el consumo es demasiado alto y el capital disminuye, y por debajo aumenta.

3. *Senda estable*: la trayectoria, creciente, que lleva al estado estacionario.

4. *Equilibrio*: hay tres. El origen es inestable y el punto $(\bar{k}, 0)$ es estable pero indeseable, pues todo se invierte y nada se consume. El tercero, $(k^*, c^*)$, el punto 0, es un *punto de silla*.

Existe una única trayectoria estable (*saddle path*) que converge al punto de silla. Las trayectorias por encima agotan el capital en un tiempo finito (sobreconsumo), tras lo cual el consumo tendría que caer bruscamente a cero, lo que incumple la ecuación de Keynes-Ramsey y la restricción de recursos. Las de abajo llevan a $k \to \bar{k}$ y $c \to 0$ (sobreahorro) e incumplen la condición de transversalidad. Así, $c$ es una *variable de salto*, que se sitúa de inmediato sobre la senda estable, y $k$ una *variable predeterminada*.

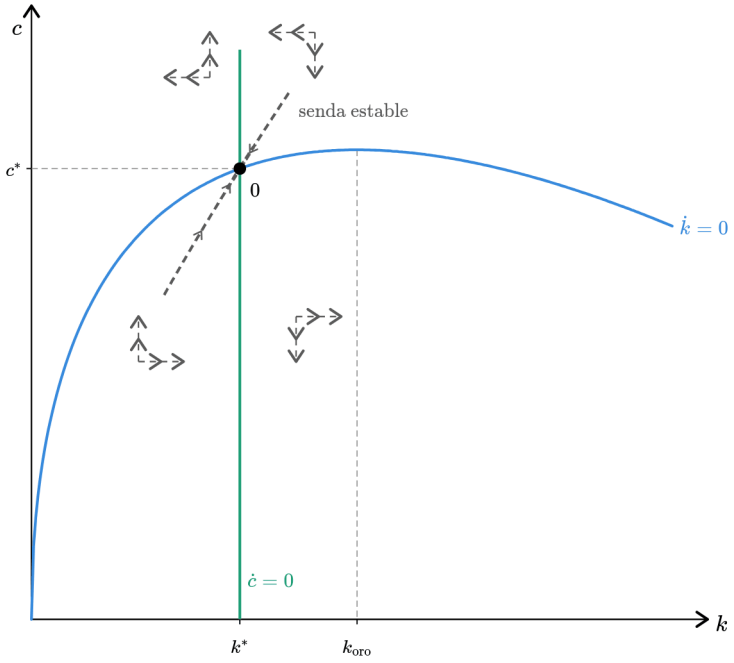



Figura A43.6: Diagrama de fases del modelo de Ramsey-Cass-Koopmans y senda estable hacia el estado estacionario.

La economía puede converger desde la región de desacumulación ($k > k^*$, $c > c^*$) o desde la de crecimiento ($k < k^*$, $c < c^*$). En esta, el capital crece a tasa decreciente, y el tipo de interés baja gradualmente hasta $\rho$ y el crecimiento del consumo, hasta cero. La forma de la senda estable depende de $\theta$:

- $\theta$ *elevado* (fuerte deseo de alisar): la senda estable discurre muy cerca del *locus* $\dot{k} = 0$, porque los hogares prefieren un consumo alto desde el principio aunque $k$ apenas aumente.

- $\theta$ *reducido*: la senda estable es más empinada, porque los hogares toleran un consumo variable y lo recortan al principio para acumular deprisa.

Por ello la velocidad de convergencia es tanto mayor cuanto menor es $\theta$[5]. Con $\alpha \approx 1/3$ y valores plausibles del resto de parámetros se obtienen velocidades muy superiores al 2 % observado, y, como en Solow, solo un concepto amplio de capital ($\alpha \approx 2/3$, como en el apartado 2.5.1) reconcilia el modelo con la evidencia.

**Estática comparativa**: se analizan tres perturbaciones no anticipadas:

- *Caída de la tasa de descuento* ($\downarrow \rho$): es el *shock* más relevante en RCK, lo más parecido a un aumento de $s$ en Solow. Gráficamente, en el plano $k$-$c$ (figura A43.7), solo se desplaza el *locus* $\dot{c} = 0$, porque el $\dot{k} = 0$ depende de la tecnología y no de las preferencias. El *locus* $\dot{c} = 0$ pasa a la derecha, de $(\dot{c} = 0)_0$ a $(\dot{c} = 0)_1$, y el capital estacionario, de $k_0^*$ a $k_1^*$.

[5] Log-linealizando el sistema en torno al estado estacionario, la velocidad de convergencia es el valor absoluto de la raíz característica negativa de su matriz jacobiana, $\frac{1}{2}\left[\sqrt{(\rho-n)^2 + 4\,\frac{1-\alpha}{\theta}(\rho+\delta)\left(\frac{\rho+\delta}{\alpha} - (n+\delta)\right)} - (\rho - n)\right]$ (Barro y Sala-i-Martin, 2004), decreciente en $\theta$ y en $\alpha$. Con $\theta$ bajo supera a la de Solow para la misma parametrización.

En el momento del cambio el consumo salta hacia abajo, de $c_0^*$ a $c_{0'}$ (del punto 0 al $0'$), pues los agentes previsores se sitúan sobre la nueva senda estable ahorrando más. A partir de ahí, $k$ y $c$ crecen hasta el punto 1, con valores estacionarios superiores a los originales. Por tanto, a corto plazo el consumo cae y a largo plazo capital y consumo aumentan, el patrón de Solow con $k < k_{\text{oro}}$.

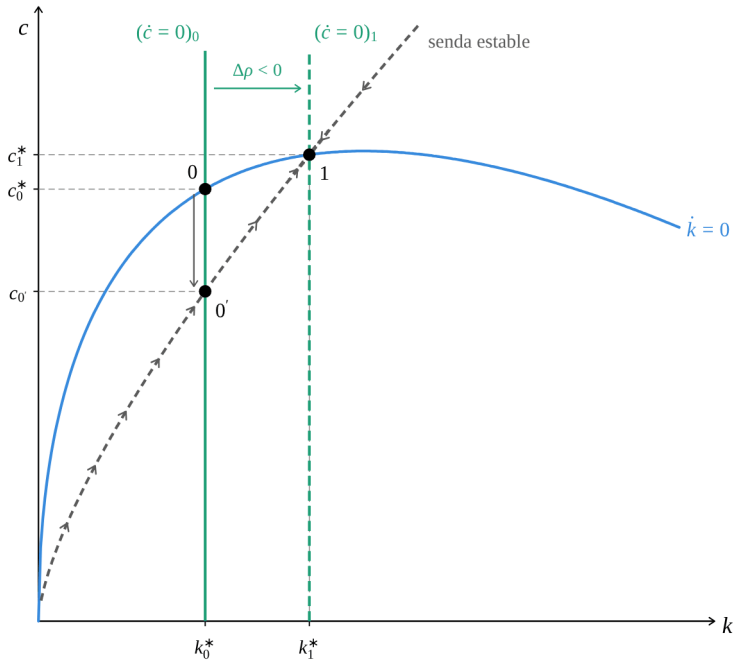



Figura A43.7: Caída no anticipada de la tasa de descuento en el modelo de Ramsey-Cass-Koopmans. El consumo salta a la nueva senda estable ($0 \to 0'$) y la economía converge al nuevo estado estacionario (1).

▪ *Mejora tecnológica permanente* ($\uparrow A$): afecta a ambos *loci*. El *locus* $\dot{c}=0$ se desplaza a la derecha, pues $df/dk$ aumenta para todo $k$, y el $\dot{k}=0$, hacia arriba, pues para cada $k$ hay más producción y más consumo compatible con un capital constante. El consumo salta de inmediato sobre el nuevo *saddle path*, hacia arriba o hacia abajo según la magnitud relativa de los desplazamientos, y a largo plazo $k^*$ y $c^*$ son ambos mayores.

▪ *Mejora tecnológica transitoria*: si el aumento de $A$ se revierte en $t_1$, los agentes lo anticipan. Por ello, en $t_0$ el consumo salta sobre una trayectoria intermedia, que no es el *saddle path* del estado transitorio ni el del permanente, y que garantiza que en $t_1$ la economía se encuentre exactamente sobre el *saddle path* original. Solow no puede generar este resultado, característico de RCK, en el que los agentes incorporan la naturaleza temporal del *shock* desde el primer momento.

**Bienestar**: dado que se cumplen las condiciones del primer teorema fundamental de la economía del bienestar (competencia perfecta, ausencia de externalidades, mercados completos), el equilibrio descentralizado es Pareto-eficiente y coincide con la solución del planificador benevolente. Esa solución impone $g_c = (1/\theta)[df/dk - \delta - \rho]$ y en el equilibrio descentralizado $r = df/dk - \delta$, de modo que las ecuaciones de Euler coinciden y la asignación es la misma. A diferencia de Solow, donde nada garantiza una tasa de ahorro eficiente, en RCK el mercado genera endógenamente

la tasa de ahorro óptima, que corresponde a la *regla de oro modificada*, $df/dk|_{k^*} = \rho + \delta$, no a la regla de oro. En consecuencia, no puede darse la ineficiencia dinámica del exceso de ahorro.

Con impuestos distorsionantes, externalidades o competencia imperfecta, en cambio, la equivalencia se rompe y el equilibrio descentralizado deja de ser eficiente, lo que deja margen a la política económica en los modelos de crecimiento endógeno.

**Extensiones**:

- *Horizonte finito*: Dorfman, Samuelson y Solow (1958) formularon el *teorema de la autopista* para el modelo multisectorial de von Neumann, y Cass (1966) lo demostró para este modelo agregado. Según el teorema, cuando el final del horizonte queda muy lejos, la senda óptima se mantiene casi todo el tiempo en las proximidades del estado estacionario y solo al acercarse el final se desvía hacia el capital terminal. Si ese capital es nulo, la senda arranca por encima del *saddle path*, porque el capital ha de agotarse en el último periodo.

- *Progreso técnico*: con progreso técnico aumentador del trabajo a la tasa $g_A$, el modelo se escribe en unidades de trabajo efectivo y la condición de transversalidad exige $\rho > n + (1-\theta)g_A$. La ecuación de Keynes-Ramsey pasa a ser $g_{\tilde{c}} = (1/\theta)[df/d\tilde{k} - \delta - \rho - \theta g_A]$, y el capital estacionario cumple $df/d\tilde{k}|_{\tilde{k}^*} = \rho + \delta + \theta g_A$. El crecimiento eleva así el tipo de interés exigido, porque los agentes, que saben que serán más ricos, descuentan más el futuro. En el estado estacionario el consumo per cápita crece a la tasa $g_A$, como en Solow, y la tasa de ahorro es $s^* = \alpha(g_A + n + \delta)/(\rho + \delta + \theta g_A)$.

- *Sector público*: con gasto público $g$ por trabajador financiado con impuestos de suma fija, la restricción de recursos es $c + g + \dot{k} + (n+\delta)k = f(k)$. El *locus* $\dot{k} = 0$ se desplaza hacia abajo en $g$ y el consumo privado es expulsado uno a uno a largo plazo, sin efecto sobre $k^*$. Por la equivalencia ricardiana, el resultado es el mismo si el gasto se financia con deuda (tema A38).

Con un impuesto proporcional sobre la renta del capital a la tasa $\tau$, en cambio, el rendimiento neto es $(1-\tau)[df/dk - \delta]$ y el capital estacionario cae, $df/dk|_{k^*} = \delta + \rho/(1-\tau)$. El impuesto tiene efecto nivel, no tasa, y el consumo estacionario disminuye.

- *Oferta de trabajo endógena*: si la utilidad incluye el ocio, $u(c, 1-l)$, se añade una condición intratemporal, $-u_l/u_c = w$, y el estado estacionario determina también las horas trabajadas. El modelo así ampliado es el núcleo de los modelos del ciclo real (tema A42).

- *Política económica*: como en Solow, las políticas que alteran los parámetros, como la fiscalidad del ahorro (que actúa como $\rho$) o las que afectan a la población, tienen efectos nivel y no tasa sobre las variables per cápita. La novedad es que, sin fallos de mercado, no hay motivo para intervenir.

## 3.3. Conclusiones

**Aspectos positivos**: el modelo dota de microfundamentos al de Solow, prueba la eficiencia del equilibrio descentralizado e incorpora las preferencias intertemporales (impaciencia y suavización del consumo) que determinan cuánto se ahorra y con qué rapidez se converge.

**Limitaciones**: comparte con Solow la exogeneidad del progreso tecnológico, y la mayor complejidad analítica no altera la conclusión de que sin tecnología exógena no hay crecimiento per cápita sostenido. También predice convergencia condicional con una velocidad similar a la de Solow. El agente representativo con horizonte infinito excluye, además, la heterogeneidad y los efectos intergeneracionales que estudian los modelos de generaciones solapadas (tema A29).

## Conclusiones

Los modelos de crecimiento exógeno comparten la conclusión de que la acumulación de capital determina el nivel de la renta per cápita, pero su crecimiento sostenido solo puede proceder de un progreso técnico que los modelos no explican. Harrod y Domar plantean si el crecimiento equilibrado es estable y, con coeficientes fijos, concluyen que no lo es, pues la economía discurre por el filo de la navaja. Solow y Swan resuelven esa inestabilidad con la sustituibilidad entre factores. Con ello, el estado estacionario es estable, aunque la velocidad de convergencia que predicen duplica la observada salvo que el capital se amplíe con el humano (Mankiw, Romer y Weil). Ramsey, Cass y Koopmans endogenizan el ahorro y muestran que el equilibrio descentralizado es eficiente, sin alterar el papel exógeno de la tecnología. La predicción empírica central es la convergencia condicional, en torno al 2 % anual, según la cual cada país converge a su estado estacionario. Que el residuo de Solow absorba el 87,5 % del crecimiento de la productividad del trabajo resume el límite de estos modelos.

El crecimiento exógeno abre un bloque que sigue con los modelos que endogenizan la tecnología (tema A44) y con la contrastación empírica, donde se mide el residuo y se estiman las convergencias $\beta$ y $\sigma$ (tema A45). El modelo de Ramsey-Cass-Koopmans aplica al crecimiento el consumidor de horizonte infinito (tema A29) y, sometido a perturbaciones tecnológicas, es también la base de la teoría del ciclo real (tema A42), de modo que crecimiento y ciclo comparten un mismo modelo. Por último, la ampliación del modelo de Solow con recursos naturales plantea si el crecimiento es compatible con un recurso agotable (tema B31).

La convergencia absoluta, ausente en las muestras de 1960-1990, tiende a aparecer desde 1990 y es clara desde 2000, a la vez que las instituciones y las políticas de los países pobres se aproximan a las de los ricos (Kremer, Willis y You, 2022). Ese cambio sugiere que la falta de convergencia anterior fue un rasgo del periodo. También se discute la tasa de descuento social, que aplica la ecuación de Keynes-Ramsey al cambio climático. El Informe Stern (2006) fija por razones éticas una tasa de preferencia temporal casi nula y pide una acción inmediata, mientras que Nordhaus (2007) calibra los parámetros con los rendimientos de mercado y obtiene una senda de mitigación gradual.

## Anexo. De los niveles a las tasas de crecimiento

Los modelos de crecimiento escriben cada ecuación de comportamiento en niveles, en términos per cápita o por unidad de trabajo efectivo y en tasas de crecimiento. El paso de una forma a otra descansa en una sola herramienta. La tasa de crecimiento de una variable es la derivada respecto al tiempo de su logaritmo, porque, por la regla de la cadena, $d\ln x/dt = (1/x)(dx/dt)$:

$$g_x \equiv \frac{\dot{x}}{x} = \frac{d\ln x}{dt}, \qquad \dot{x} \equiv \frac{dx}{dt}$$

Como el logaritmo convierte los productos en sumas y las potencias en productos, tomar logaritmos y derivar respecto al tiempo da tres reglas:

1. *Producto*: si $z = xy$, entonces $\ln z = \ln x + \ln y$ y $g_z = g_x + g_y$.

2. *Cociente*: si $z = x/y$, entonces $\ln z = \ln x - \ln y$ y $g_z = g_x - g_y$.

3. *Potencia*: si $z = x^a$, entonces $\ln z = a \ln x$ y $g_z = a\, g_x$.

Una suma no sigue ninguna de ellas. Si $z = x + y$, su tasa de crecimiento es la media de las tasas de los sumandos ponderada por su peso, $g_z = \frac{x}{x+y}\, g_x + \frac{y}{x+y}\, g_y$, como en la producción del modelo Sobelow (tema A44).

El modelo de Solow-Swan (apartado 2.2) aplica las tres reglas. En la producción, $Y = AK^{\alpha}L^{1-\alpha}$ con $A$ constante, las reglas del producto y de la potencia dan $g_Y = \alpha g_K + (1-\alpha)n$. En términos per cápita, la regla del cociente da $g_y = g_Y - n = \alpha(g_K - n) = \alpha g_k$.

La acumulación del capital exige un paso más. El capital per cápita es un cociente, $k = K/L$, de modo que $g_k = g_K - n$. Como $\dot{K} = sY - \delta K$, la tasa de crecimiento del capital es $g_K = sY/K - \delta$, y, dividiendo por $L$ el numerador y el denominador, $Y/K = y/k$. Por tanto:

$$g_k = \frac{sy}{k} - (n+\delta) \quad \Rightarrow \quad \dot{k} = g_k\, k = sAk^{\alpha} - (n+\delta)k$$

El término $nk$ aparece porque, aunque $K$ no varíe, el capital per cápita cae al ritmo al que crece la población.

Con progreso técnico (apartado 2.3), el capital por unidad de trabajo efectivo es $\tilde{k} = K/(AL)$, y las reglas del cociente y del producto dan $g_{\tilde{k}} = g_K - g_A - n$. El mismo razonamiento lleva a $\dot{\tilde{k}} = s\tilde{k}^{\alpha} - (g_A + n + \delta)\tilde{k}$. Para volver a las variables per cápita, $y = A\tilde{y}$ implica $g_y = g_A + g_{\tilde{y}}$, de modo que, en el estado estacionario, con $g^*_{\tilde{y}} = 0$, la renta per cápita crece a la tasa $g_A$.

# A44. Crecimiento económico endógeno

Crecimiento económico (II). Modelos de crecimiento endógeno: rendimientos crecientes, capital humano e innovación tecnológica.

## Introducción

**Introducción común.** Los temas A43 a A45 comparten una introducción, que se desarrolla al comienzo del tema **A43**. Define el crecimiento económico como el aumento sostenido de la producción por habitante y resume la evolución de su teoría, del modelo de Solow-Swan al crecimiento endógeno y la contrastación empírica.

El modelo de Solow-Swan (1956) ofrece una explicación elegante del crecimiento a largo plazo, pero deja fuera del análisis precisamente aquello que pretende explicar. De hecho, el propio Solow (1957), en su célebre ejercicio de contabilidad del crecimiento, cuantificó el problema. Del crecimiento del producto por hora trabajada en EE. UU. entre 1909 y 1949, la mayor dotación de capital físico daba cuenta de apenas un 12,5 %. Casi nueve décimas partes (el 87,5 %, conocido como *residuo de Solow*) correspondían a un "progreso técnico" que el modelo trata como maná caído del cielo, una variable exógena que crece a una tasa dada, sin coste y sin explicación económica alguna.

Esta conclusión resulta doblemente insatisfactoria. En primer lugar, porque convierte la teoría del crecimiento en una teoría de todo aquello que no es crecimiento. El modelo explica bien la transición y la convergencia, pero el motor último del crecimiento sostenido de la renta per cápita queda fuera del modelo. En segundo lugar, porque el progreso técnico no cae del cielo, sino que es el resultado de decisiones deliberadas de inversión en educación, investigación e innovación tomadas por agentes económicos que responden a incentivos.

Los **modelos de crecimiento endógeno**, que se desarrollan desde mediados de los años ochenta, responden a esa doble carencia. Su propósito es que la renta per cápita crezca de forma sostenida gracias al funcionamiento del propio modelo y no a un progreso tecnológico supuesto desde fuera.

Esa explicación interna devuelve a la política económica un papel que el modelo neoclásico le negaba a largo plazo. La educación, la investigación, la protección de la propiedad intelectual, la competencia o el gasto público productivo pueden alterar la tasa de crecimiento de la renta per cápita, y no solo su nivel. La importancia de esta línea de trabajo la refleja el Premio Nobel de Economía de 2025, que Aghion y Howitt compartieron con Joel Mokyr "por haber explicado el crecimiento económico impulsado por la innovación".

Con ese fin, este tema estudia las tres grandes familias de modelos de crecimiento endógeno según el siguiente esquema:

- *Función de producción no neoclásica*: los modelos AK, Sobelow, de externalidades del capital y de gasto público productivo.
- *Capital humano*: el modelo de Uzawa-Lucas, con tasa de ahorro endógena.
- *Innovación tecnológica*: los modelos de I+D.

> Sobre los modelos de crecimiento exógeno (Solow-Swan y Ramsey-Cass-Koopmans), véase el tema A43. Sobre la contrastación empírica del crecimiento y la convergencia, véase el tema A45.

# 1. Función de producción no neoclásica

Esta primera familia de modelos parte de un diagnóstico preciso, según el cual los propios supuestos neoclásicos sobre la función de producción (rendimientos constantes a escala, rendimientos marginales decrecientes de cada factor y condiciones de Inada) impiden que el crecimiento se genere endógenamente. Con rendimientos marginales decrecientes que tienden a cero, la acumulación de capital se agota a sí misma, porque cada unidad adicional de capital aporta cada vez menos producto, hasta que la inversión apenas alcanza para reponer la depreciación y el crecimiento per cápita se detiene.

Una primera solución consiste, por tanto, en abandonar alguna de estas propiedades, manteniendo el resto del andamiaje de Solow y, en particular, la tasa de ahorro exógena. Así lo hacen el modelo AK, el modelo Sobelow y los de externalidades del capital. La excepción es el modelo de Barro (1990), que incorpora familias a la Ramsey.

## 1.1. Modelo AK (Rebelo, 1991)

**Origen**: lo propuso Rebelo (1991). La microfundamentación del modelo AK mediante el *aprendizaje por la práctica* procede de Arrow (1962) y Romer (1986).

**Idea**: los supuestos son idénticos a los del modelo de Solow-Swan, con la única excepción de la función de producción, que pasa a ser lineal en un único factor, el capital:

$$Y = AK$$

donde $K$ debe entenderse como capital *en sentido amplio*, es decir, no solo capital físico, sino también capital humano, conocimientos, infraestructuras y, en general, todo factor acumulable[1].

**Supuestos**:

- *Entorno*: el habitual, con tecnología constante, tasa de ahorro exógena $s$, economía cerrada y sin sector público, población que crece a la tasa $n$ y depreciación $\delta$.

- *Tecnología*: de las propiedades neoclásicas, $Y = AK$ conserva solo los rendimientos constantes a escala. Pierde los rendimientos marginales decrecientes (el capital tiene siempre un producto marginal igual a $A$) y también las condiciones de Inada, ya que ese producto marginal no tiende a cero por mucho que crezca el capital.

**Planteamiento**: como en el modelo de Solow-Swan, la renta se reparte entre consumo e inversión, $Y = C + I$, y cada ecuación de comportamiento (renta, inversión y consumo, por este orden) se escribe en niveles, per cápita y en tasas de crecimiento (tema A43, apartado 2.2 y Anexo):

$$\begin{aligned} Y = AK \quad &\Rightarrow \; y = Ak \quad &&\Rightarrow \; g_y = g_k \\ \dot{K} = sY - \delta K \quad &\Rightarrow \; \dot{k} = sAk - (n+\delta)k \quad &&\Rightarrow \; g_k = sA - (n+\delta) \\ C = (1-s)Y \quad &\Rightarrow \; c = (1-s)Ak \quad &&\Rightarrow \; g_c = g_y \end{aligned}$$

[1] La función $Y = BK^{\alpha}H^{1-\alpha}$, que combina capital físico y humano, equivale a una AK siempre que una unidad de producto pueda convertirse indistintamente en cualquiera de los dos capitales y estos se acumulen en la proporción que hace iguales sus rendimientos netos (con igual depreciación, $K/H = \alpha/(1-\alpha)$).

La renta y el consumo per cápita son proporcionales al capital per cápita y crecen a su mismo ritmo. La diferencia con el modelo de Solow-Swan está en la tasa de crecimiento del capital, que ya no depende de $k$, porque el producto medio del capital, $A$, no disminuye al acumularlo.

**Desarrollo gráfico**: en el plano capital por trabajador-tasa de crecimiento, $k$-$g_k$ (figura A44.1), las curvas de ahorro y de depreciación se representan divididas por $k$. La diferencia con el modelo de Solow es inmediata. En Solow, la curva de ahorro $sf(k)$ es cóncava (por los rendimientos decrecientes) y acaba cruzándose con la recta de depreciación $(n+\delta)k$ en un estado estacionario sin crecimiento per cápita. En el modelo AK, en cambio, ambas son rectas que parten del origen:

1. *Curva de ahorro*: recoge la inversión bruta por trabajador, $sAk$, una recta de pendiente $sA$, de modo que, dividida por $k$, es la recta horizontal $sA$.

2. *Recta de depreciación*: recoge la inversión que mantiene constante el capital por trabajador, $(n+\delta)k$, de pendiente $(n+\delta)$, y, dividida por $k$, es la recta horizontal $n+\delta$.

3. *Equilibrio*: si $sA > n+\delta$, las rectas no se cruzan y la distancia vertical entre ellas, $g_k$, es la misma cualquiera que sea el capital inicial, $k_0$. Por tanto, la economía crece indefinidamente a una tasa constante y positiva, común a todas las variables per cápita, $g_k = g_y = g_c = sA-(n+\delta)$.

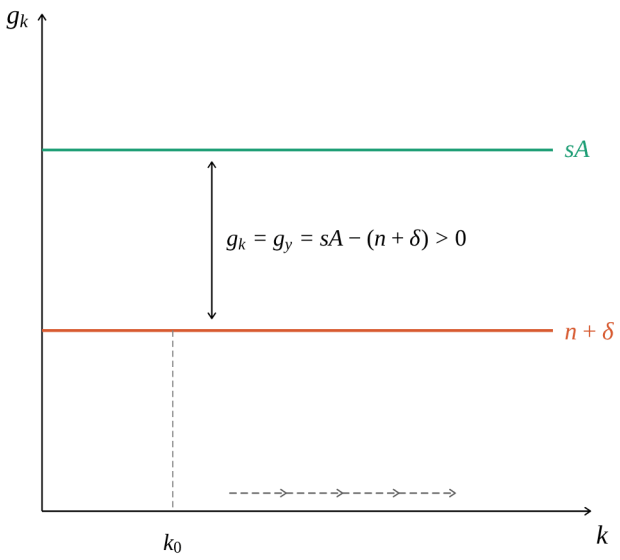



Figura A44.1: Modelo AK. Con rendimientos constantes del capital la tasa de crecimiento es positiva y permanente, sin convergencia.

**Dinámica**: las diferencias con el modelo de Solow-Swan son profundas:

1. *Crecimiento sostenido*: la renta per cápita aumenta de manera indefinida sin que haga falta un progreso técnico exógeno, rasgo que define el crecimiento endógeno.

2. *Política económica*: las políticas que eleven la tasa de ahorro o de inversión afectan a la tasa de crecimiento a largo plazo, y no solo al nivel de renta de estado estacionario, como ocurría en el modelo neoclásico. El alcance de la política económica es, por tanto, cualitativamente mayor.

3. *Sin dinámica de transición*: la economía crece desde el primer instante a un ritmo invariable, $sA-(n+\delta)$, tenga mucho o poco capital al empezar.

4. *Sin convergencia* (ni absoluta ni condicional): dos economías con parámetros idénticos crecen al mismo ritmo, partan de la renta que partan, y la distancia relativa entre ellas se mantiene para siempre.

5. *Efectos permanentes de las recesiones*: una recesión temporal tiene efectos permanentes sobre el nivel de renta. Al no existir un estado estacionario que actúe como atractor, no hay reversión a la tendencia previa.

6. *Sin ineficiencia dinámica*: dado que $s<1$, se cumple siempre que $r=A-\delta>sA-\delta=g_Y$, por lo que la economía no puede sobreacumular capital.

**Conclusiones**:

▪ *Aspectos positivos*: el modelo AK es la formulación más sencilla y transparente del crecimiento endógeno, e ilustra con nitidez su mecanismo esencial, que consiste en impedir que la productividad marginal del capital tienda a cero con la acumulación. Además, admite una microfundamentación, pues el AK es el caso $\alpha+\eta=1$ del modelo de externalidades, lo que da a su linealidad un contenido económico.

▪ *Limitaciones*: no predice convergencia, cuando la evidencia empírica, desde Barro y Sala-i-Martin (1992a) y Mankiw, Romer y Weil (1992), apoya la existencia de convergencia condicional. Tampoco distingue entre capital físico y capital humano, lo que equivale a suponer, de forma poco realista, que ambos son sustitutos perfectos en su acumulación (una unidad de producto se convierte indistintamente en uno u otro).

## 1.2. Modelo Sobelow (Jones y Manuelli, 1990)

**Origen**: procede del modelo de Jones y Manuelli (1990), bautizado por Sala-i-Martin como modelo *Sobelow* por combinar las funciones de producción de Solow y de Rebelo.

**Idea**: aislar la propiedad tecnológica sin la cual no hay crecimiento endógeno. Jones y Manuelli muestran que eliminar los rendimientos marginales decrecientes (como en el AK) basta, pero no es imprescindible. Lo indispensable es que falle la condición de Inada en el infinito, es decir, que el capital conserve una productividad marginal superior a cierta constante positiva, y el crecimiento se sostiene si ese suelo es suficientemente alto ($sA>n+\delta$).

**Supuestos**:

▪ *Entorno*: el del modelo AK.

▪ *Tecnología*: la función de producción suma un componente neoclásico y un componente lineal, $Y=AK+BK^{\alpha}L^{1-\alpha}$. Sus rendimientos a escala son constantes y los marginales del capital, decrecientes, pero la condición de Inada en el infinito falla, $\lim_{K\to\infty} f'_K=A\neq 0$.

**Planteamiento**: las ecuaciones de comportamiento son las del modelo AK, en el mismo orden, con la nueva función de producción:

$$\begin{aligned}
&Y=AK+BK^{\alpha}L^{1-\alpha} \;\Rightarrow\; y=Ak+Bk^{\alpha} \;\Rightarrow\; g_y=\varepsilon(k)\,g_k, \quad A,B>0,\ \alpha\in(0,1)\\
&\dot{K}=sY-\delta K \;\Rightarrow\; \dot{k}=sy-(n+\delta)k \;\Rightarrow\; g_k=sA+sBk^{\alpha-1}-(n+\delta)\\
&C=(1-s)Y \;\Rightarrow\; c=(1-s)y \;\Rightarrow\; g_c=g_y
\end{aligned}$$

donde $\varepsilon(k) = (Ak + \alpha Bk^{\alpha})/(Ak + Bk^{\alpha})$ es la elasticidad de la renta per cápita respecto al capital per cápita. Como la producción es una suma, su tasa de crecimiento es la media de las de sus dos componentes, $g_k$ y $\alpha g_k$, ponderada por su peso en la renta (tema A43, Anexo). Si $sA > n + \delta$, el capital crece sin límite, el peso del componente lineal tiende a uno y $g_y$ se acerca a $g_k$.

**Desarrollo gráfico**: el modelo Sobelow se representa en el plano capital por trabajador-tasa de crecimiento, $k$-$g_k$ (figuras A44.2 y A44.3):

1. *Curva de ahorro por unidad de capital*, $sA + sBk^{\alpha-1}$: recoge la inversión bruta por unidad de capital. Es decreciente, porque el componente neoclásico tiene rendimientos decrecientes, y tiende a la recta horizontal $sA$, su asíntota, porque el componente lineal no los tiene.

2. *Recta de depreciación*, $n+\delta$: recoge la inversión por unidad de capital que mantiene constante el capital por trabajador, y es horizontal porque no depende de $k$.

3. *Equilibrio*: la distancia vertical entre ambas es $g_k$. Si $sA < n + \delta$, se cortan en el estado estacionario, $k^*$ (punto 0), mientras que si $sA > n + \delta$ no se cortan y $g_k$ tiende a $sA - (n + \delta)$.

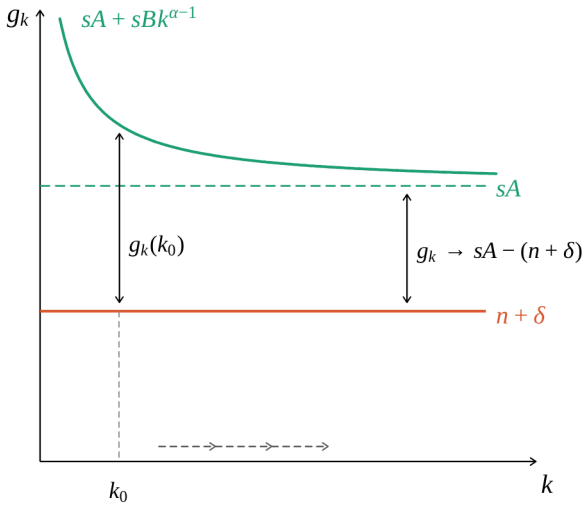



Figura A44.2: Modelo Sobelow (Jones-Manuelli). La curva de ahorro por unidad de capital no corta a la de depreciación ampliada y el crecimiento es sostenido.

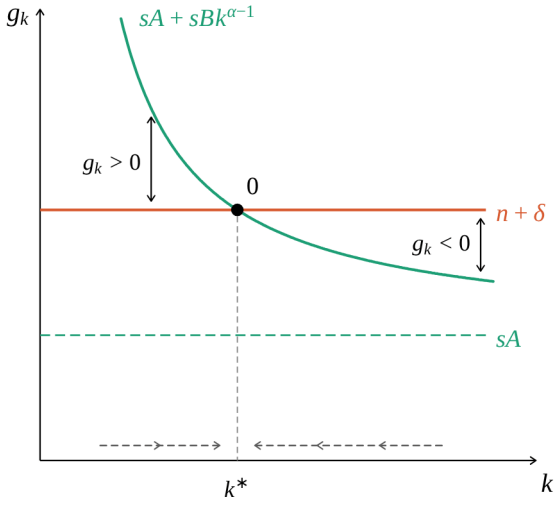



Figura A44.3: Modelo Sobelow con $sA < n + \delta$. La curva de ahorro por unidad de capital corta a la de depreciación ampliada en $k^*$ y la economía converge, como en Solow, a un estado estacionario sin crecimiento per cápita.

**Dinámica**: el comportamiento de la economía depende del signo de $sA - (n + \delta)$. Caben dos casos:

1. $sA > n + \delta$: a corto plazo la economía crece más deprisa que un modelo AK puro, pues el segundo término ($sBk^{\alpha-1}$) contribuye positivamente mientras el capital es escaso. A largo plazo, ese término se desvanece asintóticamente (por los rendimientos decrecientes del componente neoclásico) y la tasa de crecimiento converge a $g_k \to sA - (n + \delta) > 0$. La economía exhibe, así, una dinámica de transición con crecimiento decreciente hacia una tasa de largo plazo positiva (figura A44.2).

2. $sA < n + \delta$: el componente lineal no alcanza para mantener el crecimiento y, igual que en Solow, la renta per cápita acaba estancada en su nivel de estado estacionario (figura A44.3).

**Conclusiones**:

- *Aspectos positivos*: el modelo combina convergencia y dinámica transicional (herencia neoclásica) con crecimiento endógeno a largo plazo (herencia AK). Por ello, su ajuste empírico es mejor que el del modelo AK puro, que no puede dar cuenta de la evidencia de convergencia condicional.

- *Limitaciones*: sigue sin distinguir entre tipos de capital y descansa en una propiedad tecnológica (la cota inferior de la productividad marginal) que se postula sin microfundamentar.

### 1.3. Externalidades del capital (Romer, 1986; Lucas, 1988)

**Origen**: Romer (1986) y Lucas (1988) propusieron el modelo a partir del aprendizaje por la práctica (*learning by doing*) de Arrow (1962).

**Idea**: la tercera vía para escapar de los rendimientos decrecientes deja intacta la tecnología de cada empresa y añade efectos externos entre ellas. Invertir enseña a producir, y lo aprendido, al ser no rival y difícil de apropiar, no se queda en la empresa que invierte, sino que se filtra al

resto de la economía (*knowledge spillovers*). Por eso la inversión de cada empresa eleva también la producción de las demás.

**Supuestos**:

- *Entorno*: el del modelo AK.

- *Tecnología*: la función de producción de la empresa incorpora un término de externalidad $E$, en la forma $Y = AK^{\alpha}L^{1-\alpha}E^{\eta}$, con $\eta > 0$, donde $\eta$ mide la importancia de la externalidad. La empresa individual percibe rendimientos constantes en sus factores privados ($K$, $L$) y toma $E$ como dado, pero en el agregado la externalidad eleva los rendimientos del capital. Las dos formulaciones clásicas difieren en la variable que genera el desbordamiento:

  - *Visión de Romer* (1986): la externalidad procede del *stock* agregado de capital, $E = K$. En forma intensiva, $y = Ak^{\alpha+\eta}L^{\eta}$, de modo que la tasa de crecimiento depende del tamaño de la población. Según este *efecto escala*, el ritmo de crecimiento debería aumentar con la población del país, algo que los datos no confirman y que la variante siguiente elimina.

  - *Visión de Lucas* (1988): la externalidad procede del capital per cápita, $E = k$, con lo que $y = Ak^{\alpha+\eta}$ y el efecto escala desaparece. Se trata de una estilización habitual en los manuales, pues en Lucas (1988) la externalidad procede en realidad del capital humano medio de la economía, $H_a$, como se recoge en el apartado 2.2.

**Planteamiento**: tomando la visión de Lucas, las ecuaciones de comportamiento son las del modelo AK, en el mismo orden, con la nueva función de producción:

$$\begin{aligned} Y = AK^{\alpha}L^{1-\alpha}k^{\eta} &\Rightarrow y = Ak^{\alpha+\eta} \Rightarrow g_y = (\alpha+\eta)\, g_k, \quad A > 0,\ \alpha \in (0,1),\ \eta > 0 \\ \dot{K} = sY - \delta K &\Rightarrow \dot{k} = sAk^{\alpha+\eta} - (n+\delta)k \Rightarrow g_k = sAk^{\alpha+\eta-1} - (n+\delta) \\ C = (1-s)Y &\Rightarrow c = (1-s)Ak^{\alpha+\eta} \Rightarrow g_c = g_y \end{aligned}$$

La externalidad eleva el exponente del capital per cápita de $\alpha$ a $\alpha+\eta$, y con él la respuesta de la renta a la acumulación.

**Desarrollo gráfico**: el modelo de externalidades se representa en el plano capital por trabajador-tasa de crecimiento, $k$-$g_k$ (figuras A44.4 a A44.6):

1. *Curva de ahorro por unidad de capital*, $sAk^{\alpha+\eta-1}$: recoge la inversión bruta por unidad de capital, y su pendiente depende de los rendimientos del capital ampliado. Es decreciente si $\alpha+\eta < 1$ (rendimientos decrecientes), horizontal en $sA$ si $\alpha+\eta = 1$ (constantes) y creciente si $\alpha+\eta > 1$ (crecientes).

2. *Recta de depreciación*, $n+\delta$: recoge la inversión por unidad de capital que mantiene constante el capital por trabajador, y es horizontal porque no depende de $k$.

3. *Equilibrio*: la distancia vertical entre ambas es $g_k$. Si $\alpha+\eta \neq 1$, se cortan en el estado estacionario, $k^*$ (punto 0), que es estable en el primer caso e inestable en el tercero.

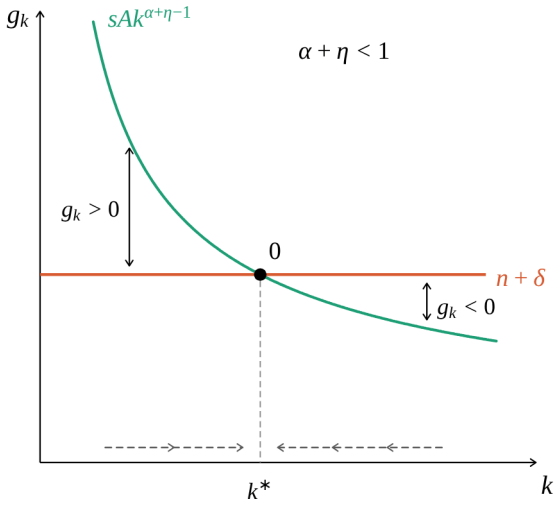



Figura A44.4: Externalidades del capital con $\alpha + \eta < 1$. La curva de ahorro por unidad de capital es decreciente, corta a la de depreciación ampliada en $k^*$ y el crecimiento per cápita a largo plazo es nulo, como en Solow.

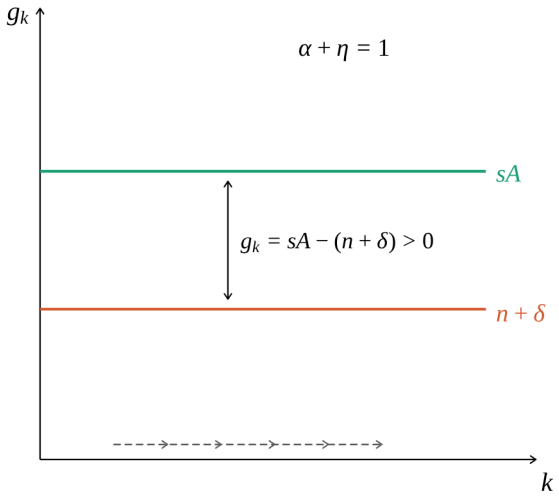



Figura A44.5: Externalidades del capital con $\alpha + \eta = 1$. La externalidad compensa exactamente los rendimientos decrecientes, la tecnología es lineal en $k$ y la economía crece a la tasa constante $sA - (n + \delta)$, como en el modelo AK.

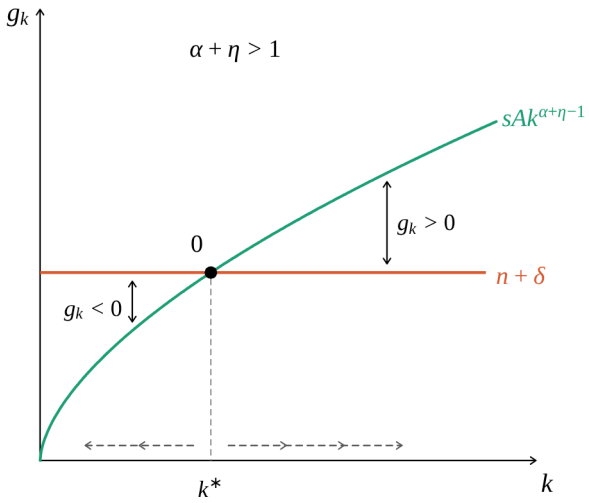



Figura A44.6: Externalidades del capital con $\alpha+\eta > 1$. La curva de ahorro por unidad de capital es creciente y el estado estacionario $k^*$ es inestable. Por encima de él la senda es explosiva y por debajo, implosiva.

**Dinámica**: el comportamiento de la economía depende críticamente del tamaño de la externalidad, es decir, de la suma $\alpha + \eta$. Caben tres casos:

1. *Externalidad pequeña* ($\alpha + \eta < 1$): el capital ampliado sigue presentando rendimientos marginales decrecientes, el ahorro por unidad de capital, $sAk^{\alpha+\eta-1}$, es decreciente en $k$ y existe un estado estacionario estable con crecimiento per cápita nulo ($g_k^* = 0$). El modelo es cualitativamente equivalente a Solow, ya que la externalidad amplifica el efecto de la acumulación de capital (el estado estacionario se sitúa más lejos), pero no genera crecimiento endógeno (figura A44.4).

2. *Externalidad media* ($\alpha + \eta = 1$): compensa exactamente los rendimientos decrecientes y el modelo se vuelve formalmente idéntico al AK, con $g_k = sA - (n + \delta)$. Hay crecimiento endógeno si $sA > n + \delta$ (figura A44.5).

3. *Externalidad grande* ($\alpha + \eta > 1$): el capital ampliado presenta rendimientos crecientes, el ahorro por unidad de capital es creciente en $k$ y el estado estacionario es inestable. Por encima de él la economía entra en una senda explosiva y por debajo, en una senda implosiva (figura A44.6). El interés empírico de este caso es limitado.

**Extensiones**: en política económica, la existencia de externalidades da margen a la intervención pública para elevar la tasa de crecimiento por dos vías:

- *Subvenciones a la inversión*: junto con las desgravaciones, aumentan el capital en sentido amplio. Se justifican porque las empresas privadas no capturan toda la externalidad y, por ello, con ahorro endógeno (como en Romer, 1986), invierten menos de lo socialmente óptimo.

- *Promoción de los efectos externos*: mediante el fomento de agrupaciones tecnológicas (*clusters*), centros de investigación o relaciones universidad-empresa que faciliten la difusión del conocimiento.

**Conclusiones**:

▪ *Aspectos positivos*: los modelos de externalidades endogenizan el progreso técnico con un mecanismo económicamente plausible (el desbordamiento del conocimiento) y sin abandonar la competencia perfecta, pues la empresa individual sigue operando con rendimientos constantes en sus factores privados.

▪ *Limitaciones*: la evidencia de *learning by doing* procede de sectores concretos, como los semiconductores, donde Irwin y Klenow (1994) estiman tasas de aprendizaje del 20 %, y no mide un desbordamiento de conocimientos de la escala necesaria para sostener el crecimiento agregado. Además, el progreso técnico sigue siendo un subproducto no buscado de la inversión, cuando en la realidad una parte sustancial de la innovación se realiza ex profeso, mediante inversión deliberada en I+D.

### 1.4. Gasto público productivo (Barro, 1990)

**Origen**: Robert Barro (1990) propuso el modelo.

**Idea**: el factor que sostiene la productividad marginal del capital privado es la provisión de servicios públicos (infraestructuras, seguridad jurídica, orden público), financiada con impuestos. El sector público es, así, a la vez un factor productivo y una distorsión, y de ese doble papel resulta un tamaño óptimo del Estado para el crecimiento.

**Supuestos**:

▪ *Entorno*: economía cerrada con familias a la Ramsey (utilidad con aversión relativa al riesgo constante, CRRA, de parámetro $\theta$ y tasa de descuento $\rho$), población constante y depreciación $\delta$. A diferencia de los modelos anteriores del apartado, la tasa de ahorro es aquí endógena.

▪ *Agentes*: el sector público provee un flujo de servicios públicos por trabajador $g$, no rival ni congestionable, financiado con un impuesto proporcional sobre la renta al tipo $\tau$ bajo presupuesto equilibrado, $g = \tau y$.

▪ *Tecnología*: la producción privada emplea capital y servicios públicos, con rendimientos constantes en el conjunto de ambos y decrecientes en el capital privado, $y = Ak^{\alpha}g^{1-\alpha}$, con $\alpha \in (0,1)$ la participación del capital, como en el tema A38 (Barro, 1990, escribe $y = Ak^{1-\alpha}g^{\alpha}$, de modo que en su notación $\tau^{*} = \alpha$). Sustituyendo la restricción presupuestaria del sector público, $g = \tau y$, y despejando, la función de producción es lineal en el capital privado, $y = \tilde{A}(\tau)\,k$, con $\tilde{A}(\tau) \equiv A^{1/\alpha}\tau^{(1-\alpha)/\alpha}$, creciente en $\tau$. Se obtiene así un modelo AK en el que la constante depende de la política fiscal.

**Planteamiento**: las familias toman $\tau$ y $g$ como dados y maximizan la utilidad intertemporal sujetas a la acumulación de capital neta de impuestos:

$$\max_{\{c\}} \int_0^{\infty} e^{-\rho t}\,\frac{c^{1-\theta}-1}{1-\theta}\,dt, \qquad \theta > 0$$

$$\text{s.a.:} \begin{cases} c + \dot{k} + \delta k = (1-\tau)Ak^{\alpha}g^{1-\alpha}, & A > 0,\ \alpha \in (0,1) \\ k_0 \text{ dado} \end{cases}$$

donde $c$ es el consumo por trabajador (la variable de control)[2], $k$ el capital privado por trabajador (la variable de estado, con $k_0$ dado), $g$ los servicios públicos por trabajador, $\tau$ el tipo impositivo

[2] En tiempo continuo se omite el argumento temporal: $c \equiv c(t)$, $k \equiv k(t)$ y $\lambda \equiv \lambda(t)$; se escribe solo en las condiciones inicial y terminal.

sobre la renta, $A$ y $\alpha$ los parámetros de la tecnología, $\theta$ el parámetro de la utilidad CRRA, $\rho$ la tasa de descuento y $\delta$ la depreciación.

**CPO**: por el principio del máximo, la ecuación de Keynes-Ramsey con el rendimiento neto del capital privado, $r = (1-\tau)\,\partial y/\partial k - \delta = (1-\tau)\,\alpha\tilde{A}(\tau) - \delta$, y la condición de transversalidad, que anula en el límite el valor descontado del capital:

$$g_c = \frac{1}{\theta}\big[\underbrace{(1-\tau)\,\alpha\tilde{A}(\tau) - \delta}_{\text{rendimiento neto del capital}} - \rho\big]$$
$$\lim_{t\to\infty} e^{-\rho t}\, c(t)^{-\theta}\, k(t) = 0$$

La ecuación de Keynes-Ramsey iguala el rendimiento neto del capital privado, una vez descontados el impuesto y la depreciación, a la tasa de descuento más el crecimiento del consumo ponderado por $\theta$. El consumo crece mientras la última unidad de capital, tras pagar el impuesto, rinde más de lo que la familia exige por posponer su consumo. El consumo crece tanto más deprisa cuanto menor es $\theta$.

**CSO**: el hamiltoniano es estrictamente cóncavo en $c$ y cóncavo en $k$ dado $g$ (la tecnología privada, $Ak^{\alpha}g^{1-\alpha}$, lo es), de modo que la CPO y la condición de transversalidad son suficientes. El crecimiento es positivo si $(1-\tau)\,\alpha\tilde{A} - \delta > \rho$ y la utilidad está acotada si $\rho > (1-\theta)g_c$.

**Equilibrio**: como la tecnología es lineal en $k$, no hay dinámica de transición, y consumo, capital y renta crecen desde el principio a la misma tasa constante $g_c$. La tasa de crecimiento depende de $\tau$ a través de dos efectos contrapuestos, ya que el impuesto reduce el rendimiento privado de la inversión (efecto *distorsión*, el factor $1-\tau$) y financia un factor que eleva la productividad del capital (efecto *productivo*, el factor $\tilde{A}(\tau)$). Maximizando $(1-\tau)\tau^{(1-\alpha)/\alpha}$ respecto a $\tau$:

$$\tau^{*} = 1 - \alpha$$

El tipo impositivo que maximiza el crecimiento iguala la elasticidad del producto respecto al gasto público (*regla de Barro*), de modo que el sector público debe tener el tamaño que corresponde a su contribución productiva. Ese tipo maximiza también la utilidad, porque, sin dinámica de transición, tanto la tasa de crecimiento como el consumo inicial dependen de $\tau$ solo a través del producto neto de impuestos por unidad de capital, $(1-\tau)\tilde{A}(\tau)$, y la utilidad de equilibrio es creciente en él.

**Desarrollo gráfico**: en el plano tipo impositivo-tasa de crecimiento, $\tau$-$g_c$ (figura A44.7), la curva $g_c(\tau)$ tiene forma de U invertida, con máximo en $\tau^{*} = 1-\alpha$ (regla de Barro). En $\tau = 0$ (sin servicios públicos no hay producción) y en $\tau = 1$ (sin rendimiento privado no hay inversión) se anula el rendimiento del capital privado después de impuestos, $(1-\tau)\,\alpha\tilde{A}(\tau)$, y la tasa de crecimiento alcanza su mínimo, $-(\delta+\rho)/\theta < 0$. A la izquierda del máximo domina el efecto productivo y un sector público mayor acelera el crecimiento, mientras que a la derecha domina el efecto distorsión y un sector público mayor lo frena.

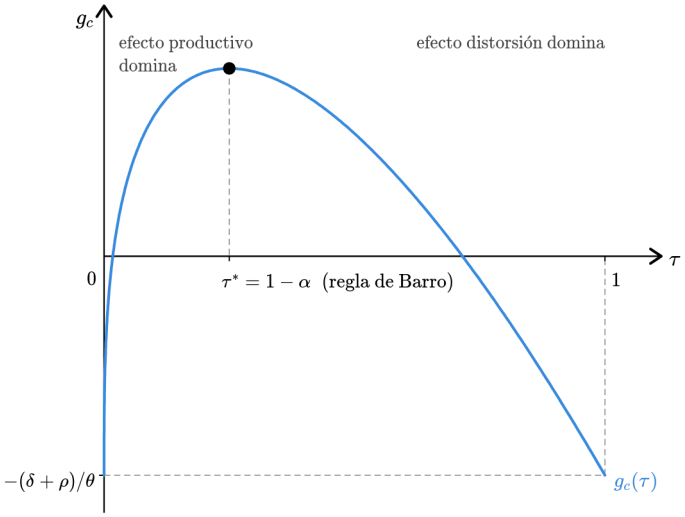



Figura A44.7: Modelo de Barro (1990). La tasa de crecimiento es una función en U invertida del tipo impositivo, con máximo en $\tau^* = 1 - \alpha$. A su izquierda un sector público mayor acelera el crecimiento y a su derecha lo frena.

**Estática comparativa**: un aumento de $A$ o una reducción de $\rho$ o $\delta$ elevan la tasa de crecimiento para todo $\tau$ (una reducción de $\theta$, allí donde la tasa es positiva), sin alterar $\tau^*$, que solo depende de la tecnología.

**Extensiones**: el modelo separa el gasto público productivo del gasto en consumo público, que no eleva $\tilde{A}$ y solo distorsiona. Con ello ofrece una lectura de la evidencia de Barro (1991), que encuentra un efecto negativo del consumo público sobre el crecimiento y no de la inversión pública (tema A45).

En política fiscal, por tanto, la composición del gasto importa más que su volumen. La financiación con impuestos de suma fija en lugar de proporcionales elimina el efecto distorsión y eleva la tasa de crecimiento, aunque el tamaño óptimo del sector público sigue siendo $g/y = 1 - \alpha$ (condición de eficiencia productiva $\partial y/\partial g = 1$). Con impuestos proporcionales, por el contrario, el equilibrio descentralizado es ineficiente porque las familias no internalizan que su inversión, al elevar la recaudación, financia servicios públicos para todos (tema A38). Si los servicios públicos son congestionables (el servicio que prestan a cada empresa depende de $g/k$, Barro y Sala-i-Martin, 1992b), el tamaño óptimo es el mismo. En ese caso, sin embargo, el impuesto sobre la renta pasa a ser el instrumento eficiente, porque internaliza la congestión que cada unidad de capital privado provoca.

**Conclusiones**:

- *Aspectos positivos*: el modelo de Barro es la referencia sobre la relación entre tamaño del sector público y crecimiento, y fundamenta la distinción entre gasto productivo e improductivo que utiliza la política fiscal orientada al crecimiento.

- *Limitaciones*: la no rivalidad de los servicios públicos es un supuesto fuerte, porque la congestión es la regla en las infraestructuras. Además, el resultado depende de que el gasto sea un flujo y no un *stock* de capital público, que introduciría dinámica de transición, y la evidencia sobre la relación entre tamaño del Estado y crecimiento es débil y muy sensible a la muestra.

## 1.5. Conclusiones

Los modelos de esta sección muestran con la mayor sencillez que el crecimiento sostenido exige que el rendimiento del capital en sentido amplio no decrezca hasta cero y que, con ahorro exógeno, se mantenga por encima de $(n + \delta)/s$. La tasa de ahorro, y con ella la política económica (en Barro, el tipo impositivo que financia el gasto público productivo), determina entonces la tasa de crecimiento y no solo el nivel de renta.

Esa tasa de ahorro sigue siendo, sin embargo, exógena (salvo en Barro, 1990), de modo que los modelos no explican por qué se ahorra. Además, la externalidad del capital de Romer (1986) implica un efecto escala que la evidencia no respalda, y ninguno de ellos distingue el capital humano del físico ni explica cómo se genera el conocimiento, cuestiones que abordan, respectivamente, los apartados 2 y 3.

# 2. Capital humano con tasa de ahorro endógena: el modelo de Uzawa-Lucas (1965; 1988)

**Origen**: procede del modelo de Uzawa (1965), recuperado y popularizado por Lucas (1988).

**Idea**: la segunda familia de modelos aborda la principal limitación común del bloque anterior, la ausencia de distinción entre capital físico y capital humano. El modelo introduce esta distinción de manera explícita y, además, endogeniza la tasa de ahorro mediante hogares optimizadores a la Ramsey.

## 2.1. Supuestos

**Entorno**: la economía es cerrada, sin sector público, y las familias son propietarias de los factores de producción. Por simplicidad la población es constante ($n = 0$) y se normaliza a la unidad, de modo que las variables agregadas y per cápita coinciden.

**Agentes**: los hogares maximizan una utilidad intertemporal CRRA con dos variables de control, el consumo $c$ y la fracción de tiempo $u$ dedicada a la producción, y dos variables de estado, $K$ y $H$.

**Tecnología**:

▪ *Factores*: se distingue entre capital físico ($K$) y capital humano ($H$), ambos factores rivales (lo que se emplea en una actividad no está disponible para la otra), que se acumulan mediante tecnologías distintas:

- *Capital físico*: se acumula a partir de la producción no consumida, neta de depreciación.
- *Capital humano*: es un factor que se distribuye entre dos sectores. Una fracción $u \in (0, 1)$ se dedica a la producción de bienes finales ($H_Y = uH$) y la fracción restante, al sector educativo ($H_H = (1 - u)H$)[3].

▪ *Sector de bienes finales*: produce con una tecnología Cobb-Douglas que combina capital físico y capital humano, $Y = AK^{\alpha}(uH)^{1-\alpha}$. El capital físico se acumula a partir de la producción no consumida, $c + \dot{K} + \delta K = Y$.

▪ *Sector educativo*: acumula capital humano con una tecnología específica de productividad $B$. En general, esta tecnología podría emplear ambos capitales, con una participación $\alpha_H$ del capital físico. Uzawa y Lucas, no obstante, parten de que, comparada con la producción de bienes, la educación hace un uso mucho más intensivo del capital humano ($\alpha_H < \alpha$), y se quedan con el caso extremo en que solo emplea ese factor ($\alpha_H = 0$):

$$\dot{H} = B(1 - u)H - \delta H$$

El supuesto tiene consecuencias, porque convierte la tecnología educativa en un AK encubierto (lineal en el factor acumulable), y de ahí nace precisamente el crecimiento endógeno.

[3] Con la noción de capital humano, elaborada por la escuela de Chicago entre finales de los cincuenta y los sesenta, el trabajo, antes visto como no acumulable, pasa a ser un factor acumulable.

## 2.2. Modelo

**Hogares**: propietarios de los dos capitales, eligen el consumo, $c$, y la fracción del capital humano que dedican a producir, $u$.

▪ *Planteamiento*: maximizan la utilidad intertemporal sujetos a las leyes de acumulación de ambos tipos de capital:

$$\max_{\{c,u\}} \int_0^\infty e^{-\rho t} \frac{c^{1-\theta}-1}{1-\theta}\, dt, \qquad \theta > 0$$

$$\text{s.a.:} \begin{cases} c + \dot{K} + \delta K = AK^\alpha (uH)^{1-\alpha}, & A > 0,\ \alpha \in (0,1) \\ \dot{H} = B(1-u)H - \delta H \\ K_0,\ H_0 \text{ dados} \end{cases}$$

donde $c$ es el consumo y $u$ la fracción del capital humano dedicada a la producción de bienes (las variables de control), $K$ el capital físico y $H$ el capital humano (las variables de estado, con $K_0$ y $H_0$ dados), $A$ y $\alpha$ los parámetros de la tecnología del sector final, $B$ la productividad del sector educativo, $\theta$ el parámetro de la utilidad CRRA, $\rho$ la tasa de descuento y $\delta$ la tasa de depreciación.

El hamiltoniano en valor corriente tiene dos multiplicadores de coestado, $\lambda$ para el capital físico y $\mu$ para el capital humano:

$$\mathcal{H} = \frac{c^{1-\theta}-1}{1-\theta} + \lambda\left[AK^\alpha (uH)^{1-\alpha} - c - \delta K\right] + \mu\left[B(1-u)H - \delta H\right]$$

▪ *CPO*: por el principio del máximo de Pontryagin, se obtienen las cuatro condiciones siguientes, a las que se suman las dos de transversalidad[4], $\lim_{t\to\infty} e^{-\rho t}\lambda K = \lim_{t\to\infty} e^{-\rho t}\mu H = 0$:

$$\frac{\partial \mathcal{H}}{\partial c} = 0 \ \Rightarrow\ c^{-\theta} = \lambda \tag{1}$$

$$\frac{\partial \mathcal{H}}{\partial u} = 0 \ \Rightarrow\ \lambda(1-\alpha)AK^\alpha u^{-\alpha} H^{1-\alpha} = \mu B H \tag{2}$$

$$\frac{\partial \mathcal{H}}{\partial K} = \rho\lambda - \dot{\lambda} \ \Rightarrow\ \lambda\left[\alpha AK^{\alpha-1}(uH)^{1-\alpha} - \delta\right] = \rho\lambda - \dot{\lambda} \tag{3}$$

$$\frac{\partial \mathcal{H}}{\partial H} = \rho\mu - \dot{\mu} \ \Rightarrow\ \mu(B-\delta) = \rho\mu - \dot{\mu} \tag{4}$$

Las condiciones (1) y (3) dan la ecuación de Keynes-Ramsey habitual, $g_c = (1/\theta)(r-\rho)$ con $r = \alpha AK^{\alpha-1}(uH)^{1-\alpha} - \delta$. La condición (2) es la de asignación del tiempo e iguala el valor de una unidad de capital humano en la producción de bienes (su productividad marginal, valorada en $\lambda$) con su valor en la educación (la productividad $B$, valorada en $\mu$). Por último, la condición (4) es la ecuación de Euler del capital humano, cuyo precio sombra decrece a la tasa $B-\delta-\rho$, el rendimiento de la educación neto de depreciación y de la tasa de descuento[5].

[4] La condición de transversalidad es la holgura complementaria de las restricciones de no negatividad de los *stocks* de capital llevada al límite (tema A43, apartado 3.2).

[5] La condición (4) muestra que, con $\alpha_H = 0$, el precio sombra del capital humano evoluciona con independencia de la tecnología de bienes, que es lo que aísla la tasa de crecimiento de $A$ y $\alpha$.

▪ *CSO*: el término $uH$ impide que el hamiltoniano sea conjuntamente cóncavo, de modo que se supone que se cumple la condición de suficiencia de Arrow, con la que las condiciones de Pontryagin y las de transversalidad son suficientes[6].

**Empresas**: resuelven el programa habitual de los entornos dinámicos neoclásicos.

▪ *Planteamiento*: alquilan capital físico y capital humano tomando como dados sus precios, $\max_{K,\,uH} \; AK^{\alpha}(uH)^{1-\alpha} - (r+\delta)K - w_H\,uH$, donde $w_H$ es la remuneración de cada unidad de capital humano empleada en producción.

▪ *CPO*: contratan cada factor hasta que su productividad marginal iguala su coste real, de donde $r = \alpha A K^{\alpha-1}(uH)^{1-\alpha} - \delta$ para el capital físico y $w_H = (1-\alpha)AK^{\alpha}(uH)^{-\alpha}$ para el capital humano empleado en producción. Con rendimientos constantes a escala, ambas remuneraciones agotan el producto y el beneficio es nulo.

▪ *CSO*: la tecnología Cobb-Douglas es cóncava en $(K, uH)$ y estrictamente cóncava en cada factor, luego las CPO caracterizan el máximo.

**Equilibrio**: en una senda de crecimiento equilibrado la ratio $K/H$ y la fracción $u$ son constantes, de modo que $r$ es constante. Diferenciando la condición (2) respecto al tiempo (con $K/H$ y $u$ constantes) e igualando con las condiciones (3) y (4) se obtiene la condición de arbitraje entre los dos capitales, $r = B - \delta$. Según esta condición, el rendimiento del capital físico iguala al del capital humano, que es la productividad del sector educativo neta de depreciación. Sustituyendo en Keynes-Ramsey, todas las variables crecen a la misma tasa constante:

$$g_c^* = g_K^* = g_H^* = g_Y^* = \frac{1}{\theta}(B - \delta - \rho)$$

y la fracción de tiempo dedicada a producir se obtiene de la ley de acumulación del capital humano, $g_H^* = B(1-u^*) - \delta$:

$$u^* = 1 - \frac{g_H^* + \delta}{B} = 1 - \frac{B - \delta - \rho + \theta\delta}{\theta B}$$

Para que la renta per cápita crezca hace falta $B > \delta + \rho$, es decir, que el sistema educativo rinda lo bastante como para cubrir la depreciación del capital humano (el "olvido" de conocimientos, $\delta$) y la impaciencia de los hogares ($\rho$). De todos los parámetros, solo $B$ empuja al alza el crecimiento a largo plazo, y por eso la educación hace aquí de motor. La lógica es la del AK, pues la tecnología educativa es lineal en el factor acumulable, pero el modelo no exige suponer que capital físico y humano sean sustitutos perfectos.

Con $\theta = 1$ (utilidad logarítmica), $u^* = \rho/B$, y los agentes más impacientes producen más y estudian menos. Por último, la solución descentralizada coincide con la del planificador benevolente, porque, al no existir externalidades ni bienes públicos, se cumple el primer teorema fundamental de la economía del bienestar.

**Desarrollo gráfico**: el equilibrio del modelo se representa en el plano ratio de capitales-rendimientos, con $\omega \equiv K/H$ en el eje horizontal (figura A44.8):

1. *Rendimiento del capital físico*, $r = \alpha A \omega^{\alpha-1} u^{1-\alpha} - \delta$: decreciente en $\omega$, por los rendimientos decrecientes del capital físico.

[6] La utilidad es estrictamente cóncava en $c$, pero el producto $uH$ impide aplicar la condición de Mangasarian, que exige la concavidad conjunta del hamiltoniano en controles y estados. La condición de Arrow solo exige la concavidad del hamiltoniano maximizado en las variables de estado, $(K, H)$.

2. *Rendimiento del capital humano*, $B-\delta$: recta horizontal que no decrece con la acumulación, la recta AK que sostiene el crecimiento.

3. *Tasa de descuento*, $\rho$: la impaciencia de los hogares, una recta horizontal porque no depende de $\omega$.

4. *Equilibrio*: en $\omega^*$ se igualan los rendimientos de ambos capitales, $r = B-\delta$, y la distancia entre $B-\delta$ y $\rho$ es $\theta g_Y^*$.

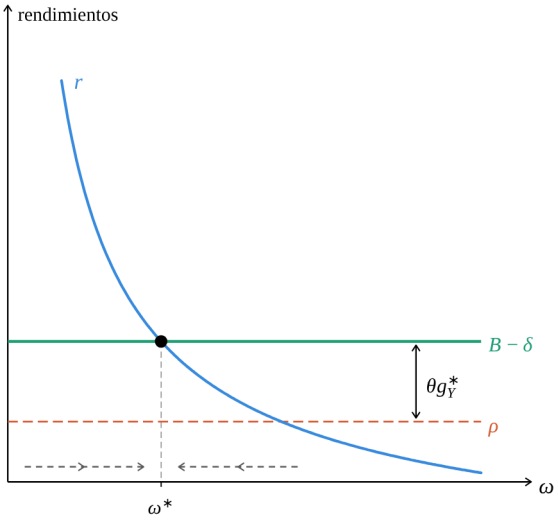



Figura A44.8: Modelo de Uzawa-Lucas. El rendimiento del capital humano, $B-\delta$, no decrece con la acumulación y sostiene el crecimiento. En $\omega^*$ se igualan los rendimientos de ambos capitales y la economía crece a $g_Y^* = (B-\delta-\rho)/\theta$.

**Estática comparativa**: la tasa de crecimiento aumenta con la productividad educativa ($B$) y disminuye con la impaciencia ($\rho$), la tasa de olvido ($\delta$) y, si es positiva, con el deseo de alisar el consumo ($\theta$). La fracción $u^*$ se mueve en sentido contrario a $g_Y^*$ para cambios en $\rho$ y, siempre que $\rho > (\theta-1)\delta$ (lo que se cumple, p. ej., con utilidad logarítmica o sin olvido), también en $B$. Una mayor productividad educativa reduce entonces $u^*$ (conviene dedicar más tiempo a estudiar), mientras que una mayor impaciencia lo eleva (los agentes impacientes prefieren producir y consumir hoy). Los parámetros de la tecnología de bienes ($A$, $\alpha$), en cambio, no afectan a la tasa de crecimiento de largo plazo, solo al nivel de la senda y a la ratio $\omega^*$.

**Extensiones**: Lucas (1988) añade una *externalidad del capital humano*, $Y = AK^{\alpha}(uH)^{1-\alpha}H_a^{\zeta}$, donde $H_a$ es el capital humano medio de la economía, que cada agente toma como dado. Dicho de otro modo, la productividad de un trabajador crece con la cualificación de quienes le rodean. Por ello, con $\zeta > 0$ la solución de mercado dedica menos tiempo a la educación que la del planificador y crece por debajo del óptimo.

Rebelo (1991) muestra que el resultado se mantiene si la educación usa también capital físico ($\alpha_H > 0$), siempre que la tecnología educativa presente rendimientos constantes en el conjunto de los factores acumulables. Por su parte, Lucas interpreta el modelo como explicación de las diferencias persistentes de crecimiento entre países y de las migraciones de trabajadores cualificados hacia los países con más capital humano.

En política educativa, en consecuencia, toda mejora permanente de la productividad del sistema educativo ($B$) eleva permanentemente la tasa de crecimiento. Con la externalidad de Lucas, además, la subvención a la educación acerca $u$ a su valor óptimo (tema A25, apartado 3.1.3).

## 2.3. Dinámica

La representación habitual de la dinámica usa dos variables estacionarias, la ratio de capitales, $\omega$, y la ratio consumo-capital $\chi \equiv c/K$, junto con $u$. En el plano $(\omega, \chi)$ el sistema tiene un único estado estacionario $(\omega^*, \chi^*)$, que es un punto de silla. La variable predeterminada es $\omega$ y la de salto, $\chi$ (con $u$). Por ello, la economía se sitúa de inmediato sobre la senda estable y converge a lo largo de ella, mientras $K$, $H$ y $c$ crecen a la tasa $g_Y^*$ en el límite.

Los procesos de ajuste surgen cuando un *shock* aleja la ratio $\omega$ de su valor estacionario. Con $\theta > \alpha$, el caso empíricamente relevante, son los siguientes:

1. *Ratio inferior a la estacionaria*: el caso $\omega < \omega^*$ se da, por ejemplo, tras una destrucción de capital físico (como un conflicto armado que arrasa las infraestructuras pero no el conocimiento de la población). El rendimiento del capital físico es entonces alto, los agentes dedican más tiempo a producir ($u > u^*$) y la economía crece transitoriamente por encima de su tasa estacionaria mientras reconstruye el capital físico[7].

2. *Ratio superior a la estacionaria*: si $\omega > \omega^*$, como ocurre tras una fuga de cerebros que reduce el capital humano, el rendimiento del capital físico es bajo y los agentes estudian más ($u < u^*$) mientras se reconstruye el capital humano.

El modelo presenta, por tanto, una dinámica de transición no trivial, a diferencia del AK, aunque la convergencia es hacia una tasa de crecimiento común y no hacia un nivel de renta. Así, dos economías con los mismos parámetros y distinta dotación inicial crecen a la misma tasa a largo plazo pero mantienen su diferencia de nivel.

## 2.4. Conclusiones

**Aspectos positivos**: el modelo genera crecimiento endógeno distinguiendo entre capitales con tecnologías de acumulación diferenciadas y hace de la asignación del tiempo entre producir y aprender la decisión que determina el crecimiento. Presenta, además, dinámica transicional (lo que mejora su ajuste empírico frente al AK), y su solución descentralizada es óptima. El modelo dota así de microfundamentos a la relación entre educación y crecimiento y explica por qué economías con la misma tecnología de producción de bienes pueden crecer a ritmos distintos de forma permanente si difieren en la productividad de su sistema educativo o en sus preferencias.

**Limitaciones**: no predice convergencia entre economías. Sus implicaciones de política, aunque potentes, descansan de manera crítica en el supuesto de que la tecnología educativa presenta rendimientos constantes en los factores acumulables (en el caso básico, que emplea solo capital humano), supuesto que la evidencia sobre los rendimientos de la educación no confirma. El modelo tampoco distingue entre el capital humano que produce y el que innova, distinción que introducen los modelos de ideas.

[7] Así se interpreta habitualmente la rápida reconstrucción de Alemania y Japón tras la Segunda Guerra Mundial.

# 3. Innovación tecnológica: modelos de I+D con competencia imperfecta

La tercera familia de modelos da el paso que las anteriores evitaban, al modelizar la innovación como una actividad económica deliberada, para lo que resulta imprescindible abandonar la competencia perfecta. Bajo competencia perfecta, el teorema de Euler garantiza que la retribución de los factores rivales (capital y trabajo) a su productividad marginal agota exactamente el valor de la producción, de modo que no queda renta alguna con la que financiar la I+D.

Además, la tecnología tiene una estructura de costes peculiar. Producir una idea exige costes fijos elevados, pero reproducirla tiene un coste marginal casi nulo, lo que genera economías de escala y hace que los bienes tecnológicos tiendan a ser producidos por empresas con poder de mercado. Los modelos de innovación introducen, por tanto, competencia imperfecta y empresas que invierten explícitamente en progreso tecnológico a cambio de rentas de monopolio protegidas por patentes.

Dentro de esta familia se distinguen dos enfoques según la naturaleza de la innovación. La innovación es *horizontal* en Romer (1990), donde aumenta el número de variedades, y *vertical* en el enfoque schumpeteriano de Aghion y Howitt (1992), donde mejora la calidad de los productos existentes. La crítica del efecto escala de Jones (1995) cierra la sección.

## 3.1. Aumento del número de variedades (Romer, 1990)

**Origen**: Romer (1990) propuso el primer modelo en el que la innovación es el resultado de una inversión deliberada en I+D.

**Idea**: en este modelo, la innovación consiste en el aumento del número $N$ de variedades de bienes de capital intermedios disponibles para la producción.

### 3.1.1. Supuestos

**Entorno**: el análisis es de equilibrio general con horizonte infinito y, por simplicidad, población constante ($n = 0$).

**Agentes**: hay tres clases. Los productores de bienes finales operan en competencia perfecta, cada inventor de un bien intermedio disfruta, gracias a una patente perpetua, del monopolio de la variedad que ha creado, y los consumidores optimizan a la Ramsey.

**Tecnología**: el bien final se produce con trabajo y con las $N$ variedades de bienes intermedios disponibles, $Y = AL^{1-\alpha}\sum_{j=1}^{N} x_j^{\alpha}$, con $\alpha \in (0,1)$. Inventar una variedad nueva exige un coste fijo $F$, y producirla después, un coste marginal constante normalizado a uno.

### 3.1.2. Modelo

La tasa de crecimiento de equilibrio resulta de la interacción entre lo que deciden los inventores, los productores de bienes finales y los consumidores.

**Productores de bienes finales**: operan en competencia perfecta.

▪ *Planteamiento*: maximizan su beneficio tomando como dados el salario y los precios de los intermedios, con el bien final como numerario:

$$\max_{L,\{x_j\}} \; A\,L^{1-\alpha}\sum_{j=1}^{N} x_j^{\alpha} - wL - \sum_{j=1}^{N} p_j x_j$$

donde el primer término es la producción, $Y$, $A$ la productividad, $L$ el empleo, $w$ el salario, $N$ el número de variedades, $x_j$ la cantidad empleada del bien intermedio $j$, $p_j$ su precio y $\alpha \in (0,1)$ el exponente de cada intermedio.

▪ *CPO*: el productor final contrata trabajo y cada bien intermedio hasta que su productividad marginal iguala su precio, $\partial Y/\partial L = w$ y $\partial Y/\partial x_j = p_j$. De ahí se obtienen la demanda de trabajo y la de cada bien intermedio, creciente en la productividad $A$ y el empleo $L$ y decreciente en el precio $p_j$, $x_j^d = L\,(\alpha A/p_j)^{1/(1-\alpha)}$. La demanda de cada intermedio es la isoelástica a la que se enfrenta cada inventor al fijar su precio.

▪ *CSO*: con $\alpha < 1$ la tecnología es estrictamente cóncava en $L$ y en cada $x_j$, de modo que las CPO definen las demandas de factores.

**Inventores**: incurren en un coste fijo $F$ para inventar una nueva variedad y obtienen a cambio una patente perpetua que les confiere el monopolio de su producción. Su toma de decisiones se analiza por inducción hacia atrás. Primero se determina el precio que fijarán una vez inventada la variedad y, después, si compensa incurrir en el coste de inventarla.

▪ *Planteamiento*: en la decisión de precio, el monopolista elige el precio que maximiza su beneficio por periodo frente a la demanda del productor final, $\max_{p_j}\,(p_j-1)\,x_j^d(p_j)$, con el coste marginal normalizado a uno. En la de participación, compara el valor de la patente con su coste, $F$.

▪ *CPO*: con una demanda isoelástica, de elasticidad $1/(1-\alpha)$, el monopolista fija su precio aplicando un margen (*mark-up*) fijo al coste marginal, $p_j = 1/\alpha > 1$. Como todas las variedades comparten la misma demanda y el mismo coste, tienen el mismo precio, se venden en igual cantidad y dejan idéntico beneficio.

▪ *CSO*: la demanda isoelástica, con elasticidad $1/(1-\alpha) > 1$, garantiza que el precio con margen constante maximiza el beneficio del monopolista.

▪ *Solución*: en la decisión de participación, con una patente que no caduca y un beneficio $\Pi$ por periodo constante y común a todas las variedades, inventar vale $\Pi/r$ en valor presente. Como la actividad inventora está abierta a cualquiera, se crean variedades hasta que ese valor coincide con el coste fijo, $\Pi/r = F$, es decir, hasta que la última variedad rinde, descontada, lo mismo que cuesta crearla[8].

**Consumidores**: optimizan a la Ramsey, como en el modelo de Ramsey-Cass-Koopmans (tema A43, apartado 3.2), con población constante.

▪ *Planteamiento*: maximizan la utilidad CRRA descontada, $\int_0^{\infty} e^{-\rho t}\,(c^{1-\theta}-1)/(1-\theta)\,dt$, sujetos a $c + \dot{a} = w + ra$ y a la condición de no-Ponzi, donde $a$ son los activos per cápita.

▪ *CPO*: dan la condición de Keynes-Ramsey estándar, $g_c = (1/\theta)(r-\rho)$, según la cual el consumo crece mientras el rendimiento del ahorro, que aquí financia las invenciones, supere a la impaciencia.

[8] Mientras $\Pi/r > F$, entran nuevos inventores y se acelera la creación de variedades. El proceso eleva la demanda de fondos y el tipo de interés de equilibrio hasta restaurar la igualdad $\Pi/r = F$.

- *CSO*: el hamiltoniano es estrictamente cóncavo en $c$ y lineal en $a$, y con la transversalidad las CPO son suficientes.

**Equilibrio**: sustituyendo $p_j = 1/\alpha$ en la demanda se obtiene $x_j^d = A^{1/(1-\alpha)}\alpha^{2/(1-\alpha)}L$ para toda variedad $j$. Introduciendo esta cantidad común en la función de producción, la producción agregada resulta proporcional al número de variedades:

$$Y = A^{1/(1-\alpha)}\alpha^{2\alpha/(1-\alpha)}L \cdot N \qquad \Rightarrow \qquad g_Y = g_N$$

La tasa de crecimiento del PIB coincide con la tasa de creación de variedades. Intuitivamente, al incorporar continuamente nuevas variedades de capital, la I+D evita que la productividad marginal del capital físico existente vaya disminuyendo. De este modo, la expansión horizontal del abanico de bienes intermedios es el mecanismo que genera el crecimiento endógeno y desempeña aquí el mismo papel que la linealidad de la tecnología en el modelo AK.

Combinando la condición de Keynes-Ramsey con la de libre entrada se obtiene la tasa de crecimiento de equilibrio, común a todas las variables:

$$g_Y^* = g_y^* = g_N^* = g_c^* = f(A^+, L^+, \theta^-, F^-, \rho^-, \alpha)$$

**Estática comparativa**: la tasa de crecimiento aumenta con la productividad ($A$) y el tamaño de la economía ($L$), y disminuye con el coste de inventar ($F$), la impaciencia ($\rho$) y, si es positiva, con el deseo de alisar el consumo ($\theta$).

**Dinámica**: como la producción es proporcional a $N$, no existe una fase de ajuste distinta del crecimiento equilibrado, y todas las variables avanzan desde el inicio al mismo ritmo, igual que en el AK.

**Extensiones**: en la formulación original de Romer el coste de inventar no es un coste fijo en bienes sino trabajo (capital humano) dedicado a la investigación. Esta versión se escribe con trabajo homogéneo y con una tecnología del bien final en la que, como antes, $\alpha$ es el exponente de cada intermedio:

$$Y = L_Y^{1-\alpha}\int_0^N x_i^\alpha \, di$$

donde $L_Y$ es el trabajo empleado en la producción del bien final, $x_i$ la cantidad de la variedad $i$ y $N$ el número de variedades. La versión coincide con el modelo de Romer (1990) con un único tipo de trabajo ($\beta = 0$, $H \to L$) y con el $\alpha$ de Romer igual a $1-\alpha$. La fuerza de trabajo se reparte entre producción e investigación, $L = L_Y + L_A$, y las nuevas variedades se producen con la *función de producción de ideas*

$$\dot{N} = \nu L_A N$$

lineal en el *stock* de conocimiento $N$, porque las ideas son no rivales y cada investigador se apoya en todas las existentes (*standing on the shoulders of giants*). La productividad de los investigadores es $\nu > 0$ ($\delta$ en Romer, 1990). La tasa de creación de variedades es entonces $g_N = \nu L_A$, y la asignación de trabajo se determina por el arbitraje salarial. En ese arbitraje, el salario en investigación es el valor de los diseños que un investigador produce, $w_A = \nu N P_N$ con $P_N = \Pi/r$, y debe igualar al salario en producción, $w_Y = (1-\alpha)Y/L_Y$. Con utilidad logarítmica ($\theta = 1$) el resultado es

$$L_A^* = \frac{L - \Lambda\rho/\nu}{1+\Lambda}, \qquad g_Y^* = \nu L_A^* = \frac{\nu L - \Lambda\rho}{1+\Lambda}, \qquad \Lambda \equiv \frac{1}{\alpha}$$

(Romer, 1990, ecuaciones 12 y 13, con $\beta = 0$), de modo que la tasa de crecimiento es proporcional al trabajo dedicado a investigar y crece con el tamaño de la economía $L$, lo que da la forma explícita del efecto escala, que se critica en el apartado 3.3. Ambas versiones, con coste fijo $F$ o con sector de I+D, comparten la estructura de tres sectores y la conclusión de que la producción es proporcional a $N$.

#### 3.1.3. Análisis normativo e implicaciones de política económica

El poder de mercado impide que el equilibrio descentralizado sea Pareto-eficiente. El margen que los inventores cargan sobre el coste marginal encarece los bienes intermedios, los productores finales usan menos de los que convendría a la sociedad y, con ello, se investiga menos y se crece por debajo del óptimo. Las soluciones se ordenan, de menor a mayor eficacia, como sigue.

1. *Solución de mercado*: sin intervención, es la que arroja el menor crecimiento.

2. *Subvención a la I+D*: al abaratar el coste efectivo $F$ de inventar, acelera la aparición de variedades, pero deja intacta la distorsión de las cantidades, pues el precio de monopolio sigue vigente y los productores finales continúan adquiriendo pocos intermedios.

3. *Subvención a la compra de bienes intermedios*: ataca la distorsión donde nace, porque el precio neto que pagan los productores finales baja hasta el competitivo. En la versión con coste fijo $F$, la subvención reproduce la asignación del planificador social y sitúa la economía en el óptimo (con sector de I+D subsiste, además, la externalidad del conocimiento sobre la investigación futura).

#### 3.1.4. Conclusiones

**Aspectos positivos**: el modelo de Romer (1990) es el primero que trata la innovación como el resultado de una inversión deliberada guiada por incentivos de mercado, integrando en equilibrio general la competencia imperfecta, las patentes y la no rivalidad de las ideas. Asimismo, sus implicaciones normativas (subvencionar la I+D y el uso de tecnología) han sido influyentes.

**Limitaciones**: su ajuste a la evidencia empírica estadounidense es pobre, porque predice un efecto escala (como Romer, 1986) que los datos refutan. Además, no predice convergencia y carece de dinámica de transición.

### 3.2. Enfoque schumpeteriano (Aghion y Howitt, 1992)

**Origen**: Philippe Aghion y Peter Howitt (1992), junto con Grossman y Helpman (1991), formalizaron la intuición de Schumpeter sobre la *destrucción creativa* con los instrumentos de la economía industrial moderna. Por esta línea de trabajo, Aghion y Howitt compartieron con Joel Mokyr el Premio Nobel de Economía de 2025, concedido "por haber explicado el crecimiento económico impulsado por la innovación".

**Idea**: a diferencia del modelo de Romer (1990), donde la innovación es horizontal (las nuevas variedades coexisten con las anteriores), en el enfoque schumpeteriano la innovación es vertical. La calidad de cada producto mejora escalón a escalón a lo largo de una "escalera de calidad" (*quality*

*ladder*). Cada innovación exitosa, además, vuelve obsoleta la generación anterior y destruye las rentas de monopolio del productor incumbente al que desplaza.

De esta estructura nace la tensión central del enfoque. La innovación crea valor (productos de mayor calidad, cuyo excedente se apropia en parte el consumidor), pero también destruye valor (las rentas del incumbente). Del balance de ambos efectos depende que la tasa de innovación de mercado sea inferior o superior a la socialmente óptima.

### 3.2.1. Supuestos

**Entorno**: economía cerrada, tiempo continuo, población constante $L$ y agentes neutrales al riesgo con tasa de descuento $r$ (de modo que el tipo de interés es constante e igual a $r$).

**Agentes**: como en Romer (1990), son de tres tipos. Los productores del bien final actúan en competencia perfecta, un productor del bien intermedio es monopolista mientras posee la tecnología de mejor calidad y los investigadores compiten por obtener la siguiente innovación.

**Tecnología**: el bien final se produce con el bien intermedio de la generación vigente $t$, $Y = A_t x^{\alpha}$, donde $A_t$ es la calidad de la generación $t$ y $x$ la cantidad de intermedio, producida uno a uno con trabajo. Cada innovación multiplica la calidad por $\gamma > 1$, $A_{t+1} = \gamma A_t$ (escalera de calidad). Si se dedican $L_A$ unidades de trabajo a investigar, las innovaciones llegan según un proceso de Poisson con intensidad $\lambda_I L_A$, de modo que el tiempo hasta la siguiente es aleatorio con media $1/(\lambda_I L_A)$. El innovador obtiene una patente que le confiere el monopolio del intermedio hasta que la siguiente innovación lo desplace. El trabajo se reparte entre producir el intermedio e investigar, $L = x + L_A$.

### 3.2.2. Modelo

**Monopolista del intermedio**: el problema se resuelve en dos pasos, y en el primero decide el monopolista de la generación $t$.

- *Planteamiento*: elige la cantidad de intermedio que maximiza su beneficio dado el salario $w_t$:

$$\max_x \ \Pi_t = p_t(x)\, x - w_t x, \qquad p_t(x) = \alpha A_t x^{\alpha-1}$$

donde $x$ es la cantidad de intermedio (la variable de decisión), $\Pi_t$ el beneficio del monopolista de la generación $t$, $p_t(x)$ la demanda inversa del intermedio, $w_t$ el salario, $A_t$ la calidad de la generación vigente y $\alpha$ el parámetro de la tecnología del bien final (la demanda del intermedio es la productividad marginal del sector final).

- *CPO*: del programa del monopolista, $\alpha^2 A_t x^{\alpha-1} = w_t$, de donde la cantidad y el beneficio son funciones decrecientes del salario en unidades de productividad, $\tilde{w}_t \equiv w_t/A_t$, es decir, $x = \tilde{x}(\tilde{w}_t)$ y $\Pi_t = A_t\, \tilde{\Pi}(\tilde{w}_t)$, con $\tilde{x}' < 0$, $\tilde{\Pi}' < 0$.

- *CSO*: el beneficio del monopolista es estrictamente cóncavo en $x$ para $\alpha < 1$.

**Investigadores**: en el segundo paso, deciden cuánto trabajo dedicar a la investigación comparando su valor esperado con el salario.

- *Condición de arbitraje*: iguala el salario al valor esperado de investigar:

$$w_t = \lambda_I V_{t+1}, \qquad V_{t+1} = \frac{\Pi_{t+1}}{r + \lambda_I L_{A,t+1}}$$

donde $V_{t+1}$ es el valor de la siguiente innovación, es decir, sus rentas de monopolio descontadas al tipo de interés más la tasa de llegada de la innovación que la destruirá. Dividiendo por $A_t$ y usando $A_{t+1} = \gamma A_t$, la condición se escribe en términos estacionarios:

$$\tilde{w}_t = \frac{\lambda_I \gamma \, \tilde{\Pi}(\tilde{w}_{t+1})}{r + \lambda_I L_{A,t+1}}$$

La condición es de equilibrio de libre entrada en la investigación (no de optimización) y, con $\tilde{\Pi}' < 0$, tiene solución única en el estado estacionario.

**Equilibrio**: en el estado estacionario $L_{A,t} = L_A$ y $\tilde{w}_t = \tilde{w}$ son constantes y se determinan con dos ecuaciones que se representan en el plano $(L_A, \tilde{w})$ (figura A44.9):

- *Arbitraje en investigación* (curva RA): $\tilde{w} = \lambda_I \gamma \tilde{\Pi}(\tilde{w})/(r + \lambda_I L_A)$, de pendiente negativa en el plano $(L_A, \tilde{w})$, porque cuanto más se investiga, más corta es la vida esperada de cada monopolio y menor el valor de innovar, de modo que el salario que la investigación puede pagar es menor.

- *Equilibrio del mercado de trabajo* (curva LL): $L = \tilde{x}(\tilde{w}) + L_A$, creciente, porque un salario mayor reduce la demanda de trabajo del monopolista y libera trabajo para la investigación.

Su intersección, única porque RA es decreciente y LL creciente, determina el esfuerzo investigador $L_A^*$ y, con él, la tasa media de crecimiento de la economía, que es el producto de la frecuencia de las innovaciones por su tamaño:

$$g_Y^* = \lambda_I L_A^* \ln \gamma$$

El crecimiento es estocástico (el logaritmo de la renta crece a saltos de tamaño $\ln \gamma$ en instantes aleatorios) y su tasa media es constante, de modo que se trata de crecimiento endógeno sostenido por la destrucción creativa.

**Estática comparativa**: un aumento de $\lambda_I$ o de $\gamma$ desplaza hacia arriba la curva RA, porque eleva, para cada $L_A$, el salario que la investigación puede pagar, y con ello eleva $L_A^*$ (de $\text{RA}_0$ a $\text{RA}_1$ y del punto 0 al 1). Un aumento de $L$ desplaza la curva LL hacia la derecha, libera trabajo para la investigación y eleva también $L_A^*$, mientras que un aumento de $r$ desplaza RA hacia abajo y reduce $L_A^*$.

Por tanto, la tasa de crecimiento crece con la productividad de la investigación ($\lambda_I$), con el tamaño de las innovaciones ($\gamma$) y con la oferta de trabajo ($L$, efecto escala). En cambio, decrece con el tipo de interés ($r$), que reduce el valor presente de las rentas futuras, y con la intensidad de la competencia de los rivales, porque una mayor capacidad de competir con el monopolista titular reduce $\tilde{\Pi}$ y el valor de innovar. Los determinantes institucionales operan a través de estos parámetros:

1. *Educación superior*: reduce el coste de investigar (eleva $\lambda_I$) y aumenta la capacidad de absorción de nuevas tecnologías.

2. *Protección de la propiedad intelectual*: una mayor protección de las patentes eleva los costes de imitación, lo que prolonga la vida esperada del monopolio y eleva $V$.

3. *Grado de competencia*: trabajos posteriores matizan la predicción básica de que la competencia frena la innovación. En particular, la evidencia para el Reino Unido documenta una relación en forma de U invertida entre competencia e innovación, en la que niveles moderados de competencia la estimulan (para "escapar de la competencia") mientras que niveles muy altos la desincentivan[9]. Sin embargo, estudios posteriores con otras muestras no siempre confirman esa relación.

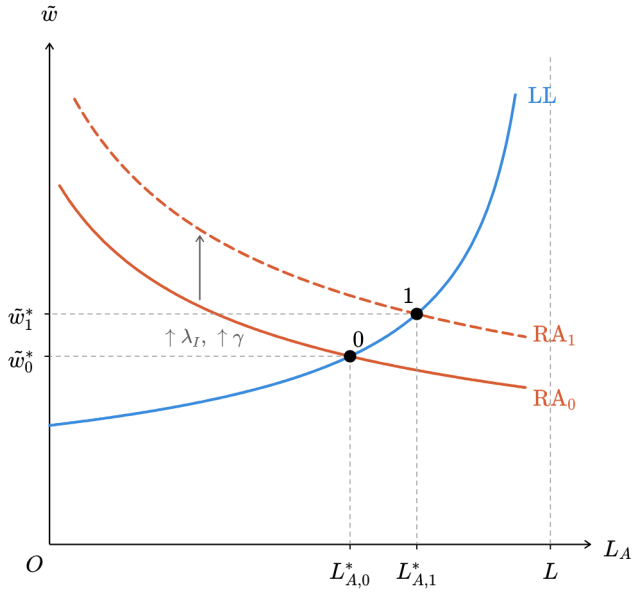



Figura A44.9: Modelo de Aghion y Howitt (1992). La curva de arbitraje en investigación, RA, decreciente, y la de equilibrio del mercado de trabajo, LL, creciente, fijan $L_A^*$ y $\tilde{w}^*$ (punto 0), y un aumento de $\lambda_I$ o de $\gamma$ desplaza RA hacia arriba y eleva $L_A^*$ (de 0 a 1). Un aumento de $L$ desplaza LL hacia la derecha y uno de $r$ desplaza RA hacia abajo.

**Extensiones**: el equilibrio descentralizado puede innovar de más o de menos respecto al planificador por la acción de tres externalidades:

- *Desbordamiento intertemporal*: cada innovación eleva la base sobre la que se construyen las siguientes, y el innovador no lo internaliza, lo que tiende a la subinversión.

- *Apropiación de negocio* (*business stealing*): el innovador no internaliza la pérdida de rentas que inflige al incumbente desplazado, lo que tiende a la sobreinversión.

- *Apropiabilidad incompleta*: el innovador no captura todo el excedente del consumidor, lo que tiende también a la subinversión.

El signo neto es ambiguo, a diferencia de la subinversión inequívoca del modelo de Romer (1990). El modelo schumpeteriano se ha extendido a muchos sectores y a la acumulación de capital, y Grossman y Helpman (1991) obtienen resultados análogos con escaleras de calidad en un continuo de bienes.

En política de innovación, en consecuencia, la subvención a la I+D solo es deseable si los efectos de desbordamiento y de apropiabilidad incompleta dominan al de *business stealing*. Por su parte, la política de competencia debe buscar el nivel intermedio de la U invertida, y la protección de

[9] Véase Aghion, Bloom, Blundell, Griffith y Howitt (2005).

patentes tiene un óptimo interior, porque una protección excesiva prolonga monopolios y frena la difusión.

#### 3.2.3. Conclusiones

**Aspectos positivos**: el enfoque schumpeteriano describe la innovación tal y como se observa en la realidad industrial (mejoras de calidad que desplazan a productos anteriores) e incorpora de forma natural la rotación empresarial y la obsolescencia. Genera, además, predicciones ricas y contrastables sobre la relación entre estructura de mercado e innovación.

**Limitaciones**: el resultado normativo es ambiguo y depende de la calibración. Además, las versiones básicas comparten el efecto escala y la ausencia de convergencia de los demás modelos de innovación.

### 3.3. La crítica del efecto escala y el crecimiento semiendógeno (Jones, 1995)

**Origen**: la crítica se debe a Charles Jones (1995) y se dirige a los modelos de innovación de primera generación.

**Idea**: en el modelo de Romer (1990) y en el de Aghion-Howitt, duplicar la población dedicada a investigar duplica la tasa de crecimiento, y en eso consiste el efecto escala. Sin embargo, el número de investigadores en las economías avanzadas se ha multiplicado desde 1950 sin que la tasa de crecimiento de la productividad haya aumentado, lo que refuta la predicción.

**Razonamiento**: la solución de Jones es suponer rendimientos decrecientes en la producción de ideas, $\dot{N} = \nu L_A N^{\phi}$, con $\nu$ la productividad de la investigación y $\phi < 1$, de modo que las ideas son cada vez más difíciles de encontrar. En ese caso la tasa de crecimiento de largo plazo es $g_y = n/(1-\phi)$, que depende solo de la tasa de crecimiento de la población, exógena, y no de la fracción de recursos dedicada a la investigación, que solo afecta al nivel de la renta.

El modelo es *semiendógeno*, porque la tasa de crecimiento se determina dentro del modelo pero no responde a la política económica, como en Solow. A diferencia de Jones, los modelos schumpeterianos de segunda generación (Young, 1998; Peretto, 1998; Aghion y Howitt, 1998; Dinopoulos y Thompson, 1998) eliminan el efecto escala suponiendo que el número de variedades crece con la población y diluye la investigación por producto. Así preservan la endogeneidad de la tasa de crecimiento respecto a la política.

El debate sobre el efecto escala es la principal frontera empírica de la teoría del crecimiento endógeno, ya que decide si las políticas de I+D elevan permanentemente la tasa de crecimiento o solo el nivel de renta.

**Conclusiones**:

- *Aspectos positivos*: la solución de Jones concilia los modelos de crecimiento basados en la I+D con una tasa de crecimiento estable pese al aumento del esfuerzo investigador, y conserva su mecanismo central, la acumulación de ideas no rivales. La hipótesis de que las ideas son cada vez más difíciles de encontrar ha recibido, además, apoyo empírico (tema A45, apartado 3.1.5).

- *Limitaciones*: el crecimiento de largo plazo vuelve a depender de una variable exógena, la tasa de crecimiento de la población, de modo que, si esta deja de crecer, el crecimiento per cápita tiende a desaparecer.

## Conclusiones

Los modelos de crecimiento endógeno dan cuenta de un aumento continuado de la renta per cápita sin apoyarse en un progreso técnico exógeno, y lo consiguen por tres vías. La primera impide que el rendimiento del capital en sentido amplio caiga hasta cero, con una tecnología lineal (AK), con una cota inferior a la productividad marginal (Sobelow), con externalidades o con gasto público productivo. La segunda, la de Uzawa y Lucas, hace del capital humano un factor acumulable y de la asignación del tiempo entre producir y aprender la decisión que fija el crecimiento. La tercera trata la innovación como una inversión deliberada, que exige competencia imperfecta y rentas protegidas por patentes, ya sea por aumento de variedades (Romer, 1990) o por mejoras de calidad que destruyen las rentas anteriores (Aghion y Howitt, 1992). Como las condiciones iniciales importan, las diferencias de renta pueden perpetuarse. Además, la política económica (educación, I+D, propiedad intelectual) afecta a la tasa de crecimiento y no solo al nivel. Frente a ello, la mayoría no predice la convergencia condicional, y el efecto escala que Jones (1995) refuta conduce a los modelos semiendógenos.

El crecimiento endógeno toma como punto de partida el ahorro óptimo de Ramsey y nace de los límites del modelo exógeno, que deja sin explicar el residuo de Solow (tema A43). Su contrastación, de la contabilidad del crecimiento a la desaceleración de la productividad, ha favorecido los modelos semiendógenos (tema A45). Por su parte, la disyuntiva entre la eficiencia dinámica que crean las patentes y la estática que destruyen reaparece en la justificación del Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio, el TRIPS (tema B34).

Bloom, Jones, Van Reenen y Webb (2020) documentan una caída sostenida de la productividad investigadora, que apoya la hipótesis semiendógena de que las ideas son cada vez más difíciles de encontrar. De esa hipótesis, Jones (2022) deduce que, si la población mundial deja de crecer, el crecimiento basado en ideas puede agotarse. El efecto de la inteligencia artificial está por determinar. Aghion, Jones y Jones (2019) muestran que automatizar la producción de ideas podría acelerar el crecimiento, aunque las tareas no automatizadas pueden frenarlo. Acemoglu (2025), en cambio, estima un aumento de la productividad total de los factores (PTF) de no más del 0,66 % en diez años.

## Anexo. La función de producción CES y la condición de Inada

**Origen**: Arrow, Chenery, Minhas y Solow (1961) introdujeron la función de elasticidad de sustitución constante (CES).

**Idea**: la Cobb-Douglas es el caso $\sigma = 1$ de la familia CES, $Y = A\left[\alpha K^{\psi} + (1-\alpha)L^{\psi}\right]^{1/\psi}$ con $\sigma = 1/(1-\psi)$. La elasticidad de sustitución decide si se cumplen las condiciones de Inada y, con ello, si el crecimiento se agota.

**Razonamiento**: en forma intensiva, la función se escribe $f(k) = A[\alpha k^{\psi} + (1-\alpha)]^{1/\psi}$. Se distinguen dos casos:

- $\sigma > 1$ *(*$0 < \psi < 1$*)*: la productividad marginal del capital tiende, cuando $k \to \infty$, a $A\alpha^{1/\psi} > 0$ y no a cero, de modo que se incumple la condición de Inada en el infinito. La curva de ahorro $sf(k)/k$ tiende a una constante positiva y, si $sA\alpha^{1/\psi} > n + \delta$, hay crecimiento endógeno con dinámica de transición, exactamente como en el modelo Sobelow, del que la CES es una versión suave.

- $\sigma < 1$ *(*$\psi < 0$*)*: la productividad marginal del capital tiende a cero, pero la del capital en el origen está acotada, $\lim_{k\to 0} f'(k) = A\alpha^{1/\psi}$, de modo que se incumple la condición de Inada en el origen. Si $sA\alpha^{1/\psi} < n + \delta$, no existe estado estacionario con capital positivo y la economía se vacía de capital desde cualquier punto de partida.

La elasticidad de sustitución gobierna, por tanto, la cota inferior de la productividad marginal del capital que exige el crecimiento endógeno (apartado 1.2).

**Conclusiones**:

- *Aspectos positivos*: generaliza la Cobb-Douglas sin imponer una elasticidad de sustitución unitaria y reúne en una sola familia el caso neoclásico y los dos incumplimientos de la condición de Inada. Ofrece así un marco único, y estimable, para discutir cuándo el crecimiento puede ser endógeno.

- *Limitaciones*: la evidencia sitúa $\sigma$ por debajo de la unidad en la mayoría de las estimaciones, lo que hace de la CES una fundamentación poco probable del crecimiento endógeno y sitúa la explicación en las ideas y el capital humano (temas A43 y A45).

# A45. Evidencia empírica sobre el crecimiento

Evidencia empírica sobre el crecimiento económico y la distribución de la renta entre los factores de producción. Contabilidad del crecimiento, con especial referencia a la productividad total de los factores. Convergencia económica internacional.

## Introducción

> **Introducción común.** Los temas A43 a A45 comparten una introducción, que se desarrolla al comienzo del tema **A43**. Define el crecimiento económico como el aumento sostenido de la producción por habitante y resume la evolución de su teoría, del modelo de Solow-Swan al crecimiento endógeno y la contrastación empírica.

Las teorías del crecimiento (temas A43 y A44) solo valen lo que valgan frente a los datos. Kaldor (1961) resumió el crecimiento de las economías avanzadas en seis **hechos estilizados**, entre ellos el crecimiento sostenido de la renta per cápita y la estabilidad de las participaciones del trabajo y del capital en la renta, que cualquier modelo debía reproducir. Desde entonces, la **contabilidad del crecimiento** descompone el crecimiento en la aportación de los factores y el residuo de Solow, y el análisis de la **convergencia** pregunta si los países pobres crecen más deprisa que los ricos.

La respuesta ha cambiado con los datos y los métodos, que hoy permiten comparar la renta de casi todos los países durante décadas con bases como la Penn World Table. Baumol (1986) encontró convergencia entre los países industrializados, pero no en el conjunto del mundo, y los trabajos de Barro y Sala-i-Martin (1992a) y de Mankiw, Romer y Weil (1992) mostraron que los países convergen hacia su propio estado estacionario (convergencia condicional) y no hacia uno común. La evidencia apunta también a **clubes de convergencia**, grupos de países que convergen entre sí según sus condiciones iniciales.

Piketty documentó el aumento de la desigualdad en las economías avanzadas desde los años ochenta y devolvió la distribución de la renta al centro del debate. Desde entonces, la participación del trabajo en la renta ha caído en muchas de ellas, lo que cuestiona uno de los hechos de Kaldor. China, que desde la reforma y apertura de 1978 creció a una media próxima al 10 % anual durante más de tres décadas, protagoniza el mayor episodio de convergencia de la historia económica, y su paso de la agricultura a la industria y los servicios ilustra la **transformación estructural** que acompaña al desarrollo.

Este tema repasa la evidencia empírica sobre el crecimiento económico:

- *Predicciones teóricas y contabilidad del crecimiento*: las del modelo neoclásico y la descomposición del crecimiento entre la acumulación de factores y el progreso técnico.

- *Herramientas*: las del análisis de la convergencia y de la distribución de la renta.

- *Evidencia empírica*: sobre el crecimiento, la convergencia, la distribución de la renta y la transformación estructural.

> Los modelos de crecimiento exógeno se desarrollan en el tema A43 y los de crecimiento endógeno, en el tema A44.

# 1. Predicciones teóricas y contabilidad del crecimiento

El modelo de Solow-Swan fija las hipótesis que contrasta el análisis empírico, y la contabilidad del crecimiento mide la aportación de los factores y el residuo que el modelo atribuye al progreso técnico, la productividad total de los factores.

## 1.1. Las predicciones del modelo de Solow-Swan

El punto de partida del estudio empírico del crecimiento es el modelo canónico de Solow-Swan (1956) con progreso tecnológico exógeno. El modelo considera un entorno sin imperfecciones en el que la función de producción agregada es de tipo Cobb-Douglas con progreso técnico aumentador del trabajo, $Y = K^{\alpha}(AL)^{1-\alpha}$, y en el que los agentes ahorran de acuerdo con una tasa de ahorro exógena y constante.

Definiendo el capital por unidad de trabajo efectivo como $\tilde{k} = K/AL$, la ecuación fundamental del modelo resulta:

$$\dot{\tilde{k}} = s\tilde{k}^{\alpha} - (g_A + n + \delta)\tilde{k}$$

donde $g_A$ es la tasa de progreso técnico, $n$ el crecimiento de la población y $\delta$ la depreciación. En el estado estacionario ($\dot{\tilde{k}} = 0$), las tasas de crecimiento de las variables en unidades de trabajo efectivo son nulas ($g^*_{\tilde{k}} = g^*_{\tilde{y}} = 0$) y las variables per cápita crecen al ritmo del progreso tecnológico ($g^*_k = g^*_y = g^*_c = g_A$).

De este resultado se desprenden cinco predicciones fundamentales, que constituyen las hipótesis a contrastar por el análisis empírico:

1. *Acumulación de capital*: determina en gran medida el nivel de renta per cápita, pero no explica su crecimiento a largo plazo, que solo puede proceder del progreso tecnológico exógeno.

2. *Automatismo*: cualquier economía llega por sí sola a su estado estacionario, sin que el sector público intervenga, aunque nada garantiza que el estado estacionario alcanzado coincida con el de la regla de oro.

3. *Alcance de la política económica*: la política económica puede situar la renta de largo plazo en un nivel más alto (efectos de nivel), pero no acelerar de forma permanente su crecimiento (los efectos sobre la tasa son solo transitorios, durante la transición entre estados estacionarios).

4. *Exogeneidad del progreso técnico*: bajo rendimientos constantes a escala y competencia perfecta, el teorema de Euler implica que pagar a cada factor su productividad marginal absorbe toda la producción, y nada sobra para costear desde dentro el progreso tecnológico. Por tanto, el progreso técnico ha de suponerse exógeno, y empíricamente se mide como residuo.

5. *Convergencia condicional*: el modelo anticipa que cada país se aproxima a su propio estado estacionario, no que el crecimiento de los pobres supere siempre al de los ricos (*convergencia absoluta*, apartado 2.1).

## 1.2. Contabilidad del crecimiento

La **contabilidad del crecimiento** mide empíricamente qué parte del aumento observado del PIB procede de una mayor dotación de factores productivos y qué parte del progreso técnico. Admite un enfoque primal, basado en cantidades, y otro dual, basado en precios. Se completa con los problemas de medición que comparten ambos, con la contabilidad del desarrollo, que compara países, y con la estimación econométrica de la función de producción.

### 1.2.1. Enfoque primal

**Versión con capital físico** (Solow, 1957): se parte de una función de producción $Y = Af(K, L)$ con rendimientos constantes a escala, en la que $A$ es la productividad total de los factores (PTF), neutral en el sentido de Hicks. La tasa de crecimiento de la PTF se denota $g_{\text{PTF}}$ y, con una Cobb-Douglas, es $1 - \alpha$ veces la del progreso técnico aumentador del trabajo. Tomando logaritmos y derivando respecto al tiempo se obtiene la descomposición fundamental:

$$g_Y = g_{\text{PTF}} + \alpha g_K + (1 - \alpha) g_L$$

donde $\alpha = A\,(\partial f/\partial K)\,K/Y$ y $(1 - \alpha) = A\,(\partial f/\partial L)\,L/Y$ son, bajo competencia perfecta, las participaciones de las rentas del capital y del trabajo en la renta nacional, observables en la contabilidad nacional (en las economías avanzadas, $\alpha$ se sitúa en torno a un tercio). El *residuo de Solow*, es decir, la tasa de crecimiento de la PTF, se obtiene entonces como diferencia:

$$g_{\text{PTF}} = g_Y - \alpha g_K - (1 - \alpha) g_L$$

Precisamente porque se obtiene como residuo, la PTF debe interpretarse con cautela, pues es "una especie de medida de nuestra ignorancia" (Abramovitz, 1956) y puede recoger no solo el progreso tecnológico en sentido estricto, sino también factores organizativos, institucionales, culturales o simples errores de medición. Con datos de EE. UU. entre 1909 y 1949, Solow (1957) atribuyó al residuo el 87,5 % del crecimiento del producto por hora trabajada.

**Versión con capital humano** (Jorgenson y Griliches, 1967): una extensión natural sustituye el trabajo bruto $L$ por el capital humano $H = hL$, que ajusta la cantidad de trabajo por su calidad. La participación del capital humano se aproxima por el peso de la masa salarial en la renta nacional, y el residuo pasa a ser $g_{\text{PTF}} = g_Y - \alpha g_K - (1 - \alpha) g_H$. Al imputar a los factores parte de lo que antes era residuo (la mejora de la calidad del trabajo), esta corrección tiende a reducir la contribución estimada de la PTF.

### 1.2.2. Enfoque dual

El enfoque dual parte de la identidad de la renta bajo rendimientos constantes, $Y \equiv rK + wH$. Diferenciando y reordenando se obtiene una expresión del residuo en función del crecimiento de los precios de los factores:

$$g_{\text{PTF}} = \alpha g_r + (1 - \alpha) g_w$$

donde $g_r$ y $g_w$ son las tasas de crecimiento del rendimiento del capital y del salario. La intuición es que, con rendimientos decrecientes, acumular un factor abarata su remuneración, de modo que una subida duradera y simultánea de ambos precios delata progreso técnico. Así, si el crecimiento

se debiera únicamente a la acumulación de capital, los rendimientos decrecientes reducirían $r$ y elevarían $w$, mientras que si se debiera únicamente a la acumulación de capital humano ocurriría lo contrario. La ventaja práctica del enfoque, además, es que la medición de los precios de los factores (tipos de interés, salarios) suele ser más sencilla y fiable que la de sus cantidades (*stock* de capital).

**Discrepancias primal-dual**: si la medición contable es correcta y la tecnología presenta rendimientos constantes, ambos enfoques deberían arrojar el mismo resultado. Sin embargo, en el "milagro" de los tigres asiáticos difieren radicalmente (apartado 3.1.2), y las causas de la discrepancia se han buscado en la medición del *stock* de capital y en el tratamiento de los impuestos sobre el capital.

### 1.2.3. Problemas de medición

La contabilidad del crecimiento es tan fiable como la medición de sus insumos, y la del capital es la más problemática. El *stock* de capital físico no se observa, sino que se construye por el método del inventario permanente, $K_t = (1-\delta)K_{t-1} + I_t$, acumulando la inversión a precios constantes y restando una depreciación supuesta. Depende, por tanto, de la tasa de depreciación elegida y del *stock* de partida, $K_0$, el punto débil del método, aunque su peso se pierde con el paso del tiempo. Además, la mejora de la calidad de los bienes de capital (un ordenador de hoy no es el de 1990) obliga a deflactar la inversión con índices de precios hedónicos, y la contribución del capital cambia sustancialmente según se haga o no.

La heterogeneidad de los bienes de capital, muy distintos entre sí, complica su agregación en una sola magnitud, problema que alimentó la controversia de Cambridge (Joan Robinson). En la práctica se usan dos índices, el de volumen de servicios de capital (IVSC), ponderado con el coste de arrendamiento, y el índice capital-riqueza (ICR), ponderado con el de adquisición. El *servicio del capital* tampoco coincide con el *stock*, porque el grado de utilización varía con el ciclo, lo que hace que la PTF medida sea procíclica aunque la tecnología no lo sea.

En el trabajo, las horas trabajadas son la medida idónea, aunque su evolución mezcla varios componentes (participación laboral, paro y jornada por trabajador). El capital humano se aproxima por los años de escolarización, que ignoran la experiencia y la calidad de la educación, distinta en cada país, o por el peso de la masa salarial, que confunde calidad del trabajo con poder de negociación.

Las participaciones factoriales se toman de la contabilidad nacional, que exige repartir la renta mixta de los autónomos entre trabajo y capital y supone competencia perfecta para identificar participaciones con elasticidades. Cada una de esas decisiones desplaza crecimiento entre los factores y el residuo, lo que explica que las estimaciones de la contribución de la PTF varíen tanto entre estudios.

### 1.2.4. Contabilidad del desarrollo (Hall y Jones, 1999; Caselli, 2005)

**Origen**: Robert Hall y Charles Jones (1999) desarrollaron la contabilidad del desarrollo, y Francesco Caselli (2005) la sintetizó.

**Idea**: la contabilidad del crecimiento descompone la variación de la renta de un país en el tiempo. La *contabilidad del desarrollo* (*development accounting*), en cambio, aplica la misma lógica a las diferencias de renta entre países en un momento dado. Mide qué parte de la brecha entre un país rico y uno pobre se debe a que el primero tiene más capital físico y humano por trabajador, y qué parte a que los usa con mayor productividad.

**Razonamiento**: con $Y = K^{\alpha}(AH)^{1-\alpha}$ y $H = hL$, donde $h$ es el capital humano por trabajador (obtenido de los años de escolarización con rendimientos mincerianos), la renta por trabajador se escribe

$$y = \left(\frac{K}{Y}\right)^{\alpha/(1-\alpha)} h\,A$$

Hall y Jones usan la ratio capital-producto en lugar del capital por trabajador porque en el estado estacionario aquella es proporcional a la tasa de inversión, $K/Y = s/(n + g_A + \delta)$, y no depende de $A$. Así se evita atribuir al capital lo que es consecuencia de la productividad. Con datos de 127 países, la ratio de renta por trabajador entre los cinco países más ricos y los cinco más pobres es de 31,7. Esa ratio se descompone en un factor 1,8 de capital físico, 2,2 de capital humano y 8,3 de productividad (Hall y Jones, 1999, cuadro V). Por tanto, la PTF explica la mayor parte de la brecha.

Caselli (2005) formaliza la medida de éxito de los factores, success $= \mathrm{var}(\ln y_{KH})/\mathrm{var}(\ln y)$, donde $y_{KH}$ es la renta que se observaría si todos los países tuvieran la misma $A$, y la sitúa en torno a 0,4. Dicho de otro modo, los factores explican el 40 % de la dispersión de la renta y la PTF, el 60 %.

**Conclusiones**:

- *Aspectos positivos*: el resultado desplaza la pregunta del crecimiento desde la acumulación hacia la productividad, y esta hacia sus determinantes últimos (instituciones, tecnología, asignación de recursos entre empresas).

- *Limitaciones*: la PTF sigue siendo un residuo, como en la contabilidad del crecimiento, y el reparto es sensible a la elasticidad de sustitución supuesta y a la medición del capital humano, que, como muestra Caselli, pueden alterar sustancialmente el reparto.

### 1.2.5. Enfoque econométrico

**Origen**: la estimación econométrica de funciones de producción agregadas se remonta a Cobb y Douglas (1928) y a la estimación de la función CES de Arrow, Chenery, Minhas y Solow (1961), mientras que su versión moderna utiliza datos de empresa.

**Idea**: la contabilidad del crecimiento impone las participaciones factoriales observadas como elasticidades, lo que exige competencia perfecta y rendimientos constantes. El enfoque econométrico, sin embargo, estima la función de producción directamente, $\ln Y_t = \ln A_0 + g_{\mathrm{PTF}} t + \alpha \ln K_t + \alpha_L \ln L_t + \varepsilon_t$, de modo que las elasticidades y los rendimientos a escala son resultados y no supuestos, y el progreso técnico se identifica como tendencia.

**Razonamiento**: la estimación permite contrastar la hipótesis $\alpha + \alpha_L = 1$ y comparar la elasticidad estimada del capital con su participación en la renta. Una elasticidad superior a la participación es evidencia de rendimientos crecientes o de externalidades del capital, o bien de poder de mercado, pues con un margen sobre el coste marginal las participaciones en la renta infravaloran las elasticidades (Hall, 1988).

**Conclusiones**:

▪ *Aspectos positivos*: el enfoque econométrico es el complemento necesario de la contabilidad del crecimiento cuando los supuestos de esta son dudosos. Además, con datos de empresa se ha convertido en la vía principal para medir la dispersión de productividad y la mala asignación de recursos (*misallocation*, Hsieh y Klenow, 2009), que explica una parte sustancial de las diferencias de PTF entre países.

▪ *Limitaciones*: sus problemas son tres:

- *Endogeneidad de los factores*: las empresas eligen los factores conociendo su productividad, lo que sesga las estimaciones. Para corregirlo, la literatura ha propuesto estimadores de función de control con datos de empresa.
- *Colinealidad*: la del capital con la tendencia.
- *Forma funcional*: los resultados son sensibles a la especificación elegida (Cobb-Douglas frente a la función de elasticidad de sustitución constante, CES, o la translog).

### 1.2.6. Conclusiones

**Aspectos positivos**: la contabilidad del crecimiento es una técnica sencilla, transparente y con requisitos de información relativamente modestos, que proporciona descomposiciones comparables entre países y periodos. Constituye el primer paso natural de cualquier análisis empírico del crecimiento, ya que permite acotar dónde hay que buscar sus fuentes (en la acumulación o en la productividad) antes de investigar las causas últimas.

**Limitaciones**:

▪ *Medición de los factores*: el residuo hereda los problemas de medición (*stock* y heterogeneidad del capital, horas y calidad del trabajo).

▪ *Variables omitidas*: no existe consenso sobre si los factores incluidos son los únicos relevantes (tierra, recursos naturales, capital público, capital organizativo).

▪ *Ausencia de causalidad*: la contabilidad identifica las fuentes *próximas* del crecimiento, pero no sus causas *fundamentales*. Constatar que un país creció acumulando capital, por ejemplo, no explica por qué acumuló capital. Esta limitación motiva los estudios de determinantes profundos del apartado 3.1.3.

## 1.3. Conclusiones

La contabilidad del crecimiento separa la aportación de los factores de la de la PTF, pero esta se obtiene como residuo y hereda los errores de medición del capital, del trabajo y de las participaciones factoriales. En las diferencias de renta entre países, la contabilidad del desarrollo atribuye a la PTF la mayor parte de la brecha, y el enfoque econométrico contrasta los supuestos que la contabilidad impone. Ninguna de esas vías identifica, sin embargo, las causas fundamentales del crecimiento. El apartado 3.1 aplica la contabilidad a la evidencia sobre el crecimiento.

# 2. Herramientas para el análisis de la convergencia y la distribución

El análisis empírico del crecimiento se completa con dos herramientas: el análisis de convergencia y el de la distribución de la renta.

## 2.1. Análisis de convergencia

El análisis de convergencia nació para poner a prueba unas familias de modelos de crecimiento frente a otras. Los modelos neoclásicos, con rendimientos marginales decrecientes del capital, predicen convergencia (condicional), mientras que los endógenos de tipo AK, con rendimientos constantes del capital ampliado, no predicen ninguna. Se distinguen varios conceptos de convergencia.

**Convergencia en el modelo de Solow-Swan**: el concepto puede ilustrarse gráficamente a partir del modelo neoclásico, partiendo de la tasa de crecimiento del capital per cápita, $g_k = sf(k)/k - (n + \delta)$.

Gráficamente, la convergencia se representa en el plano capital per cápita-tasa de crecimiento, $k$-$g_k$ (figuras A45.1 y A45.2):

1. *Curva de ahorro por unidad de capital*, $s\,f(k)/k$: recoge la inversión bruta por unidad de capital y es decreciente, por los rendimientos decrecientes del capital. Con tasas de ahorro distintas hay una curva por país, $s^R f(k)/k$ y $s^P f(k)/k$.

2. *Recta de depreciación*, $n+\delta$: recoge la inversión por unidad de capital necesaria para mantener $k$ constante y es horizontal.

3. *Equilibrio*: el corte de ambas determina el estado estacionario, $k^*$ (punto 0 en la figura A45.1), o $k^{*P}$ y $k^{*R}$ si las curvas difieren. La distancia vertical entre ellas es la tasa de crecimiento de cada economía, $g_k^P$ para el país pobre, que parte de $k_0^P$, y $g_k^R$ para el rico, que parte de $k_0^R$.

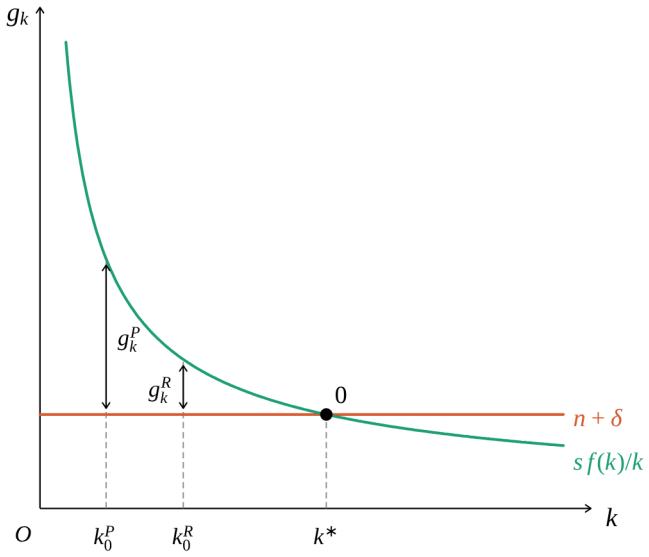



Figura A45.1: Convergencia absoluta. La tasa de crecimiento está correlacionada negativamente con el nivel inicial de renta per cápita.

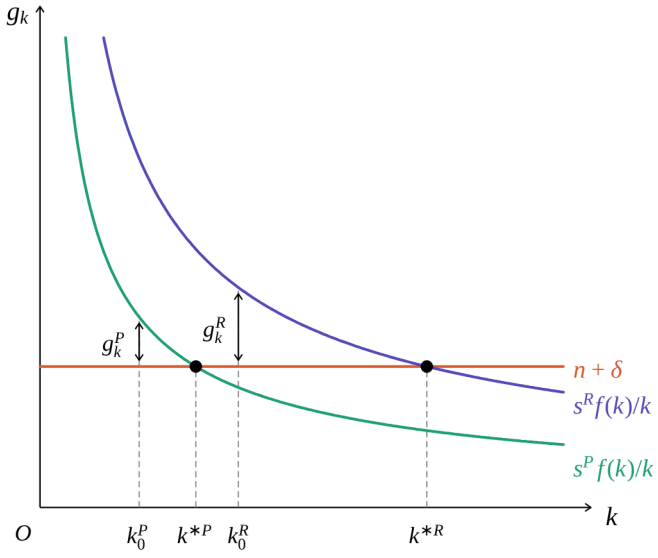



Figura A45.2: Convergencia condicional. Cada economía converge hacia su propio estado estacionario, determinado por sus parámetros.

Caben dos casos:

1. *Parámetros iguales*: si las economías comparten los mismos parámetros estructurales ($s$, $n$, $\delta$, $\alpha$ y el nivel tecnológico), comparten también las mismas curvas y el mismo estado estacionario. En tal caso, cuanto menor sea el *stock* de capital inicial, mayor será la tasa de crecimiento, de modo que los países pobres crecen más rápido y se produce $\beta$-convergencia absoluta (figura A45.1). Así ocurrió tras la Segunda Guerra Mundial con Alemania y Japón. Ambos países diferían de los aliados esencialmente en su *stock* de capital, destruido por la guerra, y crecieron sistemáticamente más rápido.

2. *Parámetros distintos*: basta con que un parámetro varíe de una economía a otra (p. ej. la tasa de ahorro) para que sus estados estacionarios se separen, y entonces el más pobre no tiene por qué crecer más. En la figura A45.2, el país rico, todavía lejos de su estado estacionario, puede crecer más que el pobre, ya próximo al suyo. De ahí que la hipótesis relevante sea la convergencia condicional (apartado 3.2), una correlación parcial negativa entre crecimiento y renta inicial, una vez se controla por el estado estacionario de cada país.

**$\beta$-convergencia**: recoge la relación inversa entre el nivel inicial de renta y la tasa de crecimiento posterior, que es mayor cuanto mayor es la distancia al estado estacionario. Su contraste empírico habitual es la regresión:

$$g_y^j = \beta_0 - \beta \ln y_0^j + \gamma Z^j + \varepsilon^j$$

donde $y_0^j$ es la renta inicial del país $j$ y $Z^j$ un vector de variables estructurales que aproximan su estado estacionario. Un coeficiente $\beta$ positivo indica convergencia. Ha de ser, además, menor que $1/T$ si $g_y^j$ es la tasa media anual de un periodo de $T$ años, para evitar adelantamientos sistemáticos. La especificación determina el tipo de convergencia contrastada:

- *Sin controles* ($Z = 0$): se contrasta la convergencia absoluta.
- *Con controles* $Z$: se contrasta la convergencia condicional.

Alternativamente, puede estimarse la versión no lineal del contraste, en la que $\beta = (1 - e^{-\lambda T})/T$, lo que permite obtener directamente la velocidad de convergencia $\lambda$.

**$\sigma$-convergencia**: exige que la dispersión de la renta per cápita entre países disminuya con el tiempo, $\sigma_t > \sigma_{t+T}$. La relación entre ambos conceptos es asimétrica, ya que la $\beta$-convergencia absoluta es condición necesaria pero no suficiente para la $\sigma$-convergencia. La razón es que un país pobre puede "adelantar" al rico y seguir creciendo más deprisa, en cuyo caso habría $\beta$-convergencia (el inicialmente pobre creció más) y, sin embargo, la dispersión podría estar aumentando.

**$\gamma$-convergencia**: mide la movilidad ordinal mediante la varianza de los cambios de posición en la clasificación (*ranking*) de rentas (índice de concordancia de Kendall). Si los países cambian sus posiciones relativas a lo largo del tiempo, existe convergencia en términos de orden, es decir, la jerarquía internacional de rentas no es fija.

**Velocidad de convergencia**: $\lambda$ mide el ritmo al que una economía cierra la brecha con su estado estacionario. En el modelo de Solow-Swan, la aproximación log-lineal en torno al estado estacionario arroja:

$$\lambda \approx (1 - \alpha)(g_A + n + \delta)$$

El resultado tiene una lectura empírica importante. Si se toma $\alpha \approx 1/3$ y valores plausibles para el resto de parámetros, el modelo básico arroja una velocidad cercana al 4 % anual, muy por

encima del 2 % que miden, con notable regularidad, los estudios empíricos (Barro y Sala-i-Martin, 1992a). Sin embargo, el desajuste desaparece en la versión con capital humano de Mankiw, Romer y Weil (1992). Si el capital incluye el físico y el humano, su participación conjunta ronda dos tercios y la velocidad teórica baja al entorno del 2 % anual que se observa. A ese ritmo, cerrar la mitad de la distancia al estado estacionario lleva unos 35 años ($t_{1/2} = \ln 2/\lambda$), lo que hace de la convergencia un proceso lento.

**Problemas econométricos**: las regresiones de convergencia de corte transversal presentan tres dificultades:

- *Heterogeneidad no observada*: si los determinantes del estado estacionario omitidos de $Z$ están correlacionados con la renta inicial, la estimación de $\beta$ está sesgada. De hecho, las estimaciones con datos de panel y efectos fijos por país, que permiten a cada economía un estado estacionario propio, arrojan velocidades de convergencia sensiblemente superiores al 2 % (Islam, 1995).

- *Endogeneidad de los regresores*: la inversión o la escolarización responden al propio crecimiento, lo que obliga a recurrir a variables instrumentales.

- *Falacia de Galton* (Quah, 1993): un coeficiente $\beta$ positivo es compatible con una dispersión constante o creciente, por simple regresión a la media. En consecuencia, la $\beta$-convergencia debe completarse con el análisis de la $\sigma$-convergencia y de la dinámica de toda la distribución.

## 2.2. Análisis de la distribución personal de la renta

> La medición de la desigualdad y sus fundamentos en la economía del bienestar se desarrollan en el tema A24, apartado 1.4, y su uso en la medición del desarrollo, la pobreza y la desigualdad, en el tema B28, apartado 1.2.

El análisis empírico del crecimiento se completa con el de la distribución de la renta, pues el mismo crecimiento agregado es compatible con evoluciones distributivas muy distintas. Para medirla se usan, sobre todo, cinco instrumentos.

**Curva de Lorenz** (Lorenz, 1905): representa la proporción acumulada de la renta total (eje de ordenadas) que percibe cada proporción acumulada de la población, ordenada de menor a mayor renta (eje de abscisas). La diagonal de 45 grados representa la equidistribución perfecta, y cuanto más se aleja la curva de la diagonal, mayor es la desigualdad. Su principal limitación es que solo proporciona una ordenación parcial, porque cuando dos curvas de Lorenz se cruzan no permite decir qué distribución es más desigual.

**Coeficiente o índice de Gini** (Gini, 1912): sintetiza la curva de Lorenz en un escalar, definido como el cociente entre el área comprendida entre la curva y la diagonal y el área total bajo la diagonal. Toma valores entre 0 (equidistribución) y 1 (concentración total de la renta en un individuo). El índice es la medida más utilizada, si bien, cuando las curvas de Lorenz se cruzan, dos distribuciones muy distintas pueden arrojar el mismo Gini. Además, el índice es más sensible a las transferencias en el centro de la distribución que en las colas.

**Índice de Theil** (Theil, 1967): medida basada en el concepto de entropía de la teoría de la información, $T = \frac{1}{H}\sum_{h=1}^{H} \frac{y^h}{\bar{y}} \ln \frac{y^h}{\bar{y}}$, donde $H$ es el número de individuos, $y^h$ la renta del individuo $h$ y $\bar{y}$ la renta media. Su gran ventaja es la *descomponibilidad aditiva*, es decir, la desigualdad total puede descomponerse exactamente en la suma de la desigualdad dentro de los grupos y la desigualdad entre grupos (p. ej. entre regiones o entre niveles educativos). Esa propiedad resulta especialmente útil para el análisis de las fuentes de la desigualdad.

**Índice de Atkinson** (Atkinson, 1970): a diferencia de las medidas anteriores, puramente estadísticas, es una medida normativa, construida a partir de una función de bienestar social con un parámetro explícito de aversión a la desigualdad ($\varepsilon$). Se define como $A_\varepsilon = 1 - y_e/\bar{y}$, donde $y_e$ es la renta "equivalente igualmente distribuida", aquella que, repartida por igual, generaría el mismo bienestar social que la distribución observada. Se interpreta como la fracción de la renta total que la sociedad estaría dispuesta a sacrificar a cambio de la equidistribución.

**Cuentas nacionales distributivas** (Piketty, Saez y Zucman, 2018): constituyen la frontera reciente de la medición. Frente a las encuestas de hogares, que infrarrepresentan las rentas altas y no agotan la renta nacional, las *Distributional National Accounts* asignan el 100 % de la renta nacional de la contabilidad nacional a los individuos. Garantizan así la coherencia entre los agregados macroeconómicos y los datos distributivos, permiten el análisis de la distribución antes y después de impuestos y transferencias, y son la base metodológica del *Informe sobre la Desigualdad Mundial* (*World Inequality Report*).

Gráficamente, la curva de Lorenz y el coeficiente de Gini se representan en el plano proporción acumulada de la población-proporción acumulada de la renta, $p$-$L(p)$ (figura A45.3):

1. *Diagonal de equidistribución* (45°): recoge el caso en que cada proporción de la población percibe la misma proporción de la renta.

2. *Curva de Lorenz*, $L(p)$: recoge la proporción acumulada de la renta que percibe cada proporción acumulada $p$ de la población. Es convexa y queda por debajo de la diagonal porque la población se ordena de menor a mayor renta.

3. *Coeficiente de Gini*: $G = A/(A + B)$, donde $A$ es el área entre la diagonal y la curva y $B$, el área bajo la curva.

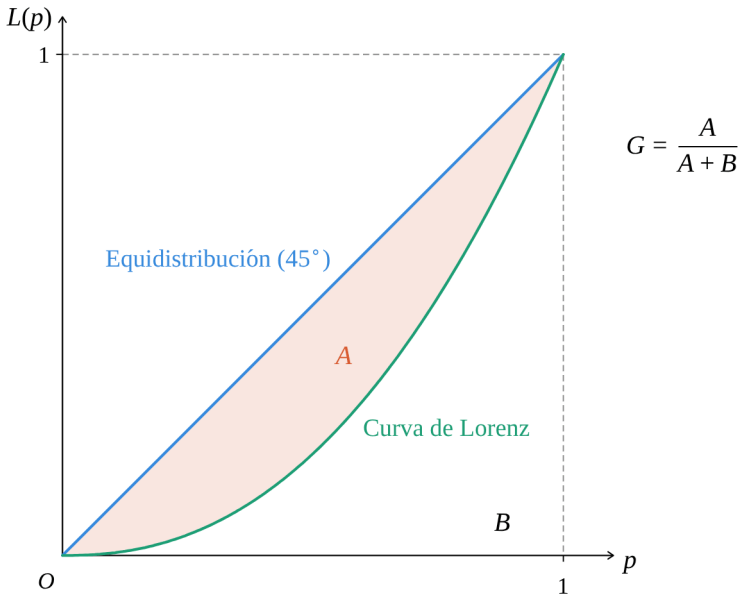



Figura A45.3: Curva de Lorenz y coeficiente de Gini como medidas de la desigualdad en la distribución de la renta.

## 2.3. Conclusiones

El análisis de convergencia traduce en contrastes las predicciones del modelo de Solow-Swan. La $\beta$-convergencia es condicional o absoluta según se controle o no por el estado estacionario, y debe completarse con la $\sigma$-convergencia, porque un $\beta$ positivo es compatible con una dispersión creciente. Las medidas de desigualdad van, por su parte, de la ordenación parcial de la curva de Lorenz a la descomposición del índice de Theil y al juicio normativo del de Atkinson. La sección 3 aplica estas herramientas a la evidencia sobre la convergencia y la distribución de la renta.

# 3. Evidencia empírica

La evidencia empírica se ordena en cinco ámbitos: el crecimiento y sus fuentes, la convergencia, la distribución personal de la renta, la distribución factorial y la transformación estructural.

## 3.1. Evidencia empírica sobre el crecimiento

La evidencia sobre el crecimiento abarca los hechos estilizados, la contribución de la PTF y de la acumulación de factores, los determinantes profundos, el enfoque institucional y la desaceleración de la productividad.

### 3.1.1. Estudios generales: hechos estilizados

Kaldor (1961) sintetizó la evidencia disponible en seis regularidades empíricas, entendidas como promedios históricos y no como constantes válidas en todo momento:

1. *Renta per cápita*: su tasa de crecimiento es aproximadamente constante (en EE. UU., en torno al 2 % anual).

2. *Capital per cápita*: su tasa de crecimiento es aproximadamente constante.

3. *Ratio capital-producto*: $K/Y$ es aproximadamente constante (lo que se deriva de las dos anteriores si ambas tasas de crecimiento coinciden).

4. *Tasa de retorno*: la del capital es aproximadamente constante.

5. *Diferencias entre países*: las tasas de crecimiento de la productividad difieren de forma importante y persistente (del orden del 2 % al 5 %).

6. *Participaciones factoriales*: las del capital y del trabajo en la renta nacional son aproximadamente constantes.

Los hechos primero a cuarto y sexto son coherentes con el modelo neoclásico con progreso técnico exógeno y tecnología Cobb-Douglas, y de hecho contribuyeron a fijarlo como modelo canónico. El quinto, en cambio, solo encaja en él como fenómeno transitorio, asociado a la dinámica de transición hacia estados estacionarios distintos. El sexto es hoy el más controvertido, porque desde los años ochenta, y con especial intensidad desde comienzos de la década de 2000, la participación del trabajo ha caído de forma apreciable en EE. UU. y otros países avanzados.

Jones y Romer (2010) actualizaron la lista con seis "nuevos hechos de Kaldor", adaptados a las teorías modernas del crecimiento:

1. *Extensión del mercado*: ha aumentado mucho (globalización, urbanización).

2. *Aceleración del crecimiento*: durante milenios, el crecimiento de la población y de la renta per cápita se ha acelerado, desde tasas casi nulas hasta las del último siglo.

3. *Variabilidad de las tasas*: la dispersión de las tasas de crecimiento entre países aumenta con la distancia a la frontera tecnológica.

4. *Papel de la PTF*: la acumulación de factores (capital físico y humano) explica menos de la mitad de las diferencias de renta per cápita entre países, por lo que la PTF explica la mayor parte.

5. *Capital humano*: su dotación por trabajador ha aumentado con fuerza.

6. *Premio salarial*: ese aumento del capital humano relativo no ha ido acompañado de una caída de su precio relativo (el premio salarial de la cualificación no ha caído), lo que sugiere una demanda creciente de cualificación sesgada por la tecnología.

Los tres primeros hechos motivan los modelos de crecimiento endógeno basados en ideas (tema A44), y el cuarto resume los resultados de la contabilidad del desarrollo (apartado 1.2.4).

### 3.1.2. Estudios sobre la PTF y la acumulación de factores

La cuestión de si el crecimiento se debe principalmente a la acumulación de factores o a la PTF dista de estar resuelta, y los resultados dependen de la metodología, la muestra y el periodo.

**Los factores dominan**: tres estudios clásicos atribuyen a la acumulación de factores la mayor parte del crecimiento o de las diferencias de renta entre países:

- *Denison (1985)*: la mayor parte del crecimiento de EE. UU. durante el periodo 1929-1982.

- *Maddison (1991)*: una conclusión similar (en torno a dos tercios) para seis países avanzados (Alemania, EE. UU., Francia, Japón, Países Bajos y Reino Unido) entre 1913 y 1987.

- *Mankiw, Romer y Weil (1992)*: el modelo de Solow con capital físico explica en torno al 60 % de las diferencias internacionales de renta ($R^2 = 0{,}6$), proporción que se eleva a cerca del 80 % cuando se amplía con capital humano ($R^2 = 0{,}8$).

**La crítica de Jorgenson y Griliches** (1967): sostuvieron que el residuo de Solow es en gran medida un artefacto de la mala medición de los factores. Corrigiendo el capital por su utilización y por la calidad de los bienes de equipo, y el trabajo por su composición (educación, edad, sexo), la contribución de la PTF al crecimiento estadounidense (1945-1965) se reducía de casi la mitad a apenas un 3 %. La réplica de Denison y el debate posterior, sin embargo, fijaron la posición intermedia que hoy se acepta, según la cual una parte sustancial del residuo es medición, pero no toda.

**El debate sobre los tigres asiáticos**: Alwyn Young (1995) mostró que el crecimiento de Hong Kong, Singapur, Corea y Taiwán entre 1966 y 1990 se explicaba sobre todo por la acumulación de factores (inversión, participación laboral, educación). El crecimiento de la PTF no era excepcional (casi nulo en Singapur). Este resultado, anticipado en trabajos previos del propio Young, había llevado a Krugman (1994) a comparar el crecimiento de los tigres asiáticos con el soviético y a predecir el agotamiento del crecimiento asiático.

La réplica llegó desde el enfoque dual (apartado 1.2.2), cuya PTF, muy superior a la primal, sugería que las series de capital de Young sobrestimaban la acumulación. Young (1998) sostiene, en cambio, que, corregidos los errores de cálculo y de método, las estimaciones duales no se alejan tanto de las primales. El episodio ilustra por qué los dos enfoques son complementarios y por qué la pregunta "acumulación o PTF" sigue abierta.

**La PTF domina**: otros tantos estudios llegan a la conclusión contraria:

- *Solow (1957)*: el residuo del 87,5 % (apartado 1.2.1).

- *Hall y Jones (1999)*: la PTF explica la mayor parte de las diferencias internacionales de renta. La discrepancia con Mankiw, Romer y Weil se debe a la medición del capital humano. Mankiw, Romer y Weil usan la tasa de escolarización secundaria como aproximación de la tasa de inversión en capital humano (un flujo), con un efecto muy pronunciado sobre la renta. Hall y Jones, en cambio, convierten los años de escolarización (un *stock*) mediante los rendimientos salariales de la educación (ecuación de Mincer), que arrojan diferencias de capital humano mucho menores entre países.

- *Caselli (2005)*: el reparto entre factores y PTF de la contabilidad del desarrollo, con una parte de los factores mayor dentro de la OCDE (en torno al 60 %) que fuera de ella (algo más de un tercio).

**Las economías en convergencia**: en las economías europeas que se aproximaron a la frontera tecnológica en las décadas posteriores a la Segunda Guerra Mundial, la contribución de la PTF al crecimiento fue sensiblemente superior a la de EE. UU., reflejo de un intenso proceso de modernización y absorción de tecnología.

### 3.1.3. Estudios sobre determinantes profundos

Una segunda línea de investigación abandona la contabilidad y busca directamente los determinantes del crecimiento mediante regresiones de corte transversal entre países:

1. *Primeras regresiones*: los trabajos fundacionales de esta literatura, que condicionan a la renta inicial, encuentran como factores positivamente correlacionados con el crecimiento la escolarización primaria y secundaria y la tasa de inversión. Como factores negativos encuentran el consumo público (que no debe confundirse con la inversión pública), la inflación, las distorsiones del precio relativo de los bienes de inversión, el crecimiento de la población y la inestabilidad política (revoluciones, golpes de Estado, asesinatos políticos).

2. *Límites extremos*: Levine y Renelt (1992) aplican el *análisis de límites extremos* (*extreme bounds analysis*) de Leamer, según el cual una variable solo es robusta si su coeficiente mantiene el signo y la significatividad para cualquier combinación de variables de control. Con este criterio, la inmensa mayoría de las más de cincuenta variables identificadas por la literatura como "determinantes" del crecimiento no son robustas a cambios de especificación, y solo la renta inicial y la tasa de inversión sobreviven al test. En consecuencia, no cabe depositar demasiada confianza en las regresiones de crecimiento para identificar determinantes.

3. *Dos millones de regresiones*: Sala-i-Martin (1997) propone un criterio menos exigente que el todo-o-nada de Levine-Renelt, basado en la distribución completa de los coeficientes estimados. Con él, en torno a una veintena de las casi sesenta variables examinadas aparecen sólida y sistemáticamente correlacionadas con el crecimiento (entre ellas, variables geográficas, institucionales y de apertura).

### 3.1.4. El enfoque institucional del crecimiento

La búsqueda de las causas fundamentales (frente a las próximas) culmina en el enfoque institucional, cuyo trabajo de referencia es el de Acemoglu, Johnson y Robinson (2001). La dificultad central de esta literatura es la endogeneidad, porque los países ricos pueden permitirse buenas instituciones, de modo que la correlación entre instituciones y renta no prueba causalidad. Ante esa dificultad, la aportación de los autores es una estrategia de identificación mediante variables instrumentales. La ecuación estimada es del tipo:

$$\ln y_i = \beta_0 + \beta_1 R_i + \mathbf{X}_i'\gamma + \varepsilon_i$$

donde $y_i$ es la renta per cápita actual del país $i$ y $R_i$ es un indicador de calidad institucional (la protección frente al riesgo de expropiación), que se instrumenta con la mortalidad de los colonos europeos en los siglos XVII-XIX. La lógica del instrumento es histórica. Donde la mortalidad era alta, los europeos no se asentaron y establecieron instituciones *extractivas*, diseñadas para expoliar recursos, mientras que donde era baja se asentaron y replicaron instituciones *inclusivas*, con protección de la propiedad. Esas instituciones persistieron tras la independencia y determinan la renta actual.

Los resultados atribuyen a las instituciones un efecto causal de primer orden sobre la renta per cápita. El trabajo posterior de los mismos autores (la "reversión de la fortuna", 2002) refuerza la conclusión, al mostrar que las zonas colonizadas más ricas en 1500 son hoy relativamente más pobres. Ello es incompatible con explicaciones basadas en efectos geográficos invariables en el tiempo y apunta a las instituciones como causa fundamental del desarrollo.

### 3.1.5. La desaceleración de la productividad

**Origen**: la *paradoja de la productividad* de Solow (1987), "la era del ordenador se ve en todas partes menos en las estadísticas de productividad", y el debate sobre el *estancamiento secular* reabierto por Gordon (2016) y Summers (2014).

**Idea**: el crecimiento de la PTF en las economías avanzadas no ha sido constante. Fue elevado entre 1920 y 1970 (en torno al 2 % anual en EE. UU.) y cayó hasta aproximadamente un tercio de ese ritmo entre 1973 y 1995. Repuntó transitoriamente entre 1995 y 2004 con la difusión de las tecnologías de la información y la comunicación (TIC). Desde mediados de la década de 2000 ha vuelto a desacelerarse, hasta tasas del orden del 0,5 % anual en EE. UU. y próximas a cero en buena parte de Europa (Gordon, 2016). De ahí que el primer hecho de Kaldor (crecimiento constante de la renta per cápita) sea, visto a esta escala, una aproximación.

**Razonamiento**: las explicaciones se agrupan en tres:

- *Pesimista* (Gordon): las grandes invenciones de la segunda revolución industrial (electricidad, motor de combustión, saneamiento) tuvieron efectos únicos que las TIC no igualan, y vientos de cara (demografía, educación estancada, desigualdad, deuda) reducen el crecimiento tendencial.

- *Medición*: la producción digital, gratuita o de calidad creciente, está infravalorada en el PIB, aunque los principales estudios concluyen que la mala medición no explica la mayor parte de la desaceleración, que es común a países con distinta intensidad digital.

- *Difusión y asignación*: la productividad de las empresas de la frontera ha seguido creciendo mientras la del resto se estanca, lo que apunta a fallos de difusión, a la caída del dinamismo empresarial y a la concentración. En la misma línea, las ideas parecen cada vez más difíciles de encontrar (Bloom, Jones, Van Reenen y Webb, 2020), de modo que mantener la tasa de crecimiento de la productividad exige un esfuerzo investigador cada vez mayor.

La desaceleración es el hecho estilizado que la teoría del crecimiento debe explicar hoy. Sus lecturas tienen implicaciones opuestas. Así, si es tecnológica y permanente, las políticas de oferta tienen poco recorrido, mientras que, si es de difusión y asignación, la inversión pública en I+D, la formación y la política de competencia pueden revertirla. La evidencia contra el efecto escala de los modelos de primera generación (Jones, 1995) y la caída de la productividad

investigadora favorecen los modelos semiendógenos, en los que sostener el crecimiento exige un esfuerzo investigador creciente (tema A44).

**Conclusiones**:

▪ *Aspectos positivos*: el debate muestra que el crecimiento tendencial de la productividad no es constante y ordena sus posibles causas en tres hipótesis contrastables, de las que se siguen recomendaciones de política opuestas.

▪ *Limitaciones*: las tres explicaciones no se excluyen y la evidencia no permite aún ponderarlas. Además, el diagnóstico descansa en la PTF, un residuo sensible a los problemas de medición del apartado 1.2.3.

## 3.2. Evidencia empírica sobre convergencia

La evidencia sobre la convergencia se resume en los puntos siguientes:

1. *La ausencia de convergencia mundial*: en los datos de la *Penn World Table* (Summers y Heston, 1991), tomado el mundo en su conjunto, los países pobres no crecen sistemáticamente más que los ricos, y no aparece convergencia $\beta$ ni $\sigma$. Esa falta de convergencia se leyó al principio como un argumento a favor del crecimiento endógeno.

2. *La contrarrevolución neoclásica* (Barro, Sala-i-Martin y Mankiw-Romer-Weil, 1990-1992): replicó que la predicción del modelo neoclásico había sido siempre la convergencia condicional. A su juicio, contrastar la absoluta equivale a ignorar que cada país tiene sus propios parámetros estructurales y, por tanto, su propio estado estacionario.

▪ *Mankiw, Romer y Weil* (1992): sin controles, el $\beta$ estimado es positivo (0,34, con el crecimiento acumulado de 1960-1985 como variable dependiente, no la tasa media anual) y significativo entre los 22 países de la OCDE. En la muestra amplia de 98 países (1960-1985), en cambio, no hay convergencia absoluta, con un $\beta$ incluso ligeramente negativo ($-0{,}094$, con la convención de signo de la regresión del apartado 2.1). La semejanza estructural de los países de la OCDE explica la diferencia. Cuando se añaden como controles la escolarización, la tasa de inversión y el crecimiento demográfico, el coeficiente es positivo y significativo en las tres muestras (las dos anteriores y una intermedia de 75 países). La convergencia condicional avanza entonces a un 2 % anual, aproximadamente, la velocidad coherente con el modelo de Solow con capital humano.

▪ *Barro y Sala-i-Martin* (1991-1992): recurren a las regiones de un mismo país, cuyos parámetros (ahorro, demografía, tecnología, gobierno e instituciones) son muy parecidos, lo que permite contrastar la convergencia absoluta en condiciones casi de laboratorio. En 73 regiones europeas, los estados de EE. UU., los *Länder* alemanes, las prefecturas japonesas y las comunidades autónomas españolas, todas las muestras arrojan $\beta$-convergencia absoluta, y lo hacen a un ritmo sorprendentemente parecido, cercano al 2 % anual (unos 35 años para recortar a la mitad la brecha). Tanta uniformidad ha convertido ese 2 % en poco menos que un hecho estilizado.

3. *Sesgo de selección y clubes de convergencia*:

▪ *Sesgo de selección* (De Long, 1988): la convergencia que Baumol (1986) había encontrado entre los países hoy ricos era un artefacto de selección, pues la muestra incluía solo a los que habían acabado convergiendo. De hecho, con una muestra definida *ex ante* por la renta de 1870, la convergencia desaparece.

- *Clubes de convergencia* (Quah, 1993, 1996): el análisis de la dinámica de la distribución va más lejos. La distribución mundial de la renta per cápita no se comprime hacia un centro, sino que tiende a una forma bimodal, los *picos gemelos*, con un club de países ricos y otro de pobres y un centro que se vacía. Ello sugiere *convergencia de clubes*, es decir, los países convergen hacia el estado estacionario de su grupo y no hacia uno común, porque el estado estacionario depende no solo de sus parámetros estructurales sino también de sus condiciones iniciales (p. ej. su capital humano). El resultado es coherente con los modelos de trampas de pobreza y de umbrales (tema B28, apartado 2.1.2).

4. *La convergencia desde 2000*: la evidencia reciente ha cambiado el diagnóstico. Desde finales de los años noventa los países en desarrollo han crecido sistemáticamente más que los ricos, con lo que la convergencia absoluta, ausente en las muestras de 1960-1990, tiende a aparecer desde 1990 y es clara desde 2000 (Kremer, Willis y You, 2022). Las instituciones y las políticas de los países pobres han convergido también hacia las de los ricos. El resultado sugiere que la no convergencia de las décadas anteriores fue un fenómeno del periodo y no una propiedad del crecimiento.

5. *El caso de China*: constituye el mayor episodio de convergencia de la historia económica. Desde la reforma y apertura de 1978, la economía china creció a una media próxima al 10 % anual durante más de tres décadas, con el ingreso en la OMC (2001) como punto de inflexión de su integración comercial. En la década de 2010 se moderó hasta el entorno del 7 %. Su PIB agregado medido en paridad de poder adquisitivo supera al de EE. UU. desde mediados de esa década. Su PIB per cápita en paridad de poder adquisitivo, en cambio, se situaba en 2025 en torno a un tercio del estadounidense (y en torno a la sexta parte a tipos de cambio de mercado). Por tanto, la convergencia, aunque muy intensa, dista de haberse completado.

### 3.3. Evidencia empírica sobre distribución de la renta: desigualdad y crecimiento

La evidencia sobre la distribución personal de la renta se ocupa de la relación entre desigualdad y crecimiento, en ambos sentidos, y del nivel y la evolución de la desigualdad.

**Curva de Kuznets** (1955): la hipótesis clásica sobre la relación entre crecimiento y desigualdad sostiene que esta describe una U invertida a lo largo del desarrollo. Aumenta en las primeras fases, cuando la población se desplaza de la agricultura de baja productividad y renta homogénea a la industria de alta productividad y renta dispersa. Disminuye después, cuando la mayoría ya está en el sector moderno y operan la educación y la redistribución.

La evidencia de corte transversal de los años sesenta y setenta apoyó la hipótesis, pero la de series temporales largas la ha desmentido. Así, Piketty (2014) muestra que la caída de la desigualdad entre 1914 y 1970 en los países ricos se debió a las guerras, la inflación y la fiscalidad, no a una ley del desarrollo. Muestra también que desde 1980 la desigualdad ha vuelto a crecer sin que el desarrollo se detuviera. Milanovic (2016) propone en su lugar *ondas de Kuznets* sucesivas, impulsadas por las revoluciones tecnológicas y la globalización.

**Desigualdad global** (Lakner y Milanovic, 2016): la *curva del elefante* representa el crecimiento de la renta real entre 1988 y 2008 para cada percentil de la distribución mundial. Muestra ganancias elevadas para las clases medias de los países emergentes (Asia) y estancamiento para las clases medias y bajas de los países ricos (el "valle" de la trompa). Para el 1 % superior mundial (la punta de la trompa), las ganancias son también elevadas, aunque inferiores a las del centro de la distribución. La desigualdad entre países, ponderada por población, ha caído por la convergencia de China e India, mientras que la desigualdad dentro de los países ha aumentado.

El resultado neto es una ligera reducción de la desigualdad global desde 2000, la primera desde la Revolución Industrial.

**Desigualdad y crecimiento**: la causalidad inversa, de la desigualdad al crecimiento, se ha estudiado con regresiones de corte transversal. Los principales resultados son los siguientes:

- *Efecto negativo*: Alesina y Rodrik (1994) y Persson y Tabellini (1994) lo encuentran y lo explican por la economía política (mayor desigualdad, mayor demanda de redistribución distorsionante). Otros trabajos (Galor y Zeira, 1993) añaden las restricciones de crédito que impiden a los pobres invertir en capital humano.

- *Efecto dependiente del desarrollo*: Barro (2000) matiza que el efecto es negativo en los países pobres y positivo en los ricos.

- *Duración del crecimiento*: otros estudios muestran que la desigualdad acorta la duración de los episodios de crecimiento.

La evidencia es frágil por los problemas de medición de la desigualdad y de identificación, pero el consenso actual es que la desigualdad elevada no favorece el crecimiento.

**Estado de la desigualdad**: el *Informe sobre la Desigualdad Mundial 2022* (Chancel, Piketty y otros) ofreció una descripción muy completa de la desigualdad contemporánea. La mitad inferior de la distribución obtiene menos del 15 % del ingreso nacional (con la excepción de Europa) y apenas el 2 % de la riqueza, mientras que el 10 % superior concentra más de la mitad del ingreso mundial (52 %) y el 76 % de la riqueza. Oriente Medio y el Norte de África (MENA) es la región más desigual del mundo. Los niveles actuales de desigualdad se aproximan a los de comienzos del s. XX. La edición de 2026 confirma esta descripción en lo esencial, pues el 10 % superior percibe más renta que el 90 % restante y posee tres cuartas partes de la riqueza mundial.

En paralelo, durante los últimos cuarenta años la riqueza del sector privado ha crecido mientras la del sector público caía. La pandemia de COVID-19 agravó esta tendencia, y durante ella los gobiernos se endeudaron con el sector privado por magnitudes del orden del 10-20 % del PIB.

Según Piketty, la ratio riqueza-renta ha aumentado sin que cayera la tasa de retorno del capital. Ello, a su juicio, eleva la participación del capital en la renta nacional, igual al producto de ambas (la "primera ley fundamental del capitalismo"), y explica el crecimiento de la desigualdad desde el final del sistema de Bretton Woods. A largo plazo, la ratio riqueza-renta tiende al cociente entre la tasa de ahorro neto y la tasa de crecimiento (la "segunda ley"), de modo que un menor crecimiento la eleva para un ahorro dado. Por último, si el rendimiento del capital supera de forma persistente al crecimiento de la renta ($r > g_Y$), la riqueza heredada tiende a crecer más deprisa que la renta y a concentrarse, que es la tesis central de Piketty. Esa acumulación de riqueza se ha visto alimentada, según el autor, por el incremento del valor inmobiliario, la competencia por bienes posicionales y la ventaja fiscal del capital sobre el trabajo.

## 3.4. Evidencia empírica sobre la distribución factorial: la participación del trabajo

**Origen**: el sexto hecho estilizado de Kaldor (1961) y la literatura que documenta su quiebra desde los años ochenta.

**Idea**: la distribución de la renta entre los factores es distinta de la distribución personal, pues mide qué fracción de la renta nacional remunera al trabajo y qué fracción al capital. La estabilidad de esa fracción, en torno a dos tercios para el trabajo en las economías avanzadas, es

el hecho estilizado de Kaldor que la función Cobb-Douglas reproduce y que la contabilidad del crecimiento usa como ponderación.

**Razonamiento**: la evidencia de las últimas cuatro décadas ha puesto en cuestión esa constancia. La participación del trabajo ha caído en la mayoría de las economías avanzadas y emergentes desde principios de los años ochenta, en torno a cinco puntos del PIB en el conjunto de la OCDE, y la literatura ha propuesto las siguientes explicaciones complementarias:

- *Abaratamiento del capital*: una primera explicación atribuye la caída al menor precio relativo de los bienes de capital, singularmente los ligados a las TIC. Con una elasticidad de sustitución entre capital y trabajo superior a la unidad, ese menor precio induce la sustitución de trabajo por capital (Karabarbounis y Neiman, 2014).

- *Deslocalización*: para EE. UU., la caída se ha atribuido a la deslocalización de las fases intensivas en trabajo.

- *Empresas superestrella*: Autor, Dorn, Katz, Patterson y Van Reenen (2020) la atribuyen a la concentración del mercado en empresas *superestrella* de márgenes altos y participación del trabajo baja.

- *Otros factores*: otros trabajos apuntan al debilitamiento del poder de negociación de los trabajadores y a la medición. En cuanto a la medición, una parte de la caída obedece a cómo se imputan la renta mixta y la de la vivienda, y a la capitalización de los productos de la propiedad intelectual, antes tratados como gasto intermedio (Koh, Santaeulàlia-Llopis y Zheng, 2020). El FMI y la Organización Internacional del Trabajo (OIT) documentan la caída con datos comparables desde 1990, con la excepción parcial de algunos países europeos.

La caída de la participación del trabajo conecta la teoría del crecimiento con la de la desigualdad. Si la renta del capital está más concentrada que la del trabajo, un desplazamiento de la distribución factorial hacia el capital eleva la desigualdad personal, como en el mecanismo de Piketty. Además, la caída obliga a revisar dos supuestos de la teoría, la elasticidad de sustitución unitaria de la Cobb-Douglas y la competencia perfecta que identifica participaciones con elasticidades.

**Conclusiones**:

- *Aspectos positivos*: la literatura, desde Karabarbounis y Neiman (2014), documenta en la mayoría de los países y sectores la quiebra de un hecho estilizado que la teoría daba por constante. Además, la literatura conecta esa quiebra con la desigualdad personal y con los supuestos de la función de producción.

- *Limitaciones*: las explicaciones son complementarias y la evidencia no permite aún ponderarlas.

## 3.5. Evidencia empírica sobre la transformación estructural

**Origen**: Kuznets (1966) y Chenery y Syrquin (1975) desarrollaron el estudio empírico de la transformación estructural.

**Idea**: el marco neoclásico es unisectorial y no captura la *transformación estructural*, la reasignación de la actividad económica entre sectores que acompaña al crecimiento y presenta con él una fuerte correlación. Este proceso opera a través de dos canales:

- *Demanda*: el patrón de consumo cambia con la renta, y las preferencias se sesgan primero hacia los productos industriales y después hacia los servicios. Capturar este efecto exige funciones de utilidad no homotéticas, en las que los consumos relativos dependan del nivel de renta.

- *Oferta*: el progreso técnico es diferenciado, es decir, algunos sectores se vuelven más productivos que otros a lo largo del tiempo. Con demandas poco sustituibles, el empleo se desplaza hacia los sectores de productividad estancada (la *enfermedad de los costes* de Baumol, 1967).

**Razonamiento**: desde Kuznets (1966) y Chenery y Syrquin (1975), la evidencia empírica muestra tres patrones regulares a lo largo del desarrollo:

- *Agricultura*: su peso en el empleo y el valor añadido decrece con el PIB per cápita (ley de Engel, según la cual la elasticidad-renta de la demanda de alimentos es inferior a la unidad).

- *Manufacturas*: su peso tiene forma de U invertida (*hump-shaped*), creciendo en las primeras fases del desarrollo y decayendo después. Rodrik (2016) documenta que el máximo se alcanza hoy a niveles de renta y con pesos del empleo menores que en los países que se industrializaron antes (*desindustrialización prematura*).

- *Servicios*: su peso aumenta de manera sostenida y, en particular, el del sector público crece con la renta (ley de Wagner).

La transformación estructural es a la vez causa y consecuencia del crecimiento. Por una parte, la reasignación del trabajo desde la agricultura hacia sectores de mayor productividad es una fuente de crecimiento agregado. En esa línea, McMillan y Rodrik (2011) concluyen que el cambio estructural explica una parte sustancial del crecimiento asiático y que ha actuado como freno en América Latina y África, donde el trabajo se ha desplazado hacia servicios informales de baja productividad. Por otra, la enfermedad de los costes de Baumol implica que, a medida que los servicios de productividad estancada absorben el empleo, la tasa de crecimiento agregado tiende a reducirse, lo que enlaza con la desaceleración de la productividad.

**Conclusiones**:

- *Aspectos positivos*: los tres patrones son regularidades sólidas, comunes a países y periodos, y explican parte de las diferencias de crecimiento entre regiones, algo que el marco unisectorial no captura.

- *Limitaciones*: los patrones son correlaciones entre el peso de cada sector y la renta per cápita, y la evidencia no permite separar con precisión el canal de la demanda del de la oferta. Además, la productividad de los servicios es difícil de medir, lo que condiciona la lectura de la enfermedad de los costes.

## Conclusiones

El crecimiento económico procede tanto de la acumulación de factores como de la mejora de la productividad, pero la contribución de cada fuente es difícil de precisar por los problemas de medición. Los enfoques primal y dual arrojan resultados distintos según la metodología y el periodo, y los estudios de determinantes son frágiles ante cambios de especificación (Levine y Renelt, 1992). Aun así, de esta literatura se extrae una lección de política económica, pues una buena política no garantiza más crecimiento, pero una mala (inestabilidad, inflación, consumo público improductivo) sí lo daña con claridad. La contabilidad del desarrollo y el enfoque institucional han desplazado además el foco hacia la productividad y sus causas fundamentales. La convergencia condicional es lenta, en torno al 2 % anual, y la absoluta a escala mundial solo aparece desde finales de los años noventa. Por último, ni la curva de Kuznets ni la constancia de las participaciones factoriales de Kaldor resisten la evidencia de las últimas décadas.

Las hipótesis que contrasta la literatura empírica proceden del modelo de Solow-Swan y de su versión ampliada con capital humano (tema A43). A su vez, la evidencia ha orientado la teoría, porque la ausencia de convergencia absoluta y el peso de la PTF motivaron los modelos de crecimiento endógeno, y la caída de la productividad investigadora favorece hoy los semiendógenos (tema A44). Por su parte, los clubes de convergencia son la versión empírica de las trampas de pobreza de la economía del desarrollo (tema B28), y el índice de Atkinson descansa en un juicio sobre la aversión a la desigualdad, el de la función de bienestar social (tema A24).

La desaceleración de la productividad tiene explicaciones rivales. Gordon (2016) la atribuye al agotamiento de las grandes invenciones, mientras que la creciente dificultad de encontrar ideas (Bloom, Jones, Van Reenen y Webb, 2020) y los fallos de difusión dejan margen a la política de innovación y de competencia. Sigue abierta también la relación entre convergencia y desigualdad. Desde finales de los años noventa las economías emergentes crecen más que las avanzadas (Kremer, Willis y You, 2022) y la desigualdad global ha empezado a caer (Lakner y Milanovic, 2016). Dentro de los países, en cambio, la concentración de la riqueza (Piketty, 2014) y la caída de la participación del trabajo (Autor et al., 2020) plantean si el crecimiento futuro será compartido.