Lecciones de economía general e internacional

Lecciones de economía, vol. 1 · Comercio Internacional

Arturo Lorenzo

Edición web 1.0 · noviembre de 2026

Tabla de contenidos

Nota **9**

B5. Teoría del comercio internacional **10**

Introducción . . . 10

1. Antecedentes y fundamentos de la teoría neoclásica: teorías clásicas sobre el comercio internacional . . . 12

1.1. Teoría de la ventaja absoluta (Smith, 1776) . . . 12

1.2. Teoría de la ventaja comparativa (Ricardo, 1817) . . . 13

1.3. Teoría de las demandas recíprocas (Mill, 1848) . . . 15

1.4. Conclusiones . . . 15

2. Teoría neoclásica del comercio internacional de bienes y servicios . . . 17

2.1. Determinación de la relación real de intercambio: las curvas de oferta neta (Marshall, 1879; Edgeworth, 1894) . . . 17

2.2. Enfoque de equilibrio general: la reformulación de la ventaja comparativa (Haberler, 1936) . . . 21

2.3. Enfoque de equilibrio general: las ventajas dotacionales (Heckscher, 1919; Ohlin, 1933; Stolper y Samuelson, 1941) . . . 24

2.4. Teoremas derivados del modelo HOS . . . 27

2.5. Modelo de factores específicos (Samuelson, 1971, y Jones, 1971) . . . 31

2.6. Conclusiones . . . 34

3. Revisiones de la teoría tradicional: levantamiento de supuestos . . . 35

3.1. Revisiones relativas al entorno: competencia imperfecta y distorsiones domésticas . . . 35

3.2. Revisiones por el lado de la oferta: bienes y tecnología . . . 36

3.3. Revisiones por el lado de la oferta: factores . . . 38

3.4. Revisiones por el lado de la demanda . . . 39

Conclusiones . . . 41

Anexos . . . 42

A.1. Demostraciones formales de los teoremas de la teoría neoclásica . . . 42

A.2. Formalización del modelo ricardiano . . . 44

A.3. Modelo ricardiano con un continuo de bienes (Dornbusch, Fischer y Samuelson, 1977) . . . 44

A.4. Reversión de las intensidades factoriales (Minhas, 1962) . . . 46

B6. Nueva teoría del comercio internacional **48**

Introducción . . . 48

1. Fundamentos de la nueva teoría . . . 50

1.1. Punto de vista empírico . . . 50

1.2. Punto de vista teórico . . . 52

1.3. Conclusiones . . . 53

2. Nueva teoría del comercio internacional . . . 55

2.1. Modelos de competencia monopolística en economía abierta . . . 55

2.2. Modelos de oligopolio en economía abierta . . . 64

2.3. Equilibrio integrado, modelos gravitatorios y nueva economía geográfica . . . 68

2.4. Conclusiones . . . 72
3. Novísima teoría del comercio internacional . . . 73
3.1. Modelo de competencia monopolística y heterogeneidad de empresas (Melitz, 2003) . . . 73
3.2. Extensiones al modelo de heterogeneidad de empresas . . . 77
3.3. Otras aportaciones de la novísima teoría . . . 80
3.4. Conclusiones . . . 81
Conclusiones . . . 82

B7. Política comercial I: instrumentos y efectos **83**
Introducción . . . 83
1. Justificación de la política comercial . . . 85
1.1. Argumentos en favor del librecambismo . . . 85
1.2. Argumentos en favor del proteccionismo . . . 85
1.3. Conclusiones . . . 87
2. Barreras arancelarias: el arancel como instrumento y sus efectos . . . 89
2.1. El arancel como instrumento . . . 89
2.2. Análisis en equilibrio parcial: efectos sobre la industria del bien importable . . . 92
2.3. Análisis en equilibrio general: efectos sobre la economía en su conjunto . . . 96
2.4. Conclusiones . . . 106
3. Barreras no arancelarias y otros instrumentos tradicionales . . . 108
3.1. Instrumentos tradicionales distintos del arancel . . . 108
3.2. Otras barreras no arancelarias . . . 114
3.3. Soluciones a las barreras no arancelarias . . . 117
3.4. Conclusiones . . . 117
Conclusiones . . . 119
Anexo. Arancel en un país pequeño en el plano de los bienes . . . 120

B8. Política comercial II: política estratégica **122**
Introducción . . . 122
1. Justificación de la política comercial moderna . . . 124
1.1. Política comercial estratégica . . . 124
1.2. Política de promoción de las exportaciones . . . 125
1.3. Conclusiones . . . 126
2. La política comercial estratégica: objetivos, instrumentos y efectos . . . 127
2.1. Objetivos e instrumentos . . . 127
2.2. Vertiente ofensiva . . . 127
2.3. Vertiente defensiva . . . 135
2.4. Vertiente de la I+D . . . 141
2.5. Conclusiones . . . 143
3. La política de promoción exterior: objetivos, instrumentos y efectos . . . 144
3.1. Objetivos, instrumentos e instituciones . . . 144
3.2. Efectos de la promoción de exportaciones: el modelo de Melitz (2003) . . . 146
Conclusiones . . . 148

B9. Comercio internacional y crecimiento **149**
Introducción . . . 149
1. Efectos del crecimiento económico sobre el comercio internacional . . . 151
1.1. Supuestos . . . 151
1.2. País pequeño . . . 151
1.3. País grande . . . 155
1.4. Efectos en función de la causa del crecimiento económico . . . 159

1.5. Conclusiones . . . 161
2. Efectos del comercio sobre el crecimiento (I): de los clásicos a la novísima teoría . . 162
2.1. Pensamiento clásico y neoclásico . . . 162
2.2. Nueva teoría . . . 163
2.3. Novísima teoría . . . 166
2.4. Conclusiones . . . 168
3. Efectos del comercio sobre el crecimiento (II): crecimiento endógeno y evidencia empírica . . . 169
3.1. Modelos de crecimiento endógeno en economía abierta . . . 169
3.2. Evidencia empírica sobre los efectos del comercio internacional en el crecimiento 176
Conclusiones . . . 178

B10. Teoría de la integración económica **179**
Introducción . . . 179
1. Efectos estáticos de la integración económica . . . 181
1.1. Creación y desviación del comercio (Viner, 1950) . . . 181
1.2. Modelo canónico neoclásico de integración económica . . . 183
1.3. Extensiones al modelo canónico . . . 189
1.4. Conclusiones . . . 192
2. Efectos dinámicos de la integración económica . . . 193
2.1. Efecto economías de escala: costes medios decrecientes (Corden, 1972) . . . 193
2.2. Efectos procompetitivos y de variedad: modelo de Krugman (1979b) . . . 195
2.3. Efectos sobre el crecimiento: variedades (Romer, 1990) y desbordamientos (Lucas, 1988) . . . 196
2.4. Efecto selección: modelo de Melitz (2003) . . . 196
2.5. Conclusiones . . . 197
3. El proceso de integración económica: retos y dificultades . . . 198
3.1. Costes . . . 198
3.2. Profundización de la integración: el mercado común (MacDougall, 1960) . . . 201
3.3. Regionalismo y multilateralismo . . . 204
Conclusiones . . . 205
Anexo. Evidencia empírica y medición de la integración . . . 206
A.1. Medición . . . 206
A.2. Resultados . . . 206

B11. Balanza de pagos **208**
Introducción . . . 208
1. Fundamentos de la balanza de pagos: concepto y medición . . . 210
1.1. Fundamento teórico: el sector exterior en el flujo circular de la renta . . . 210
1.2. Concepto, base normativa y elementos a contabilizar . . . 211
1.3. Principios contables . . . 212
1.4. Medición . . . 214
1.5. Conclusiones . . . 216
2. Estructura contable de las cuentas internacionales . . . 217
2.1. Estructura de la balanza de pagos . . . 217
2.2. Otras cuentas internacionales . . . 222
2.3. Relación con las cuentas nacionales . . . 223
2.4. Conclusiones . . . 224
3. Interpretación económica de las cuentas internacionales . . . 225
3.1. Factores que afectan de forma global a las cuentas internacionales . . . 225
3.2. Análisis de cada cuenta . . . 227
3.3. Posición de inversión internacional y hoja de balance . . . 230

3.4. Otras presentaciones y evidencia empírica . . . 231
3.5. Conclusiones . . . 233
Conclusiones . . . 235
Anexo. Registro de operaciones en la balanza de pagos . . . 236

B12. Ajuste de la balanza de pagos **237**
Introducción . . . 237
1. Teorías sobre el ajuste de la cuenta corriente . . . 239
1.1. Enfoques de tipo de cambio real . . . 239
1.2. Extensiones al análisis de elasticidades . . . 242
1.3. Enfoques de la renta . . . 245
1.4. Conclusiones . . . 252
2. Teorías basadas en el desequilibrio en la cuenta financiera: enfoque activos . . . 253
2.1. Modelo monetario (Johnson, 1958, 1972) . . . 253
2.2. Modelo de cartera (McKinnon, 1969) . . . 255
2.3. Conclusiones . . . 258
3. Enfoque intertemporal y sostenibilidad del déficit y de la deuda exterior . . . 259
3.1. Modelo intertemporal de la balanza de pagos (Sachs, 1981; Obstfeld y Rogoff, 1995) . . . 259
3.2. Desarrollos posteriores del enfoque intertemporal . . . 264
3.3. Análisis de la sostenibilidad del déficit y de la deuda exterior . . . 265
3.4. Conclusiones . . . 269
Conclusiones . . . 270
Anexo. Condición de Marshall-Lerner como condición de estabilidad . . . 271

B13. Mercado de divisas **273**
Introducción . . . 273
1. Mercado de divisas: concepto, características y agentes . . . 275
1.1. Divisa y tipo de cambio . . . 275
1.2. Características del mercado de divisas . . . 277
1.3. Agentes que participan en el mercado de divisas . . . 278
1.4. Conclusiones . . . 279
2. Instrumentos y operaciones . . . 280
2.1. Instrumentos . . . 280
2.2. Operaciones . . . 284
2.3. Conclusiones . . . 294
3. Eficiencia y regulación del mercado de divisas . . . 295
3.1. La hipótesis de eficiencia (Bachelier, 1900; Samuelson, 1965; Fama, 1970) . . . 295
3.2. Contraste de la hipótesis de eficiencia: la paradoja del margen *forward* o *puzzle* de Fama (Fama, 1984) . . . 296
3.3. La situación de la regulación en el mercado de divisas . . . 297
Conclusiones . . . 299
Anexo . . . 300
A.1. Cómo operan los creadores de mercado . . . 300
A.2. Obtención de la paridad *cubierta* de intereses (PCI) . . . 300

B14. Determinación del tipo de cambio **302**
Introducción . . . 302
1. Teorías tradicionales: teorías basadas en los modelos de flujos de la balanza de pagos 304
1.1. Modelos basados en el ajuste de la balanza de pagos . . . 304
1.2. Modelos basados en la paridad del poder adquisitivo de la moneda . . . 306
1.3. Modelos basados en la paridad de intereses . . . 310

1.4. Conclusiones . . . 313
2. Desarrollos posteriores: teorías basadas en los modelos de activos (*stocks*) de la balanza de pagos . . . 314
2.1. Antecedentes de los modelos monetarios: la teoría cuantitativa del dinero y el mecanismo flujo-especie (Hume, 1752) . . . 314
2.2. Modelos monetarios . . . 315
2.3. Modelo de cartera (McKinnon y Oates, 1966; Branson, 1968, 1977) . . . 323
2.4. El contraste empírico: la paradoja de Meese-Rogoff y el tipo de cambio real de equilibrio . . . 326
2.5. Conclusiones . . . 326
3. Desarrollos recientes: expectativas, microestructura e intermediación financiera . . . 328
3.1. Expectativas racionales: noticias y burbujas . . . 328
3.2. Expectativas heterogéneas: chartistas, fundamentalistas y caos . . . 329
3.3. Microestructura del mercado de divisas (Evans y Lyons, 2002) . . . 330
3.4. Desconexión, primas de riesgo e intermediación financiera . . . 331
3.5. Conclusiones . . . 332
Conclusiones . . . 334

B15. Regímenes cambiarios **335**
Introducción . . . 335
1. Análisis comparado de los regímenes cambiarios teóricos . . . 337
1.1. El problema de la elección del régimen de cambio . . . 337
1.2. Lado de la oferta . . . 338
1.3. Lado de la demanda: perturbaciones y eficacia de las políticas . . . 339
1.4. Diferencias en el ajuste externo . . . 346
1.5. Conclusiones . . . 348
2. Análisis comparado de los regímenes cambiarios observados . . . 350
2.1. Regímenes cambiarios de vinculación fuerte (*hard pegs*) . . . 350
2.2. Regímenes cambiarios intermedios o de vinculación débil (*soft pegs*) . . . 352
2.3. Regímenes cambiarios flotantes . . . 354
2.4. Conclusiones . . . 356
3. Intervención y regulación en los mercados de cambio . . . 358
3.1. Intervención . . . 358
3.2. Regulación . . . 365
Conclusiones . . . 368

B16. Integración monetaria **369**
Introducción . . . 369
1. Teoría de las áreas monetarias óptimas: la integración monetaria como coste . . . 371
1.1. Antecedentes . . . 371
1.2. El coste de renunciar al tipo de cambio . . . 371
1.3. Determinantes de la optimalidad . . . 374
1.4. Conclusiones . . . 383
2. Desarrollos posteriores: los beneficios de la integración monetaria . . . 385
2.1. Beneficios macroeconómicos . . . 385
2.2. Beneficios microeconómicos . . . 390
2.3. Beneficios geopolíticos . . . 391
2.4. Conclusiones . . . 392
3. El proceso de construcción de una unión monetaria: retos y soluciones . . . 393
3.1. Unión monetaria incompleta: modelo de equilibrios múltiples (De Grauwe, 2011) . . . 393
3.2. Soluciones a las uniones monetarias incompletas . . . 396

Conclusiones . . . 399

B17. Crisis de balanza de pagos **400**
Introducción . . . 400
1. Modelos de primera generación: modelo de Krugman (1979a) y Flood y Garber (1984). El problema de la incompatibilidad . . . 402
1.1. Supuestos . . . 402
1.2. Modelo . . . 402
1.3. Contagio: los vínculos comerciales . . . 406
1.4. Evidencia empírica: experiencia histórica . . . 407
1.5. Conclusiones . . . 409
2. Modelos de segunda generación: modelo de equilibrios múltiples (Obstfeld, 1994). Profecías autocumplidas . . . 410
2.1. Supuestos . . . 410
2.2. Modelo . . . 410
2.3. Contagio: los comportamientos de rebaño . . . 412
2.4. Evidencia empírica: experiencia histórica . . . 413
2.5. Conclusiones . . . 414
3. Modelos de tercera generación: modelo de crisis gemelas (Krugman, 1999). Efectos de hoja de balance . . . 416
3.1. Supuestos . . . 417
3.2. Modelo . . . 417
3.3. Contagio: los vínculos financieros . . . 420
3.4. Evidencia empírica: experiencia histórica . . . 421
3.5. Conclusiones . . . 423
Conclusiones . . . 425
Anexo. Contagio puro (Calvo y Reinhart, 1996) . . . 426

B18. Globalización **427**
Introducción . . . 427
1. La nueva globalización económica y financiera y las cadenas globales de valor . . . 430
1.1. Características de la nueva globalización . . . 430
1.2. Situación actual de los flujos . . . 430
1.3. Cadenas globales de valor . . . 434
1.4. Conclusiones . . . 438
2. Determinantes y efectos de los movimientos internacionales de factor trabajo . . . 439
2.1. El punto de partida teórico: la sustituibilidad neoclásica (Mundell, 1957) . . . 439
2.2. Determinantes de los flujos internacionales de trabajadores . . . 439
2.3. Efectos de las migraciones internacionales . . . 442
2.4. Conclusiones . . . 449
3. Determinantes y efectos de los movimientos internacionales de capital . . . 450
3.1. Determinantes . . . 450
3.2. Efectos . . . 454
3.3. Conclusiones . . . 456
Conclusiones . . . 457

B19. Coordinación de políticas económicas internacionales **458**
Introducción . . . 458
1. Fundamentos teóricos: interdependencia y transmisión internacional . . . 460
1.1. Interdependencia, efecto desbordamiento y canales de transmisión . . . 460
1.2. Canal de la renta y canal monetario: modelo IS-LM-BP . . . 461
1.3. Canal de la inflación: modelo de oferta y demanda agregadas . . . 465

1.4. Canal de las expectativas: la fragilidad de una unión monetaria incompleta (De Grauwe, 2011) . . . 467
1.5. Conclusiones . . . 468
2. Justificación teórica de la coordinación . . . 470
2.1. Perspectiva de la teoría de juegos . . . 470
2.2. Perspectiva de los bienes públicos globales . . . 475
2.3. Los límites de la coordinación . . . 477
2.4. Conclusiones . . . 478
3. Aspectos prácticos de la coordinación internacional . . . 479
3.1. Características . . . 479
3.2. El G20, la OCDE y otros foros internacionales . . . 479
3.3. Experiencia y perspectivas de la coordinación . . . 482
3.4. Problemas *de facto*: retos y dificultades de la coordinación . . . 486
Conclusiones . . . 487

Nota

Este PDF es un extracto de *Lecciones de economía general e internacional* (Lecciones de economía, vol. 1), de Arturo Lorenzo, con el bloque «Comercio Internacional» (temas B5-B19). 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.

B5. Teoría del comercio internacional

Teoría del comercio internacional (I). La teoría ricardiana de la ventaja comparativa. Determinación de la relación real de intercambio. El modelo de factores específicos. El Modelo Heckscher-Ohlin-Samuelson (H-O-S); teoremas derivados del modelo y extensiones.

Introducción

Por qué comercian los países, qué bienes y servicios se intercambian y cómo afecta ese intercambio a sus economías son las tres preguntas que intenta contestar la **teoría del comercio internacional**, cuyo objeto son las transacciones entre residentes de países distintos. La relevancia de esas preguntas no ha dejado de crecer, porque el comercio de bienes y servicios (exportaciones más importaciones) equivale hoy a más de la mitad del PIB mundial y las cadenas globales de valor han hecho que buena parte de los bienes que se consumen se produzca en varios países.

La teoría del comercio internacional tiene dos vertientes[1], una distinción clásica que ya aparece en Marshall (*The Pure Theory of Foreign Trade*, 1879) y que Jagdish Bhagwati (*The Pure Theory of International Trade: A Survey*, 1964)[2] sistematizó. La **teoría pura** se ocupa del lado real de los intercambios y estudia qué ganancias reportan, qué patrón de comercio generan, qué efectos tiene la protección, cómo se coordinan las políticas entre países y cómo se mueven los factores de un país a otro. La **teoría monetaria**, en cambio, introduce el dinero y estudia cómo se ajusta una balanza de pagos desequilibrada, de qué depende el tipo de cambio y cómo se comparan entre sí los regímenes cambiarios. Cada una admite dos enfoques, el positivo, que da cuenta de cuánto se comercia, en qué dirección y a qué precios, y el normativo, que juzga si el comercio mejora o empeora el bienestar.

Históricamente, la teoría pura ha atravesado tres etapas:

1. *Teoría tradicional*: nace con los clásicos, con la crítica de Adam Smith (1776) al mercantilismo y la ventaja comparativa de David Ricardo (1817), y se completa con los neoclásicos y el modelo de Heckscher-Ohlin-Samuelson. Explica el **comercio interindustrial** por las diferencias de tecnología o de dotaciones relativas de factores entre países.

2. *Nueva teoría*: surge a finales de los años setenta con los trabajos de Krugman y Helpman, entre otros, ante la evidencia de que buena parte del comercio se produce entre países similares y dentro de una misma industria. Introduce la competencia imperfecta y los rendimientos crecientes a escala para explicar el **comercio intraindustrial**.

[1] Krugman y Obstfeld distinguen de modo análogo entre comercio internacional (transacciones reales) y finanzas internacionales (dimensión monetaria), aunque en la práctica la frontera entre ambas no es nítida.

[2] Bhagwati, J. (1964), "The Pure Theory of International Trade: A Survey", *The Economic Journal*, 74(293), pp. 1-84.

3. *Novísima teoría*: nace con Melitz (2003) y otros autores y supone un cambio de perspectiva, pues deja de preguntarse por qué comercian los países para preguntarse por qué comercian las *empresas*, y explica por qué solo las más productivas llegan a exportar.

Este tema recorre en orden cronológico la teoría tradicional del comercio internacional:

- *Antecedentes clásicos*: se estudian las teorías de Smith, Ricardo y Mill.

- *Teoría neoclásica*: el tema expone las teorías de Marshall y Haberler y el modelo de Heckscher-Ohlin-Samuelson (HOS), con sus teoremas, y el modelo de factores específicos.

- *Revisiones de la teoría tradicional*: la última sección recoge las revisiones que levantan sus supuestos.

En todos sus modelos el comercio surge porque cada bien tiene un precio relativo de autarquía distinto en cada país, y lo que cambia entre ellos son las causas de esa disparidad.

1. Antecedentes y fundamentos de la teoría neoclásica: teorías clásicas sobre el comercio internacional

Las teorías clásicas sentaron los fundamentos de la teoría neoclásica. Smith explicó el comercio por la ventaja absoluta y Ricardo por la ventaja comparativa, mientras que Mill completó el análisis con las demandas recíprocas, que determinan la relación real de intercambio.

1.1. Teoría de la ventaja absoluta (Smith, 1776)

Origen: la principal motivación que llevó a Smith a contribuir a la teoría comercial fue refutar las ideas mercantilistas sobre el librecambio[3]. Para el mercantilismo, una economía crecía por causas monetarias más que reales, de modo que acumular oro y riqueza debía guiar la política económica. Esa acumulación exigía vender al exterior más de lo que se compraba, y el superávit comercial se buscaba con medidas proteccionistas. En consecuencia, esta escuela de pensamiento consideraba el comercio exterior un *juego de suma cero*, es decir, lo que un país ganaba con el intercambio, otro lo perdía necesariamente.

Idea: frente al mercantilismo, Smith rechaza el proteccionismo y ve en el libre comercio, tanto interior como exterior, un *juego de suma positiva*, y su argumento descansa en el concepto de *ventaja absoluta*.

Razonamiento: este concepto puede ilustrarse a través de un conocido ejemplo en el que Inglaterra y Portugal producen paño y vino, inicialmente en autarquía:

Tabla B5.1: Producción por hora de trabajo en el ejemplo de la ventaja absoluta.

	Paño (metros/h)	Vino (litros/h)
Inglaterra	2	4
Portugal	1	8

Un país posee ventaja absoluta en un bien cuando lo produce con un coste unitario, medido en horas de trabajo, inferior al de cualquier otro país o, lo que es lo mismo, con un producto medio del trabajo más alto. Ese país puede entonces destinar todos sus recursos al bien en cuestión y vendérselo después a los demás países, en un intercambio que beneficia a todas las partes.

En el ejemplo, la ventaja absoluta en paño corresponde a Inglaterra (2 metros por hora frente a 1) y la del vino a Portugal (8 litros por hora frente a 4). A cada país le conviene, así, volcarse por completo en su bien y conseguir el otro comerciando.

El *precio relativo internacional* o *relación real de intercambio* (*terms of trade*, RRI)[4] equivale a la inversa del tipo de cambio real[5]. Siempre que la RRI se encuentre estrictamente entre los precios relativos de autarquía de cada país, intervalo beneficioso para ambos, les convendrá especializarse y comerciar. Smith resume esta idea en una conocida frase, según la cual "todo

[3] De acuerdo con Landreth y Colander (2002), un cuarto de *La riqueza de las naciones* (1776) está dedicado a refutar las ideas comerciales proteccionistas del mercantilismo.

[4] La RRI es el cociente entre los precios de exportación y de importación, expresados ambos en moneda nacional: $\text{RRI} = p_X/p_Y$, es decir, las unidades de importaciones que se obtienen por unidad exportada ($\text{RRI} = Y/X$). Véase Gandolfo (2014, p. 94) y Bajo y Monés (2000, p. 474).

[5] Esa equivalencia no implica causalidad y solo es exacta en el modelo de trueque con especialización completa, en el que el precio de las exportaciones coincide con el nivel de precios interno y el de las importaciones con el exterior. Fuera de ese caso, ambas magnitudes pueden divergir, pues el tipo de cambio real compara índices generales de precios y la RRI solo precios de exportación e importación.

padre de familia prudente tendrá como máxima nunca producir en su casa un bien que sea más barato comprar".

En Smith conviven, además, dos argumentos de naturaleza distinta a favor del libre comercio. Uno es estático, el de la ventaja absoluta, que reasigna recursos dados hacia sus usos más productivos. El otro es dinámico, el de la extensión del mercado, según el cual la apertura amplía el mercado y, con él, la división del trabajo. Además, la apertura da empleo a recursos que en autarquía quedarían ociosos y eleva la productividad a lo largo del tiempo. Este segundo argumento, menos formalizado, anticipa la literatura moderna sobre comercio y crecimiento (véase el tema B09).

Conclusiones:

- *Aspectos positivos*: constituye una primera aproximación a la explicación de las causas del comercio internacional y del patrón de comercio. La causa del comercio es la posibilidad de ganancias mutuas en un juego de suma positiva, pues la especialización que permite el comercio aumenta el producto total. El patrón de comercio (qué produce cada país) se explica, a su vez, por las diferencias tecnológicas absolutas.

- *Limitaciones*: no explica por qué países con ventaja absoluta en todos los sectores participan en el comercio internacional, ni cómo se fija la relación real de intercambio, y solo admite la especialización total por considerar una tecnología de proporciones fijas.

1.2. Teoría de la ventaja comparativa (Ricardo, 1817)

Origen: David Ricardo, en sus *Principios de economía política y tributación* (1817), amplía la visión de Smith al introducir el concepto de *ventaja comparativa*. La ventaja comparativa resuelve la primera de las limitaciones anteriores, pues explica que un país sin ventaja absoluta alguna participe, con beneficio, en el comercio internacional.

Idea: la ventaja comparativa descansa en el concepto de *coste de oportunidad*[6]. El coste de oportunidad de un bien en términos de otro hace referencia al número de unidades del "otro bien" que podrían haberse producido con los recursos destinados a producir una unidad del bien "de referencia".

La ventaja comparativa aplica el concepto de coste de oportunidad al ámbito internacional como criterio de producción e intercambio. Un país tiene ventaja comparativa en la producción de un bien si el coste de oportunidad de producir ese bien (en términos de otros bienes) en ese país es menor que en otros países.

Razonamiento: la ventaja comparativa se puede calcular de dos formas equivalentes:

1. *Intrapaís*: como el coste relativo o la productividad relativa de la fabricación de los dos bienes en el mismo país. Esto significa calcular una ratio de los costes unitarios absolutos o de las productividades medias de cada bien.

2. *Interpaís*: como el coste relativo o la productividad relativa de la fabricación del mismo bien en cada país. De nuevo, esto significa calcular la ratio de los costes unitarios absolutos o de las productividades medias del mismo bien en ambos países.

[6] Esta es la interpretación moderna basada en ideas de la escuela austriaca. El concepto de coste de oportunidad apareció después de Ricardo en la literatura.

El concepto se puede ilustrar mediante una variante del ejemplo anterior, en la que se supone una mejora tecnológica en Inglaterra, como la derivada de su industrialización durante los siglos XVIII y XIX, que le otorga ventaja absoluta en la producción de ambos bienes:

Tabla B5.2: Producción por hora de trabajo y coste de oportunidad en el ejemplo de la ventaja comparativa.

	Paño (metros/h)	Vino (litros/h)	C.O. del paño (en vino)	C.O. del vino (en paño)
Inglaterra	6	12	2	1/2
Portugal	1	8	8	1/8

El razonamiento procede en tres pasos:

1. *Ventaja absoluta doble*: Inglaterra produce ahora más paño y más vino por hora de trabajo que Portugal.

2. *Costes de oportunidad*: sin embargo, el del paño es menor en Inglaterra (2 litros de vino por metro) que en Portugal (8 litros), y el del vino es menor en Portugal (1/8 de metro de paño por litro) que en Inglaterra (1/2 metro).

3. *Ganancias mutuas*: por tanto, aunque Inglaterra tiene ventaja absoluta en la producción de ambos bienes, los dos países ganan con el comercio siempre que la RRI internacional se sitúe, de nuevo, estrictamente entre sus precios relativos de autarquía. Cada país se especializará totalmente en el sector en el que tenga ventaja comparativa, e industrias de países tecnológicamente menos avanzados en términos absolutos podrán competir internacionalmente si tienen ventaja comparativa.

En conclusión, lo relevante no es la productividad de cada una de las industrias (las diferencias tecnológicas absolutas), sino el coste de oportunidad de fabricar una cosa u otra (las diferencias tecnológicas relativas).

Extensiones: Dornbusch, Fischer y Samuelson (1977) generalizan el modelo a un continuo de bienes ordenados según la ventaja comparativa. La tecnología y la demanda determinan juntas el salario relativo, y este fija el bien frontera que separa los bienes que produce cada país. Con costes de transporte surge, además, una franja endógena de bienes no comerciables (Anexo A.3).

Conclusiones:

- *Aspectos positivos*: la causa del comercio sigue siendo la misma que en Smith, la existencia de ganancias mutuas en un juego de suma positiva. El patrón de comercio (qué produce cada país) pasa, en cambio, a explicarse por las diferencias tecnológicas relativas, lo que extiende las ganancias del comercio incluso a los países sin ventaja absoluta alguna. La ventaja comparativa se convertirá, desde entonces, en la piedra angular de toda la teoría pura del comercio.

- *Limitaciones*: sigue sin determinar la relación de precios internacionales, que queda simplemente acotada entre los precios de autarquía, y sigue considerando la especialización total, consecuencia de la tecnología lineal de un solo factor.

1.3. Teoría de las demandas recíprocas (Mill, 1848)

Origen: John Stuart Mill partió del germen contenido en la obra de su padre, James Mill (1821). Sobre esa base, definió con precisión el concepto de relación real de intercambio y dio una explicación teórica de las causas que la determinaban. Lo hizo primero en 1844 y, de forma definitiva, en sus *Principios de economía política* (1848)[7][8].

Idea: según Mill, la relación real de intercambio depende de la fuerza relativa de la demanda de productos importados de cada país, puesto que esta, a su vez, determina la oferta de productos exportados con la que pagarlos. Estas demandas se conocen como *demandas recíprocas*, pues la demanda de importaciones de un país constituye, simultáneamente, su oferta de exportaciones.

Razonamiento: el equilibrio internacional se alcanza en aquella RRI para la cual la cantidad que cada país desea importar coincide exactamente con la cantidad que el otro desea exportar.

De este planteamiento se derivan dos implicaciones de interés:

1. *Determinación de la RRI dentro de los límites ricardianos*: la relación real de intercambio de equilibrio se sitúa en algún punto del intervalo definido por los precios relativos de autarquía, y su posición concreta depende de la intensidad relativa de las demandas recíprocas.

2. *El reparto de las ganancias*: de lo anterior se sigue que el reparto no tiene por qué ser igualitario. En particular, un país pequeño cuya demanda apenas influye en los precios mundiales tenderá a comerciar a una RRI próxima a los precios de autarquía del país grande, con lo que capturará la mayor parte de las ganancias del intercambio. Este resultado se conoce en la literatura como la "importancia de ser poco importante".

Conclusiones:

- *Aspectos positivos*: los historiadores del pensamiento económico han elogiado a Mill por sus aportaciones teóricas al comercio internacional, pues cerró, conceptualmente, el problema que Ricardo había dejado abierto, la determinación de la relación real de intercambio.

- *Limitaciones*: sin embargo, aunque ofrece una definición precisa de la RRI y una teoría verbal de su determinación, su aportación no proporciona un instrumento analítico que permita obtenerla. Esa formalización habrá de esperar a las curvas de oferta neta de Marshall (apartado 2.1).

1.4. Conclusiones

Los autores clásicos habían explicado tres cuestiones fundamentales, por qué comercian los países, cuál es su patrón de comercio y qué determina la relación real de intercambio. Ricardo extendió las ganancias del comercio a los países sin ventaja absoluta, y Mill explicó la RRI, que Ricardo solo acotaba entre los precios de autarquía.

Sus planteamientos adolecían, sin embargo, de dos carencias:

- *Supuestos restrictivos*: consideraban, fundamentalmente, una tecnología lineal, de proporciones fijas y coste unitario constante, de manera que la especialización siempre era completa.

[7] Mill abordó la cuestión de cómo se distribuirían las ganancias del comercio y llegó a una solución notablemente acertada, sobre todo si se considera que no empleó técnica matemática alguna y que la noción de elasticidad aún no se había formulado. Véase Landreth y Colander (2002, p. 178).

[8] En palabras de Mill: "las cantidades demandadas por cada país de los artículos que importa de su vecino son exactamente suficientes para pagarse mutuamente". Véase Landreth y Colander (2002, p. 178).

- *Falta de formalismo*: sus argumentos eran esencialmente verbales y numéricos, sin un aparato analítico que permitiera derivar los resultados con generalidad.

La revolución marginalista permitió superar ambas carencias y plantear modelos más próximos a la realidad, con varios factores de producción y especialización parcial. Surgieron así dos líneas de aportaciones, las que siguieron un enfoque de equilibrio parcial, debidas a Alfred Marshall, y las que siguieron un enfoque de equilibrio general, que ofrecen una panorámica más completa.

2. Teoría neoclásica del comercio internacional de bienes y servicios

La teoría neoclásica parte de las curvas de oferta neta, en equilibrio parcial con Marshall y en equilibrio general con Edgeworth. Siguen, ya en equilibrio general, la reformulación de la ventaja comparativa de Haberler y las ventajas dotacionales del modelo Heckscher-Ohlin-Samuelson, y la sección termina con los teoremas derivados del modelo y con el modelo de factores específicos, su versión a corto plazo.

2.1. Determinación de la relación real de intercambio: las curvas de oferta neta (Marshall, 1879; Edgeworth, 1894)

Origen: la primera aportación destacada se debe a Alfred Marshall (1879), que presenta las curvas de oferta neta (*offer curves*), formalización del concepto de demanda recíproca de Mill que permite obtener la RRI de equilibrio por vía gráfica. Francis Edgeworth (1894) las incorporó después a su teoría de los valores internacionales, con un enfoque que puede considerarse de equilibrio general, pues las combina con las curvas de indiferencia que había introducido en 1881.

Idea: la *curva de oferta neta* es el lugar geométrico de las combinaciones de exportaciones e importaciones que un país desea intercambiar a cada relación real de intercambio internacional, es decir, de las exportaciones que entregaría a cambio de cada volumen de importaciones[9].

2.1.1. Supuestos

Entorno: se consideran dos bienes, X exportable e Y importable, con libre movilidad intersectorial de los factores y pleno empleo[10].

Agentes: no se considera el destino que el sector público da a los ingresos arancelarios[11].

Mercados: hay competencia perfecta en todos los mercados, nacionales e internacionales, y no hay costes de transporte.

Sector exterior: no existen pagos ni deuda al margen de los flujos de mercancías, de modo que la balanza comercial está equilibrada en todo momento y $Y = (p_X/p_Y) \cdot X$.

2.1.2. Modelo

Planteamiento: las curvas de oferta neta pasan por el origen[12] y son simétricas entre sí, pues los ejes de exportaciones e importaciones de un país son, respectivamente, los de importaciones y exportaciones del otro[13]. Presentan además dos rasgos:

[9] Pueden obtenerse de dos formas. Marshall partió de los excesos de oferta y demanda nacionales de cada bien para cada RRI, y James Meade, en equilibrio general, de las tangencias de la RRI con la frontera de posibilidades de producción y con las curvas de indiferencia social.

[10] Este supuesto, junto con la posibilidad de producir ambos bienes en el país, permite que exportaciones e importaciones varíen en el mismo sentido, por ejemplo, que caigan ambas al introducirse un arancel.

[11] En general, la recaudación arancelaria puede distribuirse entre los consumidores o gastarse íntegramente en el bien importable, íntegramente en el exportable o en una combinación de ambos (Gandolfo, 2014, p. 229).

[12] En el origen, el país no es ni oferente neto ni demandante neto de ninguno de los bienes, puesto que no hay excesos en el mercado nacional (Gandolfo, 2014, p. 48).

[13] La simetría de las curvas no exige que los países tengan las mismas preferencias o la misma tecnología, pues las curvas simplemente recogen la disposición a comerciar de cada país.

1. *Ambigüedad*: pueden presentar un tramo elástico de pendiente positiva y otro inelástico de pendiente negativa[14]. En el *tramo elástico*, cuantas más importaciones se deseen, más exportaciones se requieren para adquirirlas, de modo que la curva es creciente y se califica de normal. En el *tramo inelástico*, por el contrario, la RRI resulta tan favorable que pueden importarse más unidades aunque las exportaciones caigan, caso que suele calificarse de anómalo[15]. La elasticidad relevante, cuyo valor determina tanto la forma de la curva como la estabilidad del equilibrio, es la de la demanda de importaciones respecto de la RRI.

2. *Convexidad*: siempre que la frontera de posibilidades de producción (FPP) sea cóncava[16], las curvas de oferta neta serán convexas en el plano exportaciones-importaciones de cada país. En el tramo elástico, la convexidad implica que, ante un aumento de la RRI, aumentan tanto las exportaciones como las importaciones, si bien estas últimas proporcionalmente más que las primeras.

Como caso límite, una curva perfectamente elástica (creciente y recta) describe a una economía dispuesta a comerciar cualquier volumen a la RRI dada. Así ve un país pequeño la curva del resto del mundo, pues la oferta neta del país pequeño no basta para alterar la RRI.

Equilibrio: se parte de precios relativos de autarquía distintos en cada país, pues de lo contrario no habría comercio[17], con $p_X^B/p_Y^B > p_X^A/p_Y^A$, de donde resulta un exceso de oferta del bien Y en B y un exceso de demanda en A. El intercambio solo tendrá lugar si la RRI internacional se sitúa estrictamente entre ambos precios de autarquía, $\text{RRI}^* \in (p_X^A/p_Y^A,\ p_X^B/p_Y^B)$. El equilibrio se alcanza en el corte de ambas curvas, con la balanza comercial equilibrada tanto en el conjunto del mundo como en cada país, y presenta tres propiedades:

- *Unicidad*: exige que las curvas se corten una sola vez, lo que queda garantizado si ambas son elásticas en todo su recorrido.

- *Estabilidad*: exige el cumplimiento de la *condición de Marshall-Lerner*, es decir, que la suma de las elasticidades, en valor absoluto, de las demandas de importaciones de ambos países supere la unidad. En términos cambiarios, partiendo de una balanza comercial equilibrada y con ofertas infinitamente elásticas, equivale a exigir que las depreciaciones reales mejoren la balanza comercial.

- *Reparto de las ganancias*: los precios de equilibrio definen dos *triángulos de comercio* congruentes, cuyos catetos son las exportaciones e importaciones de cada país. El reparto depende de dónde quede la RRI dentro del intervalo de participación. Así, cuanto más próxima al precio de autarquía de un país, menor será su ganancia, en línea con Mill.

En el equilibrio se alcanza un óptimo paretiano en el que las ganancias del intercambio están agotadas. El mecanismo de precios vacía todos los mercados y, por la ley de Walras, que se cumple a cualquier RRI, basta con que se vacíe uno para que se vacíe el otro.

Desarrollo gráfico: la determinación de la RRI se representa en el plano bien X-bien Y (figura B5.1):

[14] Nada garantiza *a priori* que las curvas de oferta neta sean siempre crecientes, ni siquiera cuando ambos bienes son normales (Gandolfo, 2014, p. 438). Entre ambos tramos hay un punto de elasticidad unitaria.

[15] Estos casos anómalos pueden darse aun cuando, en el mercado nacional, las curvas de oferta y demanda de bienes tengan buen comportamiento (Gandolfo, 2014, p. 439).

[16] La FPP es cóncava siempre que existan rendimientos decrecientes a escala en la producción de ambos bienes, factores específicos o distinta intensidad factorial.

[17] Como la tecnología, las preferencias y las dotaciones factoriales difieren entre países, es poco plausible que coincidan sus precios relativos de autarquía.

1. *Curva de oferta neta de A*, CON^A: recoge las exportaciones de X que A ofrece a cambio de cada volumen de importaciones de Y. Sale del origen con pendiente igual a su precio relativo de autarquía, p_X^A/p_Y^A, porque con volúmenes de comercio muy pequeños el país solo intercambia a su coste de oportunidad interno. Es convexa, porque la FPP es cóncava.

2. *Curva de oferta neta de B*, CON^B: simétrica de la anterior, recoge las exportaciones de Y que B ofrece a cambio de importaciones de X y sale del origen con pendiente p_X^B/p_Y^B, por la misma razón.

3. *Equilibrio*: el corte de ambas curvas fija los volúmenes de comercio, X^* e Y^*, y la RRI de equilibrio, RRI^*, pendiente del radio vector que pasa por él.

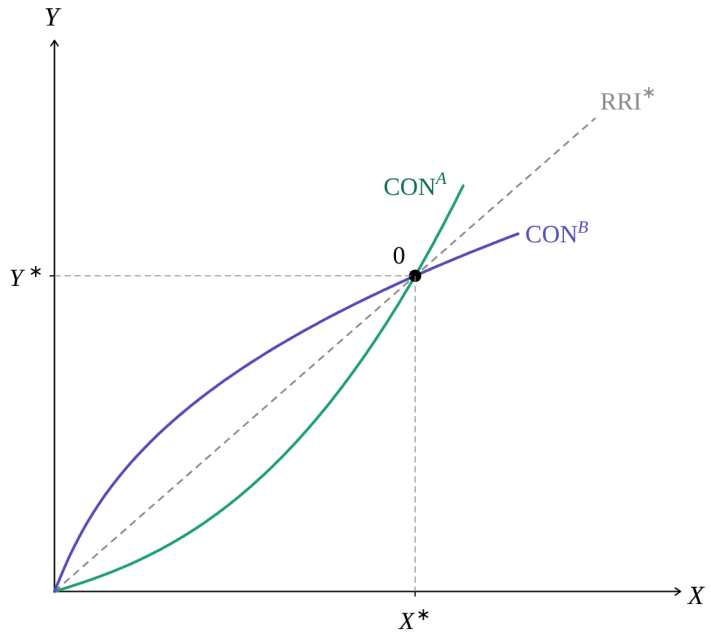

Figura B5.1: Curvas de oferta neta (Marshall-Edgeworth). La relación real de intercambio de equilibrio se determina en la intersección de ambas curvas.

Estática comparativa:

> Los instrumentos de política comercial y sus efectos se desarrollan en los temas B07 y B08, y el crecimiento empobrecedor, en el tema B09, apartado 1.3.2.

▪ *Arancel en un país pequeño*: al no poder influir sobre la RRI, la curva de oferta neta del país se contrae hacia el origen mientras la del resto del mundo permanece inalterada. Aumenta la producción del bien importable protegido, se reducen tanto exportaciones como importaciones y el Estado recauda el arancel. La curva de indiferencia social alcanzada queda así más próxima al origen, y la pérdida de bienestar es inequívoca, pues el país soporta íntegramente el coste de la distorsión sin ganancia alguna a cambio.

▪ *Arancel en un país grande*: al pesar su demanda en el mercado mundial, la contracción de su curva altera la RRI de equilibrio. Gráficamente, en el plano bien X-bien Y (figura B5.2), el arancel de A ($t > 0$) hace girar su curva de oferta neta hacia dentro, de CON_0^A a CON_1^A, y el

equilibrio pasa del punto 0 al 1. El comercio se reduce, de X_0 e Y_0 a X_1 e Y_1, y la RRI mejora para A, de RRI_0 a RRI_1.

Operan así dos efectos contrapuestos. El *efecto distorsión* reduce el bienestar como en el caso anterior, mientras que el *efecto relación real de intercambio* lo eleva, pues la caída del precio internacional del bien importable traslada parte de la carga a los productores extranjeros. De su combinación resulta la existencia teórica de un *arancel óptimo* positivo, aquel que maximiza la diferencia entre ambos[18]. El efecto sobre el bienestar mundial es, en cambio, inequívocamente negativo, porque la ganancia de RRI de un país es exactamente la pérdida del otro, mientras que la pérdida de eficiencia es común.

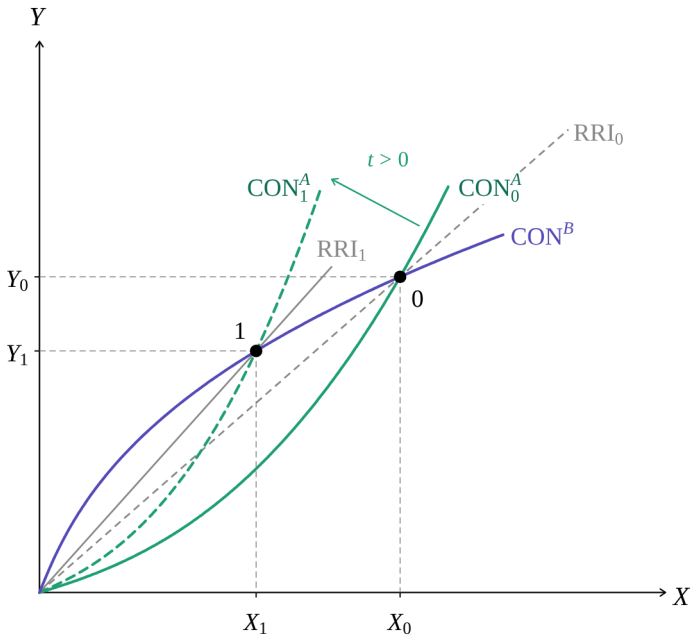

Figura B5.2: Efecto de un arancel sobre las curvas de oferta neta. La curva del país que lo impone gira hacia dentro y, si el país es grande, mejora su relación real de intercambio.

▪ *Corolario*: en ambos casos, si hay dos factores de producción, se aplica además el teorema de Stolper-Samuelson. El factor que el sector protegido emplea con mayor intensidad ve crecer, pues, su retribución real. En el país grande ello exige que el precio interno del importable suba pese a la mejora de la RRI, lo que equivale a descartar la paradoja de Metzler (1949). En ella, el precio mundial del importable cae tanto que su precio interno desciende y el efecto distributivo se invierte.

▪ *Perturbación tecnológica*: una mejora que permita producir ambos bienes con mayor eficiencia desplaza, salvo que su sesgo sea ultra-anticomercial, la curva de oferta neta hacia la derecha. El desplazamiento significa que el país está dispuesto a exportar más por cada volumen de importaciones. Si el país es grande, ello deteriora su relación real de intercambio.

[18] La existencia del arancel óptimo no constituye, sin embargo, una recomendación práctica de política comercial, pues se trata de una política de empobrecimiento del vecino. Esa política invita a la represalia y, generalizada, suele dejar a todos los países en peor situación que el libre comercio.

2.1.3. Conclusiones

Aspectos positivos: completa el análisis clásico al permitir determinar la RRI a partir de la interacción de las curvas, admite una función de producción no lineal e ilustra visualmente las ganancias del comercio.

Limitaciones: la versión marshalliana[19] sigue un enfoque de equilibrio parcial, centrado en determinar la RRI al margen de las interrelaciones con el resto de la economía.

2.2. Enfoque de equilibrio general: la reformulación de la ventaja comparativa (Haberler, 1936)

Origen: Gottfried von Haberler (1936; edición alemana de 1933) reformuló la ventaja comparativa en términos de coste de oportunidad, apoyándose en una tecnología no lineal con dos factores. Su contribución es doble. Representa el lado de la oferta mediante la frontera de posibilidades de producción, cuya pendiente mide el coste de oportunidad, e ilustra, junto con el lado de la demanda, los efectos de la apertura en un modelo reducido de equilibrio general.

Idea: la ventaja comparativa se debe tanto a la tecnología como a las preferencias, pues ambas determinan el precio relativo de autarquía, cuya comparación con el internacional fija el patrón de comercio.

2.2.1. Supuestos

Entorno: la economía es pequeña, con competencia perfecta y libre comercio sin restricciones ni costes de transporte.

Agentes: los consumidores son propietarios de los factores y de la tecnología, con función de utilidad continua, creciente, cuasicóncava y diferenciable.

Tecnología: se producen dos bienes comerciables, X exportable e Y importable, ambos normales, con dos factores (K y L). Las productividades marginales son positivas y decrecientes, con rendimientos constantes a escala, y la intensidad factorial difiere entre bienes, requisito sin el cual la FPP no resultaría cóncava. Los factores son homogéneos, perfectamente divisibles y de dotación dada.

2.2.2. Modelo

Planteamiento: el modelo combina la oferta y la demanda de la economía:

- *Oferta*: la FPP recoge las cantidades máximas de bienes que el país puede producir dada su tecnología y dotación[20]. Es cóncava por las propiedades de la función de producción, y su pendiente en el plano (X, Y) es, en valor absoluto, la relación marginal de transformación (el coste de oportunidad de X en términos de Y).

- *Demanda*: las curvas de indiferencia social representan combinaciones de igual utilidad, tanto mayor cuanto más alejadas del origen.

[19] A diferencia de Edgeworth, Marshall no explicita las curvas de indiferencia social, de modo que su análisis, aunque descansa en ideas de equilibrio general, es de equilibrio parcial o, a lo sumo, de equilibrio general incompleto.

[20] Véase Nicholson y Snyder (2012, p. 429).

Equilibrio:

▪ *Autarquía*: exige la coincidencia de las pendientes de la FPP, de la curva de indiferencia y de la restricción presupuestaria[21]:

$$\text{RMT}_{YX}^{\text{aut}} = \text{RMS}_{YX}^{\text{aut}} = p_X^{\text{aut}}/p_Y^{\text{aut}}$$

En autarquía, el precio relativo viene determinado por las condiciones interiores de oferta y demanda y, en competencia perfecta, no existen excesos de oferta ni de demanda, por lo que se consume exactamente lo que se produce.

▪ *Apertura*: la economía pequeña toma los precios exteriores, que en general difieren de los de autarquía, con lo que la triple tangencia se rompe y producción y consumo se separan:

$$\text{RMT}_{YX}(P^1) = p_X^*/p_Y^* = \text{RMS}_{YX}(C^1), \qquad P^1 \neq C^1$$

Si el bien X es relativamente más caro en el exterior, el país produce más X y menos Y, y la diferencia entre producción y consumo se cubre mediante exportaciones e importaciones. A diferencia del modelo ricardiano, la especialización no tiene por qué ser completa, al no ser lineal la tecnología. El país alcanza una curva de indiferencia superior a la de autarquía, si bien la teoría solo garantiza la mejora del bienestar y no el aumento del consumo de ambos bienes, pues el efecto sustitución podría reducir el del bien exportable, ahora relativamente más caro.

Desarrollo gráfico: los efectos de la apertura se representan en el plano bien X-bien Y (figura B5.3):

1. *Frontera de posibilidades de producción*, FPP: recoge las cantidades máximas de ambos bienes que permiten la tecnología y la dotación. Es cóncava, porque el coste de oportunidad de X crece a medida que se produce más, al diferir la intensidad factorial entre bienes, y su pendiente es la relación marginal de transformación.

2. *Curvas de indiferencia social*, CIS_0 y CIS_1: representan un bienestar tanto mayor cuanto más alejadas del origen.

3. *Rectas de precios*: la de autarquía, de pendiente $p_X^{\text{aut}}/p_Y^{\text{aut}}$, y la internacional, más inclinada, de pendiente p_X^*/p_Y^*.

4. *Equilibrio*: en autarquía, producción y consumo coinciden en $P^0 = C^0$, sobre la CIS_0. Con apertura, la producción pasa a P^1 y el consumo a C^1, sobre la CIS_1. La ganancia queda reflejada en el paso de la CIS_0 a la CIS_1, y el intercambio, en el triángulo de comercio, cuya hipotenusa es la recta de la RRI internacional y cuyos catetos son las exportaciones, $X_1^P - X_1^C$, y las importaciones, $Y_1^C - Y_1^P$.

[21] La restricción presupuestaria de la economía es $p_X^{\text{aut}} X + p_Y^{\text{aut}} Y = \text{RN}$, donde RN es la renta nacional.

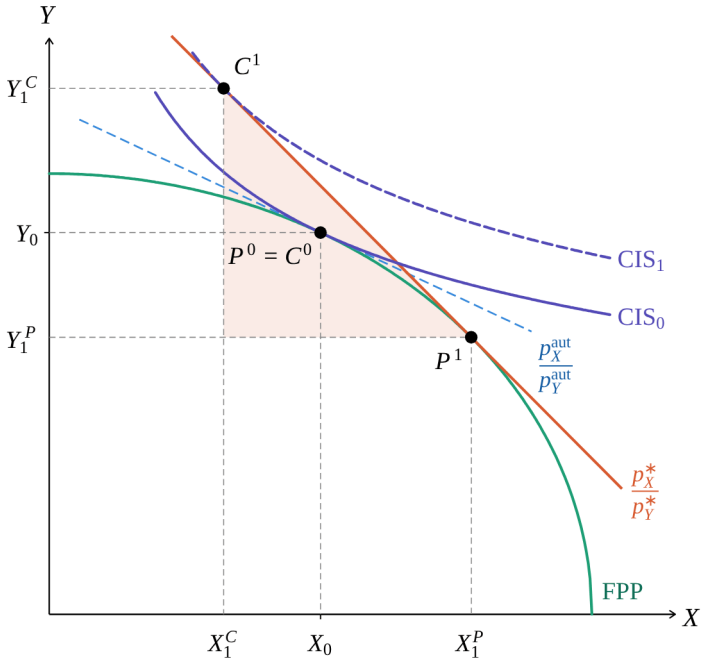

Figura B5.3: Reformulación de Haberler (1936). Con costes de oportunidad crecientes la frontera de posibilidades de producción es cóncava y la especialización es incompleta.

Estática comparativa:

▪ *Arancel*: eleva el precio relativo interno del bien importable por encima del internacional, lo que genera una *doble distorsión*. En la producción, los productores se desplazan a lo largo de la FPP hacia el bien protegido, con lo que la economía deja de maximizar el valor de su producción a precios mundiales. En el consumo, los consumidores igualan su relación marginal de sustitución al precio interno distorsionado y no a la RRI. El resultado es la contracción del triángulo de comercio y una curva de indiferencia más próxima al origen. Esta descomposición constituye la principal aportación del marco respecto del anterior, pues permite separar ambas fuentes de pérdida.

▪ *Perturbación tecnológica*: desplaza la FPP hacia afuera, de forma neutral o sesgada hacia el sector favorecido. Para la economía pequeña, que toma la RRI como dada, la ganancia de bienestar es inequívoca, y el volumen de comercio aumentará en la medida en que la mejora esté sesgada hacia el sector exportador.

2.2.3. Conclusiones

Aspectos positivos: incorpora las decisiones de consumo, muestra que las ganancias del comercio proceden tanto de la especialización en la producción como del intercambio en el consumo y que las preferencias intervienen en la determinación del precio relativo de autarquía. Además, su marco de equilibrio general reducido (FPP más curvas de indiferencia social) se convertirá en el instrumental estándar sobre el que se construye el modelo de Heckscher-Ohlin-Samuelson.

Limitaciones: no explica por qué difieren los costes relativos entre países. Si el trabajo es el único factor, como en Ricardo, las diferencias solo pueden proceder de la productividad, pero en

la realidad el comercio se ve influido también por las dotaciones de recursos. Responder a esa cuestión será la aportación del modelo de las ventajas dotacionales.

2.3. Enfoque de equilibrio general: las ventajas dotacionales (Heckscher, 1919; Ohlin, 1933; Stolper y Samuelson, 1941)

Origen: el modelo se debe a los economistas suecos Eli Heckscher (1919) y Bertil Ohlin (1933) y, posteriormente, a Paul Samuelson (Stolper y Samuelson, 1941; Samuelson, 1948b), que formalizó matemáticamente sus intuiciones.

Idea: las diferencias de ventaja comparativa proceden de las distintas dotaciones relativas de factores y de las distintas intensidades factoriales de cada bien. Frente a la tradición ricardiana, que la localizaba en las diferencias tecnológicas, el modelo supone tecnologías idénticas y desplaza la explicación hacia los recursos, de donde su denominación alternativa de teoría de las *proporciones factoriales*.

2.3.1. Supuestos

Entorno: existen dos países que comercian libremente, sin costes de transporte y en competencia perfecta en todos los mercados, con dotaciones absolutas de factores fijas. El país A es relativamente abundante en capital y el B en trabajo, abundancia que admite doble definición:

- *En términos físicos*: $(K^A/L^A) > (K^B/L^B)$[22].
- *En términos de precios*: $(r/w)^A < (r/w)^B$ en autarquía.

Agentes: las preferencias son homotéticas[23] e idénticas entre países, con función de utilidad continua, creciente, cuasicóncava y diferenciable.

Tecnología: ambos países emplean la misma tecnología, con propiedades neoclásicas de buen comportamiento[24], lo que hace que la especialización completa no sea necesaria con carácter general. Los factores son homogéneos, divisibles, perfectamente móviles dentro de cada país e inmóviles internacionalmente. El bien X es relativamente intensivo en capital y el Y en trabajo, y esa intensidad es irreversible (un cambio en los precios relativos de los factores no puede invertir la clasificación).

De ello se sigue que A tendrá ventaja comparativa en X y B en Y. Ambos países son idénticos salvo en su dotación relativa, de modo que cualquier diferencia en su comportamiento comercial ha de atribuirse necesariamente a esa única asimetría.

[22] La abundancia se define comparando ratios, no cantidades absolutas: EE. UU., con 80 millones de trabajadores y 200 millones de acres, tiene una ratio trabajo-tierra de 2/5, y R. U., con 20 millones de trabajadores y 20 millones de acres, de 1/1, de modo que R. U. es abundante en trabajo aunque tenga menos trabajadores.

[23] Con preferencias homotéticas, la cesta consumida mantiene sus proporciones cuando cambia la renta y solo las alteran los precios relativos. El supuesto choca con la ley de Engel (1857), pero resulta útil en los modelos de equilibrio general.

[24] Al tener la función de producción rendimientos constantes a escala, la senda de expansión de cada bien es un rayo desde el origen. Si, además, las isocuantas unitarias de X e Y se cortan una sola vez, para cualquier ratio de precios de los factores la ratio capital-trabajo empleada en X es siempre superior a la empleada en Y. Esa condición excluye la reversión de las intensidades factoriales (Gandolfo, 2014, p. 65).

2.3.2. Modelo

El **teorema de Heckscher-Ohlin** establece que un país exportará el bien cuya producción es intensiva en el factor que le resulta relativamente abundante.

Planteamiento: los países presentan un *sesgo de producción*, es decir, sus FPP están sesgadas hacia el eje del bien que usa intensivamente su factor abundante, de modo que ante una misma RRI adoptan decisiones de producción distintas[25]. El mapa de curvas de indiferencia es, sin embargo, común, por ser las preferencias idénticas, y cualquier rayo desde el origen corta todas las curvas en puntos de igual pendiente, por ser homotéticas.

Equilibrio:

- *Autarquía*: cada país se sitúa en su equilibrio con triple tangencia, de modo que A queda relativamente especializado en X y B en Y, y el precio relativo de X resulta menor en A que en B[26].

- *Apertura*: si la RRI queda comprendida entre ambos precios de autarquía, los países participan. En ausencia de costes de transporte la RRI es común, lo que produce una convergencia de los precios relativos. Así, el precio de X sube en A, donde era barato, y baja en B, donde era caro. Esa convergencia de los precios de los bienes es el mecanismo que transmitirá sus efectos a los precios de los factores, como muestran los teoremas de Stolper-Samuelson y de igualación del precio de los factores.

De ahí resultan tres conclusiones:

1. *Oferta*: se cumple el teorema, pues cada país se especializa en el bien intensivo en su factor abundante, sin que la especialización llegue a ser completa[27].

2. *Demanda*: ambos consumen X e Y en la misma proporción, sobre el mismo rayo desde el origen, en puntos que solo difieren según su renta. En la figura, ambos países tienen la misma renta a precios internacionales y consumen en el mismo punto C.

3. *Bienestar*: los triángulos de comercio son iguales y ambos países alcanzan curvas de indiferencia superiores a las de autarquía.

Desarrollo gráfico: el teorema de Heckscher-Ohlin se representa en el plano bien X-bien Y (figura B5.4):

1. *Fronteras de posibilidades de producción*, FPP^A y FPP^B: la de A, sesgada hacia el eje de X, y la de B, hacia el eje de Y, cada una hacia el bien intensivo en el factor abundante del país.

2. *Curvas de indiferencia social*, CIS_0 y CIS_1: un mapa común a ambos países, por ser las preferencias idénticas y homotéticas.

3. *Recta de la RRI*, RRI^*: común a ambos países y tangente a las dos fronteras.

4. *Equilibrio*: en autarquía, cada país se sitúa en la tangencia de su frontera con la CIS_0 (puntos 0^A y 0^B). Con apertura, A produce en P^A y B en P^B, y ambos consumen en el punto C, sobre la CIS_1.

[25] Que las FPP se crucen en la figura es arbitrario.
[26] Véase Gandolfo (2014, p. 71).
[27] La especialización incompleta no es un supuesto del modelo, sino el resultado habitual de la FPP cóncava que generan dos factores con intensidades distintas. Solo si la RRI queda fuera del rango de pendientes de la FPP se especializa el país por completo.

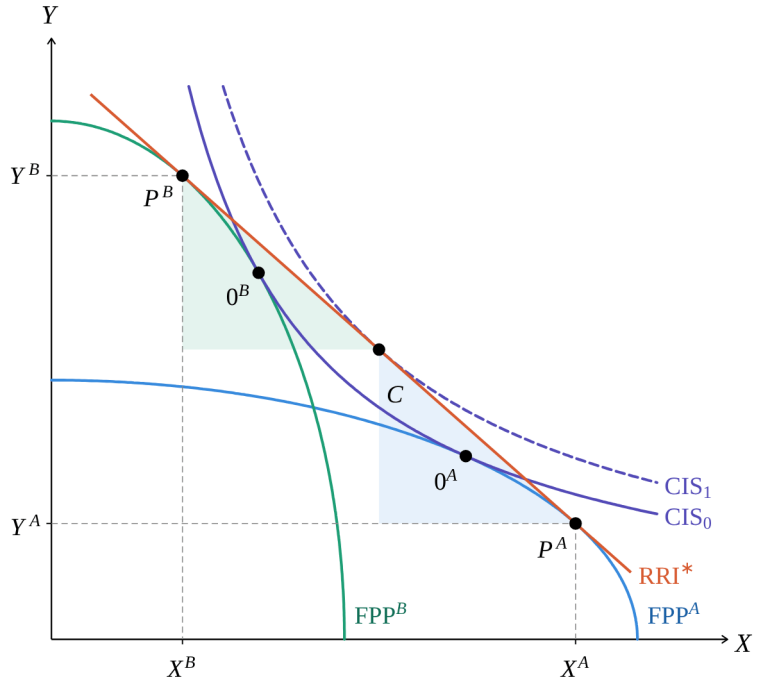

Figura B5.4: Modelo Heckscher-Ohlin-Samuelson. Cada país exporta el bien intensivo en su factor abundante, y el gráfico muestra las ganancias del comercio con especialización incompleta.

Estática comparativa: los efectos de un cambio en los precios relativos y en las dotaciones de factores los recogen los teoremas de Stolper-Samuelson y de Rybczynski.

2.3.3. Conclusiones

Aspectos positivos: el modelo es una de las contribuciones neoclásicas más influyentes, y la idea de que el comercio viene determinado en gran medida por las diferencias de recursos constituye una de las teorías centrales de la economía internacional.

Limitaciones: su contraste empírico ha generado una literatura propia. En 1953, Wassily Leontief contrastó el modelo con la tabla *input-output* estadounidense de 1947 y obtuvo el resultado opuesto al previsto. Estados Unidos, relativamente abundante en capital, exportaba bienes relativamente intensivos en trabajo, mientras que la producción nacional que sustituía a sus importaciones era relativamente intensiva en capital. Entre las explicaciones propuestas figuran la omisión del capital humano (Kenen, 1965), la proximidad del periodo a la Segunda Guerra Mundial, las diferencias de productividad del trabajo y la posible reversión de la intensidad factorial (apartado 3.3.2).

La literatura posterior refinó el diagnóstico, sobre todo mediante el contraste de la versión de Heckscher-Ohlin-Vanek (HOV, apartado 3.3.1):

- *Leamer (1980)*: mostró que, con comercio no equilibrado, el contraste estaba mal planteado y que con el test correcto los datos de 1947 resultaban compatibles con la teoría.

▪ *Bowen, Leamer y Sveikauskas (1987)*: en el primer contraste completo del modelo HOV, con doce factores y veintisiete países, encontraron, en cambio, que las predicciones de signo se cumplían poco más que por azar.

▪ *Trefler (1995)*: documentó el *comercio perdido* (*missing trade*), es decir, que el contenido factorial del comercio observado es muy inferior al previsto. El ajuste mejora, no obstante, al admitir diferencias tecnológicas y sesgo doméstico del consumo.

▪ *Davis y Weinstein (2001)*: mostraron que, al admitir diferencias tecnológicas entre países, la falta de igualación del precio de los factores y los costes del comercio, el modelo HOV acierta en el signo y, aproximadamente, en la magnitud del comercio de servicios factoriales.

2.4. Teoremas derivados del modelo HOS

El modelo dio lugar a tres teoremas que lo completan: Stolper y Samuelson (1941), igualación del precio de los factores (Samuelson, 1948b) y Rybczynski (1955), a los que se añade la relación de reciprocidad que vincula al primero con el tercero.

Junto al propio teorema de Heckscher-Ohlin, estos teoremas forman las cuatro proposiciones centrales del modelo de proporciones factoriales. Dos describen las consecuencias del comercio, y dos, la estática comparativa ante cambios en los precios y en las dotaciones.

2.4.1. Teorema de Stolper y Samuelson (1941)

Origen: lo formularon Wolfgang Stolper y Paul Samuelson, en *Protection and Real Wages* (1941).

Idea: si se producen ambos bienes, un aumento del precio relativo de un bien eleva la retribución real unitaria del factor utilizado intensivamente en su producción y reduce la del otro factor.

Razonamiento: el aumento del precio relativo de X que sigue a la apertura en el país A expande el sector X y contrae el Y. Como los factores se mueven libremente dentro del país, el sector Y libera trabajo y capital en la proporción en que los empleaba, intensiva en trabajo. El sector X, intensivo en capital, emplea a los precios vigentes de los factores menos trabajo por unidad de capital que el Y, por lo que, para absorber el capital liberado, solo necesita una parte del trabajo liberado. En el mercado de factores sobra entonces trabajo y aumenta la demanda relativa de capital, de modo que la retribución del capital sube y el salario baja.

Extensiones:

▪ *Efecto magnificación* (Jones, 1965): el incremento del precio nominal del factor beneficiado es proporcionalmente mayor que el del precio del bien, $\hat{r} > \hat{p}_X > \hat{p}_Y > \hat{w}$, lo que garantiza que su ganancia sea real y no solo nominal (Anexo A.1.1).

▪ *Oposición al libre comercio*: el teorema explica por qué determinados grupos se oponen a la apertura dentro de cada país. Así, en las economías avanzadas, relativamente abundantes en capital y trabajo cualificado, la apertura beneficia a estos y perjudica al trabajo poco cualificado, mientras que en los países en desarrollo ocurre lo contrario.

▪ *Políticas compensatorias*: siendo positivas las ganancias agregadas del comercio, existe en principio una redistribución que deja a todos mejor que en autarquía. Este resultado procede del teorema de las ganancias del comercio (Samuelson, 1962), no del de Stolper-Samuelson, que identifica a quiénes habría que compensar.

Conclusiones:

- *Aspectos positivos*: muestra que el comercio, además de generar ganancias agregadas, las reparte de forma desigual entre factores y produce así ganadores y perdedores identificables dentro de cada país.

- *Limitaciones*: supone que los factores se desplazan sin coste entre sectores y que ambos bienes siguen produciéndose, por lo que describe el ajuste a largo plazo. En cambio, a corto plazo, con factores específicos, ganadores y perdedores se definen por sectores y no solo por factores.

2.4.2. Teorema de igualación del precio de los factores (Samuelson, 1948b)

Origen: lo formuló Paul Samuelson, en *International Trade and the Equalisation of Factor Prices* (1948b).

Idea: en ausencia de especialización completa y bajo los supuestos del modelo, el comercio internacional iguala los precios relativos y absolutos de los factores, pese a que estos sean internacionalmente inmóviles (Anexo A.1.2).

Razonamiento: la ley del precio único, la irreversibilidad factorial y la tecnología idéntica con rendimientos constantes igualan, por este orden, los precios de los bienes, los precios relativos de los factores y sus retribuciones absolutas.

Extensiones:

- *Sustituibilidad*: el libre comercio de bienes es un sustitutivo perfecto de la libre movilidad internacional de factores, pues el intercambio de bienes encierra un intercambio implícito de los factores incorporados en ellos.

- *Reciprocidad de la sustitución*: Robert Mundell (1957) demostró que, a la inversa, la libre movilidad de factores sustituye también al libre comercio de bienes.

Conclusiones:

- *Aspectos positivos*: pone de relieve que el comercio de bienes actúa sobre los precios de los factores aunque estos no crucen fronteras, pues los bienes incorporan los servicios de los factores con que se producen.

- *Limitaciones*: la evidencia muestra, sin embargo, la coexistencia de comercio y movimientos de factores, así como la persistencia de amplias diferencias salariales internacionales. Ambos hechos responden al incumplimiento de varios supuestos, en particular la ausencia de costes de transporte, la identidad tecnológica y la ausencia de especialización completa.

> Los movimientos internacionales de factores se desarrollan en el tema B18, apartados 2 (trabajo) y 3 (capital).

2.4.3. Teorema de Rybczynski (1955)

Origen: lo formuló Tadeusz Rybczynski (1955).

Idea: manteniendo constantes los precios relativos de bienes y factores, un aumento de la dotación de un factor eleva la producción del bien intensivo en él y reduce la del otro bien.

Razonamiento: al mantenerse constantes los precios, los requerimientos unitarios de factores por unidad de producto no varían. El factor adicional solo puede emplearse, por tanto, expandiendo el sector que lo usa intensivamente, y esa expansión exige a su vez cantidades del otro factor que únicamente pueden proceder de la contracción del sector rival.

Gráficamente, en el plano trabajo-capital (figura B5.5), la caja de Edgeworth de producción recoge el reparto de las dotaciones entre sectores:

1. *Caja*: su anchura es la dotación de trabajo, $\bar{L}$, y su altura la de capital. El empleo del sector X se mide desde O_X y el del sector Y desde el origen opuesto, O_Y^0.

2. *Rayos de intensidad factorial*, $(K/L)_X$ y $(K/L)_Y$: salen de cada origen con la pendiente que fijan los precios de los factores, más inclinada la de X por ser intensivo en capital.

3. *Equilibrio*: el corte de ambos rayos (punto 0) emplea todo el trabajo y todo el capital. Con rendimientos constantes, la producción de cada bien es proporcional a la distancia entre su origen y el punto a lo largo de su rayo.

Un aumento de la dotación de capital ($\Delta K > 0$) eleva el origen del sector Y, de O_Y^0 a O_Y^1, y su rayo se traslada en paralelo, de modo que el equilibrio pasa del punto 0 al 1. El alargamiento del rayo de X y el acortamiento del de Y indican que la producción de X crece más que proporcionalmente al capital y la de Y cae.

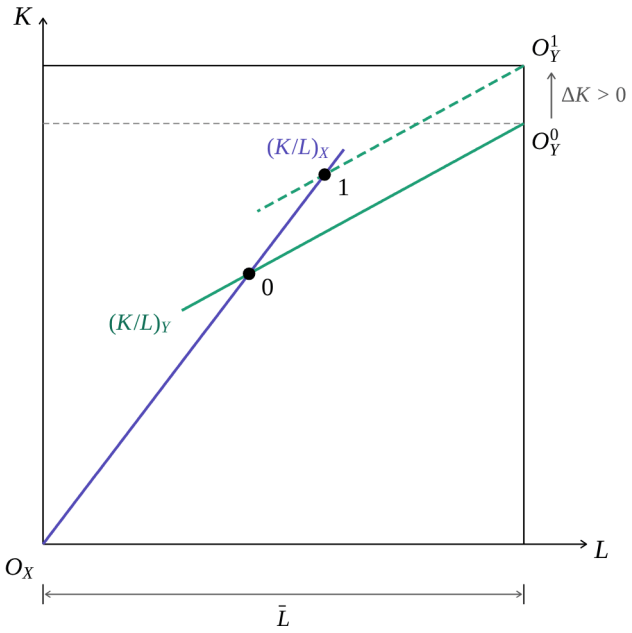

Figura B5.5: Teorema de Rybczynski en la caja de Edgeworth de producción. Con precios constantes, un aumento del capital (de O_Y^0 a O_Y^1) expande más que proporcionalmente el sector X, intensivo en capital, y contrae el sector Y (de 0 a 1).

Extensiones:

▪ *Efecto magnificación* (Jones, 1965): la producción del sector en expansión crece más que proporcionalmente al aumento del factor, $\hat{X} > \hat{K} > 0 > \hat{Y}$ (Anexo A.1.3).

▪ *Aplicaciones*: permite analizar los efectos de los movimientos migratorios o de la acumulación de capital sobre la estructura productiva. Constituye, además, la base analítica de la *enfermedad holandesa*, en la que el auge de un sector intensivo en recursos naturales contrae la producción del resto de sectores comerciables.

Conclusiones: la acumulación de un solo factor tiene efectos sesgados sobre la estructura productiva, y no meramente expansivos, pues, a precios constantes, el crecimiento de un sector se produce necesariamente a costa del otro.

▪ *Aspectos positivos*: traduce los cambios de las dotaciones en cambios de la estructura productiva sin necesidad de especificar la demanda, y es la base de las aplicaciones anteriores y del crecimiento por acumulación de factores (tema B09, apartado 1.4.1).

▪ *Limitaciones*: supone constantes los precios de los bienes, lo que solo vale para un país pequeño. En un país grande, en cambio, si crece el factor intensivo en el bien exportable, la expansión de ese sector deteriora la relación real de intercambio y puede llegar a empobrecer al país (tema B09, apartado 1.3.2).

2.4.4. Relación de reciprocidad de Samuelson (1953)

Origen: Paul Samuelson (1953) formuló esta relación.

Idea: los teoremas de Stolper-Samuelson y de Rybczynski se vinculan, pues el efecto de un aumento del precio de un bien sobre la retribución del factor intensivo equivale al efecto de un aumento de la dotación de ese factor sobre la producción del bien.

Razonamiento: la relación de reciprocidad lo expresa formalmente:

$$\frac{\partial r}{\partial p_X} = \frac{\partial X}{\partial K}$$

donde r es la retribución del capital, p_X el precio del bien X, intensivo en capital, K la dotación de capital y X la cantidad producida de ese bien. La relación vale para cualquier par bien-factor. Por ejemplo, para el trabajo y el bien Y, $\partial w/\partial p_Y = \partial Y/\partial L$.

Conclusiones:

▪ *Aspectos positivos*: basta conocer uno de los dos teoremas para deducir el otro, ya que ambos son las dos caras de una misma matriz de derivadas cruzadas.

▪ *Limitaciones*: relaciona solo efectos marginales en torno al equilibrio y descansa en los mismos supuestos que los dos teoremas, por lo que comparte sus límites.

2.5. Modelo de factores específicos (Samuelson, 1971, y Jones, 1971)

Origen: lo desarrollaron Paul Samuelson (1971) y Ronald Jones (1971), y Michael Mussa (1974) lo extendió al ajuste hacia el largo plazo.

Idea: el modelo de los factores específicos, también llamado de Ricardo-Viner, puede verse como el modelo HOS llevado al corto plazo, un periodo en que la movilidad dentro del país no alcanza a todos los factores. Así, mientras que el trabajo puede desplazarse libremente entre sectores, el capital y la tierra están "atrapados" en sus usos actuales.

2.5.1. Supuestos

Entorno:

- *Bienes*: un país pequeño produce dos, el exportable X (p. ej. manufacturas) y el importable Y (p. ej. alimentos).

- *Arancel*: protege el bien Y frente al sector exterior, de modo que $p_Y = p_Y^*(1+t)$.

- *Otras imperfecciones*: no hay ninguna, salvo el arancel.

- *Dotaciones sectoriales*: no son necesariamente iguales, por lo que tampoco lo son las productividades marginales de ambos sectores. Esa diferencia determina la posición inicial de las curvas de demanda de trabajo.

- *Dotación de trabajo*: equivale a la oferta de trabajo y se considera constante y exógena en todo momento (no existe crecimiento poblacional), por lo que el análisis puede centrarse en la demanda de trabajo de ambos sectores.

Tecnología: la tecnología es no lineal y neoclásica de buen comportamiento en cada sector, $X = f(K, L_X)$, con $f'_{L_X} > 0$ y $f''_{L_X} < 0$, e $Y = g(T, L_Y)$, con $g'_{L_Y} > 0$ y $g''_{L_Y} < 0$.

Mercados: hay competencia perfecta en el mercado de bienes y servicios y en el mercado de factores productivos. Por lo tanto, los factores se retribuyen de acuerdo con el valor de su productividad marginal, $p_X \cdot f'_{L_X} = w$ y $p_Y \cdot g'_{L_Y} = w$. Junto al trabajo, factor móvil entre sectores, existen dos factores específicos, capital y tierra, totalmente inmóviles. Cada uno de ellos se usa solo en un sector (p. ej. la tierra en la producción de alimentos y el capital en la de manufacturas).

2.5.2. Modelo

Planteamiento: el trabajo se desplaza entre sectores hasta igualar su retribución, con pleno empleo,

$$p_X \cdot f'_{L_X} = w = p_Y \cdot g'_{L_Y}, \qquad L_X + L_Y = \bar{L}$$

donde $p_Y = p_Y^*(1+t)$ es el precio interior del bien protegido y $\bar{L}$ la dotación total de trabajo.

Equilibrio: ambas condiciones determinan el salario y el reparto del empleo, y cada factor específico percibe el residuo del valor de la producción de su sector, $r_K = (p_X X - wL_X)/K$ y $r_T = (p_Y Y - wL_Y)/T$.

Desarrollo gráfico: en el plano trabajo-salario (figura B5.6), la anchura del gráfico es la dotación total de trabajo, $\bar{L}$, y cada sector mide su empleo desde un origen opuesto, O_X y O_Y:

1. *Demanda de trabajo del sector X*, $p_X \cdot f'_{L_X}(L_X)$: coincide con el valor de la productividad marginal del trabajo, porque la condición de primer orden del problema de la empresa exige que el salario nominal sea igual a ese valor. Es decreciente, por la productividad marginal decreciente, y convexa, supuesto habitual que cumplen tecnologías como la Cobb-Douglas (la mera concavidad de la función de producción solo garantiza que la curva sea decreciente).

2. *Demanda de trabajo del sector Y*, $p^*_Y(1+t) \cdot g'_{L_Y}(L_Y)$: tiene la misma forma, medida desde O_Y, y valora la productividad al precio interior del bien protegido.

3. *Equilibrio*: con protección, el cruce de ambas curvas (punto 0) fija el salario w_0, idéntico en ambos sectores, y el empleo, L^0_X en X y el resto en Y.

Estática comparativa: si uno de los bienes se encarece en términos relativos, la retribución real por unidad del factor específico de su sector aumenta y la del factor específico del otro sector disminuye. Para el factor móvil, a diferencia de los específicos, el resultado es ambiguo. Como el factor móvil se desplaza entre sectores hasta que el salario se iguala, su salario nominal sube porcentualmente menos que el precio del bien encarecido. Su poder de compra cae, por tanto, medido en ese bien y aumenta medido en el otro.

La eliminación del arancel ($\Delta t < 0$) disminuye el precio interno del bien Y y la curva de demanda de trabajo de ese sector se desplaza hacia abajo, hasta $p^*_Y \cdot g'_{L_Y}(L_Y)$:

- *Flexibilidad salarial*: el equilibrio pasa del punto 0 al 1. El salario nominal de la economía disminuye, de w_0 a w_1, y el trabajo se reasigna desde el sector Y hacia el sector X, de L^0_X a L^1_X.

- *Rigidez salarial*: el salario no baja y, por tanto, no amortigua la contracción de la demanda de trabajo del sector importador. Con w_0 fijo, el sector Y emplea hasta el punto $1'$ y surge desempleo.

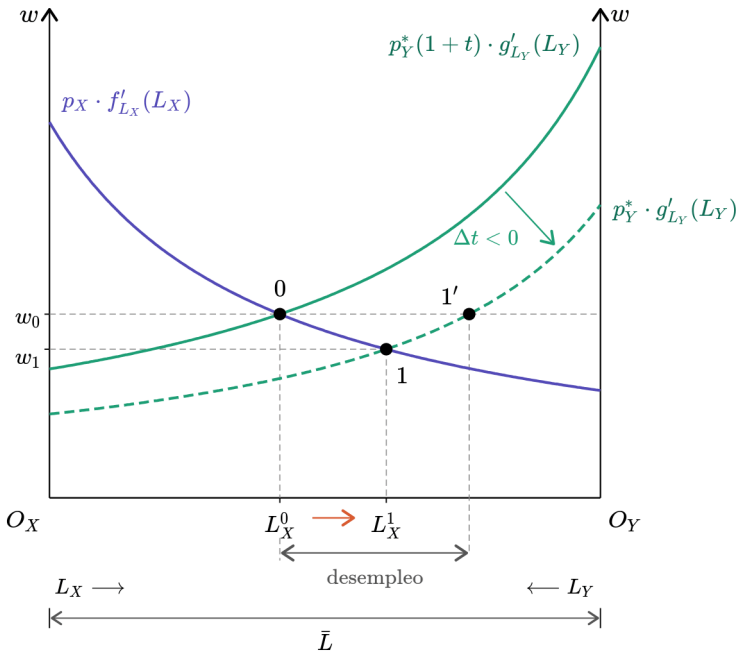

Figura B5.6: Modelo de factores específicos. La eliminación del arancel desplaza hacia abajo la demanda de trabajo del sector Y. Con salarios flexibles el salario cae y el trabajo se reasigna hacia X ($0 \to 1$), mientras que con salario rígido surge desempleo ($1'$).

Con salarios flexibles, el resultado de rentas presenta un efecto magnificación sobre los factores específicos (Jones, 1971). El salario varía según una media ponderada de los precios de los bienes, $\hat{w} = \gamma\,\hat{p}_X + (1-\gamma)\,\hat{p}_Y$, con $0 < \gamma < 1$, y el precio de cada bien, según una media de las retribuciones de sus factores, $\hat{p}_Y = \theta_{LY}\,\hat{w} + \theta_{TY}\,\hat{r}_T$. Con la eliminación del arancel ($\hat{p}_Y < 0 = \hat{p}_X$) resulta

$$\hat{r}_T < \hat{p}_Y < \hat{w} < 0 < \hat{r}_K$$

donde γ depende del peso y de la elasticidad de la demanda de trabajo de cada sector y θ_{jY} es la participación del factor j en el coste de Y. La renta de la tierra, factor específico de Y, cae más que el precio de Y, el salario cae menos (de w_0 a w_1 en la figura) y la renta del capital sube.

Por ello, la liberalización da lugar a ganadores y perdedores claramente identificables por sectores, y no solo por factores, lo que explica que las resistencias políticas al libre comercio se organicen, con frecuencia, sectorialmente. Esta situación se puede ver agravada si existen, además, rigideces internas.

2.5.3. Conclusiones

Aspectos positivos: completa el modelo HOS con un horizonte de corto plazo más realista para analizar la liberalización, pues capta a la vez los efectos distributivos por sectores y, si los salarios son rígidos, el desempleo de ajuste.

Limitaciones: la especificidad de los factores es transitoria, por lo que sus resultados son esencialmente de corto plazo. A medida que los factores específicos se reasignan entre sectores, el análisis converge hacia el del modelo HOS (Mussa, 1974).

2.6. Conclusiones

La teoría neoclásica supera las dos carencias de los clásicos. Por un lado, las curvas de oferta neta de Marshall y Edgeworth dan a las demandas recíprocas de Mill un instrumento para obtener la relación real de intercambio. Por otro lado, la frontera de posibilidades de producción de Haberler, cóncava, da cuenta de la especialización incompleta y hace intervenir las preferencias en el precio de autarquía. Haberler deja sin explicar, sin embargo, por qué difieren los costes relativos entre países.

El modelo HOS lo atribuye a las dotaciones, pues con tecnologías y preferencias comunes cada país exporta el bien intensivo en su factor abundante. Sus teoremas derivan de ahí los efectos de los cambios de precios y de dotaciones. El comercio redistribuye la renta entre factores (Stolper-Samuelson) e iguala sus precios aunque no crucen fronteras (Samuelson, 1948b), y el aumento de una dotación cambia de forma sesgada la estructura productiva (Rybczynski). A corto plazo, con factores específicos, la apertura favorece a los factores específicos del sector exportador y perjudica a los del importador. Esos resultados descansan en supuestos exigentes, y la paradoja de Leontief puso en duda su ajuste a los datos. La sección 3 levanta esos supuestos en el entorno, la oferta y la demanda.

3. Revisiones de la teoría tradicional: levantamiento de supuestos

Los supuestos neoclásicos son, en muchos sentidos, restrictivos. Por ello, diferentes desarrollos posteriores enriquecen las contribuciones anteriores al levantar algunos de sus supuestos[28]. Estas revisiones se agrupan en tres bloques, las relativas al entorno, que incluyen sus implicaciones de política económica, las relativas al lado de la oferta y las relativas al lado de la demanda.

3.1. Revisiones relativas al entorno: competencia imperfecta y distorsiones domésticas

La teoría tradicional presupone un entorno sin distorsiones. Levantar ese supuesto afecta tanto a la explicación positiva del comercio, cuando los mercados no son competitivos, como a su lectura normativa, cuando existen distorsiones internas que el libre comercio no corrige.

3.1.1. Competencia imperfecta

En ausencia de competencia perfecta, los precios relativos no reflejarán necesariamente los costes relativos, pues incorporarán márgenes sobre el coste marginal que pueden diferir entre sectores y entre países. Por lo tanto, la dirección de los flujos comerciales ya no se guía necesariamente por los costes comparativos.

En consecuencia, que un país tenga abundancia relativa de capital no implica necesariamente que el bien intensivo en capital le resulte relativamente más barato que en el otro país, de modo que el teorema de Heckscher-Ohlin puede no cumplirse. El tratamiento sistemático de las estructuras de mercado imperfectamente competitivas constituirá, precisamente, el punto de partida de la nueva teoría del comercio internacional (tema B06).

3.1.2. Teoría de las distorsiones domésticas (Bhagwati y Ramaswami, 1963)

Origen: la ausencia de distorsiones en la teoría tradicional justifica implícitamente el *laissez faire* (la no intervención en política comercial). Cuando estas existen (externalidades, rigideces salariales, fallos en los mercados de factores), la prescripción librecambista debe matizarse, y Jagdish Bhagwati y V. K. Ramaswami (1963) estudian qué instrumento las corrige mejor.

Idea: la política óptima (*first best*) es la que ataca la distorsión en su origen (p. ej. una subvención a la producción o al empleo), y el arancel constituye, a lo sumo, un instrumento de *second best*.

Razonamiento: el óptimo de una economía abierta exige que se igualen la relación marginal de transformación interna, la relación a la que el comercio transforma un bien en otro (la RRI, en el país pequeño) y la relación marginal de sustitución de los consumidores. Una distorsión interna, como una externalidad en la producción, separa la primera de las otras dos. Un impuesto o una subvención a la producción, o al uso del factor si la distorsión está en su mercado, vuelve a igualarlas. El arancel también corrige la producción, pero al precio de introducir una distorsión nueva en el consumo, pues los consumidores pasan a igualar su relación marginal de sustitución a un precio interno distinto de la RRI. Por ello el arancel puede incluso dejar al país peor que con libre comercio.

[28] Los intentos de extender el teorema de Heckscher-Ohlin y el de igualación del precio de los factores al caso de múltiples bienes, factores y países han generado una literatura inmensa (Gandolfo, 2014, p. 456).

Extensiones: Harry Johnson (1965) y Bhagwati (1971) generalizaron el resultado en la *teoría general de las distorsiones*, según la cual cada distorsión se corrige con un impuesto o una subvención dirigidos directamente a su origen. De ahí que el arancel solo sea la política óptima ante una distorsión exterior, el poder de mercado del país en el comercio (arancel óptimo, apartado 2.1.2).

Conclusiones:

- *Aspectos positivos*: ordena los instrumentos de política según la distorsión que corrigen y muestra que la mayoría de los argumentos a favor de la protección justifican, en realidad, una intervención interna y no comercial.

- *Limitaciones*: exige identificar la distorsión y medir su magnitud, y supone que las subvenciones pueden financiarse sin coste, cuando recaudar los ingresos que exigen introduce nuevas distorsiones.

3.2. Revisiones por el lado de la oferta: bienes y tecnología

Por el lado de la oferta, las revisiones afectan a la homogeneidad de los bienes, a la tecnología y al número de bienes, con la cadena de ventajas comparativas, mientras que las relativas a los factores se tratan después.

3.2.1. Relativas a la homogeneidad de los bienes: modelo neo-HOS (Falvey, 1981)

Origen: Rodney Falvey (1981) desarrolla un modelo que se aleja del modelo HOS únicamente en dos supuestos.

Idea: las diferencias de dotaciones factoriales explican también el comercio intraindustrial de variedades de distinta calidad.

Supuestos:

- *Entorno*: dos países, dos bienes y dos factores, en competencia perfecta. Un bien está verticalmente diferenciado en dos calidades, inferior y superior. La calidad de cada variedad aumenta con su intensidad de capital, de modo que la superior es intensiva en capital.

- *Mercado de factores*: el capital es específico de cada industria (puede moverse entre las distintas calidades de un mismo bien, pero no entre sectores).

Razonamiento: si sucede que $(r/w)^A < (r/w)^B$, entonces el país A será relativamente abundante en capital y, en consecuencia, contará con ventaja comparativa en el bien de calidad superior. Así, exportará este último e importará el de calidad inferior, lo que da lugar a comercio intraindustrial, en el que ambos países exportan e importan, simultáneamente, variedades del mismo bien.

Conclusiones:

- *Aspectos positivos*: llega a la misma conclusión que el teorema de Heckscher-Ohlin, pero considera la diferenciación vertical de productos en el ámbito internacional. Permite así explicar el comercio intraindustrial con supuestos muy próximos a los del modelo HOS original, sin abandonar la competencia perfecta ni los rendimientos constantes.

- *Limitaciones*: la principal es que la diferenciación vertical genera, en la práctica, importantes costes fijos (diseño, calidad, marca) que impiden mantener en el tiempo la estructura de competencia perfecta propia del modelo HOS.

3.2.2. Relativas a la tecnología

Las explicaciones tradicionales de los patrones de comercio, ya sea en los modelos ricardianos o en el modelo HOS, se basan en la existencia de rendimientos constantes a escala. Sin embargo, buena parte de la producción moderna se caracteriza por la presencia de rendimientos crecientes.

Los rendimientos crecientes pueden ser de dos tipos con implicaciones muy distintas:

- *Economías de escala externas*: asociadas a la concentración geográfica de una industria y no al tamaño de la empresa, resultan compatibles con la competencia perfecta y dan lugar a patrones de comercio potencialmente arbitrarios, dependientes de la historia.

- *Economías de escala internas*: en cambio, son incompatibles con la competencia perfecta y exigen modelizar mercados imperfectos. La nueva teoría del comercio internacional introduce el papel de los rendimientos crecientes a escala y de otras dimensiones tecnológicas del coste como causa autónoma del comercio, independiente de la ventaja comparativa (tema B06, apartado 1.2.2).

3.2.3. Relativas al número de bienes: cadena de Jones o cadenas de ventajas comparativas (Jones, 1956; Deardorff, 1979)

Origen: Ronald Jones (1956) extiende el análisis a n bienes, con dos factores y dos países.

Idea: la *proposición de la cadena de Jones* parte de ordenar los bienes según su intensidad en el factor relativamente abundante del país, de más a menos intensivos. Establece que, en esa ordenación, los que el país exporta en términos netos quedan todos por delante de los que importa.

Razonamiento: la ventaja comparativa del país en cada bien es tanto mayor cuanto más intensivo es el bien en su factor abundante, de modo que decrece a lo largo de la cadena. Por tanto, la dotación factorial determina no ya qué bien se exporta, sino en qué punto de la cadena de ventajas comparativas se "corta" entre exportables e importables. En el continuo ricardiano del Anexo A.3, ese corte lo fija el salario relativo.

Extensiones: la proposición no se cumple con carácter general cuando el libre comercio iguala los precios de los factores. En cambio, Deardorff (1979) mostró que, si los precios de los factores difieren entre países, no hay obstáculos al comercio y solo se intercambian bienes finales, los intercambios siguen la ordenación por intensidad factorial. La cadena resiste a los obstáculos al comercio o a los bienes intermedios por separado, pero no a ambos a la vez, y con muchos países basta con los obstáculos al comercio para romperla.

Conclusiones:

- *Aspectos positivos*: generaliza el teorema de Heckscher-Ohlin a muchos bienes y traslada su predicción a la ordenación de los bienes por intensidad factorial.

- *Limitaciones*: con más de dos factores los bienes ya no pueden ordenarse por una única intensidad, y la predicción contrastable pasa a ser el contenido factorial del comercio.

3.3. Revisiones por el lado de la oferta: factores

Las revisiones relativas a los factores levantan, una a una, tres hipótesis del modelo HOS: el número de factores (dos), la ausencia de reversión de las intensidades factoriales y la inmovilidad internacional de los factores.

3.3.1. Multiplicidad de factores (Vanek, 1968)

Origen: Jaroslav Vanek (1968) parte de los mismos supuestos que el teorema original de Heckscher-Ohlin pero considera, además, varios países y varios factores de producción.

Idea: este modelo reformula el teorema de Heckscher-Ohlin a partir del concepto de *contenido factorial* del comercio, entendido este como la cantidad de factor incorporada en los bienes exportados menos la cantidad de factor incorporada en los bienes importados.

Razonamiento: de su aportación se desprende el llamado *teorema de Heckscher-Ohlin-Vanek* (1968), según el cual cada país vende en términos netos al exterior los servicios de sus factores relativamente abundantes y compra, también en términos netos, los de sus factores relativamente escasos. Con tecnologías idénticas, precios de los factores igualados y preferencias homotéticas idénticas, el contenido factorial neto del comercio del país A es

$$\mathbf{F}^A = \mathbf{V}^A - s^A \mathbf{V}^w$$

donde $\mathbf{F}^A$ es el vector de contenido factorial neto de su comercio, $\mathbf{V}^A$ y $\mathbf{V}^w$ los vectores de dotaciones del país y del mundo, y s^A su participación en el consumo mundial. Un factor k es, por tanto, relativamente abundante si $V_k^A/V_k^w > s^A$, y entonces su contenido neto exportado es positivo. Con múltiples bienes y factores puede no ser posible predecir la dirección del comercio de cada bien, pero sí su contenido factorial neto, que la ecuación anterior liga a las dotaciones.

Conclusiones:

- *Aspectos positivos*: generaliza la teoría de las proporciones factoriales a cualquier número de bienes, factores y países y la convierte en una predicción contrastable sobre el contenido factorial del comercio, que es la base de los contrastes empíricos modernos del modelo HOS.

- *Limitaciones*: conserva los supuestos más exigentes del modelo HOS, como la identidad de tecnologías, las preferencias idénticas y homotéticas y la igualación del precio de los factores, cuyo incumplimiento explica buena parte de su escaso ajuste empírico (apartado 2.3.3).

3.3.2. Reversión de las intensidades factoriales (Minhas, 1962)

Origen: Bagicha S. Minhas (1962) estima por industrias funciones de elasticidad de sustitución constante (CES) y obtiene elasticidades distintas entre ellas. Halla además una correlación de rangos de 0,73 entre las intensidades de capital de veinte industrias de EE. UU. y Japón, a su juicio lo bastante baja para dar cabida a reversiones.

Idea: si la elasticidad de sustitución difiere entre industrias, un bien puede ser intensivo en capital a unos precios de los factores e intensivo en trabajo a otros, contra el supuesto de irreversibilidad del modelo HOS (apartado 2.3.1).

Razonamiento: con funciones CES, el logaritmo de la relación capital-trabajo de cada industria es lineal en el del precio relativo de los factores, con pendiente igual a su elasticidad de sustitución. Si las elasticidades difieren, las rectas se cortan y, al pasar el corte, se invierte la ordenación de

los bienes por intensidad factorial (figura en el Anexo, apartado A.4). El precio relativo de los bienes deja entonces de corresponderse de forma unívoca con el de los factores, y el comercio no tiene por qué igualar el precio absoluto de estos. Si los dos países quedan a uno y otro lado del corte, el teorema de Heckscher-Ohlin falla para uno de ellos, pues ambos tendrían que exportar el mismo bien.

Conclusiones:

- *Aspectos positivos*: convierte la irreversibilidad en una hipótesis contrastable y ofrece una explicación posible de la paradoja de Leontief.

- *Limitaciones*: Leontief (1964) rehízo los cálculos y concluyó que los datos no confirman a Minhas, y Ball (1966) mostró que el resultado es sensible a la inclusión de la agricultura.

3.3.3. Movilidad internacional de factores productivos (Markusen, 1983)

Origen: James Markusen (1983) revisa la sustituibilidad entre comercio y movilidad de factores que Mundell (1957) había demostrado en el modelo HOS (apartado 2.4.2).

Idea: la sustituibilidad procede del modelo de dotaciones, pero si el comercio se debe a otras causas (rendimientos crecientes, competencia imperfecta, impuestos o diferencias tecnológicas), la movilidad de factores puede aumentar el volumen de comercio, de modo que ambos resultan complementarios.

Razonamiento: en el modelo HOS, la movilidad internacional de los factores no altera las conclusiones cualitativas, porque se limita a reemplazar en parte al comercio de bienes. Si, en cambio, el país A exporta X por su ventaja tecnológica en ese bien, el comercio eleva en A la retribución del factor que X emplea con intensidad. Ese factor acude desde el exterior y, por el teorema de Rybczynski (apartado 2.4.3), expande todavía más el sector exportador, de modo que el volumen de comercio aumenta.

Conclusiones:

- *Aspectos positivos*: explica la coexistencia de comercio y movimientos de factores que muestra la evidencia y advierte de que un patrón de comercio acorde con las dotaciones puede ser consecuencia de esos movimientos, y no su causa.

- *Limitaciones*: el signo de la relación depende de la causa del comercio, que debe identificarse primero. Además, los factores se mueven solo por diferencias de retribución, sin las decisiones de localización de las empresas que estudia la teoría de la empresa multinacional (tema B18, apartado 3.1.3).

3.4. Revisiones por el lado de la demanda

Las revisiones por el lado de la demanda levantan la identidad y la homoteticidad de las preferencias, que en el modelo HOS impiden que la demanda altere el patrón de comercio (apartado 2.3.1). Sin ellas, la demanda puede invertir el patrón que predicen las dotaciones o explicar por sí misma el comercio entre países de renta parecida.

3.4.1. No identidad de las demandas y no homoteticidad de las preferencias

Sea un país A con abundancia física de capital. Cuando las demandas no son idénticas o las preferencias no son homotéticas, puede surgir un sesgo de demanda a favor del bien intensivo en capital que eleve su precio relativo interno. En virtud del teorema de Stolper-Samuelson, ello incrementa la remuneración del capital frente a la del trabajo. El sesgo de demanda puede llegar a invertir el patrón de comercio predicho por el teorema de Heckscher-Ohlin, de modo que el país exportaría el bien intensivo en su factor escaso.

3.4.2. Hipótesis de la demanda representativa (Linder, 1961)

Origen: Staffan Burenstam Linder (1961) propone una explicación del comercio de manufacturas basada en la demanda y reserva la de las dotaciones para el de productos primarios.

Idea: los países exportan los bienes para los que existe una demanda interna amplia, que es la que permite desarrollar la industria, y comercian, sobre todo, con países de nivel de renta similar, cuyas estructuras de demanda se asemejan.

Razonamiento: las empresas desarrollan sus productos para el mercado interno, cuya cercanía les da la información necesaria para concebirlos y mejorarlos, y solo después los venden fuera. La gama de exportaciones potenciales queda así acotada por la demanda interna, y el comercio entre dos países es tanto más intenso cuanto más se solapan sus estructuras de demanda, que dependen sobre todo de la renta per cápita.

Conclusiones:

- *Aspectos positivos*: esta hipótesis, de naturaleza puramente de demanda, anticipa la explicación del intenso comercio entre países desarrollados de dotaciones parecidas, que la teoría de las proporciones factoriales deja sin explicar, y el efecto mercado interno de la nueva teoría (tema B06, apartado 2.1.3).

- *Limitaciones*: es una hipótesis verbal, sin modelo formal, y su contraste empírico ha dado resultados dispares.

Conclusiones

La teoría tradicional atribuye el comercio a las diferencias de precios relativos en autarquía y discrepa sobre su causa. Smith la sitúa en los costes absolutos, y Ricardo demuestra que basta la ventaja comparativa para que el intercambio beneficie a ambos países. Ricardo deja, sin embargo, indeterminada la relación real de intercambio, que Mill fija con las demandas recíprocas y Marshall formaliza con las curvas de oferta neta. Después, Haberler expresa la ventaja comparativa en costes de oportunidad, y el modelo HOS la funda en las dotaciones. Sus teoremas (Stolper-Samuelson, igualación del precio de los factores y Rybczynski) muestran que el comercio, aunque eleve el bienestar agregado, redistribuye la renta entre factores. La paradoja de Leontief obligó a levantar supuestos, y el modelo de factores específicos traslada el conflicto distributivo al sector en el corto plazo. Aun así, queda sin explicar el comercio intraindustrial entre países de dotaciones parecidas y el papel de las multinacionales.

La teoría tradicional nace de la economía clásica, con la ventaja comparativa de Ricardo y las demandas recíprocas de Mill (tema A02). Su argumento normativo descansa en los teoremas del bienestar y en el *second best* (tema A22), y la teoría de las distorsiones domésticas explica por qué una distorsión interna se corrige mejor en su origen que con aranceles. Sus teoremas tienen largo recorrido. Stolper-Samuelson explica quién gana y quién pierde con un arancel (tema B07), Rybczynski, cómo cambia el comercio cuando se acumula un factor (tema B09), y el corolario de Mundell, por qué el comercio y las migraciones pueden sustituirse (tema B18). Por último, las limitaciones que quedan abiertas son el punto de partida de la nueva y la novísima teoría (tema B06).

La competencia de las importaciones chinas ha puesto de relieve cómo se reparten los costes del comercio. Autor, Dorn y Hanson (2013) muestran que esa competencia elevó el paro y redujo la participación laboral y los salarios en los mercados de trabajo locales de EE. UU. más expuestos. Ese resultado está más cerca de la inmovilidad del modelo de factores específicos que del ajuste sin fricciones del modelo HOS. Sigue abierto, además, el debate sobre la vigencia empírica de la ventaja comparativa. Los modelos ricardianos cuantitativos (Eaton y Kortum, 2002; Costinot, Donaldson y Komunjer, 2012) la han convertido en la base de la cuantificación del comercio. La teoría de las dotaciones, por su parte, solo se ajusta a los datos de contenido factorial si admite diferencias tecnológicas entre países (Trefler, 1995).

Anexos

A.1. Demostraciones formales de los teoremas de la teoría neoclásica

Las demostraciones de los tres teoremas enunciados en el apartado 2.4 se reúnen en este anexo para no interrumpir el argumento del cuerpo del tema.

A.1.1. Teorema de Stolper y Samuelson (1941)

El resultado se deriva de las condiciones de beneficio nulo que impone la competencia perfecta, que ligan el precio de cada bien a las retribuciones factoriales ponderadas por sus requerimientos unitarios:

$$p_i = a_{Li}\, w + a_{Ki}\, r$$

donde a_{Li} y a_{Ki} son las cantidades de trabajo y capital necesarias por unidad del bien i. Diferenciando en tasas de variación (Jones, 1965), la variación porcentual del precio de cada bien es un promedio ponderado, por las participaciones en el coste, de las variaciones porcentuales de las retribuciones de los factores:

$$\hat{p}_X = \theta_{LX}\, \hat{w} + \theta_{KX}\, \hat{r}, \qquad \hat{p}_Y = \theta_{LY}\, \hat{w} + \theta_{KY}\, \hat{r}, \qquad \theta_{ji} \in [0,1]$$

donde θ_{ji} es la participación del factor j en el coste del bien i, con $\theta_{Li} + \theta_{Ki} = 1$ y $\theta_{KX} > \theta_{KY}$ por ser X intensivo en capital.

Como una media ponderada queda siempre entre sus componentes, un aumento del precio relativo de X, intensivo en capital, solo resulta compatible con ambas condiciones si r aumenta proporcionalmente más que p_X y w menos que p_Y (cae si p_Y permanece constante). De ahí se siguen simultáneamente el resultado distributivo y el efecto magnificación de Jones (1965):

$$\hat{r} > \hat{p}_X > \hat{p}_Y > \hat{w}$$

donde el acento circunflejo denota variación porcentual. Que $\hat{r}$ supere a $\hat{p}_X$ es lo que garantiza que la ganancia del factor beneficiado sea real y no meramente nominal, pues su retribución crece por encima del precio de cualquier bien.

A.1.2. Teorema de igualación del precio de los factores (Samuelson, 1948b)

La demostración procede en cinco pasos encadenados, uno por cada supuesto que interviene:

1. *Libre comercio*: iguala internacionalmente el precio de los bienes por la ley del precio único, $p_X^A = p_X^B = p_X$ y $p_Y^A = p_Y^B = p_Y$, de modo que el precio relativo p_X/p_Y es común a ambos países.

2. *Irreversibilidad factorial*: como a cada precio relativo de los bienes le corresponde un único precio relativo de los factores, la igualdad del primero entre países arrastra la del segundo, $(w/r)^A = (w/r)^B$.

3. *Tecnología idéntica*: como, en términos relativos, los factores cuestan lo mismo en ambos países, cada bien se fabrica en ellos con la misma relación capital-trabajo óptima:

$$(K/L)_X^A = (K/L)_X^B, \qquad (K/L)_Y^A = (K/L)_Y^B$$

4. *Rendimientos constantes a escala*: bajo este supuesto, las productividades marginales dependen exclusivamente de las proporciones factoriales. Siendo estas idénticas, también lo serán las productividades marginales de cada factor en cada sector, $\text{Pmg}_{L,X}^A = \text{Pmg}_{L,X}^B$.

5. *Competencia perfecta*: el valor de la productividad marginal iguala al precio del factor, $p_X \cdot \text{Pmg}_{L,X}^A = w^A$ y $p_X \cdot \text{Pmg}_{L,X}^B = w^B$. La libre movilidad de los factores dentro de cada país iguala, además, su retribución entre sectores, de modo que $w_X^A = w_Y^A = w^A$.

Siendo iguales las productividades marginales y los precios de los bienes, el precio absoluto de los factores se iguala asimismo entre países, $w^A = w^B$ y $r^A = r^B$.

A.1.3. Teorema de Rybczynski (1955)

Demostración formal: el resultado se deriva de las condiciones de pleno empleo de ambos factores:

$$a_{KX}X + a_{KY}Y = K, \qquad a_{LX}X + a_{LY}Y = L$$

Al mantenerse constantes los precios relativos de bienes y factores, los requerimientos unitarios a_{ij} no varían, y en tasas de variación las condiciones se escriben

$$\lambda_{KX}\,\hat{X} + \lambda_{KY}\,\hat{Y} = \hat{K}, \qquad \lambda_{LX}\,\hat{X} + \lambda_{LY}\,\hat{Y} = \hat{L}, \qquad \lambda_{ji} \in [0,1]$$

donde λ_{ji} es la fracción de la dotación del factor j empleada en el sector i, con $\lambda_{KX} > \lambda_{LX}$ por ser X intensivo en capital. Cada variación de las dotaciones es así una media ponderada de las variaciones de la producción. Un aumento de K con L constante solo puede absorberse, por tanto, expandiendo la producción del bien intensivo en capital y contrayendo la del otro, que libera el trabajo adicional que el sector en expansión necesita. De ahí se sigue el efecto magnificación:

$$\hat{X} > \hat{K} > 0 > \hat{Y}$$

Demostración mediante la caja de Edgeworth: se recoge en la figura B5.5 del apartado 2.4.3.

Demostración mediante la frontera de posibilidades de producción: con más capital, la FPP se expande sobre todo en la dirección del bien X, que es el que usa ese factor de forma intensiva. Con la RRI inicial, la nueva tangencia exige producir más X y menos Y, como predice el teorema.

A.2. Formalización del modelo ricardiano

El modelo ricardiano del apartado 1.2 admite una formalización sencilla. Sea a_{Li} el requerimiento unitario de trabajo del bien i, es decir, las horas de trabajo necesarias para producir una unidad del bien (la inversa del producto medio del trabajo). Con el trabajo como único factor y tecnología lineal:

Precios de autarquía: en competencia perfecta, el precio de cada bien iguala su coste laboral unitario, $p_i = w \cdot a_{Li}$, de modo que el precio relativo de autarquía coincide con el coste de oportunidad, $p_X/p_Y = a_{LX}/a_{LY}$.

Patrón de comercio: el país A tiene ventaja comparativa en X si $a^A_{LX}/a^A_{LY} < a^B_{LX}/a^B_{LY}$. Abierta la economía, si la RRI internacional supera el coste de oportunidad interno del bien X, todo el trabajo del país A se asigna a ese sector, pues su remuneración es allí mayor. La especialización es completa precisamente porque la FPP es lineal.

Salarios relativos: el país con ventaja absoluta en ambos bienes disfrutará de salarios más altos. En concreto, su salario relativo queda acotado entre las ratios de productividad de ambos sectores (en el ejemplo del apartado 1.2, entre 1,5 y 6 veces el salario portugués), mientras que el país atrasado compensa su menor productividad con salarios menores. De aquí se desprende la refutación de dos falacias proteccionistas simétricas. Ni la competencia de los países de salarios bajos es "desleal" para el país avanzado (falacia del *pauper labor argument*), ni el país atrasado queda excluido del comercio por su menor productividad. La razón es que sus salarios se ajustan hasta hacerlo competitivo en el sector de su ventaja comparativa.

Ganancias del comercio: la vía indirecta de obtención del bien importable (producir el bien exportable e intercambiarlo) rinde más unidades que su producción directa siempre que la RRI difiera del coste de oportunidad interno. El resultado demuestra las ganancias del comercio sin recurrir a consideraciones de demanda. Con los datos del ejemplo, supóngase que la RRI internacional se fija en 4 litros de vino por metro de paño, dentro del intervalo (2, 8). Una hora de trabajo inglesa dedicada al paño obtiene entonces, vía intercambio, 24 litros de vino frente a los 12 que produciría directamente. A su vez, una hora portuguesa dedicada al vino obtiene 2 metros de paño frente al metro que produciría por sí misma. Ambos países duplican así el rendimiento de su trabajo en el bien importado.

A.3. Modelo ricardiano con un continuo de bienes (Dornbusch, Fischer y Samuelson, 1977)

Origen: Rudiger Dornbusch, Stanley Fischer y Paul Samuelson (1977) generalizan el modelo ricardiano de dos bienes a un continuo de bienes.

Idea: con muchos bienes, la ventaja comparativa los ordena en una cadena, y el salario relativo, que depende también de la demanda, decide dónde se corta entre los bienes de cada país.

Supuestos:

- *Entorno*: dos países, el nacional, A, y el extranjero, B, en competencia perfecta y sin costes de transporte.

- *Agentes*: preferencias Cobb-Douglas e idénticas en ambos países, con una fracción constante $\alpha(x)$ del gasto dedicada a cada bien.

▪ *Tecnología*: el trabajo es el único factor, con dotaciones L^A y L^B. Los bienes se indexan por $x \in [0,1]$, con requerimientos unitarios de trabajo $a_L^A(x)$ y $a_L^B(x)$, y se ordenan de forma que la ventaja comparativa de A decrezca con x.

Planteamiento: el país A produce el bien x si su coste unitario no supera al de B, $w^A a_L^A(x) \leq w^B a_L^B(x)$, es decir, si

$$\omega \equiv \frac{w^A}{w^B} \leq C(x) \equiv \frac{a_L^B(x)}{a_L^A(x)}, \qquad C'(x) < 0$$

donde ω es el salario relativo de A y $C(x)$ la curva de costes relativos, decreciente por la ordenación de los bienes. Dado ω, el bien frontera $\tilde{x}$ cumple $\omega = C(\tilde{x})$, de modo que A produce los bienes de $[0, \tilde{x}]$ y B los de $(\tilde{x}, 1]$.

Por el lado de la demanda, sea $G(\tilde{x}) = \int_0^{\tilde{x}} \alpha(x)\, dx$ la fracción del gasto mundial que recae en bienes de A. El equilibrio del mercado de trabajo de A, equivalente al de la balanza comercial, exige que su renta iguale el gasto mundial en esos bienes, $w^A L^A = G(\tilde{x})\,(w^A L^A + w^B L^B)$, de donde

$$\omega = \frac{G(\tilde{x})}{1 - G(\tilde{x})} \cdot \frac{L^B}{L^A} \equiv D\left(\tilde{x};\, \frac{L^B}{L^A}\right)$$

donde D es la demanda relativa de trabajo de A. Es creciente en $\tilde{x}$, pues una gama más amplia de bienes de A reduce las importaciones, eleva las exportaciones y, con ello, el salario relativo compatible con la balanza equilibrada.

Equilibrio: el corte de ambas curvas, $C(\tilde{x}) = D(\tilde{x}; L^B/L^A)$, determina a la vez el salario relativo y el bien frontera y, por tanto, el patrón de comercio. A diferencia del modelo de dos bienes, en el que la tecnología solo acota el salario relativo (Anexo A.2), aquí lo fija la interacción con la demanda.

Desarrollo gráfico: en el plano x-ω (figura B5.7):

1. *Curva de costes relativos*, $C(x)$: recoge, para cada bien, la productividad relativa de A, $a_L^B(x)/a_L^A(x)$, es decir, el salario relativo máximo con el que A sigue produciéndolo. Es decreciente, porque los bienes se ordenan de mayor a menor ventaja comparativa de A.

2. *Curva de demanda relativa de trabajo*, $D(x; L^B/L^A)$: creciente desde el origen, porque una gama más amplia de bienes de A eleva la demanda de su trabajo.

3. *Equilibrio*: el corte de ambas curvas (punto 0) fija el salario relativo ω_0 y el bien frontera $\tilde{x}_0$.

Estática comparativa:

▪ *Crecimiento relativo del extranjero*: un aumento de L^B/L^A desplaza hacia arriba la curva D, de $D(x; (L^B/L^A)_0)$ a $D(x; (L^B/L^A)_1)$, y el equilibrio pasa del punto 0 al 1. El salario relativo de A sube, de ω_0 a ω_1, y la gama de bienes de A se estrecha, de $\tilde{x}_0$ a $\tilde{x}_1$. La renta real de A aumenta, por la mejora de su relación real de intercambio, mientras que la renta real por trabajador de B cae.

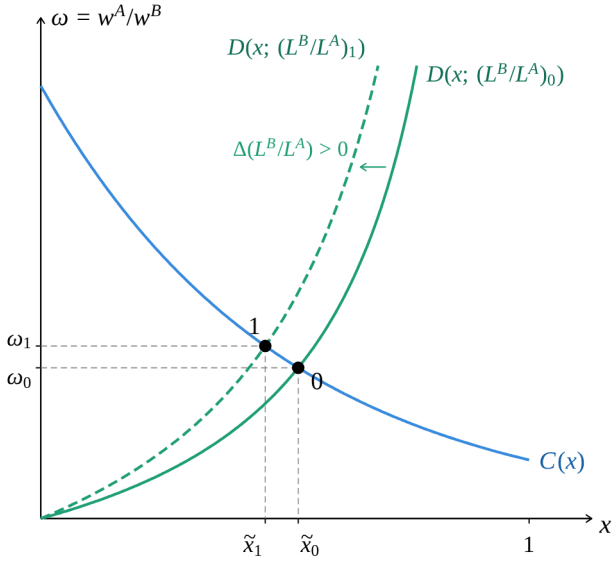

Figura B5.7: Modelo ricardiano con un continuo de bienes (Dornbusch, Fischer y Samuelson, 1977). El corte de $C(x)$ y $D(x; L^B/L^A)$ fija el salario relativo, $\omega = w^A/w^B$, y el bien frontera, $\tilde{x}$, y un aumento de L^B/L^A eleva ω y reduce $\tilde{x}$ (de 0 a 1).

- *Progreso técnico uniforme en el extranjero*: la caída proporcional de $a_L^B(x)$ desplaza $C(x)$ hacia abajo. El salario relativo de A cae, pero menos que los costes relativos, de modo que la gama de bienes de A se reduce y aun así mejora su relación real de intercambio. En consecuencia, ambos países ganan renta real.

- *Costes de transporte*: con costes de tipo *iceberg* ($\tau > 1$ unidades enviadas por unidad recibida), A exporta los bienes con $C(x) \geq \tau\,\omega$ y B los bienes con $C(x) < \omega/\tau$. Los bienes situados entre ambos umbrales no se comercian y cada país los produce para su propio mercado. La franja de bienes no comerciables surge así de forma endógena.

Conclusiones:

- *Aspectos positivos*: extiende la ventaja comparativa a cualquier número de bienes, determina el salario relativo con ayuda de la demanda y sirve de base a los modelos ricardianos cuantitativos (Eaton y Kortum, 2002).

- *Limitaciones*: conserva un único factor y la tecnología lineal, por lo que no explica la distribución de la renta dentro de cada país ni el papel de las dotaciones.

A.4. Reversión de las intensidades factoriales (Minhas, 1962)

Con una función CES en cada industria, $X_i = A_i \left[\alpha_i K_i^{\psi_i} + (1-\alpha_i) L_i^{\psi_i}\right]^{1/\psi_i}$, con $\sigma_i = 1/(1-\psi_i)$, la minimización de costes iguala la relación marginal de sustitución técnica al precio relativo de los factores, de donde

$$\ln\left(\frac{K}{L}\right)_i = \sigma_i \ln\frac{\alpha_i}{1-\alpha_i} + \sigma_i \ln\frac{w}{r}, \qquad i = X, Y$$

Gráficamente, en el plano $\ln(w/r)$-$\ln(K/L)$ (figura B5.8):

1. *Rectas de intensidad factorial*, $\ln(K/L)_X$ y $\ln(K/L)_Y$: son crecientes, porque un trabajo relativamente más caro lleva a ambas industrias a emplear más capital por trabajador, y su pendiente es la elasticidad de sustitución de cada industria. Se dibuja $\sigma_X > \sigma_Y$.

2. *Precio relativo crítico*, $(w/r)^c$: las rectas se cortan. A su derecha X es intensivo en capital, como supone el modelo HOS (apartado 2.3.1), y a su izquierda es intensivo en trabajo.

3. *Países*: A, abundante en capital, tiene en autarquía un salario relativo alto, $(w/r)^A$, a la derecha del corte, y B, abundante en trabajo, uno bajo, $(w/r)^B$, a la izquierda. A exportaría el bien intensivo en capital a sus precios, que es X, y B el intensivo en trabajo a los suyos, que también es X. Como ambos no pueden exportar el mismo bien, el teorema de Heckscher-Ohlin falla para uno de ellos.

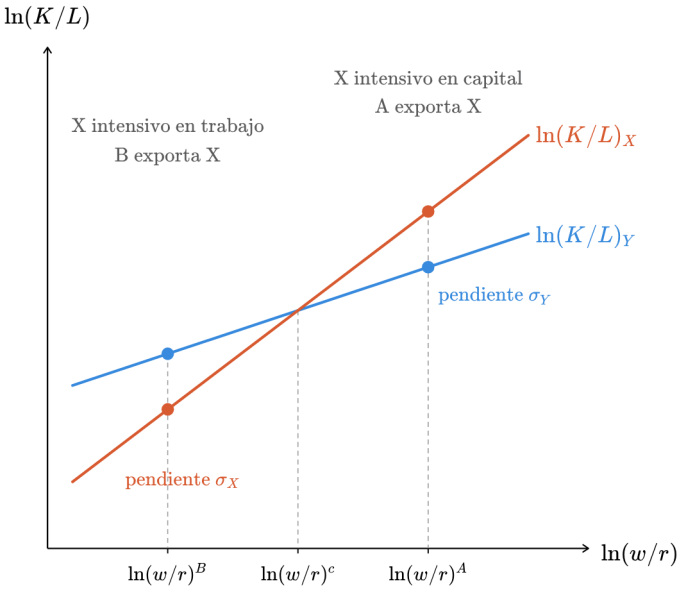

Figura B5.8: Reversión de las intensidades factoriales (Minhas, 1962). Con funciones CES, $\ln(K/L)_X$ y $\ln(K/L)_Y$ son rectas de pendiente $\sigma_X > \sigma_Y$ que se cortan en $(w/r)^c$, de modo que X es intensivo en trabajo a la izquierda del corte y en capital a la derecha. Si $(w/r)^B$ y $(w/r)^A$ quedan a uno y otro lado, ambos países exportan X y el teorema de Heckscher-Ohlin falla para uno de ellos.

B6. Nueva teoría del comercio internacional

Teoría del comercio internacional (II). Nueva teoría del comercio internacional. Especial referencia a la competencia imperfecta, los rendimientos crecientes y la heterogeneidad empresarial.

Introducción

Introducción común. Los temas B05, B06 y B09 comparten una introducción, que se desarrolla al comienzo del tema **B05**. Plantea las tres preguntas de la teoría del comercio internacional (por qué comercian los países, qué intercambian y cómo les afecta) y sus tres etapas, de la teoría tradicional a la nueva teoría y la novísima, que se pregunta por qué comercian las empresas.

La teoría tradicional explica bien el comercio entre países distintos, pero no el que se da entre economías parecidas. Desde los años sesenta, los datos mostraron que buena parte del comercio entre países industrializados era **comercio intraindustrial**, es decir, intercambio de variedades de un mismo producto, cuyo peso permitió medir el índice de Grubel y Lloyd (1975).

La **nueva teoría del comercio internacional** responde a esa evidencia con dos supuestos que la teoría tradicional excluía, las **economías de escala** y la **competencia imperfecta**. Con ellos, el comercio ya no necesita diferencias de tecnología ni de dotaciones, porque basta con que los consumidores valoren la variedad y con que producir a gran escala abarate los costes. Al combinar las economías de escala internas con los costes de transporte y la movilidad de los factores, la **nueva economía geográfica** explica además por qué la producción se concentra en unas regiones y no en otras. Krugman recibió en 2008 el Premio Nobel "por su análisis de los patrones del comercio y de la localización de la actividad económica".

La **novísima teoría** desciende del país y de la industria a la empresa. Los microdatos revelaron que exportar es una actividad excepcional. Así, de los 5,5 millones de empresas que operaban en EE. UU. en 2000, solo un 4 % exportaba, y el 10 % de los exportadores que más vendían concentraba el 96 % de las exportaciones (Bernard, Jensen, Redding y Schott, 2007). El modelo de **heterogeneidad de empresas** de Melitz (2003) explica ese patrón, porque solo las empresas más productivas pueden cubrir los costes fijos de vender en el exterior, y muestra que la apertura eleva la productividad agregada al reasignar recursos hacia ellas y expulsar del mercado a las menos productivas.

Este tema recorre la nueva teoría y la novísima teoría:

- *Fundamentos*: se estudia el punto de partida de las nuevas teorías, que comprende los hechos empíricos que ponían en cuestión las predicciones de la teoría comercial tradicional y algunas teorías precursoras de la nueva teoría.

- *Nueva teoría*: recalca el papel de las economías de escala internas en marcos de competencia monopolística y de oligopolio. Destaca también el de las economías de escala externas o de aglomeración. En la llamada *nueva economía geográfica* (modelo centro-periferia), que se estudia junto con los modelos gravitatorios, la aglomeración resulta de la interacción entre economías de escala internas, costes de transporte y movilidad de factores.

- *Novísima teoría*: se introducen sus modelos, con énfasis en los que derivan del modelo de comercio internacional con heterogeneidad de empresas de Marc Melitz.

1. Fundamentos de la nueva teoría

Diversas aportaciones posteriores a la teoría tradicional están en el origen de la nueva teoría, unas empíricas, que ponían en cuestión las predicciones de la teoría tradicional, y otras teóricas, que destacaban las economías de escala y la competencia imperfecta.

1.1. Punto de vista empírico

La teoría tradicional supone competencia perfecta y bienes homogéneos e idénticos en todos los países. Sin embargo, a mediados de los años setenta se hizo patente la necesidad de contrastar sus predicciones y de acercar la teoría a los datos, fundamentalmente para explicar el notable aumento del comercio intraindustrial.

1.1.1. Medición de la especialización y del comercio intraindustrial: índice de Balassa (1965b) e índice de Grubel y Lloyd (1975)

Cuantificar el comercio intraindustrial exige resolver antes dos cuestiones, cómo delimitar los sectores o industrias y con qué indicador medir su comercio.

Definición de sectores: lo habitual es delimitar cada industria con uno de estos dos criterios:

- *Proceso de producción*: pertenecen a una misma industria los productores que comparten fases de fabricación y que, en consecuencia, combinan los factores en proporciones parecidas (p. ej. coches y camiones).

- *Uso del producto*: forman una misma industria los bienes que el consumidor destina al mismo fin, como un libro de papel y uno electrónico.

En la práctica, las estadísticas de comercio se elaboran conforme a clasificaciones jerárquicas armonizadas, principalmente la *Clasificación Uniforme para el Comercio Internacional* (CUCI o SITC, por sus siglas en inglés) de Naciones Unidas y el *Sistema Armonizado* (SA) de la Organización Mundial de Aduanas.

Para medir los tipos de comercio se emplean dos índices de referencia, uno de especialización y otro de comercio intraindustrial.

Índice de Balassa (1965b): busca determinar los sectores débiles y fuertes de un país a través de la identificación de su *ventaja comparativa revelada* (o desventaja). Formalmente, relaciona el valor relativo de las exportaciones de un sector k de un país i con el valor relativo de las exportaciones de dicho sector k de un grupo de países de referencia:

$$\mathrm{IB}_{ik} = \frac{X_{ik}/X_i}{X_{0k}/X_0} \geq 0$$

donde $X_{ik} \geq 0$ refleja el valor total de las exportaciones del bien k realizadas por el país i y $X_i > 0$, el valor total de todas sus exportaciones. Para el grupo de países de referencia, $X_{0k} > 0$ es el valor total de sus exportaciones del bien k y $X_0 > 0$, el de todas sus exportaciones. Si $\mathrm{IB}_{ik} > 1$, el país tiene ventaja comparativa revelada en el sector y, si $\mathrm{IB}_{ik} < 1$, desventaja comparativa.

El índice "puro" es asimétrico (de 0 a 1 sin especialización y de 1 a ∞ con ella), por lo que se normaliza como $\mathrm{IB}_{ik}^N = (\mathrm{IB}_{ik} - 1)/(\mathrm{IB}_{ik} + 1) \in [-1, +1)$, que vale -1 si el país no exporta el bien.

Índice de Grubel y Lloyd (1975): mide la proporción del comercio de una industria que está "solapada", es decir, la parte de las exportaciones que se corresponde con importaciones simultáneas de productos de la misma industria:

$$\text{IGL}_k = \frac{(X_k + M_k) - |X_k - M_k|}{X_k + M_k} \in [0, 1]$$

donde IGL_k es el índice de comercio intraindustrial de la industria k y X_k y M_k son los valores de las exportaciones y las importaciones de la industria k. Cuanto más próximos entre sí estén los valores de las exportaciones y de las importaciones, más próximo estará el índice a 1 y mayor será el peso del comercio intraindustrial en el comercio total del bien considerado[1]. Para un país en el que solo se exporta o solo se importa un determinado bien, por el contrario, el IGL_k será nulo.

1.1.2. Hechos estilizados

A finales de los años setenta, los datos sobre el comercio internacional discrepaban de la teoría tradicional en cinco dimensiones:

1. *Ventaja comparativa*: los contrastes empíricos de la teoría tradicional arrojaban resultados desconcertantes, el más conocido de los cuales es la *paradoja de Leontief* (1953). Con las tablas *input-output* de EE. UU., Wassily Leontief encontró que las exportaciones estadounidenses eran relativamente más intensivas en trabajo que la producción nacional que sustituía a sus importaciones, justo lo contrario de lo que predecía el teorema de Heckscher-Ohlin para el país más abundante en capital del mundo.

2. *Peso creciente del comercio intraindustrial*: buena parte del crecimiento del comercio de posguerra se producía entre países desarrollados con dotaciones factoriales y tecnologías similares y, además, dentro de las mismas industrias (intercambio bidireccional de variedades de un mismo tipo de producto). Sin embargo, la teoría tradicional predecía justo lo contrario. Los estudios sobre la integración europea (Balassa, 1966) mostraron, además, que el ajuste a la liberalización se producía vía especialización intraindustrial, con costes de ajuste menores a los esperados.

3. *Determinantes del comercio intraindustrial*: según los estudios empíricos de la época (p. ej. Loertscher y Wolter, 1980), el comercio intraindustrial crecía con la proximidad física entre países, que abarata el transporte, y con la semejanza de sus gustos, que facilita el abastecimiento internacional. Era también mayor entre economías grandes, con más margen para diferenciar, y desarrolladas, cuyos consumidores aprecian más la variedad.

4. *Buen ajuste de los modelos gravitatorios*: estudios como el de Jan Tinbergen (1962) mostraban que estos modelos se ajustaban bien a los hechos observados. Los flujos bilaterales crecen con el tamaño económico de los socios y caen con la distancia, con independencia de sus diferencias en dotaciones factoriales.

5. *Comercio norte-sur y comercio norte-norte*: el intercambio entre países desarrollados (norte) y en desarrollo (sur) se ajustaba razonablemente al patrón interindustrial de la teoría tradicional (manufacturas por productos primarios). Representaba, sin embargo, una fracción decreciente del total frente al comercio norte-norte de carácter intraindustrial (punto 2).

[1] En general, los bienes más intensivos en capital y con mayor valor añadido tienen un IGL mayor, mientras que los intensivos en trabajo tienen un IGL menor. Por ejemplo, en EE. UU. el IGL es alto en maquinaria y productos químicos y bajo en confección, calzado y juguetes (Krugman, Obstfeld y Melitz, 2018, cap. 8).

1.2. Punto de vista teórico

Desde el punto de vista teórico, la nueva teoría se apoya en las aportaciones de varios autores de los años cincuenta y sesenta. Se distingue por dos rasgos, la importancia de las economías de escala internas (o rendimientos crecientes) y las estructuras de mercado no competitivas.

1.2.1. Antecedentes: indisponibilidad relativa, desfases tecnológicos y ciclo del producto

Indisponibilidad relativa (Kravis, 1956): según Irving Kravis, los países importan los bienes que no están disponibles en el propio país o que solo lo están a un coste mucho mayor, por la inelasticidad de su oferta interna.

Desfases tecnológicos (Posner, 1961; Hufbauer, 1966): Michael Posner y Gary Hufbauer enfatizan el papel de los desfases tecnológicos (*technology gaps*) como origen del comercio internacional. Como consecuencia de la I+D y de la iniciativa empresarial, los países innovadores producen nuevos bienes y disfrutan de un monopolio temporal hasta que otros países aprenden a producirlos (*imitation lag*). En ese lapso, los países seguidores deben importar dichos bienes y se crea comercio. Estas aportaciones, que integran los llamados *modelos neotecnológicos*, influyeron en el desarrollo posterior de las teorías del crecimiento endógeno en economía abierta (véase el tema B09).

Variabilidad de la intensidad factorial a lo largo del ciclo del producto (Vernon, 1966; Hirsch, 1967): Raymond Vernon y Seev Hirsch se preguntan por qué las innovaciones nacen sobre todo en los países desarrollados y acaban filtrándose después a las economías menos avanzadas. Su respuesta es que cada fase de la vida del producto (introducción, madurez y estandarización) pide una intensidad factorial distinta. Que un país desarrollado o en desarrollo exporte o importe el bien depende, así, de la fase en que este se encuentre.

1.2.2. Importancia de las economías de escala: externas e internas

Economías de escala externas: suelen ser el resultado de la proximidad espacial de las empresas, por lo que también se las conoce como *economías de aglomeración*. Estas economías pueden dar lugar a comercio internacional aunque no difieran las dotaciones factoriales relativas, simplemente por las externalidades positivas que se derivan de la concentración de actividad productiva en clústeres.

El coste medio de cada empresa no depende de su tamaño sino del de la industria nacional, por tres vías que ya señaló Marshall (1890), los proveedores especializados, un mercado de trabajo cualificado común y la difusión del conocimiento. Como las empresas son pequeñas, la industria puede ser competitiva y el precio iguala su coste medio, de modo que estas economías no exigen competencia imperfecta. Sus consecuencias para el comercio son tres:

1. *Dependencia de la historia*: el país que creó antes la industria y alcanzó gran escala produce más barato y conserva la ventaja, aunque otro país pudiera producir con menor coste a la misma escala. Por tanto, el patrón de especialización lo fija una ventaja inicial, a menudo accidental, y no las dotaciones.

2. *Posibles pérdidas del comercio*: si la industria se queda en el país menos eficiente, el otro puede pagar por el bien importado más de lo que le costaría producirlo en autarquía y salir perdiendo con el comercio (Graham, 1923; Ethier, 1982).

3. *Economías dinámicas*: cuando el coste cae con la producción acumulada (curva de aprendizaje), la ventaja del pionero es aún mayor, lo que fundamenta el argumento de la industria naciente (tema B08, apartado 2.3.2).

Economías de escala internas: su papel en el crecimiento del comercio de posguerra entre países desarrollados ya se había señalado antes de la nueva teoría. Lo propio de los nuevos desarrollos es incorporarlas al análisis formal como rendimientos crecientes a escala en la empresa.

Los costes medios son, en consecuencia, decrecientes a medida que aumenta la escala de producción de la empresa, por la indivisibilidad de factores o por la existencia de costes fijos. Formalmente, esta idea se introduce en los modelos a través de funciones de coste lineales con un coste fijo, $\mathrm{CT}(x) = c_0 + cx$.

Para aprovechar sus economías de escala internas, cada país puede especializarse en un número limitado de bienes, producirlos a mayor escala e intercambiarlos, de modo que la producción mundial aumenta sin sacrificar variedades. Las economías de escala internas explican así que haya comercio internacional aunque no exista ventaja comparativa.

Con todo, esto supone que habrá ganadores y perdedores. La gran empresa, al aprovechar sus economías de escala, produce con costes más bajos que la pequeña, y esta ve mermada así su competitividad. Además, como con economías de escala internas el coste marginal queda por debajo del coste medio y fijar el precio igual al coste marginal generaría pérdidas, la estructura de mercado compatible con ellas es la de competencia imperfecta.

1.2.3. Importancia de la estructura de mercado: competencia monopolística y mercados oligopolísticos

Este segundo rasgo se deriva del anterior.

Competencia monopolística: por un lado, se consideran entornos en los que un elevado número de empresas con capacidad para ejercer poder de mercado sobre su nicho o clientela, mediante la diferenciación de sus productos, puede afectar al comercio internacional. Para modelar la competencia monopolística existen dos visiones teóricas[2], la horizontal (variedades) y la vertical (calidad), que se desarrollan en el tema A18.

Mercados oligopolísticos: por otro lado, se estudia cómo la interdependencia estratégica entre empresas de distintos países puede explicar el comercio internacional.

1.3. Conclusiones

La teoría tradicional descansa en la competencia perfecta y en bienes homogéneos, y la evidencia de finales de los años setenta resultaba difícil de conciliar con ella. La paradoja de Leontief contradecía el teorema de Heckscher-Ohlin. El comercio intraindustrial crecía, además, entre países desarrollados con dotaciones y tecnologías similares, sobre todo si eran próximos, grandes y de gustos parecidos. Finalmente, los modelos gravitatorios explicaban bien los flujos con el tamaño y la distancia, sin recurrir a las dotaciones.

En el plano teórico, tras precursores como los desfases tecnológicos y el ciclo del producto, la explicación se apoya en las economías de escala. Las externas o de aglomeración generan

[2]Dentro de la visión horizontal existen, a su vez, dos enfoques. En el enfoque neohotelliano o de Lancaster (variedad de gustos), cada consumidor tiene una variedad ideal y demanda la más próxima a ella de las disponibles. Como los consumidores son diferentes entre sí, hay demanda para todas las variedades. En el enfoque neochamberliniano o de Spence-Dixit-Stiglitz (gusto por la variedad), los consumidores tienen preferencia por la variedad, de modo que un mismo consumidor prefiere consumir varias variedades al mismo tiempo.

comercio aunque las dotaciones no difieran. Por su parte, las internas permiten que cada país produzca menos bienes a mayor escala y comercie sin ventaja comparativa, y, como sitúan el coste marginal por debajo del medio, exigen competencia imperfecta. La organización industrial permitió formalizar esa combinación. La sección 2 lo hace primero con competencia monopolística y con oligopolio y después con la nueva economía geográfica.

2. Nueva teoría del comercio internacional

La nueva teoría formaliza los dos rasgos del apartado 1.2, primero con competencia monopolística. Paul Krugman (1979b) parte para ello del modelo neochamberliniano y lo extiende a la economía abierta. Siguen los modelos de oligopolio y, por último, el equilibrio integrado con dotaciones distintas, los modelos gravitatorios y la nueva economía geográfica.

2.1. Modelos de competencia monopolística en economía abierta

Los modelos de competencia monopolística se distinguen por el tipo de diferenciación del producto. Los tres primeros, desde el marco neochamberliniano hasta el efecto mercado interno de Krugman, lo hacen por variedades (**diferenciación horizontal**), y el de Shaked y Sutton, por calidades (**diferenciación vertical**).

2.1.1. Diferenciación horizontal: modelo neochamberliniano de competencia monopolística (Spence, 1976; Dixit y Stiglitz, 1977)

> Sobre el modelo de Dixit y Stiglitz en economía cerrada y su bienestar, véase el tema A18, apartado 2.2.

Origen: las ideas de Edward Chamberlin fueron reformuladas con una mayor microfundamentación por Michael Spence (1976) y, un año más tarde, por Avinash Dixit y Joseph Stiglitz (1977) en un modelo de equilibrio general con agentes representativos.

Idea: en presencia de economías de escala internas y de preferencia por la variedad, el número de variedades que sostiene una economía queda limitado por el tamaño de su mercado.

Supuestos:

- *Entorno*: existe pleno empleo de recursos.
- *Agentes*:

- *Empresas*: se modelan mediante una empresa representativa, dada la simetría del planteamiento.
    - *Variedades*: cada empresa fabrica y vende una variedad diferenciada, de modo que hay tantas variedades como empresas. Las variedades son sustitutivos imperfectos, con elasticidad de sustitución finita, $\sigma \in (1, \infty)$.
    - *Miopía*: cada empresa toma el precio de las demás como dado y no considera el efecto de sus decisiones sobre las rivales (variación conjetural nula). Su curva de ventas presenta, en consecuencia, pendiente negativa aunque poco pronunciada en el plano (X_i^d, p_i).
- *Consumidores*: preferencias idénticas y equidistribuidas, de donde se sigue que las curvas de demanda individuales, y por extensión las de ventas de las empresas, son idénticas. Los consumidores presentan además *preferencia por la variedad*, es decir, la utilidad del consumo

de n variedades supera a la de $n-1$[3], de modo que cada consumidor demanda todas las variedades existentes y adquirirá también las de nueva creación. El grado de sustituibilidad se recoge en el parámetro $\theta \in (0,1)$, cuya restricción a valores inferiores a la unidad excluye que la elasticidad de sustitución, $\sigma = 1/(1-\theta)$, sea infinita y que las variedades sean, por tanto, sustitutivos perfectos.

▪ *Tecnología*: idéntica, lo que implica la misma función de costes para todas, lineal y con rendimientos crecientes a escala, que emplea únicamente trabajo y de la que resultan economías de escala internas.

▪ *Mercado de bienes*: competencia monopolística en el mercado de bienes. Cada empresa es relativamente grande respecto de su nicho o clientela y dispone, por ello, de poder para fijar el precio de su variedad.

Consumidores: el consumidor representativo elige cuánto demanda de cada variedad.

▪ *Planteamiento*: maximiza una utilidad con elasticidad de sustitución constante (CES) sobre las n variedades sujeto a su restricción presupuestaria:

$$\max_{x_i} \; u = \left(\sum_{i=1}^{n} x_i^{\theta} \right)^{1/\theta} \qquad \text{s.a.:} \quad \sum_{i=1}^{n} p_i x_i = M_0$$

donde $x_i > 0$ es la cantidad demandada de cada variedad, $\theta \in (0,1)$ el parámetro de sustituibilidad[4], $p_i > 0$ el precio de cada variedad y $M_0 > 0$ la renta monetaria.

▪ *CPO*: de las condiciones de primer orden[5] se obtienen tres expresiones:

$$x_i = (\lambda p_i)^{1/(\theta-1)} \tag{1}$$

$$\frac{x_i}{x_j} = \left(\frac{p_i}{p_j} \right)^{1/(\theta-1)} \tag{2}$$

$$x_i^d = \left(\frac{p_i}{P} \right)^{-\sigma} \frac{M_0}{P} \tag{3}$$

donde λ es el multiplicador de Lagrange, es decir, el precio sombra de la renta, y P el índice de precios de las variedades.

La expresión (1) es la *demanda frischiana*, que relaciona la cantidad demandada con el precio y el precio sombra de la renta.

[3] Spence, Dixit y Stiglitz argumentan que la preferencia por la variedad se encuentra ya implícita en las curvas de indiferencia tradicionales cuando estas son convexas hacia el origen. La razón es que entonces se cumple el axioma de convexidad y las combinaciones equilibradas de bienes son preferidas a las combinaciones extremas. La función de utilidad con elasticidad de sustitución constante recoge formalmente esta idea, pues con ella, *ceteris paribus*, el índice de utilidad crece cuantas más variedades se consumen. En efecto, si todas las variedades tienen el mismo precio p, de (3) se deduce que la renta se divide a partes iguales, $x_i = M_0/(np)$. Sustituyendo en la función de utilidad se obtiene $u^* = n^{(1-\theta)/\theta} \cdot M_0/p$, que es creciente con el número de variedades.

[4] Por extensión, la preferencia por la variedad queda reflejada en $1-\theta$.

[5] El factor $\left(\sum_j x_j^{\theta}\right)^{(1-\theta)/\theta}$ que aparece en la utilidad marginal de cada variedad es común a todas ellas, por lo que puede absorberse en el multiplicador λ (renormalizado) para obtener (1).

La expresión (2) es la demanda relativa de dos variedades $i \neq j$, que también puede escribirse como $x_i/x_j = (p_j/p_i)^{\sigma}$[6] y refleja la dependencia de la senda de expansión del consumo respecto de los precios relativos y del grado de sustituibilidad[7].

La expresión (3) es la *demanda marshalliana*. Según ella, la cantidad demandada de cada variedad decrece con su precio relativo y crece con la renta real, $x_i = f(p_i^-, P^+, M_0^+)$.

▪ *CSO*: la función de utilidad CES es estrictamente cuasicóncava para $\theta \in (0,1)$ y la restricción presupuestaria es lineal, por lo que las condiciones de primer orden son suficientes para un máximo.

▪ *Función de demanda agregada*: dado que los consumidores son también trabajadores y son idénticos, la demanda agregada resulta de multiplicar la individual por el tamaño de mercado, $X_i^d = x_i^d \cdot L$, siendo $L > 0$ el número de trabajadores-consumidores.

Empresas: la simetría en costes permite trabajar con una empresa representativa, que fija el precio de su variedad.

▪ *Planteamiento*: elige el precio de su variedad que maximiza su beneficio, tomando como dados el salario y el índice de precios (miopía), sujeta a la demanda isoelástica de mercado que resulta del lado de los consumidores:

$$\max_{p_i} \ \pi_i = p_i x_i^s - w(l_0 + l x_i^s) \quad \text{s.a.:} \quad x_i^s = L\left(\frac{p_i}{P}\right)^{-\sigma} \frac{M_0}{P}$$

donde π_i es el beneficio, x_i^s la producción de la empresa (en equilibrio, $x_i^s = X_i^d$), $w > 0$ el salario nominal, $l_0 > 0$ los empleados fijos y $l > 0$ los empleados por unidad de producto. La función de costes, $\mathrm{CT}(x_i^s) = w(l_0 + l x_i^s)$, tiene coste medio decreciente y coste marginal constante, wl, lo que ilustra las economías de escala internas[8].

▪ *CPO*: como la producción depende del precio a través de la demanda, $\partial \pi_i/\partial p_i = x_i^s + (p_i - wl)\,\partial x_i^s/\partial p_i = 0$, y con $\partial x_i^s/\partial p_i = -\sigma x_i^s/p_i$ la condición se reduce a $(p_i - wl)/p_i = 1/\sigma$. El ingreso marginal, $p_i(\sigma-1)/\sigma$, iguala al coste marginal, wl, y el índice de Lerner es la inversa de la elasticidad de la demanda.

▪ *CSO*: la derivada del beneficio es proporcional a $\sigma wl - (\sigma-1)p_i$, positiva por debajo del precio óptimo y negativa por encima, así que este es el máximo global. Por eso se exige $\sigma > 1$, pues con $\sigma \leq 1$ el beneficio crecería siempre con el precio.

▪ *Solución*: despejando el precio se obtiene la regla de fijación de precios:

$$p_i = \frac{\sigma}{\sigma - 1} \cdot wl = \frac{1}{\theta} \cdot wl, \qquad \sigma > 1 \tag{4}$$

[6] A su vez, (2) se puede reescribir como $x_i = x_j p_j^{\sigma} \cdot p_i^{-\sigma} = k\, p_i^{-\sigma}$, es decir, como una función de demanda isoelástica del tipo $x = k \cdot p^{-\sigma}$, con $k > 0$, cuya elasticidad-precio es constante e igual, en valor absoluto, a la elasticidad de sustitución, $\varepsilon_{x_i p_i} = -\sigma$. Por ello, cuanto mayor es la preferencia por la variedad $(1-\theta)$, menor es σ y menos elástica (más rígida) es la demanda de cada variedad, con $\sigma \to 1$ en el límite $\theta \to 0$.

[7] Para obtener (3) a partir de (2) es necesario usar el gasto total en todas las variedades y un índice de precios CES del tipo $P = \left(\sum_i p_i^{\theta/(\theta-1)}\right)^{(\theta-1)/\theta}$.

[8] Aunque hay economías de escala internas, no hay economías de alcance, de modo que cada empresa fabrica una sola variedad. Además, dada la preferencia por la variedad, producir la misma que un competidor supondría perder beneficios frente a mantenerse en el propio nicho.

Según (4), la estrategia óptima consiste en aplicar un margen μ, con $1+\mu = \sigma/(\sigma-1) > 1$, sobre el coste marginal wl, con independencia de los precios ajenos, y el margen es tanto mayor cuanto menor es la elasticidad.

Equilibrio: el precio superior al coste marginal podría generar beneficios extraordinarios, pero la libre entrada los elimina. Imponiendo $\pi(p_i^*, x_i^s) = 0$ se obtiene[9]:

$$x_i^s = \frac{l_0}{l}(\sigma - 1), \quad \text{es decir,} \quad x_i^s = f(l_0^+, \sigma^+, l^-)$$

La escala de cada empresa queda determinada exclusivamente por la tecnología (l_0, l) y por la elasticidad de sustitución, por lo que, con elasticidad constante, es independiente del tamaño de mercado. El tamaño de mercado actúa por otra vía, la condición de pleno empleo $L = \sum_{i=1}^{n} l_i(x_i^s)$, con $l_i(x_i^s) = l_0 + l x_i^s$, que determina el número de empresas:

$$n = \frac{L}{l_0 + l \cdot x_i^s} = \frac{L}{l_0 \sigma}, \quad \text{es decir,} \quad n = f(L^+, l_0^-, \sigma^-) \tag{5}$$

Según (5), el número de variedades está restringido por el tamaño de mercado y es finito. La intuición es que más consumidores implican, dada su preferencia por la variedad, mayor demanda de variedades, y como esos consumidores trabajan, esa demanda encuentra la producción que la atienda.

Extensiones (economía abierta): la apertura sin costes de transporte a un país idéntico, con L^* trabajadores-consumidores, crea un mercado integrado de tamaño $L + L^*$. La escala $x_i^s = (l_0/l)(\sigma - 1)$ y el precio (4) no cambian, y por (5) el mercado integrado sostiene $(L + L^*)/(l_0\sigma)$ variedades, la suma de las de ambas autarquías. Por ello, ninguna empresa sale ni crece, surge comercio intraindustrial con un patrón indeterminado y la ganancia procede solo de la mayor variedad. Krugman (1980) parte de este caso y le añade costes de transporte.

Conclusiones:

- *Aspectos positivos*: el modelo proporciona la microfundamentación de la competencia monopolística en equilibrio general y establece con la ecuación (5) el resultado que vertebra toda la literatura posterior. Constituye, por ello, la base analítica del análisis de la apertura comercial en competencia imperfecta.

- *Limitaciones*: la elasticidad de sustitución constante fija la escala de cada empresa y el margen, de modo que un mercado mayor, también el de la apertura (Extensiones), se traduce solo en más variedades y no en empresas mayores, sin efecto escala ni efecto procompetitivo. Superar ambas restricciones exige abandonar ese supuesto, como hace Krugman (1979b).

2.1.2. Diferenciación horizontal: competencia monopolística en economía abierta (Krugman, 1979b)

Origen: el objetivo de Paul Krugman (1979b) es mostrar cómo esta estructura de mercado da lugar a comercio internacional aun cuando no haya ventaja comparativa.

Idea: lo esencial de la apertura es que amplía el mercado, pues al comerciar los países forman un único mercado integrado, más amplio que el de cada uno por separado.

Supuestos:

[9] Para obtener la cantidad ofertada de cada variedad basta con combinar la condición de libre entrada con la regla de precios (4): $\pi = \frac{1}{\theta} wl \cdot x_i^s - wl_0 - wlx_i^s = 0 \Rightarrow wlx_i^s\,(\frac{1}{\theta} - 1) = wl_0 \Rightarrow x_i^s\,\frac{1-\theta}{\theta} = \frac{l_0}{l} \Rightarrow \frac{x_i^s}{\sigma - 1} = \frac{l_0}{l}$.

▪ *Simetría*: el modelo elimina toda fuente posible de ventaja comparativa, al suponer tecnología, intensidad factorial y preferencias idénticas entre países.

▪ *Elasticidad variable*: el modelo comparte los demás supuestos de Spence-Dixit-Stiglitz, salvo que la elasticidad-precio de la demanda de cada variedad es decreciente en la cantidad demandada, $\partial|\varepsilon_{x_i p_i}|/\partial x_i < 0$. La intuición es que, al aumentar el número de variedades, cae el volumen consumido de cada una y los individuos reaccionan con mayor intensidad ante cambios de precio, de modo que el poder de mercado de cada empresa disminuye. El margen deja así de ser constante.

Planteamiento: el equilibrio se resume en dos condiciones de la empresa ya descrita, escritas en función del consumo per cápita de cada variedad, x_i, y de la ratio precio-salario, p/w, con una producción por empresa $x_i^s = Lx_i$:

$$\frac{p}{w} = \frac{|\varepsilon_{x_i p_i}|}{|\varepsilon_{x_i p_i}| - 1} \cdot l \tag{6}$$

$$\frac{p}{w} = l + \frac{l_0}{L\, x_i} \tag{7}$$

donde $|\varepsilon_{x_i p_i}| > 1$ es la elasticidad-precio de la demanda de cada variedad en valor absoluto, decreciente en x_i, y l_0, l y L son los ya definidos. La (6) generaliza la regla de precios (4), a la que se reduce con elasticidad constante, $|\varepsilon_{x_i p_i}| = \sigma$, y la (7) es la condición de libre entrada (precio igual a coste medio). En economía abierta, L pasa a ser $L + L^*$ en (7) y en el pleno empleo, $n = (L + L^*)/(l_0 + l \cdot x_i^s)$, y los consumidores incorporan las variedades extranjeras.

Desarrollo gráfico: el equilibrio del modelo se representa en el plano $(x_i, p/w)$ (figura B6.1):

1. *Curva de maximización del beneficio* (CMB): deriva de la ecuación (6) y recoge el *efecto competición*. Es creciente, porque, al aumentar la cantidad demandada de una variedad, cae su elasticidad-precio y, dada la relación inversa con el margen, sube el precio.

2. *Curva de libre entrada* (CLE), CLE_0: deriva de la ecuación (7) y recoge el *efecto escala*. Es decreciente, porque al aumentar la cantidad cae el coste medio y, con él, el precio compatible con beneficio nulo.

3. *Equilibrio*: el cruce de ambas curvas determina el equilibrio de autarquía (punto 0), con un consumo per cápita de cada variedad x_i^0 y una ratio precio-salario $(p/w)_0$.

Estática comparativa: la apertura a un país idéntico y sin costes de transporte amplía el mercado de L a $L + L^*$ y desplaza hacia abajo la curva de libre entrada, de CLE_0 a CLE_1, pues el coste fijo se reparte entre más consumidores. El equilibrio pasa del punto 0 al 1, con un consumo per cápita de cada variedad menor, x_i^1, porque el consumidor reparte su gasto entre más variedades, y una ratio precio-salario más baja, $(p/w)_1$. Con elasticidad constante, la CMB sería horizontal y la apertura solo añadiría variedades. En cambio, con elasticidad decreciente se siguen tres resultados[10]:

[10] Por ejemplo, en autarquía el país A produce cerezas, peras y manzanas, y el país B produce peras, manzanas y naranjas, de modo que existen tres empresas y tres variedades en cada país. Tras la apertura, A produce cerezas y manzanas y B produce peras y naranjas. Hay dos empresas en cada país, pero se intercambian cuatro variedades.

▪ *Número de empresas*: el menor consumo de cada variedad eleva su elasticidad-precio y comprime el margen, de modo que salen empresas hasta que el beneficio vuelve a ser nulo. Su número cae en cada país y también respecto del total combinado de las dos autarquías, pues cada país se especializa en un rango más estrecho de variedades e importa el resto.

▪ *Escala de producción*: como la ratio precio-salario baja y, por (7), es igual al coste medio por unidad de salario, $l + l_0/x_i^s$, la producción de cada superviviente, $x_i^s = (L + L^*)\, x_i^1$, supera a la de autarquía, $L\, x_i^0$. Con ello, las supervivientes aprovechan mejor las economías de escala y reducen los precios. Su expansión constituye el efecto escala, y la salida de las demás, una *racionalización* de la industria (no un efecto selección en sentido estricto, pues todas las empresas son iguales).

▪ *Número de variedades*: aunque haya menos empresas, los consumidores de los dos países pueden elegir entre más variedades que en autarquía.

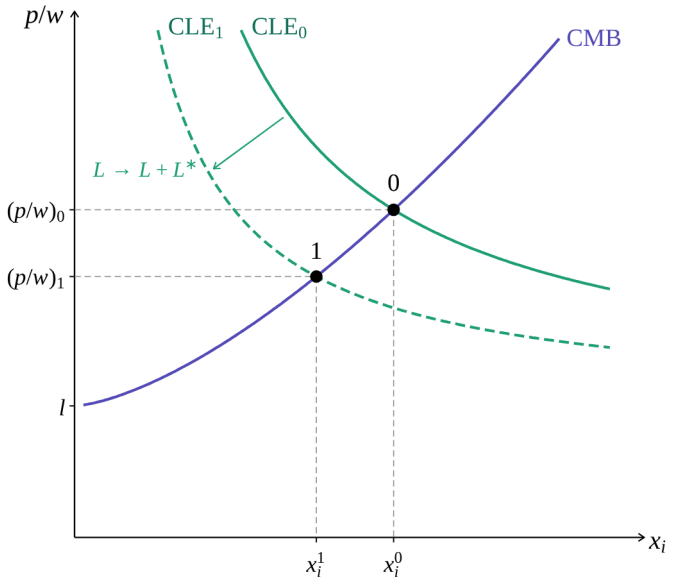

Figura B6.1: Modelo de Krugman (1979b). La apertura desplaza la curva de libre entrada (CLE) y, sobre la curva de maximización del beneficio (CMB), reduce el consumo per cápita de cada variedad y la ratio precio-salario ($0 \to 1$).

Bienestar: como la libre entrada mantiene el beneficio nulo antes y después de la apertura, la salida de empresas no supone pérdida de bienestar. La ganancia recae en los consumidores-trabajadores, cuyo salario real (w/p) aumenta y que disponen de más variedades. Basta, pues, con que existan economías de escala internas para que ambos países ganen bienestar al comerciar, aun siendo idénticos en recursos y tecnología.

Conclusiones:

▪ *Aspectos positivos*: constituye una de las obras fundacionales de la nueva teoría del comercio. Captura los elementos esenciales de los mercados con economías de escala y competencia imperfecta, y ofrece la primera explicación teórica del elevado volumen de comercio intraindustrial que se observa entre países similares.

▪ *Limitaciones*: el patrón de comercio queda indeterminado, pues al ser los países idénticos y no existir costes de transporte los precios relativos no se alteran y no cabe determinar qué país

produce qué variedades. El modelo tampoco resulta apropiado para analizar el efecto selección, ya que todas las empresas son iguales salvo en la variedad que producen, de modo que no predice cuál sobrevive. Además, el margen variable descansa en un supuesto específico sobre la demanda (elasticidad decreciente en la cantidad). La indeterminación del patrón motivó la extensión de 1980.

2.1.3. Diferenciación horizontal: el efecto mercado interno (Krugman, 1980)

Origen: Krugman (1980) introduce costes de transporte de tipo iceberg[11] para resolver la indeterminación anterior.

Idea: si una región presenta economías de escala internas en la producción de un bien, afronta costes de transporte al exportarlo y cuenta con una demanda interna fuerte, la producción tiende a concentrarse en ella. La concentración permite aprovechar a la vez las economías de escala y minimizar los costes de transporte.

Razonamiento: el resultado principal de su aportación es que el país con mayor mercado interno para un bien concentra una cuota de la producción mundial de ese bien superior a su cuota de la demanda. Ese país se convierte, por tanto, en exportador neto del bien. Este *efecto mercado interno* (*home market effect*) establece un nexo entre tamaño de mercado y exportaciones ajeno por completo a la ventaja comparativa. En las versiones con movilidad factorial, el país grande atrae además empresas y capital procedentes de las economías menores, resultado que constituye el fundamento de la nueva economía geográfica.

Conclusiones:

- *Aspectos positivos*: resuelve la indeterminación del patrón de comercio que arrastraba el modelo de 1979 y proporciona una predicción contrastable, la de que la especialización exportadora se correlaciona con el tamaño del mercado interno.

- *Limitaciones*: la evidencia empírica sobre los efectos de escala es débil, aunque sí la hay sobre el efecto selección. A ello se añade que la competencia monopolística describe la estructura de pocas industrias, pues lo habitual es el oligopolio con un número reducido de empresas.

2.1.4. Diferenciación vertical: calidad y oligopolio natural (Shaked y Sutton, 1982, 1983, 1984)

Origen: Avner Shaked y John Sutton analizan la diferenciación por calidad en contexto autárquico (1982, 1983) e internacional (1984). La exposición que sigue combina estos resultados con la versión de Sutton (1991), en la que la calidad se elige junto con un coste fijo de I+D creciente en ella.

Idea: cuando la diferenciación es por calidad y su coste recae sobre los costes fijos de I+D, la competencia en precios solo deja sobrevivir a un número finito de empresas, con independencia del tamaño del mercado.

Supuestos:

- *Agentes*:

[11] Esta modelización de los costes se debe a Paul Samuelson (1954b) y se utiliza ampliamente en economía internacional. Supone que, para que una unidad llegue a su destino, debe exportarse más de una unidad, dado que parte del bien se "derrite" por el camino (p. ej. el hielo).

- *Consumidores*: un continuo de consumidores con preferencias idénticas, todos los cuales prefieren mayor calidad, uniformemente distribuidos sobre un intervalo de renta. Presentan demandas unitarias, es decir, cada uno adquiere una unidad de su bien favorito o ninguna. Pese a compartir preferencias, su disposición a pagar difiere por el nivel de renta, de modo que, fijado un valor crítico[12], el mercado se divide con nitidez, y los bienes de calidad superior los adquieren quienes superan ese umbral.

- *Empresas*: inicialmente numerosas, cada una fabrica un producto verticalmente diferenciado[13].

Planteamiento: la conducta de las empresas se modela como un juego no cooperativo en tres etapas[14]:

1. *Entrada*: deciden si entrar, decisión que comporta una inversión en I+D que, una vez realizada, constituye un coste hundido irrecuperable.

2. *Calidad*: observan quién ha entrado y eligen qué calidad fabricar, elección que fija la cuantía de esa I+D (creciente y convexa en la calidad) y no afecta al coste variable.

3. *Precio*: con la I+D ya hundida, fijan el precio de venta atendiendo solo al coste variable, sin incurrir en la falacia del Concorde[15].

Equilibrio: a corto plazo la competencia en precios expulsa a algunas empresas. A largo plazo el proceso se repite y, si la distribución de renta es suficientemente estrecha[16], se alcanza un equilibrio de Nash perfecto en subjuegos en el que solo dos empresas cubren el mercado, lo que los autores denominan *oligopolio natural*.

Este resultado es un caso particular de la *propiedad de finitud*. Bajo diferenciación vertical, el número de empresas o calidades con cuota de mercado positiva está acotado por un máximo, independiente del tamaño de mercado y del coste de entrada. La condición es que todo consumidor elija la máxima calidad cuando se le ofrece a un precio igual a su coste marginal. A diferencia de lo que ocurre con la diferenciación horizontal, un mercado mayor no implica, por tanto, más variedades. Los autores demuestran asimismo que la propiedad de finitud es tanto más probable cuanto más recaiga el coste de la calidad sobre los costes fijos de I+D en lugar de sobre los variables.

Desarrollo gráfico: la elección de calidad de las dos empresas se representa en el plano calidad-ingresos y costes (figura B6.2):

1. *Curva de coste total*, $\mathrm{CT}(s)$: común a ambas empresas, creciente y convexa en la calidad, porque la I+D que exige cada calidad crece, y cada vez más, con ella (Planteamiento).

2. *Curva de ingreso total de la empresa 1*, $\mathrm{IT}_1(s)$: en forma de cueva. La empresa 1 fabrica la calidad baja y su ingreso alcanza el máximo en calidades bajas.

[12] Este valor crítico sería el nivel de renta que deja a un consumidor indiferente entre consumir un bien no diferenciado y un bien diferenciado verticalmente con la calidad más baja posible.

[13] Con n bienes verticalmente diferenciados, $0 < s_1 < s_2 < \cdots < s_n$, donde s_i es la calidad del bien i.

[14] La intuición que subyace a esta secuencia es que, mientras que los precios se pueden cambiar con facilidad, la elección de la calidad requiere cambios en las instalaciones, y la entrada a la industria requiere su construcción.

[15] La falacia del Concorde o falacia del coste hundido hace referencia a cómo, en ocasiones, los decisores tienen en cuenta los costes hundidos en sus decisiones. La expresión se asocia a las decisiones del Reino Unido y Francia sobre la continuación del programa del avión supersónico Concorde.

[16] A la vez, la distribución debe ser lo suficientemente "ancha" para que una empresa de calidad baja pueda abastecer al estrato inferior de la distribución de renta.

3. *Curva de ingreso total de la empresa 2*, $\mathrm{IT}_2(s)$: también en forma de cueva, más amplia y con el máximo en calidades altas. Por debajo de $\tilde{s}$ la empresa 2 no cubre costes.

4. *Equilibrio*: cada empresa elige la calidad en la que su ingreso marginal iguala al coste marginal (tangentes paralelas), s_1^* y s_2^*, y su beneficio es la distancia vertical entre ambas curvas. Ese beneficio es positivo en los dos casos y mayor para la empresa 2 ($\pi_2 > \pi_1$), y la empresa 1 pierde ingresos si se acerca a la calidad de la 2, con lo que se verifica el *principio de diferenciación máxima*.

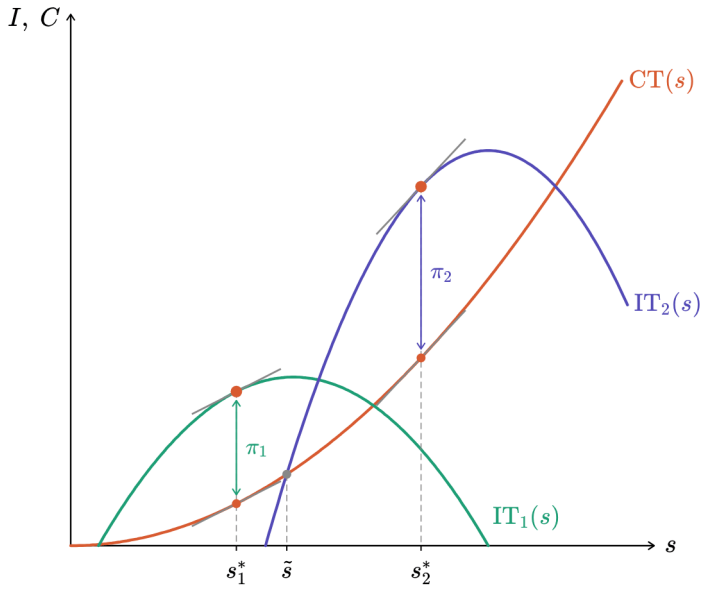

Figura B6.2: Modelo de Shaked y Sutton. El gráfico combina una curva de coste común, $\mathrm{CT}(s)$, y dos curvas de ingreso en forma de cueva, $\mathrm{IT}_1(s)$ e $\mathrm{IT}_2(s)$. Cada empresa elige la calidad en la que su ingreso marginal iguala al coste marginal (s_1^*, s_2^*) y obtiene beneficios positivos, mayores para la de calidad alta (diferenciación máxima).

Extensiones (economía abierta, 1984): surge comercio intraindustrial de calidades (comercio vertical), y los resultados difieren según el horizonte:

- *Corto plazo*: se mantiene la propiedad de finitud, de modo que el mercado integrado sostiene menos empresas que la suma de las que operaban en las dos economías cerradas y algunas deben salir.

- *Largo plazo*: se reproduce el resultado autárquico y se alcanza un duopolio natural internacional, sin que quepa determinar en qué país se sitúa cada empresa[17].

Conclusiones:

- *Aspectos positivos*: aporta nuevas intuiciones sobre la diferenciación de productos y muestra cómo los costes fijos de I+D concentran la estructura de mercado con independencia de su tamaño, resultado relevante para las industrias intensivas en investigación.

[17] Estos resultados se cumplen al margen de que las distribuciones de renta de los países considerados sean idénticas o no.

▪ *Limitaciones*: Giacomo Bonanno (1986) demuestra que el oligopolio natural no se sostiene si la variable estratégica de la última etapa es la cantidad en lugar del precio. El modelo tampoco determina la localización de las empresas en el caso internacional, limitación que comparte con los de diferenciación horizontal.

2.1.5. Conclusiones

Aspectos positivos: los modelos de competencia monopolística en economía abierta podrían servir para explicar el *dumping*, pues, con mercados segmentados, una empresa que afronte una demanda exterior más elástica y sea capaz de competir fuera cargará un precio exterior inferior al nacional.

Limitaciones: ese proceso sería, sin embargo, accidental, al depender de factores externos al modelo como la elasticidad de la demanda exterior. La razón de fondo es que la competencia monopolística no contempla la interacción estratégica. La competencia monopolística excluye además la posibilidad de comercio internacional de bienes idénticos, que sí aparece bajo otras estructuras de competencia imperfecta.

2.2. Modelos de oligopolio en economía abierta

Si se considera la interdependencia estratégica entre empresas de distintos países, el análisis se enriquece. Ello conduce a las aportaciones de la nueva teoría en mercados oligopolísticos y al concepto de *dumping* recíproco.

2.2.1. Duopolio y competencia en cantidades: el *dumping* recíproco (Brander, 1981; Brander y Krugman, 1983)

Origen: James Brander (1981) y, más adelante, James Brander y Paul Krugman (1983) estudiaron la interdependencia estratégica en el comercio internacional.

Idea: la interdependencia estratégica entre dos duopolistas situados en países distintos genera comercio intraindustrial de un bien idéntico y, con costes de transporte, *dumping* recíproco.

Supuestos:

▪ *Entorno*: dos empresas oligopolísticas situadas en países distintos compiten en ambos mercados con un bien homogéneo, con la misma información y la misma tecnología.

▪ *Agentes*:

- *Consumidores*: demanda decreciente e idéntica en ambos países, con función inversa lineal $p_j = a - b(x_{1j} + x_{2j})$, donde x_{ij} es la cantidad ofrecida por la empresa i en el mercado j y $a, b > 0$. La demanda de cada país se cubre en parte con producción nacional y en parte con importaciones.

- *Empresas*: la variable estratégica es la cantidad. Cada empresa elige cuánto vender en su país y cuánto en el extranjero, si bien todo el producto se fabrica en el país de origen, y los precios se ajustan después[18]. Las empresas son miopes, es decir, reconocen que su cantidad afecta al precio pero no que afecte a la conducta rival, de modo que las variaciones conjeturales

[18] Este supuesto es fundamental. Al competir en cantidades tomando como dadas las de la rival, los precios se ajustan para sostenerlas. Por tanto, $\partial p/\partial x$ es relativamente grande y la elasticidad percibida es relativamente pequeña.

son nulas. Se cumplen las condiciones de Hahn[19], de las que se sigue que las cantidades son *sustitutos estratégicos*[20] y que las *funciones de reacción* tienen pendiente negativa e inferior a la unidad en valor absoluto. Se sigue también que la función de beneficios es estrictamente cóncava, lo que garantiza la continuidad de las funciones de reacción.

▪ *Tecnología*: función de costes lineal e idéntica, $\mathrm{CT}(X_1) = c_0 + c(x_{11} + x_{12})$, con $c_0 > 0$ el coste fijo y $c > 0$ el coste marginal constante.

Planteamiento (sin costes de transporte): cada empresa maximiza el beneficio obtenido en ambos mercados. Para la empresa 1:

$$\max_{x_{11},x_{12}} \pi_1 = \underbrace{[a - b(x_{11} + x_{21})] \cdot x_{11}}_{\mathrm{IT}_{11}} + \underbrace{[a - b(x_{12} + x_{22})] \cdot x_{12}}_{\mathrm{IT}_{12}} - \underbrace{[c_0 + c(x_{11} + x_{12})]}_{\mathrm{CT}(X_1)}$$

donde x_{ij} es la cantidad vendida por la empresa i en el mercado j, a y b los parámetros de la demanda inversa, c_0 el coste fijo y c el coste marginal constante.

CPO: en cada mercado, el ingreso marginal se iguala al coste marginal común, $a - 2bx_{11} - bx_{21} - c = 0$ y $a - 2bx_{12} - bx_{22} - c = 0$. De ellas resultan funciones de reacción del tipo $x_{11} = \frac{a-c}{2b} - \frac{1}{2}x_{21}$, decrecientes en la cantidad de la rival por ser las cantidades sustitutos estratégicos. Su intersección da la solución simétrica $x_{11} = x_{12} = x_{21} = x_{22} = (a - c)/(3b)$.

CSO: la función de beneficios es estrictamente cóncava en las cantidades propias, $\partial^2\pi_1/\partial x_{1j}^2 = -2b < 0$, por lo que las condiciones de primer orden son suficientes.

Equilibrio: las dos empresas fabrican la misma cantidad y reparten su producción a partes iguales entre ambos mercados, igual que en el modelo lineal-simétrico de Cournot de economía cerrada. Lo nuevo es que cada productor está en un país distinto, y de ahí surge comercio intraindustrial de un bien idéntico, algo que ninguna forma de competencia monopolística puede explicar.

Desarrollo gráfico: el equilibrio del duopolio se representa en el plano (x_{11}, x_{21}) del mercado nacional de la empresa 1 (figura B6.3):

1. *Función de reacción de la empresa 1*, $\varphi_1(x_{21})$: decreciente, porque las cantidades son sustitutos estratégicos, y continua por las condiciones de Hahn. Corta el eje horizontal en $(a - c)/(2b)$ y el vertical en $(a - c)/b$.

2. *Función de reacción de la empresa 2*, $\varphi_2(x_{11})$: simétrica de la anterior, con el mismo intercepto en el eje opuesto.

3. *Equilibrio*: en el corte de ambas (punto 0), cada empresa vende $x_{11}^0 = x_{21}^0 = (a - c)/(3b)$.

Estática comparativa (costes de transporte): el planteamiento anterior, sin costes de transporte, equivale a tratar el mercado mundial como integrado, cuando lo habitual es que esté segmentado por barreras comerciales, geográficas o costes de transporte. La introducción de estos últimos, de tipo iceberg[21], produce dos efectos:

[19] El cumplimiento de las condiciones de Hahn garantiza la existencia, unicidad y estabilidad del equilibrio del modelo de Cournot.

[20] Las cantidades y las capacidades suelen ser sustitutos estratégicos, mientras que los precios suelen ser complementarios estratégicos.

[21] Formalmente, la función de costes es del tipo $\mathrm{CT}(X_i) = c_0 + c\,[x_{ii} + \tau\, x_{ij}]$, donde $\tau > 1$ es el coste de transporte de tipo iceberg, es decir, las unidades que deben enviarse desde i para que llegue una a j.

1. *Coste marginal*: abastecer el mercado exterior cuesta más, en el margen, que abastecer el nacional.

2. *Ingreso marginal*: el exterior resulta superior al nacional, lo que genera un incentivo a vender fuera. La razón es que los costes de transporte reducen la cuota de la empresa en el mercado exterior respecto de la que tiene en el propio, y con menor cuota la empresa percibe allí una demanda más elástica. Por ello, una rebaja del precio de exportación contribuye más al beneficio que una del precio nacional, siempre que se cubra el coste marginal.

Ambas empresas tienen, por tanto, incentivos para practicar *dumping*, y el resultado, el *dumping* recíproco, es un caso particular de discriminación de precios.

En el mercado nacional de la empresa 1 (figura B6.3), la función de reacción de esta no varía. La de la empresa 2, en cambio, se desplaza hacia el origen, de $\varphi_2(x_{11})$ a $\varphi_2'(x_{11})$, pues el coste de transporte ($\tau > 1$) eleva su coste marginal. El equilibrio pasa del punto 0 al 1, de modo que aumenta la cantidad vendida por la empresa nacional, de x_{11}^0 a x_{11}^1, y disminuye la de la extranjera, de x_{21}^0 a x_{21}^1, con caída neta de la cantidad total.

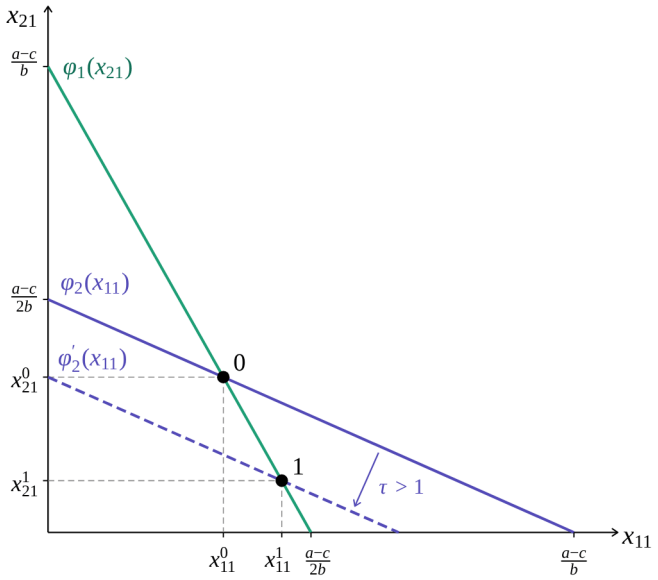

Figura B6.3: Modelo de Brander y Krugman. En el mercado de la empresa 1, los costes de transporte desplazan hacia el origen la curva de reacción de la empresa extranjera. La empresa nacional vende más, la extranjera menos y la cantidad total cae ($0 \to 1$).

Como en cada mercado rige un único precio y, por simetría, el del mercado nacional de la empresa 1 coincide con el del mercado propio de la extranjera (donde la curva de reacción de esta no varía), la empresa extranjera absorbe el coste de transporte. Por eso, el precio neto (*free on board*, FOB) que percibe por sus exportaciones es inferior al que obtiene en su propio mercado, lo que constituye *dumping*. Por simetría, ambas empresas actúan igual.

Conclusiones:

- *Aspectos positivos*: el modelo proporciona el fundamento analítico del *dumping* recíproco como fenómeno de equilibrio y no como práctica desleal.

- *Limitaciones*: el efecto sobre el bienestar no es concluyente. El comercio introduce competencia donde, en autarquía, cada empresa sería monopolista en su país, produciría menos y vendería más caro. A cambio, el *cross-hauling* despilfarra factores reales al transportar en ambos sentidos bienes idénticos, sin ampliar la gama disponible. El signo final depende de cuál de los dos efectos pese más.

2.2.2. Política comercial estratégica: orientaciones ofensivas y defensivas

La interdependencia estratégica hace posible, además, una intervención pública que incline el resultado en favor de las empresas nacionales.

> La política comercial estratégica se desarrolla en el tema B08, apartados 2.2 y 2.3, y los instrumentos tradicionales de política comercial, en el tema B07. Aquí se recogen sus resultados principales.

Orientación ofensiva con competencia en cantidades: subvenciones a la exportación (Brander y Spencer, 1985): cuando dos empresas de países distintos compiten a la Cournot en un tercer mercado, una subvención a la exportación desplaza la curva de reacción de la empresa nacional. El sector público puede comprometerse de forma creíble, cosa que la empresa no logra por sí sola. Con la subvención óptima, el equilibrio pasa al resultado de Stackelberg, con la empresa subvencionada como líder, aunque la competencia sigue siendo a la Cournot.

La producción total aumenta, el precio cae y los consumidores del tercer país ganan, mientras que el país subvencionador gana a costa de la empresa rival (*rent shifting*), como ilustra la rivalidad entre Airbus y Boeing. Se trata, con todo, de una política de empobrecimiento del vecino que invita a la represalia.

Orientación ofensiva con competencia en precios: impuestos a la exportación (Eaton y Grossman, 1986): con bienes ligeramente diferenciados y precios que son complementarios estratégicos, el instrumento óptimo se invierte. Un impuesto a la exportación hace creíble la subida del precio de la empresa nacional, que la rival secunda. Así, ambos precios suben, y el país que grava gana (beneficio de su empresa más recaudación) a costa de los consumidores del tercer país. Como subrayó Krugman, el instrumento óptimo depende del modo de competencia, que el Gobierno rara vez conoce, por lo que la política puede resultar contraproducente.

Orientación defensiva: aranceles a la importación (Eaton y Kierzkowski, 1984): en una estructura de mercado mixta, un arancel puede resultar socialmente óptimo si induce la entrada de un productor nacional del bien diferenciado.

La nueva teoría da, además, fundamento moderno al argumento de la industria naciente (Hamilton, 1791; List, 1841). Con economías de escala, internas o externas por *learning by doing*, y fallos en los mercados de capitales, la protección transitoria permite a la industria descender por su curva de aprendizaje, siempre que se cumplan los criterios de Mill y de Bastable. Aun así, el arancel no suele ser el instrumento de *first best* y la protección invita a la captura de rentas.

2.2.3. Conclusiones

Aspectos positivos: los modelos de oligopolio muestran que la interacción estratégica genera comercio intraindustrial incluso de bienes idénticos y ofrecen un marco riguroso para evaluar la política comercial cuando existen rentas oligopolísticas susceptibles de desplazamiento entre países.

Limitaciones: no existe un modelo general de oligopolio. Las empresas pueden actuar en colusión o de forma no cooperativa, y en este último caso el resultado depende de la variable de decisión, de la naturaleza de las variaciones conjeturales, de la especificación del producto y del carácter segmentado o integrado del mercado. No cabe, por tanto, un análisis general de los efectos del oligopolio sobre el comercio, sino únicamente resultados particulares. Esa dependencia de los supuestos constituye la principal debilidad práctica de la política comercial estratégica.

2.3. Equilibrio integrado, modelos gravitatorios y nueva economía geográfica

El equilibrio integrado devuelve las diferencias de dotaciones a la competencia monopolística, los modelos gravitatorios explican los flujos bilaterales por el tamaño y la distancia, y la nueva economía geográfica, las economías de aglomeración por la interacción entre economías de escala internas, costes de transporte y movilidad de factores.

2.3.1. Dotaciones distintas y comercio intraindustrial: el equilibrio integrado (Helpman y Krugman, 1985)

Origen: Elhanan Helpman y Paul Krugman (1985) integran la competencia monopolística en el modelo de Heckscher-Ohlin y devuelven a las dotaciones factoriales el papel que les habían quitado los modelos anteriores, con países idénticos.

Idea: el comercio interindustrial y el intraindustrial conviven y responden a causas distintas. Las diferencias de dotaciones explican el primero, y las economías de escala con diferenciación del producto, el segundo.

Razonamiento: la economía mundial tiene dos factores, capital y trabajo, y dos sectores, uno de bienes homogéneos con rendimientos constantes y otro de manufacturas diferenciadas, más intensivo en capital, con economías de escala internas y competencia monopolística. Si las dotaciones de los países no difieren demasiado, el comercio reproduce el *equilibrio integrado*, el que alcanzaría una economía mundial sin fronteras, con igualación de los precios de los factores. De ese equilibrio se derivan tres resultados:

1. *Comercio neto*: sigue a Heckscher-Ohlin, pues el país abundante en capital es exportador neto de manufacturas e importador neto del bien homogéneo.

2. *Comercio intraindustrial*: los dos países producen variedades distintas de manufacturas y las intercambian, de modo que el comercio de manufacturas es bidireccional. Su peso en el total, que mide el índice de Grubel y Lloyd, aumenta cuanto más se parecen las dotaciones y es total si son idénticas, el caso de Krugman (1979b).

3. *Volumen de comercio*: Helpman (1987) deduce que, entre países con dotaciones relativas parecidas, el volumen de comercio en proporción de su PIB conjunto crece con la semejanza de sus tamaños[22]. La razón es que cada país exporta la parte de su producción que compran los demás, igual al peso de estos en el PIB del grupo, y la media ponderada de esas partes es máxima cuando los tamaños se igualan.

Conclusiones:

[22] Si todos los bienes son diferenciados y cada país produce sus propias variedades, el país j, con una proporción s_j del PIB del grupo, Y, exporta la parte $1 - s_j$ de su producción. El volumen de comercio es entonces $V = \sum_j s_j(1 - s_j)\,Y$, es decir, $V/Y = 1 - \sum_j s_j^2$, máximo cuando todos los países tienen el mismo tamaño.

▪ *Aspectos positivos*: reúne en un solo modelo las dos explicaciones del comercio y da cuenta de los hechos estilizados 2 y 5 del apartado 1.1.2, pues el comercio entre países desarrollados de dotaciones parecidas es sobre todo intraindustrial y el comercio norte-sur, interindustrial. Sus predicciones sobre el índice de Grubel y Lloyd y sobre el volumen de comercio son, además, contrastables, y Helpman (1987) las encontró coherentes con los datos de catorce países industrializados.

▪ *Limitaciones*: exige que las dotaciones permitan la igualación de los precios de los factores y mantiene empresas idénticas, supuesto que abandonan Bernard, Redding y Schott (2007) (apartado 3.3.2). Su contraste empírico es, además, menos concluyente de lo que parece (apartado 2.4).

2.3.2. Modelos gravitatorios (Isard, 1954; Tinbergen, 1962; Anderson, 1979)

Origen: los modelos gravitatorios de comercio están inspirados en la ley de la gravitación universal (1687) de Isaac Newton. Nacen con el trabajo de Walter Isard (1954), si bien fueron popularizados por Jan Tinbergen (1962).

Idea: por analogía con la ley de Newton, el flujo comercial entre dos economías es proporcional a su tamaño económico e inversamente proporcional a la distancia que las separa.

Razonamiento: el modelo gravitatorio simple se puede aplicar tanto a flujos comerciales como a flujos de factores y establece que el flujo entre dos regiones guarda una relación del tipo:

$$X_{ij} = G \cdot \frac{Y_i^{\alpha} \cdot Y_j^{\beta}}{d_{ij}^{\delta}}$$

donde X_{ij} es el flujo del país o región i al j, G la "constante gravitacional", Y_i e Y_j el PIB de cada uno, d_{ij} la distancia entre ambos y α, β y δ las elasticidades respectivas. El término Y_i^{α} puede interpretarse como lo que i está dispuesto a vender a j, e Y_j^{β}, como lo que j está dispuesto a comprar a i.

La naturaleza multiplicativa de la fórmula hace que sus parámetros puedan estimarse tras una log-linealización, $\ln X_{ij} = \ln G + \alpha \ln Y_i + \beta \ln Y_j - \delta \ln d_{ij} + \varepsilon_{ij}$, donde ε_{ij} es un error de predicción sin interpretación económica.

De entre los parámetros estimados, los más relevantes son dos. Por un lado, la constante gravitacional G recoge los factores que afectan a los flujos con independencia de la distancia. Por otro lado, la elasticidad-distancia δ mide la sensibilidad de los flujos bilaterales a la distancia, de modo que un valor elevado implica que la proximidad es decisiva.

Conclusiones:

▪ *Aspectos positivos*: el principal atractivo de estos modelos es la calidad de su ajuste empírico, que los ha convertido en la herramienta de referencia de la economía internacional aplicada. Las estimaciones confirman que la proximidad geográfica es, junto con el tamaño económico, un determinante fundamental de los flujos comerciales:

- *Anderson y van Wincoop* (2004): cifran los costes totales de comercio en las economías desarrolladas en un equivalente *ad valorem* del 170 % sobre el precio en fábrica.

- *Disdier y Head* (2008): en un metanálisis sobre 1467 estimaciones, sitúan la elasticidad-distancia media en 0,9[23]. Con ese valor, duplicar la distancia reduce los flujos casi a la mitad, con un efecto más acusado en los países en desarrollo por la menor calidad de sus infraestructuras.

- *Limitaciones*: se manifiestan en dos planos:

- *Plano teórico*: los modelos adolecen de falta de microfundamentación, pues la distancia opera como una caja negra que abarca desde las infraestructuras hasta los lazos culturales. La literatura posterior ha subsanado esa carencia derivando la ecuación desde fundamentos muy distintos, ya sean preferencias Armington (Anderson, 1979), competencia monopolística (Krugman, 1980) o un marco ricardiano (Eaton y Kortum, 2002), lo que explica su ubicuidad empírica pero dificulta emplearla para discriminar entre teorías.

- *Plano empírico*: muchas estimaciones simples están sesgadas, en particular por omitir la *resistencia multilateral*, es decir, los costes de comercio de cada país con el resto del mundo (Anderson y van Wincoop, 2003).

2.3.3. Nueva economía geográfica y modelo centro-periferia (Krugman, 1991a)

Origen: la *nueva economía geográfica*, inaugurada por Paul Krugman en *Increasing Returns and Economic Geography* (1991a), parte del hecho observado de que la riqueza, la población y el empleo se distribuyen espacialmente de forma muy desigual, tanto en el ámbito internacional como en el regional.

Idea: esos desequilibrios responden a dos clases de determinantes. Los de *primera naturaleza* derivan de la distribución desigual de los recursos y de factores geográficos, como la presencia de materias primas o la proximidad a vías naturales de comunicación. Los de *segunda naturaleza* proceden de la acción humana y, señaladamente, de la estructura de mercado y de las reglas de fijación de precios.

A su vez, existen dos tipos de fuerzas. Las *centrípetas* o de aglomeración concentran la actividad en los mercados mayores y más productivos, mientras que las *centrífugas* o de dispersión, como los costes de transporte, la congestión y la demanda inmóvil, la esparcen por el territorio.

Razonamiento: el modelo considera dos regiones, un bien homogéneo producido por trabajadores inmóviles y un bien diferenciado producido en competencia monopolística por trabajadores móviles, que migran hacia la región de mayor salario real, con costes de transporte de tipo iceberg. De la interacción entre ambas fuerzas resultan dos equilibrios posibles, el simétrico, con la actividad repartida, y el de aglomeración, con toda la industria concentrada en el centro y la otra región reducida a periferia.

El equilibrio depende del grado de libertad comercial, y la transición entre uno y otro no es gradual sino abrupta. A medida que caen los costes de comercio, es decir, que aumenta la libertad comercial, φ (de 0, con costes prohibitivos, a 1, con libre comercio), se suceden tres tramos:

1. *Simetría estable*: con costes de comercio elevados ($\varphi < \varphi_S$), el equilibrio estable es el reparto simétrico de la actividad entre las dos regiones.

[23] Este valor estimado engloba, no obstante, una gran heterogeneidad de bienes. Es previsible, por ejemplo, que los bienes de construcción tengan una elasticidad-distancia mayor que la media.

2. *Equilibrios múltiples*: superado el *punto de sostenimiento*, φ_S, la aglomeración pasa a ser sostenible sin que el reparto simétrico deje de ser estable, con dos equilibrios asimétricos inestables entre ellos, y el resultado presenta histéresis y depende de los accidentes históricos.

3. *Simetría inestable*: superado el *punto de ruptura*, φ_B, cualquier perturbación del reparto simétrico desencadena la aglomeración completa.

La figura B6.4 (diagrama de *tomahawk*) los representa, con la cuota de la región 1 en la industria, λ, en ordenadas.

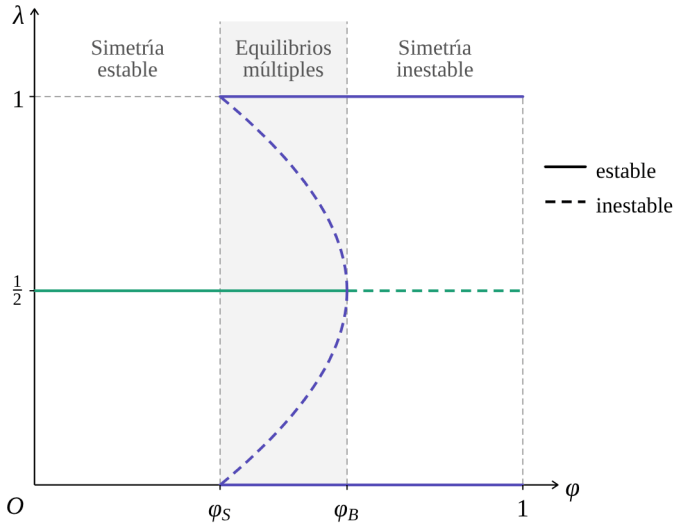

Figura B6.4: Modelo centro-periferia (Krugman, 1991a). La simetría ($\lambda = 1/2$) es estable hasta φ_B y la aglomeración ($\lambda = 0$ o 1), desde φ_S. Los equilibrios inestables van en trazo discontinuo.

Conclusiones:

▪ *Aspectos positivos*: el modelo formaliza intuiciones estructuralistas que carecían de fundamento microeconómico, explica la concentración espacial de la actividad sin apelar a diferencias exógenas y constituye la base analítica para estudiar los efectos regionales de la integración. Estas aportaciones, junto con la nueva teoría del comercio, valieron a Krugman el Premio Nobel de Economía en 2008.

▪ *Limitaciones*: son de tres tipos:

- *Movilidad del trabajo*: el supuesto de movilidad perfecta del trabajo industrial resulta poco realista en el plano internacional, dadas las restricciones migratorias, por lo que el modelo se aplica mejor a regiones dentro de un mismo país.
- *Formas funcionales*: los resultados dependen de la especificación elegida.
- *Contraste empírico*: la existencia de equilibrios múltiples dificulta contrastar el modelo.

2.4. Conclusiones

Aspectos positivos: la nueva teoría altera el resultado neoclásico según el cual, en ausencia de economías de escala y con competencia perfecta, el equilibrio es Pareto-eficiente y ninguna intervención mejora el bienestar nacional, salvo el arancel óptimo de un país grande. En cambio, con economías de escala y competencia imperfecta existen ganancias potenciales que la política industrial puede tratar de materializar[24], reasignando recursos hacia las industrias con rendimientos crecientes. En el plano empírico, la evidencia sobre el comercio entre países industrializados es coherente con que la competencia monopolística y las economías de escala expliquen el crecimiento del comercio intraindustrial, fenómeno que la teoría neoclásica no podía justificar.

Limitaciones: la intervención así justificada comporta dos costes. Uno es interno, pues la reasignación desvía recursos desde las demás industrias y genera ganadores y perdedores. El otro es externo, pues en economía abierta la política que beneficia a un país empobrece al vecino e invita a la represalia.

El respaldo empírico es, además, más débil de lo que parece. David Hummels y James Levinsohn (1995) observan la relación de Helpman (1987) entre el volumen de comercio y la semejanza de tamaño también entre países en desarrollo, a los que el modelo no debería aplicarse porque apenas tienen comercio intraindustrial, lo que cuestiona que el buen ajuste constituya validación del mecanismo teórico.

La objeción de fondo es, sin embargo, que estos modelos tratan a las empresas como idénticas, de modo que tras una apertura comercial todas pasarían a exportar. Por ello no pueden explicar tres hechos firmemente establecidos a nivel de empresa:

- *Pocas exportadoras*: solo unas pocas empresas exportan.
- *Primas del exportador*: las exportadoras son sistemáticamente mayores, más productivas y más intensivas en capital y en trabajo cualificado que el resto, y pagan salarios superiores[25].
- *Productividad agregada*: aumenta con la apertura.

[24] Autores como Paul Krugman o Joseph Stiglitz, que se suelen encuadrar en la escuela neokeynesiana, han mostrado que, con competencia imperfecta y rendimientos crecientes, existen argumentos teóricos a favor de la intervención, también en el ámbito internacional. Ello no equivale a recomendarla como regla general, y el propio Krugman, por las objeciones recogidas en el apartado 2.2.2, siguió defendiendo el libre comercio como pauta práctica.

[25] La alta productividad precede a la exportación (véase el apartado 3.4), si bien muchas decisiones para aumentarla, como la innovación, se toman con la perspectiva de exportar después.

3. Novísima teoría del comercio internacional

La novísima teoría explica el comercio por las diferencias entre empresas de una misma industria y desplaza así la unidad de análisis a la empresa (véase la introducción). Su premisa es que la evidencia a nivel de empresa (apartado 2.4) desmiente el supuesto de homogeneidad. De ahí resulta un mecanismo de ganancia del comercio distinto de los conocidos, ya que la apertura no opera únicamente reasignando recursos entre sectores o aprovechando economías de escala, sino seleccionando dentro de cada sector.

3.1. Modelo de competencia monopolística y heterogeneidad de empresas (Melitz, 2003)

Origen: el modelo de Marc Melitz se desarrolla en su artículo de 2003.

Idea: la apertura comercial selecciona dentro de cada sector al expulsar a las empresas menos productivas y desplazar la cuota de mercado hacia las más productivas, con lo que la productividad agregada aumenta sin que ninguna empresa mejore individualmente.

3.1.1. Supuestos

En su trabajo, Melitz parte del modelo de Krugman (1980), del que conserva la competencia monopolística, los rendimientos crecientes, la preferencia por la variedad y la simetría, si bien introduce tres diferencias.

Heterogeneidad: las empresas difieren en su productividad.

Costes de exportar: comprenden un coste fijo, $c_{0X} > 0$, por prospección, segmentación o promoción, que debe asumirse antes de exportar, y un coste variable unitario de tipo iceberg, $\tau > 1$.

Salario: se normaliza a la unidad en todo momento.

3.1.2. Modelo en autarquía

Consumidores: son idénticos a los del modelo de Spence-Dixit-Stiglitz (apartado 2.1.1).

- *Planteamiento*: maximizan la utilidad CES sujetos a su renta, M_0.

- *CPO*: dan la demanda isoelástica de cada variedad, $x_i = (p_i/P)^{-\sigma} \cdot M_0/P$, y la agregada, $X_i = x_i L$, que afronta cada empresa al fijar su precio.

- *CSO*: la del mismo modelo (CES estrictamente cuasicóncava).

Empresas: cada una produce una variedad, fija su precio y decide si opera según su productividad.

- *Planteamiento*: la tecnología incorpora un parámetro de productividad ψ propio de cada empresa:

$$\mathrm{CT}(\psi) = c_0 + \frac{1}{\psi}\, x^s(\psi)$$

donde $x^s(\psi)$ es la producción de la empresa (en equilibrio, la demanda agregada de su variedad, X_i), $c_0 > 0$ es el coste fijo y $1/\psi$ el coste marginal constante (la inversa de la productividad). La empresa elige el precio que maximiza su beneficio, $\pi = p\,x^s - \mathrm{CT}(\psi)$, frente a esa demanda, y opera si el beneficio máximo no es negativo.

- *CPO*: la del mismo modelo, con coste marginal $1/\psi$ (el salario es unitario), da la regla de precios:

$$p(\psi) = \frac{\sigma}{\sigma - 1} \cdot \frac{1}{\psi}$$

La regla sigue siendo un margen proporcional al coste marginal, si bien ahora las empresas más productivas fijan precios más bajos.

- *CSO*: la del mismo modelo, con $\sigma > 1$.

- *Solución*: sustituyendo la regla de precios y la demanda, el beneficio máximo en función de la productividad resulta[26]:

$$\pi_0(\psi) = B_0 \cdot \psi^{\sigma-1} - c_0, \quad B_0 > 0$$

donde B_0 recoge la facilidad de obtener beneficios en el mercado y depende del tamaño de este, L, y del índice de precios, P ($B_0 = f(L^+, P^+)$). Valores elevados de B_0 señalan menor competencia y, por tanto, mayores beneficios para las empresas establecidas.

Dado que $B_0 > 0$ y $\sigma > 1$, el beneficio es creciente en la productividad y las empresas más productivas venden además mayor cantidad[27]. Existe, en consecuencia, una *productividad de corte* $\bar{\psi}$ a partir de la cual el beneficio es no negativo y la empresa decide operar.

Equilibrio: existe un continuo de entrantes potenciales, idénticos antes de entrar, que para obtener su productividad de una distribución aleatoria $\psi \sim g(\psi)$ deben realizar una inversión previa c_{0e}, que se convierte en coste hundido y cabe interpretar como gasto en I+D[28]. La libre entrada iguala los beneficios esperados a ese coste, $\pi_e = c_{0e}$.

Conocida su realización, la empresa produce si $\psi \geq \bar{\psi}$ y abandona sin llegar a operar si $\psi < \bar{\psi}$. Este mecanismo aleatorio genera la heterogeneidad, de modo que empresas idénticas *ex ante* resultan distintas *ex post*.

Desarrollo gráfico: en el plano $(\psi^{\sigma-1}, \pi)$ (figura B6.5), el beneficio, $\pi_0(\psi) = B_0\,\psi^{\sigma-1} - c_0$, es una recta de pendiente B_0 y ordenada en el origen $-c_0$. Su corte con el eje horizontal marca la productividad de corte, $\bar{\psi}^{\sigma-1} = c_0/B_0$, punto en que la empresa alcanza el umbral de rentabilidad (*break-even*).

[26] Sustituyendo la regla de precios y la demanda en la definición del beneficio: $\pi = px^s - c_0 - \frac{x^s}{\psi} = \frac{x^s}{\psi(\sigma-1)} - c_0$. Sustituyendo a su vez la demanda agregada $x^s = X_i = L(p/P)^{-\sigma} M_0/P$ y la regla de precios, se obtiene $\pi = L\,\sigma^{-\sigma}(\sigma-1)^{\sigma-1} P^{\sigma-1} M_0 \cdot \psi^{\sigma-1} - c_0 \equiv B_0\,\psi^{\sigma-1} - c_0$.

[27] De la demanda y la regla de precios se obtiene $x^s(\psi) = L\left(\frac{p(\psi)}{P}\right)^{-\sigma} \frac{M_0}{P} \propto \psi^{\sigma}$, función creciente de la productividad porque $\sigma > 1$. En el punto de corte, $\pi = \frac{1}{\sigma-1}\frac{x^s(\bar{\psi})}{\psi} - c_0 = 0 \Rightarrow x^s(\bar{\psi}) = c_0(\sigma-1)\bar{\psi}$, que es la escala mínima para operar.

[28] De esta forma se evita que las empresas repitan indefinidamente la "lotería" para obtener mejores productividades.

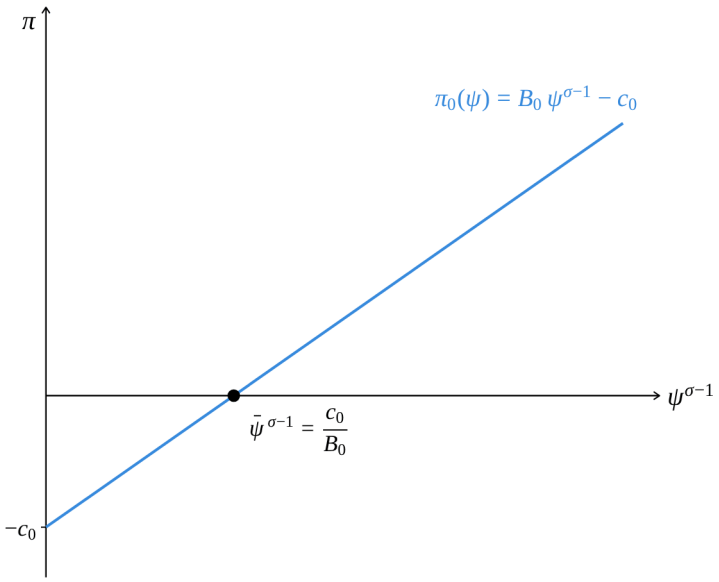

Figura B6.5: Modelo de Melitz (2003) en autarquía. La empresa opera a partir de una productividad de corte.

3.1.3. Modelo abierto: efectos del comercio internacional

Supóngase ahora que la economía comercia con un país idéntico, con el mismo salario normalizado[29]. Surgen dos productividades de corte, una que determina quién produce para el mercado interior y otra quién exporta[30], y el índice de precios pasa a recoger también las variedades importadas.

Empresas: además de producir para el mercado interior, cada una decide si exporta.

- *Planteamiento*: como el coste marginal es constante, la empresa maximiza por separado el beneficio interior y el de exportación, que tiene coste marginal τ/ψ y coste fijo c_{0X}. La empresa exporta si el beneficio de exportación no es negativo.

- *CPO*: dan en cada mercado la regla de precios de autarquía, con el coste iceberg en el de exportación:

$$p(\psi) = \frac{\sigma}{\sigma-1}\cdot\frac{1}{\psi} \qquad \text{y} \qquad p_X(\psi) = \frac{\sigma}{\sigma-1}\cdot\frac{\tau}{\psi}$$

- *CSO*: la de autarquía, en cada mercado.

- *Solución*: el beneficio máximo de cada mercado es lineal en $\psi^{\sigma-1}$:

[29] La simetría y la normalización de los salarios garantizan que todas las variables agregadas relevantes (L, M_0 y P) son idénticas entre países y que se produce la igualación de los precios de los factores. Ello permite abstraerse de las diferencias salariales relativas que se derivarían de las diferencias de productividad.

[30] Los costes de comercio son esenciales, porque sin ellos se alcanzaría el equilibrio integrado. La apertura reproduciría entonces el caso del apartado 2.1.1 (Extensiones).

- *Beneficio interior*: la apertura eleva la competencia y reduce el índice de precios respecto de la autarquía, de modo que $B_1 < B_0$ y el beneficio interior pasa a ser $\pi_1(\psi) = B_1\psi^{\sigma-1} - c_0$, con lo que la productividad mínima para operar resulta mayor que en autarquía.

- *Umbral de exportación*: el beneficio de exportar adopta la misma forma, $\pi_X(\psi) = B_X\psi^{\sigma-1} - c_{0X}$, con $B_X = B_1\tau^{1-\sigma}$. Comparando ambos umbrales:

$$\frac{\bar{\psi}_X}{\bar{\psi}_1} = \left(\frac{c_{0X}\,\tau^{\sigma-1}}{c_0}\right)^{1/(\sigma-1)} = \tau\left(\frac{c_{0X}}{c_0}\right)^{1/(\sigma-1)}$$

  Siempre que los costes de exportación sean suficientemente elevados ($\tau^{\sigma-1}c_{0X} > c_0$), el umbral exportador supera al de producción y solo las empresas más productivas exportan, pues únicamente ellas encuentran rentable pagar el coste fijo correspondiente.

Desarrollo gráfico: los efectos de la apertura se representan en el plano $(\psi^{\sigma-1}, \pi)$ (figura B6.6):

1. *Recta de beneficios interiores*: rota a la baja sobre su ordenada en el origen, $-c_0$, de $\pi_0(\psi)$ a $\pi_1(\psi)$, al reducirse su pendiente ($B_1 < B_0$), porque la apertura eleva la competencia y reduce el índice de precios.

2. *Recta de beneficios de exportación*, $\pi_X(\psi)$: recoge el beneficio de servir el mercado exterior. Es menos inclinada, con pendiente $B_X = B_1\tau^{1-\sigma}$, porque el coste iceberg encarece cada unidad exportada, y su ordenada en el origen, $-c_{0X}$, es el coste fijo de exportar.

3. *Productividades de corte*: el umbral interior sube de $\bar{\psi}_0^{\sigma-1}$ a $\bar{\psi}_1^{\sigma-1}$, y el de exportación, $\bar{\psi}_X^{\sigma-1}$, queda a su derecha.

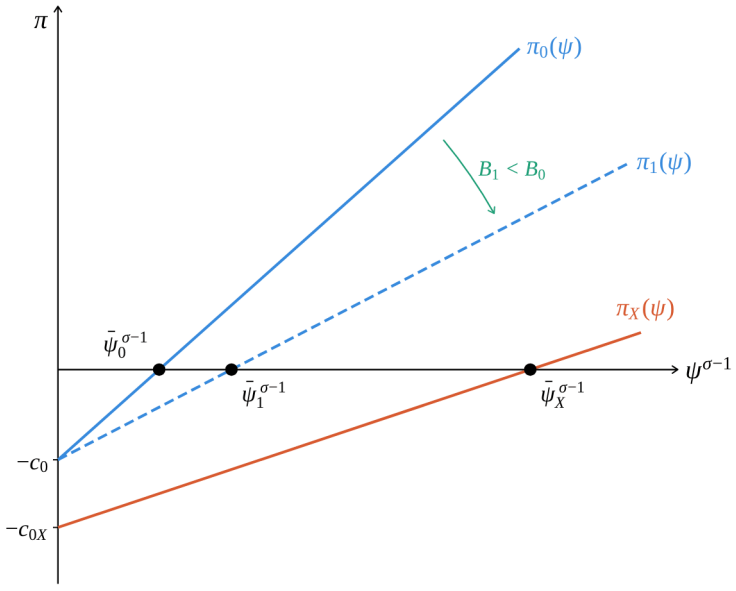

Figura B6.6: Modelo de Melitz. La apertura eleva el umbral de supervivencia y reasigna cuota de mercado hacia las empresas más productivas.

Estática comparativa: de la caída del beneficio interior y del acceso a la exportación se derivan las cuatro predicciones que constituyen la aportación del modelo:

1. *Efecto selección*: la elevación del umbral interior expulsa del mercado a las empresas menos productivas, que en autarquía sobrevivían.

2. *Efecto reasignación*: la cuota de mercado liberada por las salientes, junto con la demanda exterior, se desplaza hacia las empresas más productivas, que expanden su producción.

3. *Productividad agregada*: en consecuencia, la productividad media de la industria aumenta sin que ninguna empresa mejore su productividad individual, progreso puramente composicional que constituye el hallazgo central del modelo.

4. *Ganadores y perdedores*: las empresas exportadoras ganan cuota de mercado (y beneficio, salvo las más próximas al umbral de exportación), las que solo abastecen el mercado interior pierden beneficio y las menos productivas desaparecen. La apertura genera, por tanto, una ganancia agregada con perjuicio para una parte identificable del tejido empresarial.

3.1.4. Conclusiones

Aspectos positivos: el modelo explica los tres hechos que la nueva teoría no podía justificar (apartado 2.4), con un ajuste empírico notable.

Limitaciones: la productividad se obtiene de una distribución aleatoria exógena, de modo que el modelo no explica de dónde procede ni cómo puede mejorarse, cuestión que aborda la literatura posterior sobre inversión en productividad. La distinción entre exportar y no exportar es además binaria, sin considerar la elección entre exportar e invertir en el exterior, que desarrollarán Helpman, Melitz y Yeaple (2004). Por último, el ajuste, tanto por el margen extensivo como por el intensivo, se produce sin fricciones laborales ni costes de reasignación, lo que omite precisamente los costes de transición que el efecto selección impone sobre los trabajadores desplazados.

3.2. Extensiones al modelo de heterogeneidad de empresas

El marco de heterogeneidad ha dado lugar a extensiones que enriquecen el análisis, entre las que destacan el modelo de inversión extranjera directa (IED) y el de elección de la forma organizativa óptima.

3.2.1. Exportar o invertir: modelo de IED horizontal (Helpman, Melitz y Yeaple, 2004)

Origen: Elhanan Helpman, Marc Melitz y Stephen Yeaple (2004) propusieron el modelo.

Idea: para vender en el exterior, una empresa puede exportar o instalarse en el país de destino mediante inversión extranjera directa horizontal. La elección plantea la *disyuntiva entre proximidad y concentración* (*proximity-concentration trade-off*). Unos costes de exportación elevados incentivan la instalación en destino, pero las economías de escala internas aconsejan concentrar la producción en origen y exportar.

Supuestos: se mantienen los del modelo de heterogeneidad, al que se añade la decisión sobre la forma de acceso. La disyuntiva nace de la estructura de costes, ya que exportar comporta menores costes fijos pero mayores costes variables, mientras que invertir comporta mayores costes fijos, al duplicar la planta, pero menores costes variables, al evitarse el transporte.

Razonamiento: sean $\mathrm{CT}_X = c_{0X} + c_X \cdot x$ y $\mathrm{CT}_I = c_{0I} + c_I \cdot x$ los costes de servir el mercado exterior por una y otra vía. Simplificando con $c_{0X} = 0$ y $c_I = 0$, la empresa resulta indiferente cuando $\bar{x} = c_{0I}/c_X$. Por encima de ese umbral, el coste variable de exportar excede el coste fijo de invertir y conviene la inversión directa, y por debajo conviene exportar.

Desarrollo gráfico: la elección entre exportar e invertir se representa en el plano $(\psi^{\sigma-1}, \pi)$ (figura B6.7):

1. *Rectas de beneficios interiores y de exportación*, $\pi_1(\psi)$ y $\pi_X(\psi)$: las del modelo abierto, con ordenadas en el origen $-c_0$ y $-c_{0X}$.

2. *Recta de beneficios de la inversión directa*, $\pi_I(\psi)$: más inclinada que la de exportación, al ahorrarse el coste variable, pero con menor ordenada en el origen, $-c_{0I}$, por el mayor coste fijo.

3. *Productividades de corte*: $\bar{\psi}_1^{\sigma-1}$ separa a las empresas que salen de las que sirven solo el mercado interior, y $\bar{\psi}_X^{\sigma-1}$, a estas de las que exportan. Para productividades bajas la inversión rinde menos que la exportación, pero ambas rectas se cruzan en $\bar{\psi}_I^{\sigma-1}$, a partir de la cual la inversión pasa a ser la opción más rentable.

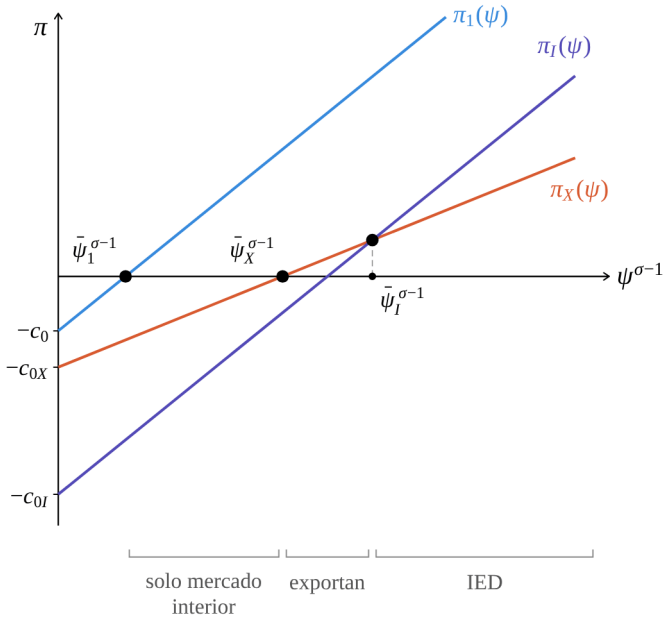

Figura B6.7: Exportar o invertir (Helpman, Melitz y Yeaple, 2004). Las empresas más productivas realizan IED, las intermedias exportan y las menos productivas sirven solo el mercado interno.

Conclusiones:

- *Aspectos positivos*: el modelo predice una ordenación completa de las empresas por productividad, en la que las menos productivas atienden solo el mercado nacional, las intermedias exportan y las más productivas invierten en el exterior. Los propios autores la contrastaron con datos de empresas estadounidenses y confirmaron que las multinacionales más productivas sustituyen exportación por inversión, especialmente en sectores con costes de transporte elevados y economías de escala de planta débiles.

▪ *Limitaciones*: el modelo se ocupa únicamente de la inversión horizontal, que replica la producción en destino, y no de la vertical, que fragmenta el proceso productivo entre países. Tampoco considera formas intermedias de acceso, como la licencia o la empresa conjunta, que en la práctica son frecuentes.

3.2.2. Elección de la forma de organización óptima (Antràs y Helpman, 2004)

Origen: Pol Antràs y Elhanan Helpman (2004) abordan dónde realizar el proceso productivo y cómo estructurarlo (qué actividades internalizar y cuáles subcontratar). La respuesta descansa en la teoría de los contratos incompletos.

Idea: al no poder especificarse contractualmente todas las contingencias, la propiedad del activo determina el poder de negociación y, con él, los incentivos a invertir de cada parte.

Supuestos: un bien final, un único factor puro (el trabajo), dos *inputs* que reflejan las funciones de la empresa (servicios centrales o estratégicos y servicios de manufactura) y dos regiones, norte y sur, con salarios inferiores en el sur.

Razonamiento: la combinación de las decisiones de propiedad y de localización genera cuatro formas organizativas:

Tabla B6.1: Formas organizativas según la propiedad y la localización de la producción.

	Subsidiaria	Empresa independiente
Región norte	Integración nacional	Externalización nacional
Región sur	IED vertical	Externalización exterior (*foreign outsourcing*)

De ahí resultan dos predicciones. La localización depende de la relación entre los diferenciales salariales y los costes fijos de cada opción y, en último término, de la productividad, con el mismo criterio del modelo de heterogeneidad. Así, solo las empresas más productivas pueden asumir los costes fijos de trasladar actividad al sur. La propiedad depende además de la intensidad en servicios centrales, de modo que en los sectores intensivos en ellos las empresas más productivas tienden a integrarse verticalmente en el sur.

Conclusiones:

▪ *Aspectos positivos*: el modelo proporciona el fundamento microeconómico de la fragmentación internacional de la producción en cadenas globales de valor (CGV). Explica, además, por qué una misma decisión de deslocalización adopta la forma de filial en unos casos y de proveedor independiente en otros. Su implicación conceptual es de alcance, pues el objeto del comercio deja de ser el bien terminado para ser la tarea, lo que difumina la distinción entre comercio de bienes y de servicios.

▪ *Limitaciones*: el marco es estático y no explica la dinámica de las cadenas una vez constituidas, ni su reconfiguración ante perturbaciones. Prescinde asimismo del papel de las instituciones del país receptor, cuya calidad contractual condiciona precisamente la viabilidad de la externalización, y del reparto del valor a lo largo de la cadena. Ese reparto puede ser muy desigual, como en el caso del iPod, en el que Apple, la empresa líder, capta buena parte del valor (Dedrick, Kraemer y Linden, 2010).

> Las cadenas globales de valor, su concepto, tipología, medición y evolución reciente se desarrollan en el tema B18 §1.3.

3.2.3. Conclusiones

Aspectos positivos: los modelos de heterogeneidad de empresas se han convertido en el marco de referencia del análisis moderno del comercio internacional. Permiten estudiar de forma unificada los márgenes extensivo (cuántas empresas y productos participan en el comercio) e intensivo (cuánto exporta cada una), las decisiones de exportación, inversión y organización, y los efectos del comercio sobre la productividad agregada. Ninguno de estos aspectos es visible, en cambio, en los modelos de empresa representativa.

Limitaciones: las extensiones heredan del modelo de Melitz la productividad exógena y el ajuste sin costes de reasignación. Además, el modelo de IED se ciñe a la inversión horizontal y el de forma organizativa es estático.

3.3. Otras aportaciones de la novísima teoría

Dos trabajos extienden el análisis de la liberalización comercial con empresas heterogéneas. Baldwin y Forslid (2010) estudian sus efectos sobre la variedad de productos y el reparto de la renta, y Bernard, Redding y Schott (2007), su interacción con la ventaja comparativa y con la remuneración de los factores.

3.3.1. Efecto antivariedad (Baldwin y Forslid, 2010)

Origen: Richard Baldwin y Rikard Forslid (2010) examinan la liberalización comercial en el marco de Melitz (2003).

Idea: con empresas heterogéneas, la liberalización redistribuye la renta y puede reducir el número de variedades consumidas, aunque, con preferencias CES, siempre eleva el bienestar.

Razonamiento: la liberalización recíproca tiene dos efectos destacados:

1. *Efecto redistributivo*: un efecto de tipo Stolper-Samuelson eleva el salario real y la remuneración de los propietarios de las empresas exportadoras, y puede reducir la de los propietarios de las que solo venden en el mercado interior.

2. *Efecto antivariedad*: la selección expulsa más variedades nacionales de las que se ganan vía importación, porque cada variedad importada, más barata, desplaza a más de una nacional. El número de variedades consumidas puede así caer, sobre todo en los países pequeños, aunque el bienestar aumenta porque la selección reduce el índice de precios.

Conclusiones:

- *Aspectos positivos*: muestra que la ganancia del comercio con empresas heterogéneas procede de la selección y no exige más variedad, y el análisis conecta con la crítica antiglobalización que asocia la globalización a la homogeneización de los productos.

- *Limitaciones*: la utilidad CES trata igual a todas las variedades, de modo que perder las locales no tiene un coste propio, y los autores admiten que el apego cultural a ellas podría atenuar o invertir la ganancia.

3.3.2. Ventaja comparativa y heterogeneidad de empresas (Bernard, Redding y Schott, 2007)

Origen: Andrew Bernard, Stephen Redding y Peter Schott (2007) integran el modelo de Melitz en el de Heckscher-Ohlin-Samuelson, con países que difieren en dotaciones relativas e industrias que difieren en intensidad factorial.

Idea: la selección de empresas que provoca la apertura es más intensa en las industrias con ventaja comparativa y amplía esa ventaja.

Razonamiento: la apertura tiene dos efectos principales:

1. *Darwinismo empresarial*: la selección, la reasignación y la subida composicional de la productividad media del modelo de Melitz son mayores en las industrias con ventaja comparativa, lo que crea diferencias de productividad de tipo ricardiano que amplían la ventaja comparativa inicial.

2. *Encarecimiento de los factores*: la expansión de las empresas exportadoras eleva la demanda de factores y su precio, con lo que las de baja productividad dejan de cubrir sus costes fijos y salen, de modo que el mercado de factores es un canal adicional del efecto selección. Además, las ganancias de productividad atenúan, e incluso pueden invertir, la caída del salario real del factor escaso que predice el teorema de Stolper-Samuelson.

Conclusiones:

- *Aspectos positivos*: une la ventaja comparativa con la heterogeneidad de empresas y explica a la vez la reasignación entre industrias y dentro de ellas.

- *Limitaciones*: hereda las del modelo de Melitz y se ciñe a dos países, dos factores y dos industrias.

3.4. Conclusiones

Aspectos positivos: la evidencia a nivel de empresa ha respaldado ampliamente el planteamiento. Andrew Bernard y Bradford Jensen (1995, 1999) documentaron para Estados Unidos las primas del exportador descritas en el apartado 2.4. Mostraron, además, que la productividad elevada precede a la entrada en los mercados exteriores, lo que apunta a la autoselección como mecanismo dominante, en línea con Melitz[31].

En cuanto a los efectos agregados, Daniel Trefler (2004) documenta para el acuerdo entre Canadá y Estados Unidos ganancias notables de productividad manufacturera. Las ganancias se deben tanto a la contracción y salida de las plantas menos productivas como, en los sectores exportadores, a mejoras dentro de las propias plantas.

Limitaciones: la propia evidencia matiza el alcance del planteamiento, pues el modelo excluye el aprendizaje por la exportación (*learning by exporting*). La revisión de Wagner (2007), con 54 estudios de datos de empresa, no lo encuentra de forma sistemática, a diferencia de la autoselección (véanse las Conclusiones del tema). A ello se añade que la liberalización genera ganadores y perdedores dentro de cada industria, con costes de ajuste transitorios sobre el empleo que Trefler documenta junto a las ganancias. Las políticas de acompañamiento en formación, movilidad y protección social se convierten, así, en complemento necesario de la apertura para sostener su viabilidad política.

[31] Las primas de Bernard y Jensen se mantienen controlando por sector y región.

Conclusiones

La nueva teoría parte de hechos que la teoría tradicional no podía acomodar, como el comercio intraindustrial entre países de dotaciones parecidas. Con economías de escala internas y competencia monopolística, Krugman (1979b, 1980) muestra que el comercio surge aunque los países sean idénticos, que amplía la variedad disponible y que concentra la producción en el mercado mayor. Los modelos de oligopolio explican el *dumping* recíproco, y la nueva economía geográfica hace depender la localización de la historia y de los costes de transporte. En cambio, la novísima teoría traslada el análisis a la empresa. En el modelo de Melitz (2003) solo las empresas más productivas cubren el coste fijo de exportar, y la apertura eleva la productividad agregada al reasignar cuota hacia ellas y expulsar a las menos productivas. Las ganancias de la intervención resultan frágiles y las de la apertura, robustas, aunque generan ganadores y perdedores dentro de cada industria.

La nueva teoría importa al comercio los modelos de la organización industrial, la competencia monopolística con preferencia por la variedad (tema A18) y el oligopolio (tema A19), para explicar lo que la teoría de las dotaciones no podía (tema B05). La competencia imperfecta abre, a su vez, la puerta a la intervención. Los beneficios extraordinarios justifican la política comercial estratégica, y los costes fijos de exportar, la promoción exterior (tema B08). Por su parte, los efectos de variedad y de selección explican las ganancias dinámicas de la integración económica (tema B10) y la reasignación de recursos hacia las empresas más productivas que vincula el comercio con el crecimiento (tema B09).

Arkolakis, Costinot y Rodríguez-Clare (2012) muestran que, en una amplia clase de modelos, bastan la participación del gasto interior y la elasticidad del comercio para medir las ganancias del comercio. Dados esos datos, los nuevos modelos no las aumentan. Melitz y Redding (2015) replican que la selección de empresas es un canal adicional de ganancia y que esos dos estadísticos dejan de bastar si se relajan los supuestos paramétricos que los hacen suficientes. Sigue abierta, además, la causalidad entre productividad y exportación. Bernard y Jensen (1999) atribuyen la prima del exportador a la autoselección, pero Atkin, Khandelwal y Osman (2017), en un experimento aleatorio con productores egipcios de alfombras, encuentran aprendizaje por la exportación (*learning by exporting*).

B7. Política comercial I: instrumentos y efectos

La política comercial (I). Instrumentos y efectos. Barreras arancelarias y no arancelarias. Otros instrumentos tradicionales.

Introducción

Las autoridades hacen **política comercial** cada vez que intervienen en los intercambios de su país con el exterior para reducir lo que pierde o aumentar lo que gana con el comercio internacional. Esa intervención puede mirar hacia dentro o hacia fuera. Mira hacia dentro la política comercial **defensiva**, que trata de frenar la entrada de importaciones y, con ella, la dependencia del exterior, sea con aranceles o con otras trabas (contingentes, barreras técnicas y administrativas, entre otras). Mira hacia fuera la política comercial **ofensiva**, que se ha entendido de dos maneras. En su versión clásica, el Estado respalda directamente a los exportadores, o a ciertos sectores que exportan, p. ej. con subvenciones. En su versión actual, en cambio, el sector público coopera con el privado[1] y ayuda por igual a todas las empresas a internacionalizarse, no solo vendiendo fuera sino también invirtiendo en otros países y, en general, dando a su actividad una dimensión global. Esta segunda versión, hoy extendida en muchos países, se conoce como política de promoción exterior o de internacionalización. Si los mercados son de competencia imperfecta y las empresas se condicionan mutuamente, la política comercial es además **estratégica**[2] cuando intenta que sean las empresas del propio país las que se queden con los beneficios extraordinarios que pueda haber en el mercado.

La importancia de la política comercial se debe a que, históricamente, ha estado en el centro de dos grandes debates. El primero, sobre su *justificación*[3], es la controversia recurrente entre libre comercio[4] y proteccionismo, que se remonta a la crítica de Smith y Ricardo al mercantilismo y a la defensa de la industria naciente de Hamilton y List. La apertura comercial beneficia a unos sectores y perjudica a otros, redistribuye la renta y puede provocar la desaparición de empresas, de modo que sus ganancias se reparten de forma desigual y generan conflictos de intereses que abren la puerta a la intervención pública. El segundo, sobre su *articulación*, es si cada país debe fijar su política comercial de forma discrecional o con arreglo a normas negociadas multilateralmente, en el marco del GATT desde 1947 y de la OMC desde 1995. Estas normas

[1] En España, según la Ley 14/2013, de apoyo a los emprendedores y su internacionalización, el fomento de la presencia exterior de las empresas y de los emprendedores forma parte de la política de internacionalización.

[2] El adjetivo procede de la nueva teoría del comercio con competencia imperfecta. La teoría pura tradicional supone competencia perfecta.

[3] Entre los argumentos en favor de la intervención destacan la redistribución de rentas, la destrucción de empresas, el arancel óptimo en el caso de un país grande o el argumento de la industria naciente. Entre los argumentos en favor del libre comercio figuran las ganancias de eficiencia del intercambio, el aprovechamiento de las economías de escala, la disciplina competitiva sobre las empresas nacionales y el riesgo de captura del regulador y de búsqueda de rentas que acompaña a toda protección.

[4] Sinónimo de librecambio: extensión del *laissez faire* al ámbito internacional.

favorecen la coordinación no solo de la política comercial, sino del conjunto de las políticas económicas. Ambos debates han recobrado actualidad con la guerra comercial entre EE. UU. y China iniciada en 2018, la crisis del sistema multilateral de solución de diferencias y la escalada arancelaria de 2025.

Este tema examina los instrumentos de la política comercial y sus efectos:

- *Justificación*: breve repaso del debate entre el proteccionismo y el libre cambio para entender por qué se recurre a la política comercial.

- *Barreras arancelarias*: el arancel como instrumento y sus efectos en equilibrio parcial y general.

- *Barreras no arancelarias y otros instrumentos tradicionales*: cuotas, subvenciones e impuestos a la exportación, restricciones voluntarias y demás barreras no arancelarias, y sus efectos.

1. Justificación de la política comercial

El debate normativo sobre la política comercial se ha articulado, históricamente, en torno a dos grandes familias de argumentos, los favorables al librecambismo y los favorables a la intervención.

1.1. Argumentos en favor del librecambismo

Desde las aportaciones de Adam Smith (1776) y David Ricardo (1817), la teoría ha defendido el libre comercio como política comercial óptima, incluso cuando se adopta de forma unilateral.

1.1.1. Argumentos relacionados con efectos estáticos

Ventaja comparativa: la especialización conforme a la ventaja comparativa y el posterior intercambio generan ganancias mutuas de bienestar, con independencia de cuál sea la fuente de esa ventaja, tecnológica o dotacional (tema B05, apartados 1.2 y 2.3).

Eficiencia económica: las barreras arancelarias y no arancelarias actúan como costes de transacción que distorsionan los precios relativos y generan pérdidas irrecuperables de eficiencia, por lo que, en ausencia de otras distorsiones, su eliminación restaura el óptimo paretiano. A ello se añade el efecto procompetitivo de las importaciones, que disciplinan el poder de mercado interno, y la mayor variedad disponible para consumidores y empresas.

1.1.2. Argumentos relacionados con efectos dinámicos

Tamaño de mercado: la apertura amplía el mercado de cada empresa y le permite aprovechar economías de escala internas, que crean comercio y ganancias aun en ausencia de ventaja comparativa. Por ejemplo, en el modelo de Krugman (1979b), cada empresa produce a mayor escala y menor coste medio, y la ganancia llega a los trabajadores como mayor salario real (tema B06, apartado 2.1.2).

Crecimiento económico: las economías de aprendizaje y los efectos desbordamiento del conocimiento asociados al libre flujo internacional de bienes e ideas pueden elevar la tasa de crecimiento, no solo el nivel de renta. La integración amplía el mercado de cada innovación y el acervo de ideas disponible para investigar (Rivera-Batiz y Romer, 1991), aunque el signo del efecto depende de la ventaja comparativa de cada país (tema B09, apartado 3.1).

1.2. Argumentos en favor del proteccionismo

Frente a ello, los argumentos en favor de la intervención pueden agruparse en dos bloques, los derivados de los fallos del mercado nacional y los derivados del propio proceso de liberalización.

1.2.1. Argumentos derivados de los fallos del mercado nacional

Los mercados nacionales de capital o de trabajo pueden presentar imperfecciones (un sistema financiero ineficiente, rigideces laborales). En ese caso, el trasvase del ahorro y de los factores desde los sectores tradicionales hacia los sectores nuevos con mayores beneficios potenciales puede no producirse, lo que restringe el crecimiento de las industrias emergentes y justificaría, en principio, su protección.

Incluso en estos casos, no obstante, el proteccionismo comercial es un instrumento de *second best*, puesto que la solución de *first best* consiste en corregir el fallo del mercado en su origen (Bhagwati y Ramaswami, 1963), sin introducir una distorsión adicional en el comercio.

1.2.2. Argumentos derivados del proceso de liberalización (I): la industria naciente (Hamilton, 1791; Raymond, 1820; List, 1841)

Origen: Alexander Hamilton enunció el argumento de la industria naciente en su *Informe sobre las manufacturas* (1791), y Daniel Raymond (1820) y Friedrich List (1841) lo desarrollaron.

Idea: el argumento sostiene que algunas industrias en etapas tempranas necesitan protección temporal frente a rivales más competitivos, pues la capitalización y el aprendizaje requieren tiempo durante el cual esas industrias aún no disponen de ventaja comparativa.

Razonamiento: para que la protección esté justificada, la industria candidata debe superar un doble filtro:

▪ *Criterio de Mill*: la protección debe ser estrictamente temporal y la industria debe ser capaz, transcurrido el periodo de aprendizaje, de competir internacionalmente sin ayuda pública[5].

▪ *Criterio de Bastable* (1891): las ganancias futuras descontadas de la industria madura deben superar los costes acumulados que la protección impone a la economía.

Conclusiones:

▪ *Aspectos positivos*: el argumento justifica una protección compatible con la ventaja comparativa a largo plazo, y los criterios de Mill y de Bastable permiten contrastar cuándo procede.

▪ *Limitaciones*: el argumento está sujeto a dos críticas:

- *Momento de la inversión*: no siempre conviene invertir hoy en industrias que tendrán ventaja comparativa en el futuro, pues puede ser preferible esperar a que se acumule el capital humano necesario.

- *Capacidad de selección*: es dudoso que el sector público, expuesto a la captura y a problemas de información, identifique mejor que los inversores privados a las industrias prometedoras[6].

[5] Por competitiva internacionalmente se entiende que, dados los precios internacionales, tenga la suficiente productividad media.

[6] Por ambas razones (momento de la inversión y capacidad de selección), en lugar de proteger cabe fomentar antes la competencia interna (liberalización de sectores, defensa de la competencia), y así las empresas nacionales maduran sin que las extranjeras las expulsen.

1.2.3. Argumentos derivados del proceso de liberalización (II): la distribución de la renta

Teorema de Stolper y Samuelson (1941): con dos bienes, dos factores y producción diversificada, si un bien se encarece en términos relativos, mejora la retribución real del factor que su producción emplea con más intensidad y empeora la del otro. En la medida en que el comercio altera los precios relativos, el reparto de las ganancias de la apertura es desigual, ya que ganan los propietarios del factor abundante y pierden los del factor escaso.

Modelo de factores específicos (Samuelson, 1971, y Jones, 1971): con un factor móvil entre sectores (el trabajo) y otros específicos de cada sector, la apertura beneficia a los propietarios de los factores específicos de los sectores exportadores. Perjudica, en cambio, a los de los sectores que compiten con las importaciones.

Efecto selección (Melitz, 2003; tema B06, apartado 3.1): la liberalización expulsa del mercado a las empresas menos productivas, lo que, pese a su ganancia agregada de eficiencia, puede generar destrucción de empresas y desempleo transitorio de recursos.

1.2.4. Argumentos derivados del proceso de liberalización (III): la balanza de pagos

Efectos sobre el ahorro público: la liberalización supone la pérdida de los ingresos arancelarios, relevantes en países con administraciones tributarias débiles, y la renuncia, en el caso de un país grande, a la ganancia potencial del arancel óptimo.

Efecto absorción interna: la apertura aumenta el consumo, la inversión y el gasto público y, con ello, la absorción interna, lo que deteriora la cuenta corriente durante la transición.

Efectos apreciatorios reales: la falta de credibilidad institucional del proceso liberalizador puede provocar el adelanto de las compras de importaciones, presiones inflacionistas y una apreciación real que erosione la competitividad exterior.

1.2.5. Evidencia empírica

La evidencia empírica ofrece apoyo matizado a ambas posiciones. De un lado, los estudios transversales que recoge el tema B09, apartado 3.2, asocian apertura y crecimiento (Sachs y Warner, 1995), si bien la robustez de esa correlación ha sido cuestionada por problemas de medición y causalidad (Rodríguez y Rodrik, 2000). De otro, la experiencia de la protección a la industria naciente es ambivalente. La protección fracasó de forma generalizada en la industrialización por sustitución de importaciones latinoamericana, pero formó parte del éxito exportador de Asia oriental. Allí, la protección selectiva se combinó con disciplina exportadora y retirada efectiva de las ayudas.

1.3. Conclusiones

Los argumentos librecambistas presentan el comercio como un juego de suma positiva. En el plano estático, la especialización según la ventaja comparativa genera ganancias mutuas y la supresión de las barreras elimina pérdidas de eficiencia. En el dinámico, el mercado ampliado permite aprovechar economías de escala, y la difusión del conocimiento puede elevar la tasa de crecimiento.

Los argumentos proteccionistas parten, en cambio, de los fallos del mercado nacional, de la industria naciente y de los ganadores, perdedores y desequilibrios que genera la apertura. Sus

condiciones son exigentes, pues frente a un fallo del mercado la protección es un instrumento de *second best* y la industria naciente solo la justifica si supera los criterios de Mill y de Bastable. La evidencia apoya con matices ambas posiciones, ya que la protección fracasó en la sustitución de importaciones latinoamericana y acompañó, con disciplina exportadora, el éxito de Asia oriental. Cuando se interviene, la política comercial recurre a **barreras arancelarias**, cuyos efectos estudia la sección 2, y a **barreras no arancelarias** (p. ej. cuotas o restricciones voluntarias), objeto de la sección 3.

2. Barreras arancelarias: el arancel como instrumento y sus efectos

Se empieza por las barreras arancelarias, pues el arancel ha sido históricamente uno de los instrumentos más extendidos.

2.1. El arancel como instrumento

Antes de estudiar los efectos del arancel conviene precisar qué es y cómo se mide la carga arancelaria, con aranceles promedio o con la tasa de protección efectiva.

2.1.1. Definición

Un **arancel** es un impuesto o gravamen que se aplica sobre mercancías que cruzan la frontera. Los aranceles pueden clasificarse según el objeto de gravamen o según la forma de aplicación.

1. *Según el objeto de gravamen*:

- *A la importación*: gravan la mercancía que entra en un país (es decir, las importaciones), con carácter general con el doble objetivo de proteger la industria nacional y conseguir recaudación. Estos aranceles son, con gran diferencia, los más habituales.

- *A la exportación*: gravan la mercancía que sale de un país (es decir, las exportaciones), generalmente con el objetivo de impulsar el abastecimiento nacional, mejorar la relación real de intercambio (RRI) o conseguir recaudación. Estos aranceles son poco frecuentes y se denominan derechos aduaneros a la exportación.

- *De tránsito*: gravan las mercancías que atraviesan el territorio nacional con destino a un tercer país. Estos aranceles fueron históricamente relevantes como fuente de recaudación, pero hoy han desaparecido casi por completo.

2. *Según la forma en que se aplica*:

- *Ad valorem*: aplican un porcentaje arancelario fijo sobre el valor de un bien, de forma que $\mathrm{CT}(x) = (1+t) \cdot p \cdot x \to p_p = (1+t)p$, con $t > 0$ la tasa fija arancelaria, p_p el precio protegido y p el precio desprotegido. Este arancel es el más común y su principal ventaja es que el nivel de protección nominal que ofrece es siempre conocido y siempre el mismo, puesto que coincide con la tarifa aplicada y es independiente del nivel general de precios. Como contrapartida, el arancel *ad valorem* exige la valoración en aduana de la mercancía, lo que encarece su administración y facilita la manipulación de las facturas (subfacturación).

- *Fijos o específicos*: aplican una cantidad monetaria arancelaria fija por unidad del bien. Así, $\mathrm{CT}(x) = (p + t_e) \cdot x \to p_p = p + t_e$, con $t_e > 0$ la cantidad monetaria fija por unidad y x la cantidad (sea de importaciones o de exportaciones). La ventaja de este arancel es que resulta más sencillo de administrar, pues no exige valorar la mercancía. Sus dos desventajas, en cambio, son que es menos equitativo (regresivo)[7] y que la protección real que ofrece varía con el nivel de precios, porque la inflación erosiona una cuantía fija (y la deflación la amplifica).

- *Mixtos*: consisten en una combinación de ambos.

[7] El arancel específico grava con la misma cuantía por unidad las variedades baratas y las caras, de modo que su equivalente *ad valorem* es mayor en las baratas, que pesan más en el consumo de los hogares de renta baja.

2.1.2. Medición (I): enfoque basado en el arancel promedio

El arancel promedio es el indicador seguido por la teoría tradicional del comercio internacional y se puede calcular de dos formas:

1. *Media simple*: cociente entre la suma de los aranceles nominales de cada categoría y el número de categorías, $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$.

2. *Media ponderada*: ponderación de los tipos arancelarios en función del valor de las importaciones de cada categoría, $\bar{t} = \sum_{i=1}^{n} \omega_i t_i$, donde ω_i es la participación de las importaciones de la categoría i en las importaciones totales.

Conclusiones:

- *Aspectos positivos*: se trata de indicadores sencillos de calcular y de comparar internacionalmente, por lo que son los utilizados habitualmente en las estadísticas de organismos como la OMC o el Banco Mundial.

- *Limitaciones*: ambos indicadores presentan sesgos conocidos. La media simple puede sobrestimar la protección si unas pocas categorías con aranceles bajos concentran el comercio del país y las muchas categorías con aranceles altos apenas comercian. La media ponderada, por su parte, puede subestimarla, ya que un arancel prohibitivo reduce mucho la importación del producto y, con ella, el peso que recibe en la media.

2.1.3. Medición (II): enfoque basado en la protección efectiva o arancel implícito

Harry Johnson (1965), Béla Balassa (1965a) y Max Corden (1966) critican el enfoque del arancel promedio porque no tiene en cuenta el comercio de bienes intermedios[8]. Por ello, el arancel nominal sobre un producto puede ser muy diferente del arancel implícito que protege a la actividad que lo produce.

La **tasa de protección efectiva** es el incremento porcentual en el valor añadido unitario de un bien con respecto a una situación en la que no existen aranceles, manteniendo el tipo de cambio y todas las variables relevantes constantes.

Formalmente[9], la tasa o grado de protección efectiva se define como:

$$t_{\text{PE}} \equiv \frac{\text{VA}_{\text{CON}} - \text{VA}_{\text{SIN}}}{\text{VA}_{\text{SIN}}} \tag{1}$$

donde VA_{CON} y VA_{SIN} son el valor añadido unitario con arancel y sin él, de modo que (1) mide el "exceso" de valor añadido que genera el arancel.

Además, operando, se puede demostrar que:

$$(t_{\text{PE}} - t) = \beta(t - t_I), \qquad \beta = \frac{\alpha}{1-\alpha} > 0 \tag{2}$$

[8] Ese comercio supone más de la mitad del comercio de bienes (el 56 % en los países de la OCDE, según Miroudot, Lanz y Ragoussis, 2009).

[9] Se supone una tecnología de coeficientes fijos (Gandolfo, 2014, p. 244). Sin aranceles, el valor añadido unitario es $\text{VA}_{\text{SIN}} = p - \alpha p_I$, con p el precio del bien final, p_I el del intermedio y $\alpha \in (0,1)$ la participación del intermedio. Normalizando $p_I = p$, $\text{VA}_{\text{SIN}} = p(1-\alpha)$. Si se gravan ambos, $\text{VA}_{\text{CON}} = (1+t)p - (1+t_I)\alpha p$. Sustituyendo en (1), $t_{\text{PE}} = \dfrac{t - \alpha t_I}{1-\alpha}$ y, operando, se llega a (2).

donde t y t_I son los aranceles nominales sobre el bien final y sobre los bienes intermedios, $\alpha \in (0,1)$ es la participación de los bienes intermedios en el valor del bien final y β crece con ella. La expresión (2) muestra cómo el exceso del arancel efectivo sobre el arancel nominal del bien final equivale al exceso del arancel nominal del bien final sobre el arancel nominal de los bienes intermedios, ponderado por β.

Ejemplo: tasa de protección efectiva

Con una participación de los bienes intermedios de $\alpha = 0{,}5$, un arancel nominal del 10 % sobre el bien final y del 5 % sobre el bien intermedio, la tasa de protección efectiva es $t_{\text{PE}} = (0{,}10 - 0{,}5 \cdot 0{,}05)/(1 - 0{,}5) = 0{,}15$, es decir, un 15 %. Con (2), $t_{\text{PE}} - t = 1 \cdot (0{,}10 - 0{,}05) = 0{,}05$, de modo que la protección efectiva sobre el valor añadido supera a la nominal.

De la expresión (2) se deduce que la tasa de protección efectiva o arancel implícito depende de:

Estructura o diseño arancelario: el arancel implícito depende positivamente del arancel nominal sobre el bien final y negativamente del arancel nominal sobre los bienes intermedios. Por ello, podría llegar a ser negativo si $\beta t_I > (1+\beta)t$. El caso de un arancel implícito negativo no es una mera disquisición teórica, ya que las estimaciones empíricas de la protección efectiva lo registran en algunos sectores.

Grado de participación: cuanto mayor sea la participación de los bienes intermedios, mayor será el grado de protección efectiva, para una misma brecha positiva entre aranceles nominales ($t > t_I$).

Grado de sustituibilidad: la expresión (2) supone coeficientes fijos, de modo que este efecto no se deduce de ella. Cuanto mayor sea el grado de sustituibilidad, mayor será la protección efectiva de la que disfruta realmente la industria, porque la empresa podrá sustituir más fácilmente el bien intermedio gravado por otros no gravados.

Conclusiones:

- *Aspectos positivos*: el enfoque mide la protección allí donde importa, sobre la actividad productiva y su valor añadido. Además, permite identificar el fenómeno de la *escalada arancelaria* (*tariff escalation*), en el que estructuras con aranceles nominales crecientes con el grado de elaboración, moderadas en apariencia, esconden tasas de protección efectiva muy elevadas sobre las fases finales de transformación. Ello dificulta la industrialización exportadora de los países en desarrollo.

- *Limitaciones*: el indicador supone, por un lado, que el país es pequeño, por lo que sus aranceles no afectan a los precios mundiales de los *outputs* ni de los *inputs*. Por otro, supone una función de producción de tipo Leontief, sin sustituibilidad entre *inputs* intermedios, lo que sesga la medición cuando las empresas pueden recomponer su cesta de insumos.

2.1.4. Conclusiones

Aspectos positivos: el arancel es un instrumento transparente, pues figura en el precio, a diferencia de la cuota (apartado 3.1.1), y en su forma *ad valorem* ofrece un nivel de protección nominal conocido.

Limitaciones: el tipo nominal no mide la protección de cada actividad, que depende también de los aranceles sobre sus bienes intermedios, y los promedios que resumen la estructura arancelaria están sesgados.

Entendido el concepto de arancel y cómo medir el grado de restricción que supone, se analizan sus efectos económicos:

1. *Enfoques*: primero, los efectos sobre la industria de un bien importable Y, en equilibrio parcial, y, después, los efectos sobre la economía en su conjunto, en equilibrio general.

2. *Bienestar*: cada unidad monetaria ganada o perdida por cada agente económico (consumidores, sector público o empresas) tiene la misma importancia[10].

3. *Tamaño del país*: en ambos casos se distingue entre un país pequeño, que no puede afectar a los precios mundiales a través de su volumen comercial, y un país grande, que sí puede.

2.2. Análisis en equilibrio parcial: efectos sobre la industria del bien importable

Metodología y herramientas: el análisis se apoya en la herramienta propia del equilibrio parcial, la cruz marshalliana de oferta y demanda de la industria del bien importable.

Supuestos: salvo en la mención a la competencia imperfecta del apartado 2.2.3, se considera un mundo neoclásico (competencia perfecta, ausencia de costes de transporte y de otras fricciones).

2.2.1. Análisis de los efectos en un país pequeño y en competencia perfecta

Desarrollo gráfico: los efectos de un arancel en un país pequeño se representan en el plano cantidad-precio del bien importable Y (figura B7.1):

1. *Curva de oferta nacional*, $S(y)$: recoge la producción nacional del bien a cada precio y es creciente, porque en competencia perfecta refleja el coste marginal de la industria, que crece con la producción.

2. *Curva de demanda nacional*, $D(y)$: recoge el consumo nacional a cada precio y es decreciente, porque mide lo que los consumidores están dispuestos a pagar por cada unidad adicional, que disminuye conforme consumen más.

3. *Precio internacional*, p_0^*: horizontal, porque el país pequeño importa cualquier cantidad a ese precio sin alterarlo.

4. *Equilibrio inicial*: el país, pequeño y abierto al exterior, toma como dado el precio internacional, de modo que $p_0 = p_0^*$. A ese precio produce y_0^s (punto 0) y consume y_0^d (punto $0'$), es decir, es importador neto del bien Y e importa una cantidad $M_0 = y_0^d - y_0^s > 0$.

[10] Este supuesto resulta fundamental para resolver los problemas de agregación y poder utilizar curvas de indiferencia social (Södersten y Reed, 1994, p. 13).

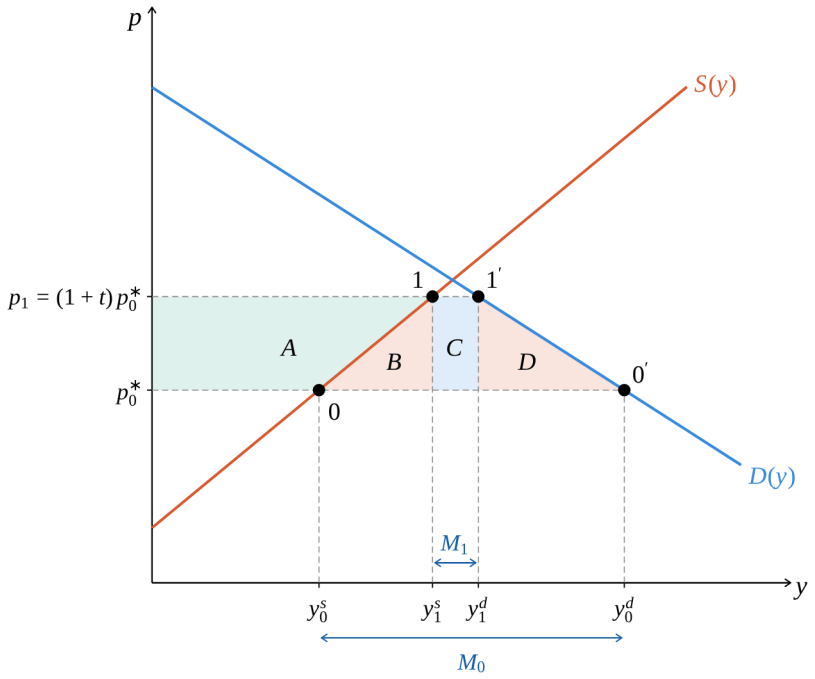

Figura B7.1: Arancel en un país pequeño (equilibrio parcial). El arancel tiene efectos producción, consumo y recaudación y genera una pérdida neta de bienestar.

Estática comparativa: se introduce un arancel *ad valorem* que eleva el precio nacional por encima del internacional, $p_1 = (1+t)p_0^* > p_0^*$, de modo que se alcanza un nuevo equilibrio dado por los puntos 1 y $1'$. El arancel afecta en cuatro dimensiones:

- *Producción (nacional)*: aumenta de y_0^s a y_1^s y el excedente del productor aumenta en el área A, pues los productores nacionales cubren parte de la demanda antes atendida con importaciones.

- *Consumo*: se reduce de y_0^d a y_1^d y el excedente del consumidor se reduce en las áreas A, B, C y D.

- *Recaudación*: el área C representa la recaudación del sector público, que vendrá dada por el producto de la cuantía monetaria del impuesto *ad valorem* por unidad del bien $(t\,p_0^*)$[11] y la cantidad importada (M_1).

- *Comercio internacional*: las importaciones se reducen debido a la mayor producción nacional y al menor consumo $(M_1 = y_1^d - y_1^s > 0)$.

El efecto neto sobre el bienestar será una pérdida de bienestar recogida en las áreas B y D $(\text{EN} = -(B+D))$. La pérdida del área B puede entenderse como la pérdida por el efecto producción (recursos nacionales empleados a un coste superior al precio mundial) y la del área D como la pérdida por el efecto consumo (consumo socialmente valioso al que se renuncia).

La magnitud y la distribución del efecto neto dependen de dos factores:

[11] En el caso de un impuesto específico, la idea es la misma, solo que la cuantía monetaria por unidad de bien importable viene dada directamente por t_e.

1. *Elasticidad-precio de la oferta y de la demanda* (+): cuanto mayor sea la sensibilidad de las cantidades al precio, mayores serán las distorsiones del arancel sobre el consumo y la producción.

2. *Cuantía del arancel* (++): las distorsiones aumentan de forma más que proporcional, pues la pérdida de eficiencia crece, aproximadamente, con el cuadrado del tipo. Por tanto, conviene repartir la imposición entre varios productos antes que concentrarla en uno solo.

Bienestar:

▪ *Pérdida de eficiencia nacional*: en un país pequeño, un arancel siempre reduce el bienestar social, puesto que se produce un exceso de gravamen[12] derivado de la fijación de un precio de mercado artificial.

▪ *Pérdida de eficiencia mundial*: *ceteris paribus*, dada la incapacidad del país para alterar la RRI, se reducirá el bienestar mundial.

▪ *Efecto redistributivo*: los consumidores transfieren a los productores el área A, como una subvención equivalente, y al sector público el área C, como recaudación. El arancel equivale, así, a la combinación de un impuesto sobre el consumo y una subvención a la producción del bien importable, ambos al mismo tipo, lo que explica su doble distorsión.

▪ *Efectos adicionales*: habría que añadir los costes de administración del impuesto y los efectos sobre el empleo[13].

2.2.2. Análisis de los efectos en un país grande y en competencia perfecta

Un país grande puede alterar la RRI, pues, al disminuir su demanda del bien importable en los mercados internacionales, el precio de sus importaciones cae.

Desarrollo gráfico: los efectos de un arancel en un país grande se representan en el plano cantidad-precio del bien importable (figura B7.2):

1. *Curvas de oferta y de demanda nacionales*, $S(y)$ y $D(y)$: recogen la producción y el consumo nacionales a cada precio, y son, respectivamente, creciente y decreciente, como en el caso anterior.

2. *Equilibrio inicial*: un país grande y abierto al exterior importa un bien. Al precio $p_0 = p_0^*$, que ahora depende de la demanda del país, produce y_0^s (punto 0) y consume y_0^d (punto $0'$), de modo que importa una cantidad $M_0 = y_0^d - y_0^s$.

[12] El exceso de gravamen es la pérdida adicional de bienestar social, medida en términos monetarios, en relación con la pérdida de bienestar social que se hubiera producido en el caso de un impuesto neutral (es decir, de suma fija) de igual recaudación.

[13] Si existe pleno empleo de recursos, el aumento de la producción nacional del bien importable exige detraer recursos de otros sectores, por lo que no genera empleo neto. Solo con recursos ociosos podría el arancel elevar el empleo nacional.

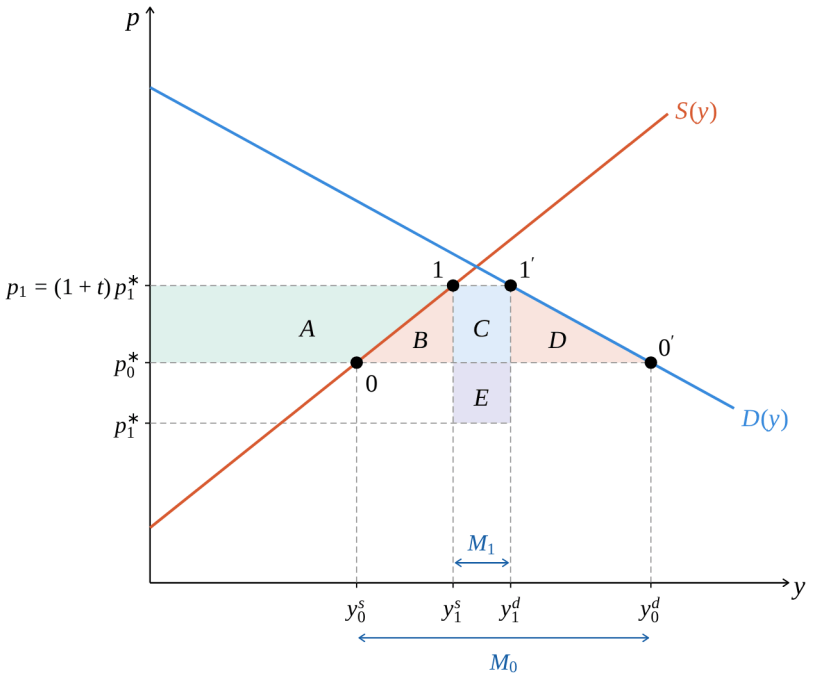

Figura B7.2: Arancel en un país grande (equilibrio parcial). La mejora de la relación real de intercambio puede compensar la pérdida de eficiencia.

Estática comparativa: en este caso, el arancel eleva el precio nacional pero, además, afecta al precio internacional. El precio internacional cae, de p_0^* a p_1^*, en la cuantía que los exportadores extranjeros absorben del arancel, y el precio nacional pasa a $p_1 = (1+t)\,p_1^*$ (puntos 1 y 1′). Los efectos son similares a los del caso anterior, con una ganancia adicional para el país que impone el arancel:

- *Consumo*: el país importador consume menos y pierde las áreas A, B, C y D. El consumo crece, en cambio, en el país exportador, cuyo mercado interno se abarata.

- *Producción*: el país importador produce más (ganancia del área A) y el exportador, menos.

- *Comercio internacional*: las importaciones del país que impone el arancel se reducen y, con ellas, las exportaciones del país exportador, que son su contrapartida.

- *Recaudación*: en el país importador nace una nueva fuente de ganancia para el sector público, el área E, que representa la parte de la carga arancelaria que asumen los exportadores extranjeros como consecuencia de la caída del precio internacional del bien importable. Así, la recaudación total es C + E. En cambio, en el país exportador la recaudación no cambia.

- *Efecto neto*: en el país importador será ambiguo y dependerá de si la ganancia de relación real de intercambio (E) es mayor o menor que la suma de la distorsión sobre la producción (B) y la distorsión sobre el consumo (D). En el país exportador se suman sus propias distorsiones internas de producción y consumo y la pérdida de relación real de intercambio (E), que coincide exactamente con la ganancia que obtiene el país importador, por lo que su efecto neto es inequívocamente negativo. Para el mundo en su conjunto, las transferencias se cancelan y solo quedan las distorsiones, de modo que la pérdida mundial es inequívoca.

Bienestar: a diferencia del país pequeño, un país grande puede obtener con el arancel ganancias potenciales de bienestar social, por lo que podría existir un arancel socialmente óptimo.

2.2.3. Mención a la competencia imperfecta

Libre comercio: en un entorno de monopolio, la apertura elimina el comportamiento monopolista, pues el precio internacional pone un techo al precio y reemplaza a la curva de demanda de autarquía. Por ello, la mejor estrategia del monopolista pasa a ser producir hasta igualar su coste marginal al precio mundial, como cualquier otra empresa.

Fijación de un arancel: el monopolista sigue viendo limitada su capacidad para aumentar el precio y, por tanto, se seguirá comportando como si estuviera en competencia perfecta internacional.

No obstante, si el arancel es prohibitivo, es decir, tan elevado que elimina completamente las importaciones, los efectos serán diferentes en función de la naturaleza de la industria:

1. *Competitiva*: una vez eliminadas las importaciones, cualquier aumento posterior del arancel no tendrá efecto.

2. *Monopolista*: un arancel prohibitivo permitirá al monopolista aumentar su precio, acercándose a su precio de máximo beneficio.

2.2.4. Conclusiones

En equilibrio parcial, el arancel reduce siempre el bienestar de un país pequeño, y en uno grande su efecto depende de que la ganancia de relación real de intercambio supere las distorsiones.

Aspectos positivos: el análisis mide con áreas de excedente lo que gana o pierde cada agente y separa la pérdida de eficiencia de la ganancia de relación real de intercambio.

Limitaciones: no considera la interrelación entre distintas industrias, que se aborda a continuación, de modo que omite la caída de las exportaciones y los efectos sobre la renta de los factores que muestra el equilibrio general.

2.3. Análisis en equilibrio general: efectos sobre la economía en su conjunto

El equilibrio general requiere herramientas propias, las curvas de oferta neta y las de indiferencia social comerciales. Con ellas se repiten los casos del país pequeño y del país grande, y se llega al arancel óptimo, a la guerra arancelaria y a las paradojas de Lerner y Metzler.

2.3.1. Instrumentos de análisis

Metodología y herramientas: la frontera de posibilidades de producción (FPP) subyace a las dos herramientas del análisis, las curvas de oferta neta o demandas recíprocas y las curvas de indiferencia social comerciales, pero queda implícita por claridad expositiva.

Curvas de indiferencia social comerciales (CISC): cada una reúne las combinaciones de importaciones y exportaciones (volúmenes de comercio) que reportan al país igual bienestar social[14]. En cuanto a su forma, estas son:

[14]Ni la homoteticidad de las preferencias ni una distribución de la renta constante garantizan por sí solas la agregación (Södersten y Reed, 1994, pp. 13-14). De ahí el supuesto de que la redistribución no es costosa.

- *Crecientes*: tienen pendiente positiva en el plano (X, Y) porque las exportaciones pueden considerarse un mal (bienes a los que se renuncia para pagar las importaciones sin incurrir en déficit comercial[15]). Por ello, para mantener constante el bienestar, cantidades adicionales de bienes importados (un bien) deben compensarse con cantidades adicionales de bienes exportados (un sacrificio).

- *Convexas*: se supone que aumentos sucesivos e idénticos de las importaciones requieren, para mantener el bienestar constante, un aumento cada vez menor de las exportaciones (es decir, la relación marginal de sustitución entre exportaciones e importaciones es decreciente)[16].

> Las curvas de oferta neta y los supuestos de este análisis se desarrollan en el tema B05, apartado 2.1. Aquí se recogen las propiedades imprescindibles para seguir el análisis del arancel.

Curvas de oferta neta (CON) o demandas recíprocas: recogen, para cada RRI internacional, las exportaciones que un país está dispuesto a entregar a cambio de cada volumen de importaciones. Pasan por el origen y se representan en orientación especular, pues el bien que exporta un país es el que importa el otro. En el plano (X, Y), presentan un tramo elástico o "normal", de pendiente positiva, y pueden presentar otro inelástico o "anómalo", de pendiente negativa, en el que la RRI es tan favorable que se puede seguir importando aunque las exportaciones disminuyan.

Si la FPP es cóncava, las curvas de oferta neta son convexas, y una curva perfectamente elástica (una recta desde el origen) describe a quien comercia cualquier volumen a la RRI dada, como el resto del mundo frente a un país pequeño.

Supuestos:

- *Entorno*: dos bienes (X, Y), exportables e importables, cuyos excesos reflejan las exportaciones y las importaciones.

- *Agentes*: se prescinde de cómo reparte el sector público los ingresos arancelarios[17].

- *Mercados*: competencia perfecta en todos los mercados, nacionales e internacionales, sin costes de transporte, con libre movilidad nacional y pleno empleo de los recursos[18].

- *Sector exterior*: la balanza comercial está siempre equilibrada, ya que no existen pagos ni deuda más allá de los flujos comerciales de mercancías y el valor se mide a través de los propios bienes, es decir, se produce trueque. Por ello, la RRI equivale a la ratio de importaciones frente a exportaciones. El país tiene ventaja comparativa en X frente al resto del mundo, de modo que exporta X e importa Y.

[15] El déficit comercial queda excluido porque, en el modelo estático de trueque, la balanza comercial equilibrada no es un objetivo, sino la restricción presupuestaria del país. Esta visión se contrapone a la mercantilista, que busca el superávit comercial.

[16] Intuitivamente, el país "prefiere la variedad", es decir, combinaciones equilibradas de volúmenes exportados e importados frente a combinaciones extremas.

[17] La recaudación puede repartirse entre los consumidores o gastarse en el bien importable, en el exportable o en una combinación de ambos (Gandolfo, 2014, p. 229).

[18] Este supuesto, junto con el hecho de que los bienes sean susceptibles de ser fabricados nacionalmente, permite ilustrar el descenso de las exportaciones a la vez que las importaciones cuando se introduce un arancel.

2.3.2. Análisis de los efectos en un país pequeño y en competencia perfecta

Desarrollo gráfico: los efectos de un arancel en un país pequeño se representan en el plano exportaciones-importaciones[19] (figura B7.3):

1. *Curva de oferta neta del país*, CON_0: recoge las exportaciones que el país entrega a cambio de cada volumen de importaciones y es normal, es decir, creciente (tramo elástico).

2. *Curva de oferta neta del resto del mundo*, CON^R: perfectamente elástica, una recta desde el origen de pendiente igual a la RRI internacional, RRI^*, porque, al ser el país pequeño, las alteraciones de su volumen comercial no afectan a la RRI internacional.

3. *Curvas de indiferencia social comerciales*, CISC_0 y CISC_1: reúnen los volúmenes de comercio con un mismo bienestar social y son crecientes y convexas.

4. *Equilibrio inicial* (punto 0): el país exporta X_0 e importa Y_0, y la CISC_0 es tangente a la relación real de intercambio internacional. El precio relativo nacional (es decir, la RRI nacional) coincide, por tanto, con la RRI internacional ($p_X^0/p_Y^0 = p_X^{*0}/p_Y^{*0} \equiv \text{RRI}^*$), de modo que se cumple la ley del precio único. Los volúmenes de comercio quedan reflejados en los dos triángulos de comercio simétricos, cuyos vértices son el origen, el punto 0 y la proyección de este sobre cada uno de los ejes (X_0 e Y_0).

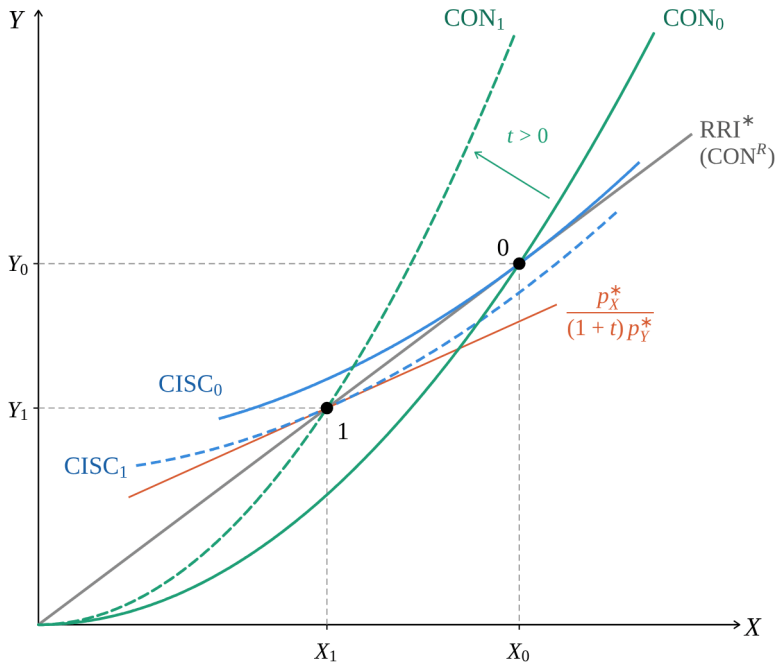

Figura B7.3: Arancel en un país pequeño (equilibrio general). El precio interior se aparta del mundial y la economía pasa a una curva de indiferencia social inferior.

[19] El análisis gráfico de Marshall-Edgeworth se realiza, por tanto, en el plano de los excesos de bienes. A diferencia de él, el de Haberler se realiza en el plano de los bienes importables y exportables, y los excesos (es decir, exportaciones e importaciones) se obtienen por diferencia.

Estática comparativa: un arancel *ad valorem* ($t > 0$) hace que el precio relativo nacional ya no coincida con la RRI internacional[20]. Concretamente, el precio relativo nacional será menor que el internacional ($p_X^{*0}/[(1+t)p_Y^{*0}] < p_X^{*0}/p_Y^{*0}$) en la cuantía del arancel.

El coste de oportunidad de comerciar con el exterior aumenta, porque a los productores les resulta más rentable producir el bien importable que seguir produciendo el exportable. Por ello, la curva de oferta neta del país se contrae, de CON_0 a CON_1, y el nuevo equilibrio se produce en el punto 1, con exportaciones X_1 e importaciones Y_1. La introducción del arancel tiene cinco efectos principales:

- *Producción*: aumenta la producción del bien importable Y (ahora protegido) y disminuye la del bien exportable X.

- *Consumo*: con bienes normales, disminuye el consumo del bien importable, ahora más caro, y el efecto sobre el del exportable es ambiguo (Anexo).

- *Recaudación*: el Gobierno recauda ingresos arancelarios equivalentes al arancel por unidad multiplicado por el volumen de importaciones.

- *Comercio internacional*: se producirá un descenso tanto de las exportaciones como de las importaciones. Por tanto, el triángulo de comercio será menor.

- *Efecto neto*: el país sufre una pérdida inequívoca de bienestar, pues la CISC alcanzada está más próxima al origen dada la nueva relación de precios nacionales. La pérdida de bienestar del mundo en su conjunto es exactamente la que sufre el propio país.

Bienestar: en un marco con dos factores de producción, el teorema de Stolper y Samuelson (1941) implica que la retribución real unitaria del factor usado intensivamente en la producción del bien importable aumentará.

2.3.3. Análisis de los efectos en un país grande y en competencia perfecta

Desarrollo gráfico: en el plano exportaciones-importaciones (figura B7.4), dos países grandes (A, B) comercian dos bienes (X, Y):

1. *Curva de oferta neta de A*, CON_0^A: recoge los intercambios que A está dispuesto a realizar a cada RRI y es normal.

2. *Curva de oferta neta de B*, CON^B: también normal. A diferencia del caso del país pequeño, restringe las posibilidades comerciales del país A, cuyas combinaciones factibles se encuentran sobre esa curva o por debajo de ella.

3. *Curvas de indiferencia social comerciales*, CISC_0^A y CISC_0^B: reúnen, para cada país, los volúmenes de comercio con un mismo bienestar social, y las dibujadas son las de libre comercio.

4. *Equilibrio inicial* (punto 0): del cruce de las curvas de oferta neta se determina la RRI internacional de equilibrio, RRI_0^*. En este punto, las CISC son tangentes entre sí y también tangentes a la RRI_0^*, y los triángulos de comercio son idénticos. Sin embargo, aunque la CISC_0^A es tangente a la CISC_0^B, no es tangente a la CON^B. Esa falta de tangencia refleja, como en el análisis de equilibrio parcial, que existen ganancias potenciales de bienestar para A realizables mediante aranceles.

[20] Se aprecia mejor en el plano de los bienes, con la FPP (véase el anexo).

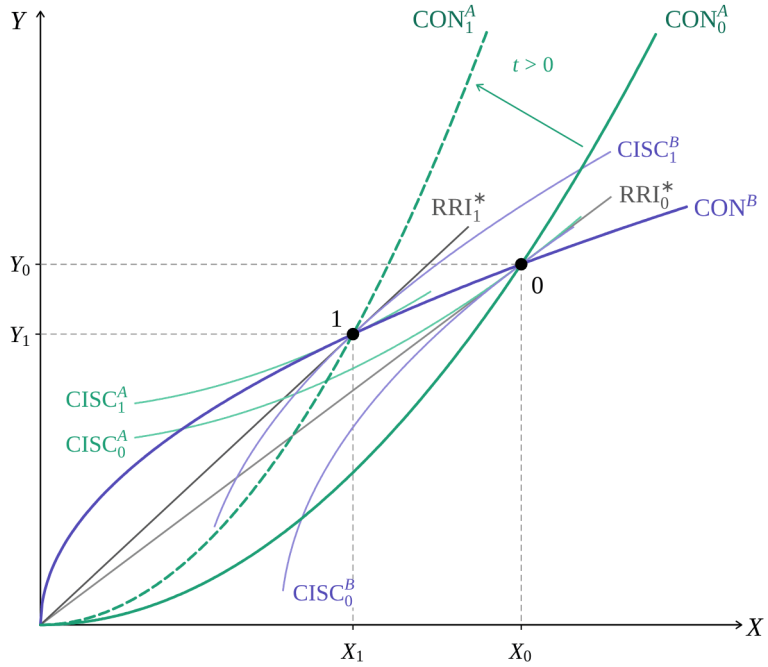

Figura B7.4: Arancel en un país grande (equilibrio general). La curva de oferta neta del país que impone el arancel gira y la relación real de intercambio mejora frente al resto del mundo. En libre comercio (punto 0) las CISC de ambos países son tangentes entre sí y a RRI_0^*, pero no a CON^B. En el punto 1 dejan de ser tangentes entre sí.

Estática comparativa: un arancel *ad valorem* ($t > 0$) sobre el bien importable Y hace que el país A ya no desee tantas importaciones del país B. Por lo tanto, la demanda recíproca o curva de oferta neta de A pivotará hacia la izquierda, de CON_0^A a CON_1^A.

Ya que el país es grande, esto tiene efectos sobre la RRI internacional y sobre la RRI nacional:

▪ *Mejora de la RRI internacional*: se debe a la conjunción de dos efectos de explotación ($\uparrow \text{RRI} = \uparrow p_X^* / \downarrow p_Y^*$):

- *Efecto monopsonio* ($\downarrow p_Y^*$): la reducción de la demanda del bien importable tendrá un efecto significativo sobre la demanda mundial de dicho bien debido al tamaño del país, lo que provoca una disminución de su precio mundial.

- *Efecto monopolio* ($\uparrow p_X^*$): al reducirse las importaciones se reducen también las exportaciones del país, pues ya no es necesario exportar tanto para mantener la balanza comercial equilibrada. Así, sube el precio mundial del bien exportable.

▪ *Aumento del precio relativo nacional del bien importable*: induce un aumento de la producción del bien importable y una disminución de la del bien exportable.

El equilibrio final se alcanzará en un nuevo punto (punto 1), en el que, de nuevo, la curva de oferta neta distorsionada del país A corta a la del país B, con una RRI internacional RRI_1^*. En este punto, las CISC de ambos países, CISC_1^A y CISC_1^B, ya no son tangentes entre sí, lo cual refleja la pérdida de bienestar mundial derivada de la imposición.

Bienestar:

- *Bienestar nacional*: el efecto sobre el bienestar del país grande es ambiguo y depende de qué efecto domine, si la mejora de la relación real de intercambio o la pérdida de eficiencia derivada de la contracción del comercio.

- *Pérdida de eficiencia mundial*: es inequívoca, pues la ganancia de bienestar del país importador que impone el arancel será siempre inferior a la pérdida que sufre el resto del mundo. Lo que uno gana en RRI es exactamente lo que el otro pierde, mientras que la pérdida de eficiencia la soportan ambos.

- *Efecto redistributivo*: además, dado que el precio relativo nacional del bien importable habrá aumentado, opera de nuevo el efecto redistributivo de Stolper-Samuelson en favor del factor utilizado intensivamente en el sector protegido.

2.3.4. Arancel óptimo y guerra arancelaria

Origen: la cuestión se remonta a Robert Torrens y John Stuart Mill, si bien el primer tratamiento riguroso se debe a Francis Edgeworth (1894).

Idea: por *arancel óptimo* se entiende aquel que maximiza el bienestar social de un país, dado su tamaño, es decir, el que iguala la ganancia marginal derivada de la mejora de la RRI con la pérdida marginal derivada del menor volumen importado.

Razonamiento: el arancel óptimo se puede obtener a través de un programa de maximización del bienestar social restringido por la curva de oferta neta del país B, que recoge el conjunto de intercambios que este está dispuesto a realizar. Concretamente, se puede demostrar[21] que este arancel es:

$$t^{A*} = \frac{1}{\varepsilon_{XY}^{B} - 1} \geq 0$$

donde t^{A*} es el arancel óptimo del país A y ε_{XY}^{B} la elasticidad de la curva de oferta neta del país B en el sentido de Marshall. Esta elasticidad mide, en valor absoluto, la respuesta de la demanda de importaciones de X de B al precio de X en unidades de Y, el bien que B exporta, es decir, a la RRI, a lo largo de su curva de oferta neta.

Existe, así, una relación inversa entre el arancel óptimo y la elasticidad de la curva de oferta neta del país extranjero, de la que se derivan cuatro casos[22]:

1. $\varepsilon_{XY}^{B} = \infty$: la CON^{B} es perfectamente elástica y el arancel óptimo es nulo ($\lim_{\varepsilon \to \infty} 1/(\varepsilon - 1) = 0$) y corresponde al caso del país pequeño.

2. $\varepsilon_{XY}^{B} \in (1, \infty)$: la CON^{B} es elástica y el arancel óptimo es positivo y moderado, tanto menor cuanto mayor sea la elasticidad[23], lo que corresponde al caso típico del país grande.

[21] La demostración puede consultarse en Gandolfo (2014). El resultado es análogo a la regla del margen del monopolista, pues el país actúa como monopolista de sus exportaciones y monopsonista de sus importaciones sobre la curva de oferta neta del rival. Fija así un "margen" (el arancel) inversamente proporcional a la elasticidad de la oferta de exportaciones del rival, $\varepsilon_{XY}^{B} - 1$.

[22] Si se representa el bienestar en función del arancel, se obtiene una curva en forma de U invertida, con el máximo en el arancel óptimo.

[23] La semejanza de esta relación inversa con la regla de Ramsey (1927) es solo formal, pues el arancel óptimo explota poder de mercado frente al exterior y no minimiza el exceso de gravamen de una recaudación dada.

3. $\varepsilon_{XY}^{B} \to 1^{+}$: el arancel óptimo tiende a infinito, aunque no llega a ser prohibitivo, pues en el óptimo sigue habiendo comercio.

4. $\varepsilon_{XY}^{B} \leq 1$: si el equilibrio se sitúa en el tramo rígido o "anómalo" de la CON^{B}, la fórmula anterior deja de ser directamente aplicable. En ese rango, al país A le conviene seguir elevando el arancel, contrayendo el comercio, hasta que el equilibrio se desplace al tramo elástico de la curva del rival, donde el óptimo interior vuelve a estar definido.

Desarrollo gráfico: sin reacción exterior, el arancel óptimo se representa en el plano exportaciones-importaciones (figura B7.5):

1. *Curva de oferta neta de B*, CON^{B}: recoge los intercambios que B está dispuesto a realizar y restringe los de A.

2. *Curvas de oferta neta de A*: la de libre comercio, CON_{0}^{A}, y la distorsionada por el arancel óptimo, CON_{m}^{A}.

3. *Curvas de indiferencia social comerciales de A*: la de libre comercio, CISC_{0}^{A}, que pasa por el punto 0, y la más alejada del origen que A puede alcanzar sobre la CON^{B}, CISC_{m}^{A}.

4. *Óptimo*: el arancel óptimo será aquel con el que la CON_{m}^{A} interseque a la CON^{B} en el punto en que la CISC_{m}^{A} es tangente a la CON^{B} (punto m).

El punto m es óptimo para A, pero no es un óptimo de Pareto mundial. A diferencia del equilibrio de libre comercio, las curvas de indiferencia comerciales de ambos países ya no son tangentes entre sí, por lo que existirían reasignaciones que mejorarían a ambos.

Si A continúa elevando el arancel, su bienestar disminuirá y, más allá del punto e', situado sobre la CISC_{0}^{A}, quedará por debajo del alcanzado en libre comercio, pese a la mejora de la RRI.

Hasta ahora se ha supuesto que el resto del mundo no responde ni adopta represalias frente al país A. Sin embargo, el punto m no constituye un equilibrio de Nash, ya que el país B tiene incentivos para tomar represalias contra A, sobre todo por ser también un país grande.

Extensiones: con reacción exterior, el arancel óptimo desemboca en una guerra arancelaria (Johnson, 1953), que la figura B7.5 recoge en tres pasos:

1. *Represalia*: si el arancel del país A desplaza el equilibrio a E' (el punto m), el país grande B podría responder imponiendo a su vez un arancel que desplace su curva de oferta neta. El equilibrio pasa entonces a E'', donde B obtiene más bienestar que en E'.

2. *Equilibrio de Nash*: la escalada podría prolongarse hasta un punto N en el que a ningún país le conviene subir su arancel. La curva de indiferencia de cada país es allí tangente a la curva de oferta neta del otro (la CISC_{N}^{A} a la CON_{1}^{B} y la CISC_{N}^{B} a la CON_{N}^{A}). En ese equilibrio, por lo general, el bienestar de los dos países resulta inferior al de partida. No obstante, Johnson (1953) mostró que uno de ellos puede salir ganando frente al libre comercio si su curva de oferta neta es suficientemente más elástica que la de su rival. El mundo en su conjunto, en cambio, pierde siempre.

3. *Subóptimo de Pareto*: el equilibrio de Nash responde a la lógica del dilema del prisionero. Con todo, según advirtió Johnson (y como recoge la figura), la guerra arancelaria no llega a suprimir por completo el comercio.

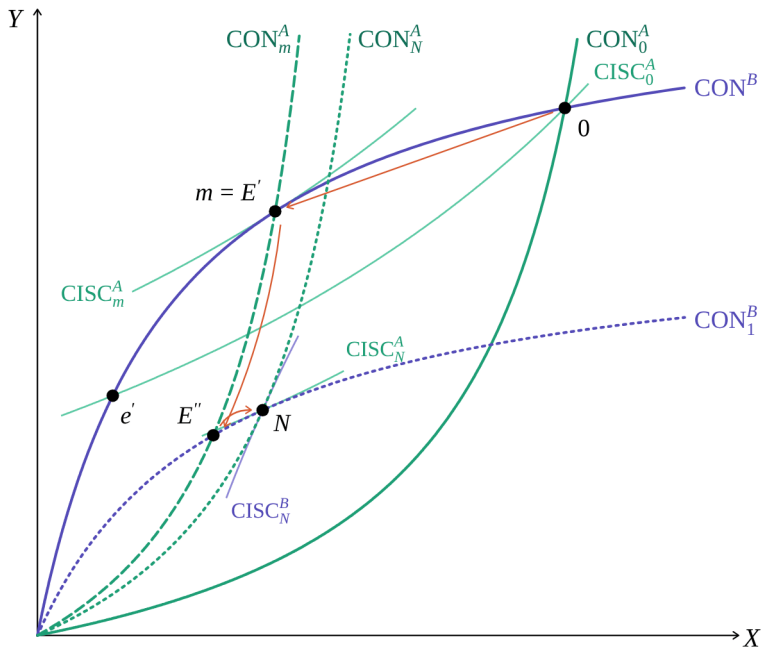

Figura B7.5: Arancel óptimo y guerra arancelaria (Johnson, 1953). El arancel óptimo lleva a A al punto m, tangencia de su CISC con la CON de B. La represalia de B (E″) y las sucesivas conducen a un equilibrio de Nash (N) subóptimo en el sentido de Pareto, sin suprimir el comercio.

Esta estructura de dilema del prisionero constituye, precisamente, la justificación teórica de la cooperación comercial multilateral. La razón es que los acuerdos negociados y vinculantes del tipo GATT-OMC permiten a los países comprometerse de forma creíble con el resultado cooperativo (aranceles bajos recíprocos), que ninguno sostendría unilateralmente.

Ejemplo (aranceles de Trump, desde 2018): la guerra comercial entre EE. UU. y China refleja esta lógica. Durante su primer mandato (2017-2021), el presidente estadounidense Donald Trump impuso unilateralmente una serie de aranceles que abarcaban numerosas industrias (p. ej. acero, aluminio, paneles solares y un amplio abanico de bienes chinos). Ello dio lugar a una guerra arancelaria con represalias por parte de China y de otros socios.

Según una estimación para 2018, los aranceles y las represalias afectaban a más de 400.000 M$ de importaciones y exportaciones, y los compradores de importaciones perdían unos 51.000 M$ anuales[24].

Los estudios sobre la guerra arancelaria iniciada en 2018 (Amiti, Redding y Weinstein, 2019) documentan que los aranceles estadounidenses de 2018 se trasladaron de forma prácticamente completa a los precios interiores de EE. UU. La carga recayó, así, sobre consumidores y empresas importadoras estadounidenses, sin la mejora de relación real de intercambio que predice la lógica del arancel óptimo, al menos a corto plazo.

En la misma línea, Fajgelbaum, Goldberg, Kennedy y Khandelwal (2020) estiman una pérdida agregada neta para la economía estadounidense una vez computadas la recaudación arancelaria

[24] Según Fajgelbaum et al. (2020), los aranceles cubrían 303.000 M$ de importaciones (el 12,7 % del total) y las represalias, 127.000 M$ de exportaciones (el 8,2 %). La pérdida de los compradores equivale al 0,27 % del PIB.

y las ganancias de los productores protegidos, junto con importantes efectos redistributivos regionales. La evidencia también confirma la lógica estratégica de las represalias, pues los aranceles de respuesta de los socios comerciales se concentraron en productos con relevancia política para el país iniciador[25].

Tras su regreso a la Casa Blanca en enero de 2025, la segunda Administración Trump amplió esta política, con aranceles generalizados a la práctica totalidad de sus socios comerciales y una fuerte escalada de represalias cruzadas con China. En febrero de 2026, el Tribunal Supremo declaró que la Ley de Poderes Económicos de Emergencia Internacional (IEEPA) no autoriza a imponer aranceles, y la Administración recurrió a otras bases legales para mantener parte de ellos.

Conclusiones:

▪ *Aspectos positivos*: el modelo formaliza el argumento de la relación real de intercambio, el único por el que un arancel puede elevar el bienestar de un país sin distorsiones internas. Precisa, además, que la cuantía del arancel óptimo depende de la elasticidad de la curva de oferta neta del resto del mundo.

▪ *Limitaciones*: la ganancia se obtiene a costa del resto del mundo, solo está al alcance de países con poder de mercado y exige conocer una elasticidad difícil de estimar. Asimismo, la represalia puede anularla, y la evidencia sobre los aranceles estadounidenses de 2018 no muestra la mejora de la relación real de intercambio que el modelo predice.

2.3.5. Corolarios: las paradojas de Lerner (1936) y Metzler (1949)

Los resultados anteriores para un país grande suponen que ambas CON son normales y que la recaudación arancelaria no se gasta de forma preferente en el bien importado. Si no se cumplen estas condiciones, se puede llegar a dos predicciones paradójicas.

Paradoja de Lerner (1936):

▪ *Origen*: Abba Lerner la formula en *The Symmetry between Import and Export Taxes* (1936), al examinar cómo influye el destino de la recaudación en los efectos del arancel.

▪ *Idea*: bajo ciertas condiciones, el arancel de un país grande no mejora su RRI internacional, sino que la empeora.

▪ *Supuestos*: el sector público gasta en el bien importado una parte muy elevada de lo que recauda con el arancel (en el límite, todo).

▪ *Razonamiento*: el gasto de la recaudación en el propio bien importado sostiene su demanda mundial y eleva su precio ($p_X^{*0}/ \uparrow p_Y^{*0} \rightarrow \downarrow$ RRI), lo que anula la lógica del arancel óptimo.

▪ *Conclusiones*:

- *Aspectos positivos*: pone de relieve que el efecto de un arancel sobre la RRI depende de cómo se gasta la recaudación y de la forma de las curvas de oferta neta, de modo que su mejora en el país grande no es un resultado general.

- *Limitaciones*: la condición que exige, un gasto de la recaudación concentrado en el bien importado, es poco plausible, por lo que su interés es sobre todo teórico.

[25] Véase Fajgelbaum et al. (2020), cit., según los cuales los trabajadores de los sectores comerciables de los condados de mayoría republicana fueron los más perjudicados por las represalias.

Paradoja de Metzler (1949):

▪ *Origen*: Lloyd Metzler la formula en 1949, al estudiar los efectos del arancel sobre la distribución de la renta.

▪ *Idea*: el arancel puede hacer que el precio relativo nacional del bien importable disminuya, de modo que, lejos de proteger a la industria nacional del bien importable, la desprotege.

▪ *Supuestos*:

- *CON*: la CON del país grande debe ser normal, pero la del resto del mundo debe ser anómala, y el equilibrio inicial debe darse en su tramo inelástico.
- *Distribución nacional*: los ingresos arancelarios, los gaste el sector público o se transfieran a los consumidores, deben destinarse en proporción suficiente al consumo del bien exportable.
- *Condición analítica*: el país que no aplica el arancel debe tener una elasticidad de su demanda de importaciones inferior a la propensión marginal del país que lo aplica a consumir el bien exportable. En el modelo, $\varepsilon_{XY}^{B} < c_X^A$, donde c_X^A es la propensión marginal del país A a gastar en X, es decir, la fracción de cada unidad adicional de renta que A destina al bien X[26].

▪ *Desarrollo gráfico*: la paradoja se representa en el plano exportaciones-importaciones (figura B7.6):

1. *Curva de oferta neta de A*, CON_0^A: recoge los intercambios que A está dispuesto a realizar a cada RRI y es normal.

2. *Curva de oferta neta de B*, CON^B: anómala, con un tramo elástico, creciente, y un tramo inelástico, decreciente, en el que B importa más aunque exporte menos.

3. *Equilibrio inicial* (punto 0): el de libre comercio, en el tramo inelástico de la CON^B, con una RRI internacional RRI_0^*.

[26] Si A gasta gran parte de la recaudación en X y la demanda de X por parte de B es muy inelástica, surge un exceso de demanda de X que eleva su precio relativo en A y abarata, en términos relativos, el bien importado Y.

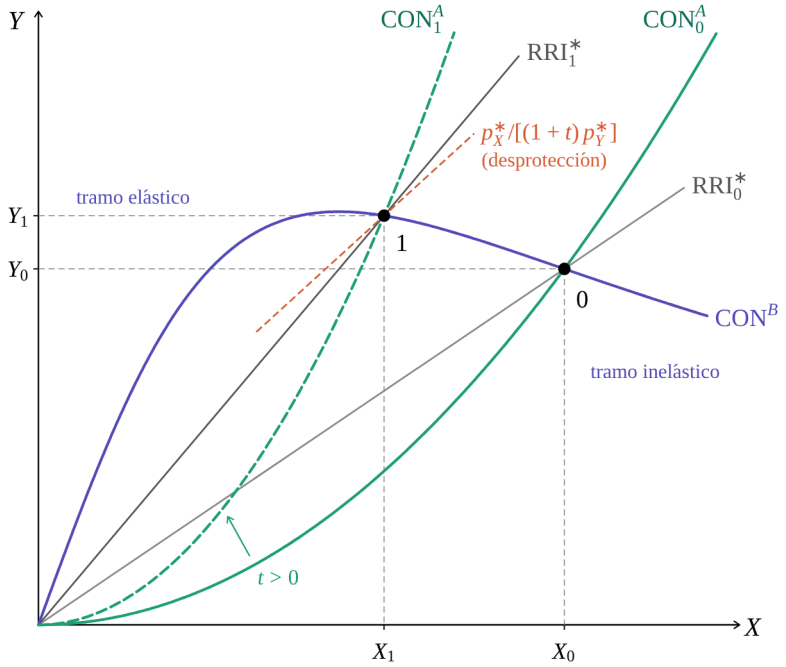

Figura B7.6: Paradoja de Metzler. Con el equilibrio inicial en el tramo inelástico de la CON de B, el arancel mejora tanto la RRI de A que el precio relativo nacional del bien importable disminuye.

▪ *Estática comparativa*: con el arancel ($t > 0$), A contrae su curva de oferta neta hacia el origen, de CON_0^A a CON_1^A, y el equilibrio pasa del punto 0 al 1, con dos efectos:

1. *Precio mundial*: el precio relativo mundial de Y (el bien que importa A) disminuye intensamente, debido al exceso de oferta en el mercado internacional tras el arancel, por lo que la RRI de A mejora, de RRI_0^* a RRI_1^*.

2. *Precio nacional*: la caída del precio mundial es tan acusada que, incluso sumado el arancel, el precio relativo nacional de Y disminuye, pues $p_X^*/[(1+t)\,p_Y^*]$ supera a RRI_0^*, con lo que, a diferencia del caso ordinario, la industria nacional queda desprotegida.

▪ *Conclusiones*:

- *Aspectos positivos*: advierte de que el arancel no protege necesariamente al sector importador y, a través del teorema de Stolper-Samuelson, de que puede perjudicar incluso al factor empleado intensivamente en ese sector, cuya renta real pretendía elevar.

- *Limitaciones*: como la paradoja de Lerner, exige condiciones poco plausibles, en particular la condición analítica anterior, por lo que se considera un caso improbable en la práctica.

2.4. Conclusiones

Del análisis del arancel se extraen cuatro implicaciones:

1. *País pequeño*: sin distorsiones internas, el arancel genera siempre una pérdida inequívoca de bienestar, tanto en equilibrio parcial como en equilibrio general.

2. *País grande*: existen ganancias potenciales vía relación real de intercambio, que fundamentan la existencia teórica de un arancel óptimo.

3. *Distorsión interna*: frente a ella, el arancel es, en el mejor de los casos, un instrumento de *second best*, pues la política óptima la ataca en su origen (apartado 1.2.1).

4. *Cooperación multilateral*: el dilema del prisionero de la guerra arancelaria justifica reglas comerciales multilaterales vinculantes, que se retoman con los acuerdos de la OMC (apartado 3.3.1).

3. Barreras no arancelarias y otros instrumentos tradicionales

Desde los años sesenta, la Ronda Kennedy[27] y la reforma del GATT que trajo consigo impulsaron la reducción de aranceles, mientras surgían instrumentos proteccionistas nuevos, distintos del arancel. En los años setenta, con la crisis del petróleo, el proteccionismo regresó bajo formas no arancelarias, fenómeno que recibe el nombre de **neoproteccionismo**. Más tarde, la Gran Recesión dio lugar a un nuevo repunte de medidas de todo tipo, también de barreras no arancelarias (BNA).

3.1. Instrumentos tradicionales distintos del arancel

Los instrumentos tradicionales distintos del arancel son las cuotas, las subvenciones y los impuestos a la exportación, y las restricciones voluntarias de exportaciones y las expansiones voluntarias de las importaciones. Las cuotas y las medidas voluntarias (RVE y EVI) actúan como barreras no arancelarias, y las subvenciones y los impuestos a la exportación son otros instrumentos tradicionales de la política comercial.

3.1.1. Cuotas

Las **cuotas** limitan cuántas unidades de un bien puede importar un país. El art. XI del GATT las prohíbe con carácter general (tema B33, apartado 1.2.1).

Ejemplo: el Acuerdo Multifibras (1974) permitió a los países desarrollados fijar cuotas bilaterales a sus importaciones de textiles y prendas de vestir de los países en desarrollo. Las cuotas rigieron hasta el 1 de enero de 2005, cuando el Acuerdo sobre los Textiles y el Vestido de la OMC las eliminó.

Desarrollo gráfico: los efectos de una cuota se representan en el plano cantidad-precio (figura B7.7), como en el caso del arancel y para un país pequeño:

1. *Curvas de oferta y de demanda nacionales*, $S(y)$ y $D(y)$: recogen la producción y el consumo nacionales a cada precio, y son, respectivamente, creciente y decreciente, como en el apartado 2.2.1.

2. *Oferta con la cuota*, $S(y) + \bar{M}$: suma a la oferta nacional, por encima del precio mundial, la cantidad que la cuota permite importar.

3. *Equilibrio inicial*: el precio del bien importado viene dado exógenamente por el precio mundial, p^*, al que se consume y_0^d (punto 0) y se importa M_0.

[27] Sexta ronda del GATT (Ginebra, 1964-1967).

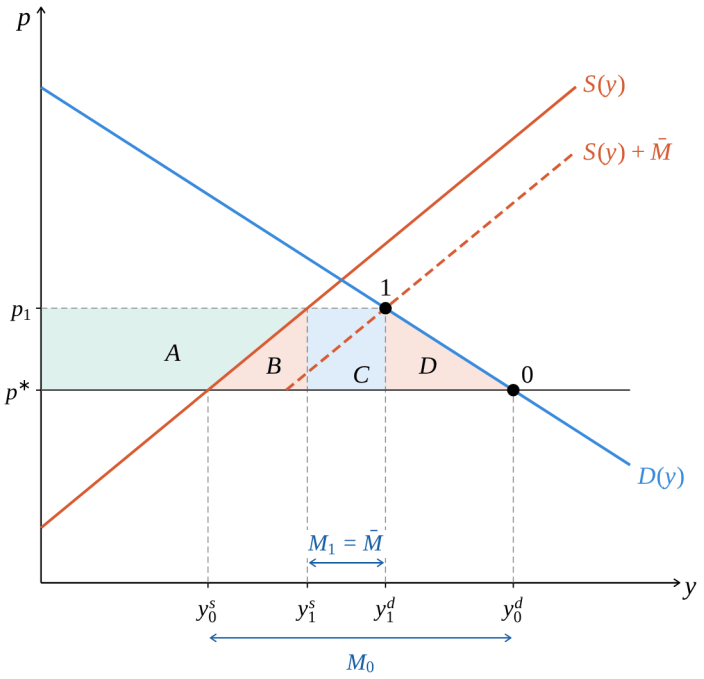

Figura B7.7: Cuota a la importación en un país pequeño. El precio interior sube hasta p_1. El productor gana A, el consumidor pierde A, B, C y D, la renta de la cuota es C y la pérdida neta, B y D.

Estática comparativa: se introduce una cuota que limita la cantidad importada al tramo $y_1^d - y_1^s$, $M_1 = \bar{M}$, y el precio interior sube hasta p_1, donde la oferta con la cuota corta a la demanda (punto 1). La cuota afecta en tres dimensiones de forma similar al caso de los aranceles:

▪ *Consumo*: se reduce de y_0^d a y_1^d. El excedente del consumidor cae en las áreas A, B, C y D, debido a la caída del consumo y al encarecimiento del bien.

▪ *Producción*: aumenta de y_0^s a y_1^s. El excedente del productor aumenta en el área A, debido al aumento de la producción protegida.

▪ *Rentas de la cuota*: surge una renta equivalente al área C que no tiene por qué ir destinada al Gobierno, sino que su destino depende de cómo se administre la cuota, lo que constituye la principal diferencia con respecto al arancel. El Gobierno del país importador puede:

- *Subastar las licencias*: con una subasta perfecta, se apropia de la totalidad de las rentas de la cuota y la equivalencia con el arancel es completa.
- *Conceder directamente los derechos de importación a los importadores nacionales*[28]: estos capturan la renta y aumentan su bienestar.
- *Conceder directamente los derechos de importación a los exportadores extranjeros*: la renta sale al exterior y el bienestar del país disminuye.

[28] La concesión directa no tiene por qué ser discrecional, pues existen varios criterios de asignación, como el orden de llegada.

- *Subdividir las licencias*: las reparte entre los importadores en proporción a las cantidades importadas antes de la cuota.

Una variante híbrida son los *contingentes arancelarios* (*tariff-rate quotas*), en los que las importaciones realizadas dentro del contingente pagan un arancel reducido y las que lo exceden soportan un arancel muy superior. Son frecuentes en el comercio agrícola desde la "arancelización" de las restricciones cuantitativas acordada en la Ronda Uruguay.

Conclusiones:

▪ *Aspectos positivos*: las cuotas representan una medida más certera que los aranceles para reducir las importaciones, pues el efecto del arancel depende de las elasticidades de la oferta y de la demanda del bien, mientras que la cuota limita directamente la cantidad. En condiciones de competencia perfecta y con oferta y demanda dadas, además, Bhagwati (1965) demostró la *equivalencia entre aranceles y cuotas*, según la cual para toda cuota existe un arancel que replica exactamente sus efectos sobre precios y cantidades.

▪ *Limitaciones*: dicha equivalencia se rompe fuera de la competencia perfecta, y con ello aparecen los costes propios de la cuota:

- *Extranjero*: si los exportadores extranjeros no operan bajo competencia perfecta y tienen capacidad para alterar los precios internacionales, frente a un arancel podrían absorber parte de él reduciendo su precio para no perder cuota de mercado. Frente a una cuota, que fija la cantidad, bajar el precio no les permite vender más, pero sí pueden elevarlo y apropiarse de parte de la renta de la cuota.

- *Nación*: a diferencia de los aranceles, la cuota puede alterar profundamente la estructura de mercado. Al fijar la cantidad importada, suprime la disciplina que impone el precio internacional, y quien solo era un monopolista potencial puede convertirse en uno efectivo. Además, si ya existe competencia imperfecta, una cuota produce una mayor pérdida de bienestar que el arancel equivalente, puesto que amplifica el poder de mercado.

- *Redistribución*: se pueden producir diferencias redistributivas relevantes en función de a quién se le asignen las licencias, incluidas conductas de captura de rentas (*rent seeking*).

- *Transparencia*: la cuota resulta más difícil de identificar para los consumidores que el arancel, puesto que no figura explícitamente en el precio, lo que la hace políticamente más atractiva y económicamente más opaca.

3.1.2. Subvenciones a la exportación

Las **subvenciones a la exportación** son pagos públicos que reciben los productores nacionales de un bien por cada unidad que exportan. La OMC las prohíbe[29], también en la agricultura desde que la Conferencia Ministerial de Nairobi (2015) acordó eliminarlas, y su aplicación puede dar lugar a derechos compensatorios (tema B33, apartados 2.3.3 y 2.3.4).

Ejemplo: la política agrícola común (PAC) de los años ochenta, cuya intervención productivista en el mercado agrícola generó excedentes que tuvieron que ser absorbidos por programas de compras públicas y subvenciones a la exportación.

[29] En este contexto, los países utilizan diferentes formas opacas de conceder subvenciones a la exportación, como condiciones de crédito favorables, seguros de riesgo (p. ej. contra posibles impagos de importadores extranjeros) o actividades promocionales (ferias, publicidad).

Desarrollo gráfico: los efectos de una subvención a la exportación se representan en el plano cantidad-precio del bien exportable (figura B7.8), para un país pequeño que subvenciona sus exportaciones *ad valorem*:

1. *Curvas de oferta y de demanda nacionales*, $S(x)$ y $D(x)$: recogen la producción y el consumo nacionales del bien exportable a cada precio, y son, respectivamente, creciente y decreciente.

2. *Equilibrio inicial* (punto 0): en libre comercio, el precio nacional es igual al precio internacional ($p_0 = p^*$) y se exporta $x_0^s - x_0^d$.

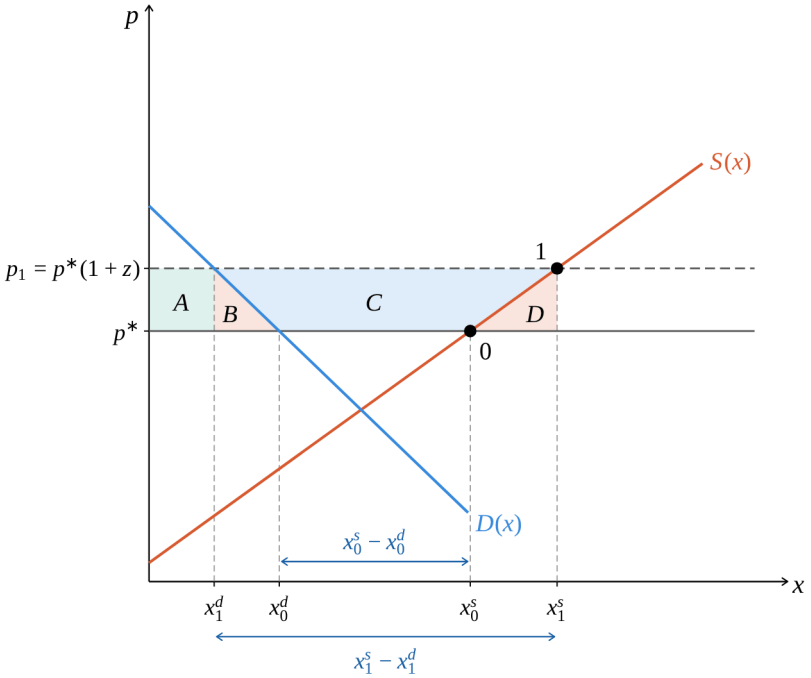

Figura B7.8: Subvención a la exportación en un país pequeño. El precio interior sube hasta $p^*(1+z)$. El consumidor pierde A y B, el productor gana A, B y C, el coste fiscal es B, C y D y la pérdida neta, B y D.

Estática comparativa: si se introduce una subvención *ad valorem* a la tasa $z > 0$, el precio percibido por los productores pasa a ser $p_1 = p^*(1+z)$, por lo que el precio interior y la cantidad ofrecida para exportar, $x_1^s - x_1^d$, aumentan (punto 1). De nuevo, se producirán efectos en varios sentidos:

- *Consumo*: disminuye de x_0^d a x_1^d, pues el precio interior sube. El excedente del consumidor se reduce en las áreas A y B.

- *Producción*: aumenta de x_0^s a x_1^s. El excedente del productor aumenta en las áreas A, B y C.

- *Coste fiscal*: a diferencia del arancel, la subvención no genera recaudación, sino un gasto público equivalente a las áreas B, C y D, que financia la brecha entre el precio percibido por los productores y el precio mundial.

▪ *Efecto neto*: la pérdida inequívoca de bienestar nacional asciende a las áreas B y D (distorsiones de consumo y de producción). En el caso de un país grande, a ello se añade el deterioro de la relación real de intercambio, pues el aumento de la oferta exportada deprime el precio mundial del bien. La subvención es, así, la imagen especular del arancel.

Conclusiones:

▪ *Aspectos positivos*: con competencia imperfecta, puede trasladar beneficios hacia la empresa nacional (Brander y Spencer, 1985), como desarrolla el tema B08, apartado 2.2.1.

▪ *Limitaciones*: en un país grande, parte de su coste presupuestario beneficia a los consumidores extranjeros, y la subvención invita a la represalia, con derechos compensatorios o con subvenciones rivales.

3.1.3. Impuestos a la exportación

Además de las importaciones, es posible gravar las exportaciones. El **impuesto a la exportación** (derecho aduanero a la exportación, apartado 2.1.1) equivale al arancel, pero aplicado a los bienes exportados. La OMC no lo prohíbe con carácter general, pues el art. XI del GATT excluye de la prohibición de restricciones los derechos, impuestos y otras cargas.

Ejemplo: en noviembre de 1995, con precios mundiales del trigo superiores a los de apoyo comunitarios, la UE gravó por primera vez en más de una década sus exportaciones de trigo para contener los precios interiores.

Desarrollo gráfico: los efectos de un impuesto a la exportación se representan en el plano cantidad-precio del bien exportable (figura B7.9), para un país pequeño:

1. *Curvas de oferta y de demanda nacionales*, $S(x)$ y $D(x)$: recogen la producción y el consumo nacionales del bien exportable a cada precio, y son, respectivamente, creciente y decreciente.

2. *Equilibrio inicial* (punto 0): al precio mundial, p^*, el país exporta $x_0^s - x_0^d$.

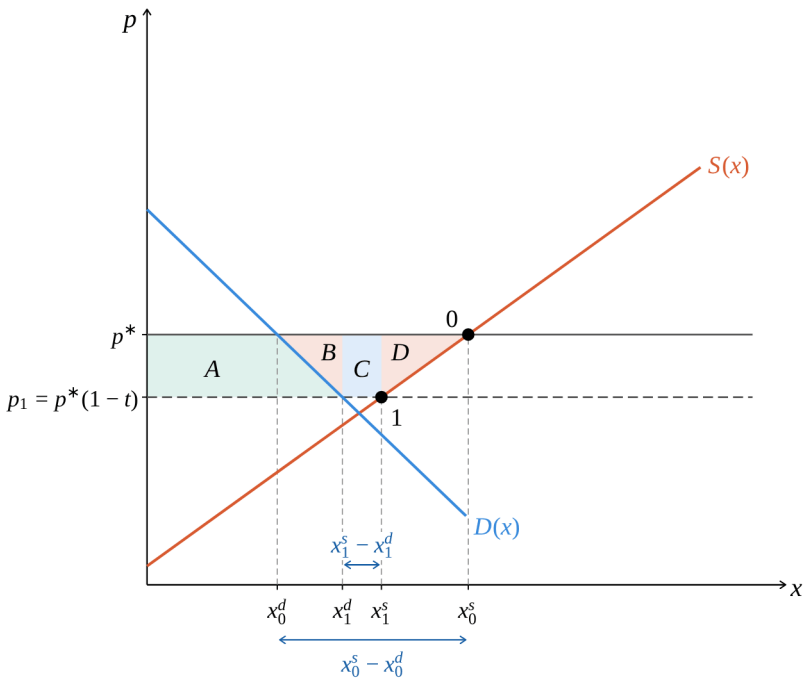

Figura B7.9: Impuesto a la exportación en un país pequeño. El precio interior cae hasta $p^*(1-t)$. El consumidor gana A, el productor pierde A, B, C y D, la recaudación es C y el coste social, B y D.

Estática comparativa: el impuesto reduce el precio que perciben los productores a $p_1 = p^*(1-t)$ (punto 1), y la exportación, a $x_1^s - x_1^d$. Sus efectos son los siguientes:

- *Consumo*: aumenta de x_0^d a x_1^d y el excedente del consumidor aumenta en el área A.

- *Producción*: se reduce de x_0^s a x_1^s y el excedente del productor se reduce en las áreas A, B, C y D.

- *Recaudación*: el Estado recauda el área C.

- *Efecto neto*: las áreas B y D representan el coste social del impuesto, una pérdida inequívoca de bienestar nacional.

Extensiones: según el *teorema de la simetría de Lerner* (1936), en equilibrio general y con balanza comercial equilibrada, gravar las exportaciones con un impuesto *ad valorem* de tipo t produce los mismos efectos reales que gravar las importaciones con un arancel *ad valorem* de tipo $t/(1-t)$. Ese arancel altera el precio relativo interior en la misma proporción que el impuesto.

Conclusiones:

- *Aspectos positivos*: por esa simetría, un país grande puede obtener con el impuesto a la exportación la mejora de la relación real de intercambio del arancel óptimo.

- *Limitaciones*: el impuesto grava el sector de ventaja comparativa y desincentiva su inversión, y, si varios exportadores lo aplican a la vez para contener sus precios interiores, el precio mundial sube y se vuelve más inestable, un problema de acción colectiva (Martin y Anderson, 2012).

3.1.4. Restricciones voluntarias de exportaciones (RVE) y expansiones voluntarias de las importaciones (EVI)

Restricciones voluntarias de exportaciones: el país exportador fija "voluntariamente" una cuota a sus exportaciones hacia un país importador.

Este tipo de medida se hizo popular en los años ochenta dado que eludía las disciplinas del GATT (medidas de "zona gris"). Sin embargo, en la Ronda Uruguay (1986-1994) se acordó ir desmantelándolas. Conviene distinguir las RVE negociadas de las restricciones unilaterales:

- *Restricciones negociadas*: el país exportador y el país importador negocian bilateralmente, con la amenaza del importador de imponer cuotas o aranceles. Esta forma es la más común (p. ej. la restricción de Japón a la exportación de automóviles hacia EE. UU. en los años ochenta).

- *Restricciones unilaterales a la exportación*: se deben a motivos internos del país exportador, sin presión del importador (p. ej. la prohibición de exportar grano que Rusia impuso en agosto de 2010, tras una sequía, para aumentar su oferta nacional y contener el precio interior). En rigor no son RVE, sino restricciones a la exportación[30].

Este tipo de medidas tiene efectos sobre los dos países:

1. *Sobre el país exportador*: la RVE reduce sus exportaciones como lo haría un impuesto a la exportación. Sin embargo, la renta que genera la restricción, que es la diferencia entre el precio, más alto, del mercado importador y el precio interno del exportador, no va al Estado, sino a las propias empresas exportadoras. Así ocurrió con los automóviles japoneses, vendidos en EE. UU. en menor número pero a precios más altos, de modo que el país exportador puede incluso salir ganando.

2. *Sobre el país importador*: los efectos son los de una cuota a la importación cuyos derechos se conceden a los exportadores extranjeros, el caso más costoso de los analizados, pues la renta de la restricción sale del país. En la figura B7.7, la renta de la cuota (área C) va a los exportadores extranjeros, de modo que el país importador pierde las áreas B, C y D.

Expansiones voluntarias de las importaciones: el país importador aumenta "voluntariamente" sus compras al exportador. Las EVI siguen los mismos criterios que las RVE y tendrán efectos análogos, pero para el caso de importaciones.

3.2. Otras barreras no arancelarias

Además de las barreras tradicionales, el comercio afronta otras de naturaleza muy diversa, como las físicas o administrativas, las técnicas o regulatorias, las restricciones a la contratación pública, las barreras fiscales y otras medidas como el *dumping* o las sanciones.

3.2.1. Barreras físicas o administrativas: *red tape*

Barreras administrativas: consisten en trámites burocráticos (*red tape*)[31] a los que se enfrentan las empresas en su actividad comercial internacional. En su versión explícita, son controles de aduana.

[30] El GATT las admite con carácter temporal para prevenir o remediar una escasez aguda de alimentos u otros productos esenciales (art. XI:2(a), tema B33, apartado 1.2.2).

[31] Término peyorativo para la burocracia excesiva o redundante que obstaculiza las decisiones de los agentes.

Controles: el principal es la inspección previa a la expedición, por la que empresas especializadas comprueban, antes del envío, aspectos de las mercancías destinadas a la exportación (p. ej. su cantidad o su clasificación aduanera).

Normas de origen: criterios técnicos que establecen de qué país procede un producto importado, dato del que dependen el régimen comercial aplicable o las medidas antidumping[32].

3.2.2. Barreras técnicas o regulatorias: SPS y TBT

Por **barreras técnicas** se entiende el conjunto de normas y regulaciones sobre aspectos concretos que, al diferir de un país a otro, obstaculizan la circulación de los bienes. Han sido especialmente relevantes históricamente en sectores punta[33]. Incluyen las medidas sanitarias y fitosanitarias (SPS) y los obstáculos técnicos al comercio (TBT), como las normas de embalaje o etiquetado.

Las barreras técnicas presentan un triple problema:

- *Empresas fabricantes*: soportan sobrecostes, porque deben producir series más cortas adaptadas a las exigencias técnicas de cada mercado de destino, con la consiguiente pérdida de eficiencia productiva.
- *Exportadores*: afrontan obstáculos a la exportación.
- *Sector público*: soporta costes de información.

3.2.3. Restricciones a la contratación pública

Las **restricciones a la contratación pública** favorecen a los proveedores nacionales frente a los extranjeros cuando se adjudican contratos públicos, un mercado que supone una fracción muy sustancial del PIB de las economías avanzadas. En el marco de la OMC, el Acuerdo sobre Contratación Pública (ACP), de carácter plurilateral, extiende los principios de no discriminación y transparencia a los contratos públicos cubiertos por los compromisos de sus partes.

3.2.4. Barreras fiscales

Las **barreras fiscales** comprenden el uso discriminatorio de la imposición interior contra los productos importados. Los ajustes fiscales en frontera, es decir, la aplicación del IVA en destino y la devolución de la imposición indirecta soportada por las exportaciones, son, en principio, neutrales. Sin embargo, la fiscalidad puede convertirse en instrumento proteccionista en dos casos:

- *Impuestos especiales discriminatorios*: cuando los impuestos especiales o las accisas se diseñan de forma que gravan, *de facto*, características propias de los productos extranjeros (p. ej. tramos impositivos definidos sobre cilindradas, graduaciones alcohólicas o categorías que la producción nacional apenas ocupa).
- *Devoluciones excesivas*: cuando las devoluciones a la exportación exceden la carga efectivamente soportada, operando como una subvención encubierta.

[32] Por ejemplo, en productos textiles, EE. UU. exige en sus acuerdos de libre comercio tres transformaciones (hilo, tejido y confección) dentro de la zona para que una prenda sea considerada originaria (regla *yarn forward*).

[33] En 1985, las barreras técnicas eran cuantitativamente más importantes que los controles aduaneros, ya que afectaban a buena parte del comercio intracomunitario. Los más perjudicados eran los sectores punteros, porque en ellos la fragmentación de los mercados suponía una desventaja considerable frente a competidores como EE. UU. o Japón.

3.2.5. Otras medidas: *dumping*, salvaguardias, medidas compensatorias, tipos de cambio múltiples, ayudas condicionadas, sanciones y embargos

Dumping: tiene lugar cuando una mercancía se exporta a otro país a un precio de exportación inferior a su valor normal en el país de origen, y el art. VI del GATT lo condena si causa o amenaza causar un daño importante a la industria nacional del país importador. Se distinguen tres tipos:

- *Dumping esporádico*: liquidación puntual de excedentes.

- *Dumping persistente*: discriminación internacional de precios sostenida, posible cuando los mercados están segmentados y la demanda exterior es más elástica.

- *Dumping predatorio*: precios bajo coste para expulsar a los rivales y explotar después el poder de mercado adquirido. Este tipo es el único inequívocamente dañino desde el punto de vista teórico.

La respuesta prevista son los *derechos antidumping*, cuya imposición exige, conforme al Acuerdo Antidumping de la OMC, probar la existencia de *dumping*, de daño a la industria nacional y de relación causal entre ambos.

Medidas de salvaguardia: se aplican como medidas de urgencia (art. XIX del GATT) cuando un aumento de las importaciones causa daño grave, o amenaza con causarlo, a la rama de producción nacional, aunque no exista práctica desleal alguna.

Su objetivo es dar a la industria un respiro para reajustarse, por lo que el Acuerdo sobre Salvaguardias de la OMC exige que sean temporales[34].

Medidas compensatorias: consisten en derechos arancelarios adicionales destinados a neutralizar el beneficio que las subvenciones concedidas por el país exportador otorgan a sus productos, en los términos del Acuerdo sobre Subvenciones y Medidas Compensatorias, que se estudia en el tema B33. Al igual que en el antidumping, su imposición exige acreditar la subvención, el daño y la causalidad.

Tipos de cambio múltiples: el Gobierno puede exigir que toda entrada o salida de divisas se tramite ante una autoridad de control de cambios, la cual fija un tipo diferente según el bien objeto del intercambio.

Ayudas condicionadas (*tied aid*): son, por lo general, ayudas al desarrollo cuyos fondos deben gastarse en productos del país donante, lo que distorsiona los flujos comerciales y encarece los proyectos financiados.

Embargos y sanciones comerciales: por razones de política exterior o de defensa, el Gobierno puede prohibir el comercio de ciertos bienes, en exportación o en importación, con ciertos países. Los embargos y las sanciones han cobrado especial relevancia tras la invasión rusa de Ucrania en 2022:

- *EE. UU.*: retiró a Rusia el trato de nación más favorecida, lo que permitió imponer nuevos aranceles y controles a la importación.

- *UE*: aprobó embargos escalonados, que comenzaron por el carbón ruso (importaciones de unos 8000 M€ anuales) y se extendieron después al petróleo, en el marco de sucesivos paquetes de sanciones.

[34] Su duración es de hasta cuatro años, prorrogable hasta un máximo de ocho (diez para los países en desarrollo). Si duran más de un año deben liberalizarse progresivamente, y su prórroga exige demostrar que la rama de producción se está reajustando.

3.3. Soluciones a las barreras no arancelarias

Para evitar las barreras no arancelarias caben dos vías, los Acuerdos de la OMC, que las disciplinan, y la integración económica y comercial, que les hace frente entre sus miembros.

3.3.1. Los Acuerdos de la OMC

Desde el nacimiento del GATT (1947) y, después, con la OMC (1995), el sistema multilateral promueve la competencia leal en el comercio internacional y combate las barreras arancelarias y no arancelarias.

Esto se ha visto reflejado en una serie de acuerdos que cabe agrupar en tres bloques: los relativos a los procedimientos aduaneros, los relativos a las barreras técnicas y los relativos a los instrumentos de defensa comercial.

> Sobre el GATT y la OMC, véase el tema B33. Sobre el Acuerdo General sobre el Comercio de Servicios (GATS) y el Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (TRIPS), véase el tema B34.

3.3.2. Integración económica y comercial

Además de estos acuerdos multilaterales, la integración comercial (áreas de libre comercio, uniones aduaneras o mercados comunes) permite hacer frente a las barreras arancelarias y no arancelarias entre sus miembros.

Esa misma vía tiene un coste, porque los procesos de integración bilateral pueden hacer proliferar las barreras no arancelarias y dar lugar a lo que Jagdish Bhagwati (1995) llamó **problema del bol de espaguetis**. Cada acuerdo de libre comercio genera nuevas normas de origen. Por ello, la proliferación de estos acuerdos, cada uno con su propio entramado de reglas superpuestas, puede terminar siendo un obstáculo al comercio al elevar los costes de transacción y de cumplimiento de las empresas.

> Sobre la teoría de la integración económica y los acuerdos comerciales preferenciales, véase el tema B10, y sobre los procesos de integración no comunitarios, el tema B35, cuyo apartado 1.2.2 trata el bilateralismo.

3.4. Conclusiones

Como el proteccionismo tiene costes económicos y también sociales, la elección del responsable de la política económica (*policy-maker*) entre proteccionismo y libre comercio puede razonarse desde el teorema del *second best* (Lipsey y Lancaster, 1956; tema A22, apartado 3.1). Según este teorema, si al menos una de las condiciones necesarias para alcanzar la optimalidad paretiana no se cumple, puede resultar deseable no perseguir el cumplimiento del resto. Por tanto, en determinados contextos puede ser preferible un *second best* (p. ej. con aranceles) a perseguir a toda costa el *first best* del libre comercio. En el mundo real no cabe determinar *a priori* qué alternativa de política económica es la ideal. En todo caso, puede convenir realizar análisis coste-beneficio que eviten las recetas uniformes (*one-size-fits-all*).

La respuesta desde el *second best* se completa hoy con dos marcos, desarrollados en el tema B21, Anexo A.1, que sitúan la disyuntiva en el terreno de la gobernanza global. Por un lado, según

el trilema de Rodrik (2011), no cabe simultáneamente hiperglobalización, soberanía nacional y democracia. Por otro lado, la tesis de la parálisis o *gridlock* (Hale, Held y Young, 2013) sostiene que las instituciones multilaterales son incapaces de producir cooperación cuando más se necesita.

Conclusiones

La política comercial parte del debate entre librecambio y proteccionismo, en el que los argumentos a favor de la intervención justifican, en general, medidas que ataquen la distorsión en su origen y no barreras al comercio. El arancel reduce siempre el bienestar de un país pequeño. En uno grande, en cambio, puede elevarlo al mejorar la relación real de intercambio, lo que fundamenta el arancel óptimo, pero la represalia conduce a una guerra arancelaria con estructura de dilema del prisionero. Las paradojas de Lerner y Metzler muestran que, en casos poco plausibles, el arancel puede empeorar la relación real de intercambio (Lerner) o desproteger al sector importador (Metzler). Las cuotas, las subvenciones y los impuestos a la exportación y las restricciones voluntarias tienen efectos parecidos a los del arancel, pero reparten las rentas de otro modo, y las barreras técnicas, administrativas y fiscales elevan los costes de transacción. Desde el *second best*, la elección entre proteccionismo y libre comercio depende de cada caso.

La política comercial se evalúa con los instrumentos de la economía del bienestar. Los excedentes miden sus efectos, y el *second best* fundamenta la regla de atacar la distorsión en su origen (tema A22). Sus efectos distributivos descansan en el teorema de Stolper-Samuelson, por el que el arancel eleva la remuneración del factor escaso (tema B05), y la versión moderna de la política comercial, con oligopolios y empresas heterogéneas, da lugar a la política comercial estratégica y a la promoción exterior (tema B08). Por último, como el arancel óptimo invita a la represalia, la cooperación entre países es necesaria. El regionalismo la busca con el riesgo de un bol de espaguetis de acuerdos (tema B10), y el multilateralismo, con las reglas de la OMC, puestas a prueba por la guerra arancelaria de 2025 (tema B33).

Amiti, Redding y Weinstein (2019) documentan que el coste de los aranceles estadounidenses de 2018 recayó en los compradores del país y no, como promete el arancel óptimo, en los exportadores extranjeros, al menos a corto plazo (apartado 2.3.4). Sigue abierto, además, si las instituciones multilaterales pueden todavía gobernar el comercio mundial, algo que Hale, Held y Young (2013) ponen en duda con su tesis de la parálisis (apartado 3.4).

Anexo. Arancel en un país pequeño en el plano de los bienes

Gráficamente, en el plano bien X-bien Y (figura B7.10), el análisis de los aranceles en un país pequeño parte de los siguientes elementos:

1. *Frontera de posibilidades de producción*, FPP: recoge las combinaciones máximas de X e Y que el país puede producir y es cóncava, porque el coste de oportunidad de cada bien crece con su producción.

2. *Recta de balance internacional*: valor de la producción a precios internacionales, con pendiente igual a la RRI, RRI^*.

3. *Curvas de indiferencia social*, CIS_0 y CIS_1: representan un bienestar tanto mayor cuanto más alejadas del origen.

4. *Equilibrio inicial*: el país consume C^0 y produce P^0 y los excesos de oferta y demanda se cubren a través del comercio exterior, cuyo volumen se refleja en el triángulo de comercio. El precio relativo nacional coincide con la RRI[35], es decir, $p_X^0/p_Y^0 = p_X^{*0}/p_Y^{*0} \equiv \text{RRI}^*$. Con competencia perfecta, sin costes de transporte y con libre comercio, los bienes deben tener los mismos precios en el mercado interior y en el exterior, es decir, se cumple la ley del precio único.

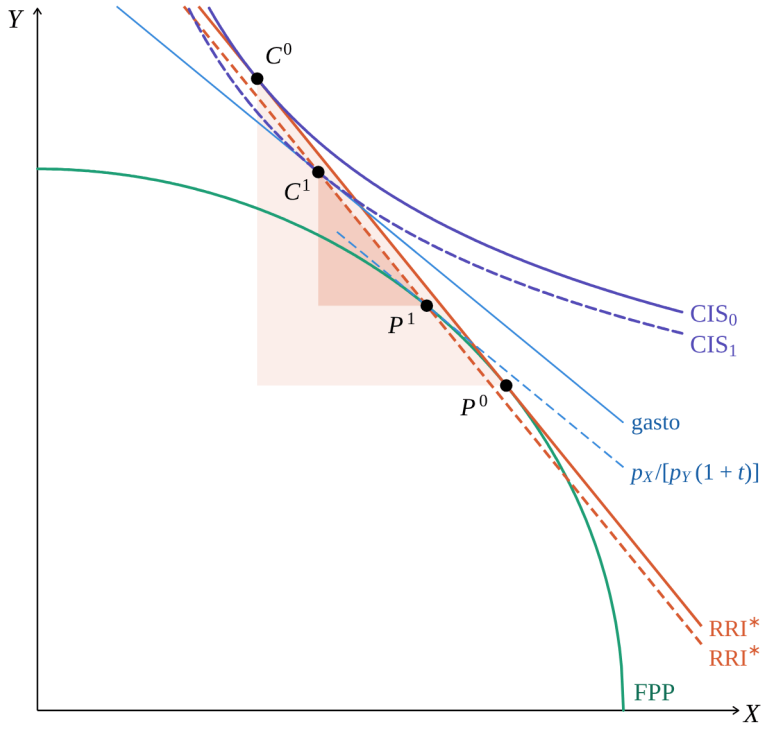

Figura B7.10: Arancel en un país pequeño en el plano de los bienes. La producción se desplaza hacia el bien protegido ($P^0 \to P^1$), el consumo se sitúa en C^1 sobre la recta de balance internacional y el bienestar cae ($\text{CIS}_1 < \text{CIS}_0$).

[35] Sobre la definición y propiedades de la relación real de intercambio, véase el tema B05.

Un arancel *ad valorem* sobre el bien importable rompe con esta equivalencia, porque eleva el precio interior del bien importable. Como el país es pequeño y el arancel no afecta a la RRI, el precio relativo nacional pasa a ser menor que la relación real de intercambio, $p_X^0/[p_Y^0(1+t)] < p_X^{*0}/p_Y^{*0} \equiv \text{RRI}^*$.

Gráficamente, ello hace que[36]:

1. *Recta de balance internacional*: se desplace en paralelo, ya que la RRI no cambia.

2. *Recta de balance nacional*: su pendiente disminuya en valor absoluto en el plano (X, Y), puesto que el precio nacional del bien importable aumenta.

Efectos: el comercio y el bienestar caen como en el apartado 2.3.2, y el plano de los bienes precisa los otros tres efectos del arancel:

▪ *Producción*: aumenta la producción del bien importable o protegido y disminuye la producción del bien exportable X, de modo que la economía se sitúa en el punto P^1, con los efectos sobre la renta de los factores vistos en el apartado 2.3.2.

▪ *Consumo*: disminuye el consumo del bien importable o protegido, puesto que ahora resulta más caro, mientras que el efecto sobre el consumo del bien exportable es ambiguo, porque el efecto sustitución lo eleva y el efecto renta, por la pérdida de bienestar, lo reduce. El consumo no se sitúa en un punto de tangencia con la nueva recta de balance nacional, sino sobre la recta de balance internacional, porque el comercio se equilibra a precios mundiales. El consumo se sitúa allí donde la recta de balance internacional corta a la recta del gasto nacional a precios nacionales, a la que es tangente la curva de indiferencia social.

▪ *Recaudación*: el Estado puede utilizar los ingresos del arancel para aumentar sus compras o redistribuirlos entre los consumidores, en cuyo caso se expande la frontera de posibilidades de consumo de estos y aumenta su gasto. En ambos casos, el arancel eleva el gasto de la economía, por lo que el equilibrio del consumo se dará en el punto C^1.

[36] La recta de balance nacional refleja el valor de la producción a precios nacionales. La recta del gasto nacional refleja el valor del gasto agregado de la economía a precios nacionales, incluida la renta arancelaria redistribuida.

B8. Política comercial II: política estratégica

La política comercial (II). La política comercial estratégica.

Introducción

Introducción común. Los temas B07 y B08 comparten una introducción, que se desarrolla al comienzo del tema **B07**. Distingue la política comercial **defensiva**, que frena las importaciones, de la **ofensiva**, que impulsa la presencia en el exterior, y resume el debate entre librecambio y protección y el marco multilateral del GATT y la OMC.

La **política comercial estratégica** nace de la nueva teoría del comercio. En los mercados de competencia imperfecta hay beneficios extraordinarios, y Brander y Spencer (1985) mostraron que una subvención a la exportación puede trasladarlos de la empresa extranjera a la nacional cuando ambas compiten en cantidades. Sin embargo, el argumento resultó frágil, porque Eaton y Grossman (1986) probaron que, si las empresas compiten en precios, la política óptima es un impuesto, y las represalias y la entrada de nuevas empresas pueden anular la ganancia. Krugman (1987a) concluyó, por ello, que el libre comercio seguía siendo una buena regla práctica, aunque hubiera dejado de ser la política óptima en teoría. El caso que inspiró estos modelos, el apoyo europeo a Airbus frente a Boeing, terminó en la disputa más prolongada y cuantiosa de la historia de la OMC.

La idea ha vuelto con fuerza bajo la forma de política industrial. La Ley CHIPS y de Ciencia de EE. UU. (2022) destinó 52.700 M$ a la investigación y la fabricación de semiconductores, y la UE respondió con el Reglamento (UE) 2023/1781, de chips, que aspira a movilizar 43.000 M€ entre fondos públicos y privados para que la UE produzca en 2030 el 20 % de los semiconductores del mundo. Ambas responden, en parte, a la lógica estratégica de asegurar a las empresas propias un lugar en un sector de rendimientos crecientes, junto a razones de seguridad del suministro.

La **política de promoción exterior** parte, en cambio, de otro fundamento. No busca desplazar rentas de un competidor concreto, sino corregir los fallos de mercado que encarecen la internacionalización, como la falta de información sobre los mercados de destino o los problemas de coordinación entre empresas, y rebaja así el umbral de productividad a partir del cual exportar resulta rentable. Por eso ayuda por igual a todas ellas, mediante agencias como el ICEX y estrategias plurianuales como la Estrategia de Internacionalización de la Economía Española 2017-2027.

Este tema trata la política comercial estratégica y la política de promoción exterior. Aunque el título solo nombra la primera, la segunda entra aquí porque es, como se ha visto, la versión actual de la política comercial ofensiva y la que hoy prefieren los países desarrollados (apartado 2.5). Ambas se estudian en tres secciones:

- *Justificación*: fundamentos de la política comercial estratégica y de la política de promoción exterior.
- *Política comercial estratégica*: objetivos, instrumentos y efectos.
- *Promoción exterior*: objetivos, instrumentos y efectos.

1. Justificación de la política comercial moderna

La política comercial estratégica se justifica por la interdependencia estratégica entre empresas, y la de promoción exterior, por razones históricas y por fallos de mercado.

1.1. Política comercial estratégica

La justificación de la política comercial estratégica depende de la orientación que se considere (defensiva u ofensiva). No obstante, en cualquiera de sus dos versiones, la justificación está relacionada con la interdependencia estratégica.

1.1.1. Vertiente ofensiva: interdependencia estratégica y credibilidad

En primer lugar, la política comercial estratégica, en su vertiente ofensiva, puede justificarse por los **comportamientos estratégicos** que mantienen aquellos agentes económicos entre los que existe interdependencia estratégica en un tercer mercado. En esta situación, las empresas extranjeras pueden estar apropiándose de beneficios extraordinarios en detrimento de las empresas nacionales. Esto da lugar a un reparto desigual de dichos beneficios y, por tanto, la autoridad fiscal nacional puede buscar el estímulo de la renta nacional (y, por ende, del empleo) mediante la redistribución de dichos beneficios.

El mecanismo de esa redistribución descansa en la **credibilidad del compromiso**. Una empresa que anuncia que expandirá su producción para disuadir a su rival puede no resultar creíble, pues la amenaza no necesariamente es su mejor respuesta ante el comportamiento del rival. El sector público, en cambio, puede alterar de forma creíble los incentivos de su empresa nacional, porque sus instrumentos (presupuestarios y normativos) son públicos, vinculantes y difíciles de revertir.

Mediante una subvención a la exportación o a la producción, un crédito blando, ayudas a la I+D o la protección del mercado doméstico como plataforma, la autoridad reduce el coste marginal (efectivo) de la empresa nacional y puede convertirla en líder estratégico. Desplaza así hacia ella beneficios extraordinarios del tercer mercado, que pueden superar el coste de la ayuda. Brander y Spencer (1985) formalizan este mecanismo, que exige competencia en cantidades (a la Cournot) y se desarrolla en el apartado 2.2.1.

1.1.2. Vertiente defensiva: interdependencia estratégica y economías de escala externas

En su vertiente defensiva, no solo deben existir comportamientos estratégicos, sino también otro tipo de fallo de mercado como, por ejemplo, economías de aglomeración que dificultan la competitividad de las empresas nacionales. En este caso, la política comercial estratégica defensiva se dirige a reequilibrar la falta de competitividad internacional en línea con el **argumento de la industria naciente**.

El argumento (Hamilton, 1791; Raymond, 1820; List, 1841) y sus criterios de validez, de Mill y de Bastable, se exponen en el tema B07, apartado 1.2.2, pero aquí interesa su versión estratégica (apartado 2.3.2).

En lugar de la protección, puede fomentarse primero la competencia, para que las compañías nacionales ganen competitividad y no las desplacen las extranjeras (p. ej. liberalizando sectores o mediante organismos de defensa de la competencia).

1.2. Política de promoción de las exportaciones

La justificación de la política de promoción de las exportaciones es doble y va más allá del fallo de mercado, pues incluye también razones históricas que llevaron al desarrollo de este tipo de políticas.

1.2.1. Razones históricas

Desde los años sesenta, la Ronda Kennedy (1964-1967) y la reforma del GATT que la siguió redujeron los aranceles de forma progresiva, al tiempo que aparecían nuevos instrumentos proteccionistas no arancelarios. La crisis del petróleo de los setenta reavivó después el proteccionismo, ahora sobre todo por vías no arancelarias, y dio origen al llamado **neoproteccionismo**.

En el nuevo paradigma comercial, la OMC ha ido prohibiendo o disciplinando progresivamente los instrumentos arancelarios y no arancelarios, pero no se ha enfrentado tan abiertamente a las políticas de promoción de exportaciones. La razón es que estas no restringen el acceso a los mercados sino que lo facilitan, de modo que ningún socio comercial resulta perjudicado de forma directa y evidente. Por ello, solo quedan sujetas a disciplina cuando adoptan la forma de subvención específica en el sentido del Acuerdo sobre Subvenciones y Medidas Compensatorias (ASMC).

Lo anterior explica, en parte, que los países hayan desplazado su intervención hacia este tipo de políticas. No obstante, también existen razones que permiten justificar la promoción exterior desde el punto de vista de los fallos de mercado.

1.2.2. Fallos de mercado: imperfecciones en los mercados de factores y costes de acceso al mercado de bienes

Además de las razones históricas, la teoría económica ofrece motivos por los cuales la política de promoción es justificable:

Imperfecciones en el mercado de capitales o de trabajo: un país puede carecer de un sistema financiero eficiente o tener un mercado de trabajo con rigideces. En ese caso, el trasvase del ahorro y de los factores productivos desde sectores tradicionales (p. ej. manufacturas maduras) hacia sectores nuevos con mayores beneficios potenciales puede no producirse adecuadamente. Por lo tanto, el crecimiento de las nuevas industrias se verá restringido por su incapacidad para obtener la financiación y la mano de obra que les permitirían generar mayores flujos de caja en el futuro, pues solo cuentan con sus reducidos flujos de caja actuales.

Estas dificultades pueden ser un argumento en favor de políticas de apoyo horizontal (que no directo) a los exportadores o a los sectores que se vean lastrados por esas dificultades. Entre esas políticas figuran las líneas públicas de financiación a la internacionalización y los programas públicos de formación de especialistas en comercio exterior.

Costes de acceso al mercado de bienes finales: en sus procesos de internacionalización, las empresas se enfrentan a barreras informacionales y legales[1]. La política de promoción de exportaciones puede justificarse por múltiples causas:

[1] Entre las barreras legales destaca el exceso de burocracia (*red tape*), es decir, la regulación excesiva, rígida o redundante que obstaculiza las decisiones de los agentes económicos.

- *Información*: la información útil para internacionalizarse tiene naturaleza de bien público, así que, si su suministro queda en manos privadas, puede resultar ineficiente (p. ej. por infraprovisión). El sector público, por su parte, puede ofrecer a las empresas un conocimiento que les costaría mucho reunir por su cuenta.

- *Incentivos*: cuando los agentes son distintos entre sí y conocen poco el mercado de destino, a los exportadores puede no compensarles tejer allí una red de contactos. El sector público puede suplir esa carencia con mecanismos que faciliten a las empresas el acceso a los mercados exteriores.

- *Implementación práctica y capital humano*: cuando una empresa carece del "saber hacer" (*know-how*) necesario, puede costarle llevar a la práctica sus proyectos en el exterior. En esa tarea puede apoyarse en lo que la Administración conoce de los mercados exteriores y de su legislación.

1.3. Conclusiones

En síntesis, la política comercial estratégica descansa en dos rasgos de los oligopolios internacionales, la interdependencia estratégica entre empresas y los beneficios extraordinarios que generan. Su variante defensiva exige, además, otro fallo de mercado, como las economías de escala externas. La promoción exterior se apoya, en cambio, en cómo ha evolucionado el marco comercial multilateral y en los fallos de mercado que entorpecen la decisión de internacionalizarse.

Que haya un fallo de mercado es, en los dos casos, un requisito previo, pero no basta para que la intervención esté justificada. También el sector público falla, y por eso cada medida debe superar un análisis coste-beneficio. Identificadas las causas, los apartados 2 y 3 estudian los objetivos, instrumentos y efectos de cada política.

2. La política comercial estratégica: objetivos, instrumentos y efectos

La política comercial estratégica se aborda a partir de sus objetivos e instrumentos, para estudiar después sus tres vertientes, la ofensiva, la defensiva y la de la I+D.

2.1. Objetivos e instrumentos

Se exponen primero los objetivos de la política comercial estratégica, que busca desviar hacia el país nacional los beneficios extraordinarios de las empresas extranjeras, y después sus dos instrumentos, el impuesto y la subvención.

2.1.1. Objetivos

Como se adelantó en el apartado 1.1, la justificación de la política comercial estratégica es la existencia de comportamientos estratégicos en los mercados oligopolísticos internacionales. Ante esta situación, el objetivo último de la política comercial estratégica es desviar los beneficios extraordinarios obtenidos por las empresas extranjeras en favor del país nacional. Esta "repatriación" de beneficios puede redundar en un mayor beneficio para las empresas nacionales (p. ej. vía subvenciones) o derivar en una mayor recaudación del sector público (p. ej. vía impuestos a la exportación). El mayor beneficio puede mejorar la competitividad de esas empresas.

2.1.2. Instrumentos

La política comercial estratégica dispone de dos instrumentos, el impuesto y la subvención. Lo habitual es que recaigan sobre las exportaciones, aunque nada impide aplicarlos a la producción. En teoría, la elección entre una y otra base (exportación o producción) carece de importancia cuando las empresas estudiadas no venden en su propio país y colocan toda su producción en un tercer mercado, porque en ese caso los resultados son idénticos. La vertiente defensiva recurre, en cambio, al arancel a la importación, y la de la I+D, al apoyo público a la investigación.

2.2. Vertiente ofensiva

Si existen comportamientos estratégicos en un mercado, estos pueden derivar de la existencia de barreras a la entrada. Lo óptimo sería, por tanto, suprimirlas, algo que a menudo resulta inviable (p. ej. cuando la I+D exige costes muy altos, como en la industria aeronáutica). Por ello, el objetivo de la política comercial estratégica no es corregir el fallo de mercado, sino tan solo capturar esos beneficios extraordinarios, por lo que se trata de una política de *second best* (en concreto, la denominada de **desplazamiento de rentas** o *rent shifting*).

> Los modelos de oligopolio en economía abierta que fundamentan esta sección se desarrollan en el tema B06.

2.2.1. Competencia a la Cournot en bienes homogéneos: subvenciones a la exportación (Brander y Spencer, 1985)

Origen: James Brander y Barbara Spencer (1985) estudian la política comercial estratégica basada en subvenciones a la exportación.

Idea: las subvenciones a la exportación pueden alterar el resultado de la interacción estratégica, dado que el sector público cuenta con credibilidad institucional suficiente como para desplazar el equilibrio hacia una solución de Stackelberg en favor de la empresa nacional.

Supuestos: los habituales del modelo canónico de Cournot:

▪ *Entorno*: duopolio con un bien homogéneo. A diferencia del modelo de Brander y Krugman (1983), ambas empresas compiten en un tercer mercado[2].

▪ *Agentes*: competencia en cantidades, con expectativas miopes. Se cumplen las condiciones de Hahn (funciones de reacción continuas, cantidades que son sustitutos estratégicos y estabilidad, $|\partial\varphi_i/\partial x_j| < 1$), de modo que las funciones de reacción tienen pendiente negativa en el plano (x_1, x_2).

Planteamiento: con una subvención z a favor de la empresa 1, los programas de maximización de las dos empresas son:

$$\max_{x_1} \ \pi_1 = p(x_1 + x_2) \cdot x_1 - \mathrm{CT}_1(x_1) + zx_1$$

$$\max_{x_2} \ \pi_2 = p(x_1 + x_2) \cdot x_2 - \mathrm{CT}_2(x_2)$$

donde $p(x_1 + x_2)$ es la función inversa de demanda del tercer mercado, que depende de la producción total $X = x_1 + x_2$, x_i la cantidad exportada por la empresa i, $\mathrm{CT}_i(x_i)$ su función de costes y z la subvención por unidad exportada.

CPO: las condiciones de primer orden son $(dp/dX)\, x_1 + p - d\mathrm{CT}_1/dx_1 + z = 0$ y $(dp/dX)\, x_2 + p - d\mathrm{CT}_2/dx_2 = 0$. Es decir, cada empresa iguala su ingreso marginal a su coste marginal, que la subvención reduce de forma efectiva para la empresa 1. De ellas se obtienen las *funciones de reacción* $\varphi_1(x_2)$, con $\partial\varphi_1/\partial x_2 < 0$ y $\partial\varphi_1/\partial z > 0$, y $\varphi_2(x_1)$, con $\partial\varphi_2/\partial x_1 < 0$.

CSO: las condiciones de Hahn garantizan la concavidad estricta de la función de beneficios de cada empresa en su propia cantidad, por lo que las condiciones de primer orden son suficientes.

Equilibrio: el equilibrio de Nash-Cournot es el par de cantidades (x_1^*, x_2^*) que resuelve a la vez $x_1 = \varphi_1(x_2; z)$ y $x_2 = \varphi_2(x_1)$, de modo que cada empresa produce su mejor respuesta a la cantidad de la otra. Como la función de reacción de la empresa 1 depende de la subvención, también lo hacen las cantidades de equilibrio, $x_1^*(z)$ y $x_2^*(z)$.

Desarrollo gráfico: los efectos de una subvención a la exportación se representan en el plano de cantidades, (x_1, x_2) (figura B8.1):

1. *Función de reacción de la empresa 1*, $\varphi_1(x_2)$: recoge la cantidad que maximiza el beneficio de la empresa 1 para cada cantidad de su rival (CPO) y es decreciente, porque las cantidades son sustitutos estratégicos.

[2] El tercer mercado aísla el desplazamiento de rentas del comercio intraindustrial (Brander y Krugman, 1983).

2. *Función de reacción de la empresa 2*, $\varphi_2(x_1)$: recoge la mejor respuesta de la empresa 2 a cada cantidad de la empresa 1 y es también decreciente, por la misma razón.

3. *Curvas de isobeneficio de la empresa 1*, π_1^0 y π_1^1: recogen el beneficio de mercado de la empresa 1 (sin contar la subvención), tanto mayor cuanto menor es la cantidad de la rival.

4. *Equilibrio*: el corte de ambas funciones de reacción es el equilibrio de Nash-Cournot inicial (punto 0), con cantidades x_1^0 y x_2^0.

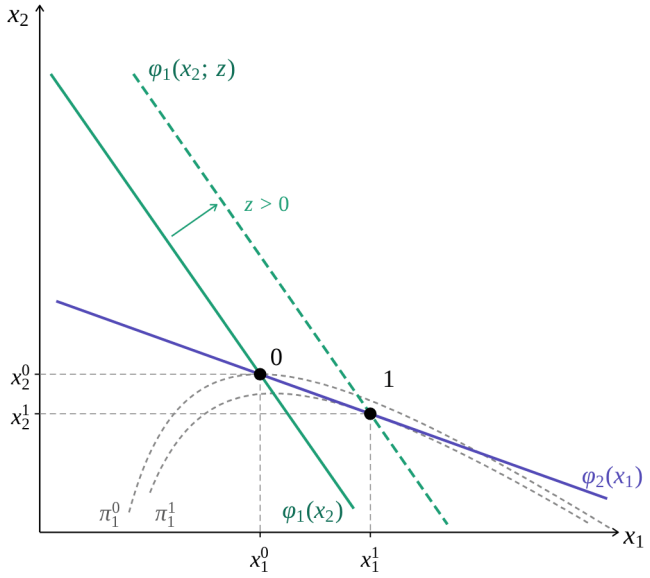

Figura B8.1: Modelo de Brander y Spencer (1985). La subvención a la exportación desplaza la curva de reacción de la empresa nacional y le permite alcanzar el resultado de Stackelberg.

Estática comparativa: la subvención a la exportación ($z > 0$) desplaza a la derecha la función de reacción de la empresa 1, de $\varphi_1(x_2)$ a $\varphi_1(x_2; z)$. Dicho de otro modo, para cualquier nivel de producción de su rival, la empresa subvencionada lanza al mercado una cantidad superior. El equilibrio pasa así del punto 0 al 1 (x_1^1, x_2^1). Se puede demostrar que la subvención aumenta x_1^* y la producción total, $X^* = x_1^* + x_2^*$, y reduce x_2^*. En consecuencia, el precio cae, π_1 aumenta y π_2 disminuye.

Bienestar: el bienestar del país que subvenciona se define como $W_s = \pi_1(z) - z \cdot x_1(z)$, donde $z \cdot x_1(z)$ es el coste presupuestario de la subvención, de modo que W_s coincide con el beneficio de mercado de la empresa 1 (sin contar la subvención). La subvención que maximiza W_s (subvención óptima) es positiva y lleva el equilibrio al punto 1, donde la curva de isobeneficio π_1^1 es tangente a $\varphi_2(x_1)$, es decir, al resultado de Stackelberg con la empresa nacional como líder[3].

En ese resultado los consumidores del tercer país ganan, porque la producción total de Stackelberg es mayor que la de Cournot, y la empresa nacional aumenta su beneficio a costa de su

[3]No se trata de una competición a la Stackelberg, sino a la Cournot con un resultado como el de Stackelberg. La distinción es relevante porque en Stackelberg la empresa líder no se sitúa sobre su propia curva de reacción $\varphi_1(x_2)$, sino que elige directamente el punto de $\varphi_2(x_1)$ en que esta es tangente a su curva de isobeneficio. Aquí, en cambio, la empresa 1 sigue sobre su curva de reacción, y la subvención desplaza $\varphi_1(x_2)$ hasta que corta a $\varphi_2(x_1)$ exactamente en ese punto de tangencia con la curva de isobeneficio correspondiente al liderazgo.

competidora[4]. La subvención es, por tanto, una política de empobrecimiento del vecino (*beggar thy neighbour*)[5].

Conclusiones:

▪ *Aspectos positivos*: el modelo de Brander-Spencer muestra cómo la credibilidad de la autoridad fiscal puede dar lugar a ganancias de bienestar para el país mediante una subvención a las exportaciones que desplaza rentas desde la empresa rival hacia la nacional. Un ejemplo clásico de este tipo de apoyo es la industria aeronáutica. Durante décadas, el sector aeroespacial comercial ha sido objeto de subvenciones a la producción con base en el argumento del fallo de mercado[6]. Así, tras los acuerdos intergubernamentales de finales de los años sesenta, en 1970 se constituyó el consorcio europeo Airbus con el objetivo de ganar cuota en el mercado internacional (es decir, un "tercer mercado") frente a la empresa estadounidense entonces hegemónica, Boeing. El apoyo recibido por Airbus consistía, esencialmente, en la financiación de la I+D.

A raíz de este apoyo, Airbus comenzó a ganar cuota de mercado en detrimento de Boeing, lo cual dio lugar a disputas comerciales entre EE. UU. y la UE. Las planteadas en el seno del GATT desembocaron en el acuerdo bilateral de 1992, que limitó el apoyo público a ambos fabricantes. Las posteriores ante la OMC (desde 2004) quedaron en suspenso con la tregua de 2021, y en julio de 2026, al vencer su plazo de cinco años, la UE prorrogó la suspensión de sus aranceles de represalia.

▪ *Limitaciones*: por un lado, el modelo pasa por alto cómo se paga la subvención a la exportación. Cuando los fondos proceden de impuestos distorsionantes, el coste de eficiencia que estos generan (el exceso de gravamen) puede neutralizar una parte del bienestar nacional ganado en el tercer mercado. Por otro, este tipo de políticas da lugar a riesgo de *dumping*, puesto que la empresa subvencionada tiene incentivos para cobrar un precio menor en el mercado extranjero que en el mercado nacional. De ahí el riesgo de represalias (derechos antidumping o, frente a la propia subvención, derechos compensatorios al amparo del ASMC).

2.2.2. Competencia a la Bertrand en bienes diferenciados: impuestos a la exportación (Eaton y Grossman, 1986)

El modelo anterior supone que la variable estratégica son las cantidades. Sin embargo, nada asegura que las empresas se comporten así y, a corto plazo, lo habitual es que compitan en precios.

Origen: Jonathan Eaton y Gene Grossman (1986) estudian el papel de los impuestos a la exportación, es decir, de los derechos aduaneros a la exportación, en un marco de competencia oligopolística en precios.

Idea: cuando las empresas compiten en precios y los precios son complementarios estratégicos, el instrumento óptimo se invierte y pasa a ser un impuesto a la exportación, que eleva los precios de ambas rivales. De este modo, el país nacional (beneficio neto de su empresa más recaudación) gana a costa de los consumidores del tercer país.

Supuestos: son similares a los de la competencia a la Cournot en un tercer mercado, salvo en tres aspectos:

[4] O a costa de las exportaciones de un tercer país, pues $\varphi_2(x_1)$ puede representar la reacción de la producción del país de destino o de las exportaciones de un tercer competidor.

[5] La idea aparece ya en *La riqueza de las naciones* (1776) de Adam Smith. En referencia a la doctrina mercantilista, Smith critica que se haya enseñado a las naciones que su interés consiste en empobrecer a todos sus vecinos (*beggaring all their neighbours*).

[6] Concretamente, en la industria aeronáutica aparecen dos tipos de barreras a la entrada, las importantes economías de escala internas y las cuantiosas inversiones en I+D.

- *Entorno*: duopolio en un tercer mercado con bienes ligeramente diferenciados. La diferenciación imperfecta hace que las empresas descarten vender a coste marginal.

- *Agentes*:

  - *Comportamiento*: competencia en precios, con expectativas miopes.

  - *Tratamiento*: los precios son complementarios estratégicos, de modo que las funciones de reacción tienen pendiente positiva en el plano (p_1, p_2)[7]. Cuando un rival aumenta su precio, parte de su demanda se desplaza hacia la otra empresa, lo que eleva el beneficio marginal de subir el propio precio. Por ello, la mejor respuesta es aumentarlo también[8].

Planteamiento: el programa de maximización con un impuesto a la exportación de tipo t sobre la empresa 1 es:

$$\max_{p_1} \ \pi_1 = (1-t)\, p_1\, x_1(p_1, p_2) - \mathrm{CT}_1[x_1(p_1, p_2)]$$

$$\max_{p_2} \ \pi_2 = p_2\, x_2(p_1, p_2) - \mathrm{CT}_2[x_2(p_1, p_2)]$$

donde p_i es el precio fijado por la empresa i, $x_i(p_1, p_2)$ su demanda, CT_i su función de costes y t el tipo del impuesto a la exportación.

CPO: las condiciones de primer orden son $\partial\pi_1/\partial p_1 = 0$ y $\partial\pi_2/\partial p_2 = 0$. Es decir, cada empresa elige el precio que iguala el ingreso marginal, neto del impuesto en el caso de la empresa 1, al coste marginal de las unidades que a ese precio vende. De ellas se obtienen las funciones de reacción $\varphi_1(p_2)$, con $\partial\varphi_1/\partial p_2 > 0$ y $\partial\varphi_1/\partial t > 0$, y $\varphi_2(p_1)$, con $\partial\varphi_2/\partial p_1 > 0$.

CSO: se supone que la función de beneficios de cada empresa es estrictamente cóncava en su propio precio, de modo que las condiciones de primer orden son suficientes.

Equilibrio: el equilibrio de Nash-Bertrand es el par de precios (p_1^*, p_2^*) que resuelve a la vez $p_1 = \varphi_1(p_2;\, t)$ y $p_2 = \varphi_2(p_1)$, en el que cada empresa fija el precio que mejor responde al de la otra. Los precios de equilibrio dependen, por tanto, del tipo del impuesto, $p_1^*(t)$ y $p_2^*(t)$.

Desarrollo gráfico: los efectos de un impuesto a la exportación se representan en el plano de precios, (p_1, p_2) (figura B8.2):

1. *Función de reacción de la empresa 1*, $\varphi_1(p_2)$: recoge el precio que maximiza el beneficio de la empresa 1 para cada precio de su rival (CPO) y es creciente, porque los precios son complementarios estratégicos.

2. *Función de reacción de la empresa 2*, $\varphi_2(p_1)$: recoge la mejor respuesta de la empresa 2 a cada precio de la empresa 1 y es también creciente, por la misma razón.

3. *Equilibrio*: el corte de ambas funciones es el equilibrio de Nash-Bertrand inicial (punto 0), con precios p_1^0 y p_2^0.

[7] Además, en este caso las curvas de isobeneficio más alejadas del origen conllevan mayores beneficios, a diferencia del caso con pendientes negativas.

[8] En la competencia en cantidades ocurre lo contrario. Cuando el rival produce más, el precio de mercado cae y el ingreso marginal de la propia empresa disminuye, por lo que su mejor respuesta es producir menos (sustituibilidad estratégica).

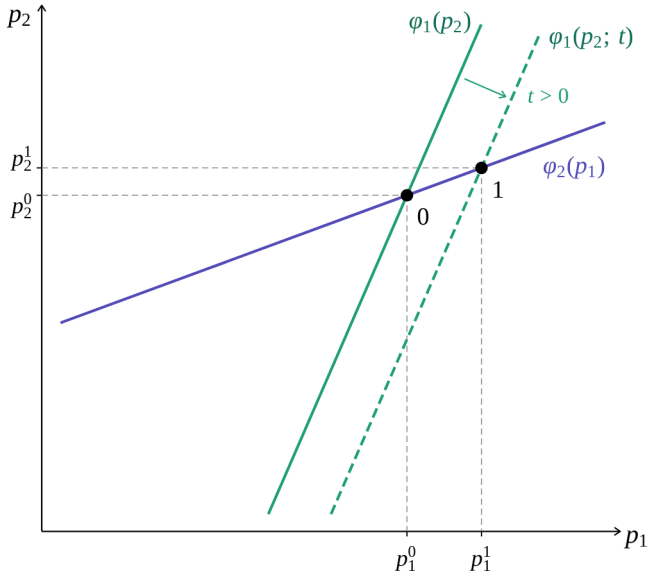

Figura B8.2: Modelo de Eaton y Grossman (1986). Con competencia a la Bertrand el instrumento óptimo es un impuesto a la exportación.

Estática comparativa: un impuesto a la exportación ($t > 0$) sobre la empresa 1 produce un desplazamiento hacia la derecha de su curva de reacción, de $\varphi_1(p_2)$ a $\varphi_1(p_2;\, t)$. La razón es que la empresa nacional traslada el impuesto a un aumento de su precio para cualquier precio fijado por su rival. Dado que los precios son complementarios estratégicos, la subida de p_1 da lugar a un aumento del beneficio marginal de la empresa 2 ($\partial^2\pi_2/\partial p_1 \partial p_2 > 0$), por lo que esta también sube su precio. Al final, ambas empresas fijan un precio superior, p_1^1 y p_2^1, y el equilibrio pasa del punto 0 al 1.

Bienestar: el bienestar del país nacional es el beneficio neto de su empresa más la recaudación, $W = \pi_1(t) + t\, p_1 x_1$, es decir, el beneficio bruto (antes del impuesto) de la empresa 1. Con el impuesto óptimo, el punto 1 corresponde a la solución de Stackelberg en precios con la empresa nacional como líder, con mayores beneficios brutos para ambas empresas y mayor bienestar para el país nacional.

Esta situación no sería posible sin intervención, porque la rival no creería una promesa de subida del precio de la empresa nacional, pues sabe que subir el precio unilateralmente lleva a perder cuota de mercado. El impuesto, en cambio, es un compromiso público que la rival sí toma como vinculante. Los consumidores del tercer país, por su parte, pierden bienestar debido a la reducción de la presión competitiva (es decir, al comportamiento cuasicolusorio de los duopolistas). En resumen, el país nacional sale ganando con esta política a expensas de los consumidores extranjeros.

Conclusiones:

- *Aspectos positivos*: el modelo permite estudiar otro tipo de instrumentos y obtener intuiciones adicionales sobre la política comercial estratégica. Hasta el s. XIX, los impuestos a la exportación fueron comunes en Europa como medio de recaudación y como forma de conservar recursos para la manufactura local, y con la liberalización comercial se eliminaron. No obstante, a pesar de su declive en Europa, se mantuvieron en muchas colonias africanas y asiáticas con un doble objetivo. Por un lado, garantizar que las exportaciones desde las colonias no se destinaran a terceros países

y, en su lugar, se dirigieran a las metrópolis (p. ej. Reino Unido, España o Francia). Por otro, obtener recaudación, como en el caso de los impuestos a las materias primas exportadas desde esas colonias. Esta herencia explica que muchos países en desarrollo todavía mantengan este tipo de impuestos con el fin de obtener ingresos adicionales para el sector público.

▪ *Limitaciones*: la principal crítica, subrayada por Paul Krugman, es que el resultado de la política comercial estratégica depende del instrumento y del modo de competencia. Así, bajo Cournot el instrumento óptimo es una subvención a la exportación y bajo Bertrand, un impuesto. Dado que el Gobierno rara vez conoce el verdadero modo de competencia de las empresas, la política puede resultar contraproducente.

2.2.3. Extensiones a los modelos anteriores: guerras comerciales estratégicas, libre entrada y consumo nacional

Por último, tres extensiones de los modelos anteriores aportan intuiciones adicionales, la reacción del Gobierno extranjero, la libre entrada de empresas y la existencia de consumo nacional.

Reacción extranjera: guerra comercial estratégica: hasta aquí se ha supuesto que el Gobierno extranjero permanece pasivo ante cualquiera de las dos políticas, aunque en la práctica los países suelen responder a la política comercial de sus rivales (p. ej. China y EE. UU.). En el plano teórico, si el Gobierno rival replica con el mismo instrumento (subvención z^* o impuesto, según el caso), su empresa ve desplazarse en paralelo su propia función de reacción. El desplazamiento se produce por igual en Brander y Spencer (1985) y en Eaton y Grossman (1986), pero lo que cambia de un modelo a otro son las consecuencias.

▪ *Brander y Spencer*: ambos países salen perdiendo, en una situación de dilema del prisionero. El bienestar de cada país (beneficio de su empresa neto del coste de la subvención) es el beneficio de mercado de su empresa, que cae porque ambas producen más y el precio baja. Queda así por debajo del alcanzado sin intervención.

▪ *Eaton y Grossman*: ambas empresas obtienen beneficios brutos (antes de impuestos) mayores y los consumidores del tercer país sufren una nueva pérdida de bienestar, señal de que el resultado queda aún más lejos del de competencia perfecta.

Gráficamente, en el plano de cantidades, (x_1, x_2), del modelo de Brander y Spencer (figura B8.3), cada función de reacción recoge la cantidad que maximiza el beneficio de una empresa dada la de su rival. Cada curva de isobeneficio reúne las combinaciones de cantidades que dan a una empresa el mismo beneficio de mercado. Las subvenciones de ambos Gobiernos ($z > 0$ y $z^* > 0$) desplazan hacia fuera las dos funciones de reacción, a $\varphi_1(x_2; z)$ y $\varphi_2(x_1; z^*)$. Así, el equilibrio pasa del punto 0 al 1, en el que ambas empresas se sitúan en curvas de isobeneficio de mercado (sin contar la subvención) peores, π_1^1 y π_2^1.

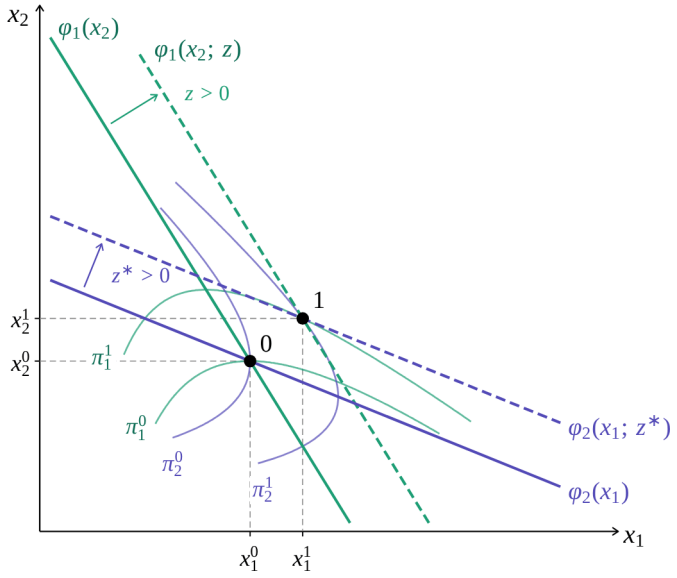

Figura B8.3: Guerra de subvenciones. Cuando ambos Gobiernos intervienen, el resultado es peor para los dos que el equilibrio sin intervención.

Endogenización del número de empresas (Horstmann y Markusen, 1986): otra línea de trabajo, en la que se inscriben varios autores, admite la libre entrada, con lo que puede haber más de dos competidores. Bajo ese supuesto, y siempre que todas las empresas sean homogéneas, los beneficios extraordinarios que genera la subvención atraen a nuevas empresas y alimentan la búsqueda de rentas, hasta que, con el paso del tiempo, esos beneficios acaban por desaparecer.

La entrada intensifica la competencia entre las empresas nacionales. La producción conjunta aumenta (aunque la de cada empresa no aumenta e incluso se reduce), el precio de las exportaciones se reduce y, con él, los beneficios. En consecuencia, la subvención se disipa en un exceso de entrada con plantas de escala ineficiente. No obstante, este resultado no es general, pues Venables (1985) obtiene que, con mercados segmentados y libre entrada, aranceles y subvenciones pueden elevar el bienestar del país que los aplica.

Existencia de consumo nacional: por último, si cada país consume también el bien en su propio territorio, la política comercial estratégica suma objetivos nuevos, porque ha de procurar que la ineficiencia del mercado doméstico, medida frente a la competencia perfecta, sea la menor posible.

2.2.4. Conclusiones

Aspectos positivos: en sentido amplio, la vertiente ofensiva de la política comercial estratégica suele asociarse a la experiencia de los cuatro tigres asiáticos (Corea del Sur, Hong Kong, Singapur y Taiwán) bajo la doctrina de la *industrialización orientada a la exportación* (IOE, *export-oriented industrialization*). No obstante, Hong Kong es una excepción parcial, pues su política comercial se aproximó más al *laissez faire*. La IOE tiene el objetivo de acelerar la industrialización de un país a través del fomento de las exportaciones, de modo que este pueda explotar su ventaja comparativa. Para ello combina la liberalización y la protección comercial. Dado que la expansión de los sectores exportadores habitualmente requería mejoras del capital humano

y de la tecnología, en la práctica la IOE se combinó con elementos de la industrialización por sustitución de importaciones.

Limitaciones: hasta ahora solo se han considerado dos diferencias respecto a los mercados de competencia perfecta, los comportamientos estratégicos derivados de la interdependencia y el poder de mercado con bienes diferenciados. Sin embargo, la consideración de elementos adicionales habituales en el mundo real (p. ej. economías de escala externas) permite justificar la política comercial estratégica en su vertiente defensiva a través del argumento de la industria naciente.

2.3. Vertiente defensiva

La vertiente defensiva se estudia a partir de cuatro argumentos, la estructura de mercado mixta, la industria naciente, el argumento centro-periferia y la protección de las importaciones como promoción de las exportaciones.

2.3.1. Estructura de mercado mixta y bienes diferenciados: aranceles a la importación (Eaton y Kierzkowski, 1984)

Origen: Jonathan Eaton y Henryk Kierzkowski (1984) estudian estructuras de mercado mixtas.

Idea: en una estructura de mercado mixta, un arancel a la importación puede resultar socialmente óptimo si induce la entrada de un productor nacional del bien diferenciado.

Supuestos:

- *Entorno*: la economía tiene dos sectores. Uno fabrica en competencia perfecta un bien homogéneo y el otro, en régimen de oligopolio, un bien diferenciado.

- *Agentes*: la demanda del bien diferenciado responde al enfoque de características (Lancaster, 1966). Cada consumidor fija un umbral o precio crítico que decide su elección entre uno y otro bien. Con un precio por debajo de ese umbral demanda el bien diferenciado y, con uno por encima, se queda con el homogéneo.

Razonamiento: con costes fijos de entrada, la apertura al comercio puede conducir incluso a un resultado inferior en el sentido de Pareto. El modelo admite, además, un caso particular, aquel en que el bien diferenciado no se produce en el país. En esa situación, gravar las importaciones con un arancel, como medida de política comercial estratégica, puede ser lo socialmente óptimo, porque empuja a una empresa nacional a entrar en el sector.

Conclusiones:

- *Aspectos positivos*: el modelo completa el análisis de los apartados anteriores al considerar una estructura de mercado mixta, con un sector competitivo y otro oligopolístico.

- *Limitaciones*: como los modelos de Brander y Spencer y de Eaton y Grossman, no estudia los efectos de equilibrio general de la política.

2.3.2. Argumento de la industria naciente: aranceles a la importación (Hamilton, 1791; List, 1841)

Origen: Alexander Hamilton (1791) y Friedrich List (1841), como se indicó en el apartado 1.1.2.

Idea: una industria nueva que opera con economías de escala externas puede necesitar protección temporal frente a rivales ya establecidos hasta alcanzar la escala y el aprendizaje que le permitan competir.

Supuestos:

- *Entorno*: existen dos países (Suiza y Tailandia) que pueden producir un mismo bien (relojes) para el mercado mundial, del que forma parte la demanda tailandesa. Además, un accidente histórico ha hecho que Suiza comenzase a producir antes.

- *Tecnología*: ambos países producen en un régimen de economías de escala externas[9], si bien su estructura de costes difiere. En Suiza los salarios son comparativamente elevados, mientras que en Tailandia son bajos.

Planteamiento: la situación inicial se representa en el plano cantidad agregada-precio y coste (figura B8.4):

1. *Curva de demanda*, $D(Y)$: refleja la demanda de relojes en el mercado internacional y es decreciente, porque a un precio menor el mercado absorbe más relojes.

2. *Curva de coste medio de Suiza*, $\text{Cme}_S(Y)$: tiene pendiente negativa, dado que las economías de escala externas hacen que el coste medio caiga conforme aumenta la producción agregada del país.

3. *Curva de coste medio de Tailandia*, $\text{Cme}_T(Y)$: también decreciente, queda por debajo de la de Suiza, puesto que Tailandia goza de una mayor competitividad salarial. El coste medio de su primera unidad es c_0.

4. *Equilibrio*: con libre entrada, el precio se iguala al coste medio de la industria, no a su coste marginal, que generaría pérdidas. Suiza abastece el mercado mundial al precio p_S, con la cantidad Y_S (punto 0). Por su mayor competitividad salarial, Tailandia podría abastecerlo al precio p_T, con la cantidad Y_T (punto 1, potencial). Sin embargo, el accidente histórico se lo impide, a pesar de su ventaja comparativa potencial, porque el coste medio de su primera unidad está por encima del precio internacional fijado por la industria suiza ($c_0 > p_S$)[10].

[9] Con economías externas, cada empresa sigue siendo precio-aceptante. Se consideran externas porque se razona en términos de industrias nacionales y no de empresas individuales.

[10] Ninguna empresa tailandesa querrá dar el primer paso, porque hay equilibrios múltiples y un fallo de coordinación impide pasar al de menor coste. Caben dos soluciones: la cooperativa (entrada simultánea, que sitúa de inmediato el coste medio por debajo del precio internacional) y la proteccionista (un precio interior superior a los costes medios iniciales hace de la entrada una estrategia dominante, y la protección se retira cuando los costes medios caen por debajo del precio internacional).

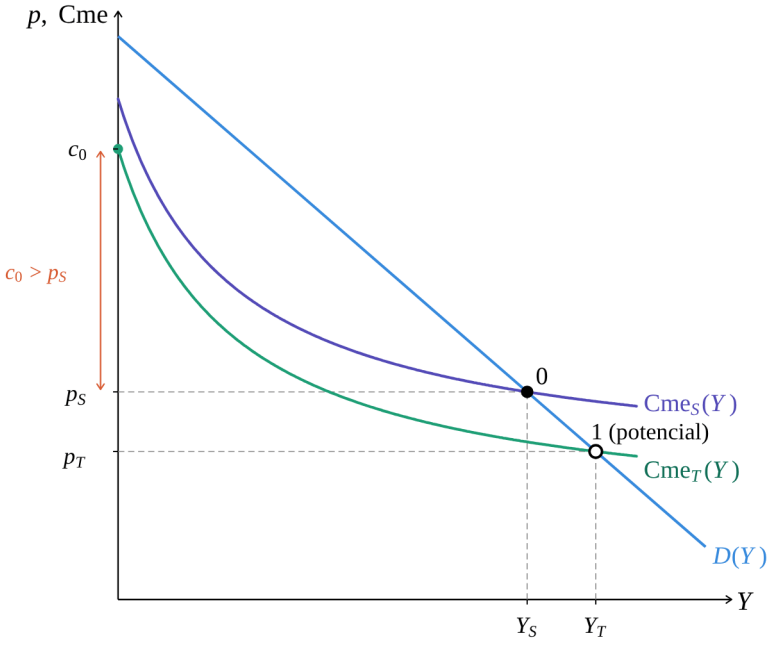

Figura B8.4: Industria naciente con economías de escala externas. Suiza abastece el mercado mundial al precio p_S (punto 0) aunque Tailandia tiene costes medios menores, porque el coste de la primera unidad tailandesa, c_0, supera p_S.

Desarrollo gráfico: los efectos de un arancel se representan en el plano bien X-bien Y (figura B8.5):

1. *Frontera de posibilidades de producción inicial*, FPP_0: recoge las combinaciones máximas de X e Y que el país puede producir antes de proteger la industria, y es cóncava, porque el coste de oportunidad de cada bien crece con su producción.

2. *Recta de la relación real de intercambio*, RRI: la internacional, que el país toma como dada. Su pendiente, en valor absoluto, es el precio relativo de X en términos de Y en el mercado mundial.

3. *Curvas de indiferencia social*, CIS_0, CIS_1 y CIS_2: representan un bienestar tanto mayor cuanto más alejadas del origen.

4. *Equilibrio inicial*: se parte de una situación de libre comercio, con la producción en P^0 y el consumo en C^0, sobre la CIS_0. En esta situación, existe un exceso de oferta del bien X que se exporta y un exceso de demanda del bien Y que se cubre con importaciones (p. ej. los relojes en el caso tailandés anterior). La diferencia esencial con la representación habitual es que el país produce una gran cantidad del bien X y una escasa cantidad del bien Y, si bien este último es susceptible de ser exportado en el futuro.

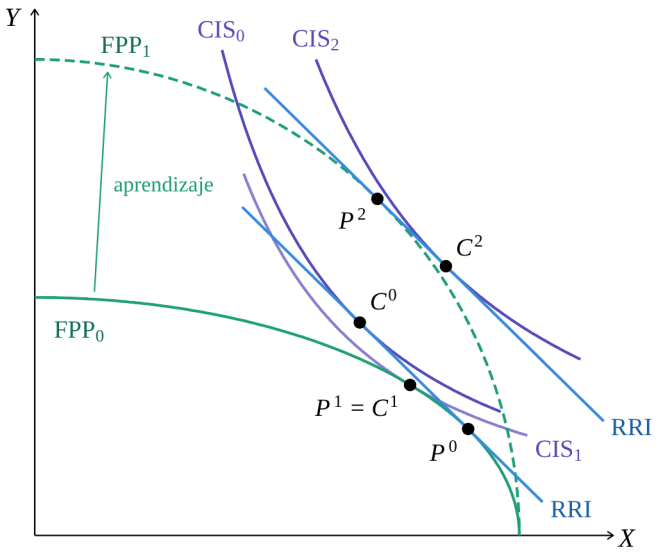

Figura B8.5: Argumento de la industria naciente. La protección temporal permite el desplazamiento de la frontera de posibilidades de producción por aprendizaje.

Estática comparativa: la protección temporal de la industria del bien Y se desarrolla en tres etapas:

1. *Arancel*: un arancel prohibitivo sobre el bien Y eleva el precio relativo nacional de este y hace que producción y consumo coincidan en un mismo punto, $P^1 = C^1$, sobre la CIS_1, es decir, se regresa a la autarquía. Esto provoca un aumento de la producción nacional del bien Y, pero reduce el bienestar total del país, pues la curva de indiferencia alcanzada está más próxima al origen.

2. *Desarrollo*: al amparo del arancel, que reserva a los productores nacionales la demanda interior, la industria del bien Y alcanza la escala con la que su coste medio queda por debajo del precio internacional (p_S en la figura B8.4), y crece gracias al progreso tecnológico que generan el efecto aprendizaje y las economías de escala dinámicas. Así, el país gana competitividad en la producción de ese bien, lo cual se traduce en una expansión sesgada hacia el bien Y de la frontera de posibilidades de producción, de FPP_0 a FPP_1.

3. *Apertura*: una vez desarrollada la industria nacional, el país vuelve a abrirse al comercio internacional. Su patrón de comercio se invierte y el país exporta ahora el bien Y (producción en P^2 y consumo en C^2). El bienestar, además, supera el de partida y alcanza la CIS_2.

Bienestar: sin intervención, el accidente histórico de la figura B8.4 produce una especialización "perversa" y una doble pérdida de eficiencia. Por un lado, existen ganancias potenciales de bienestar global desaprovechadas, ya que el mercado internacional está abastecido por Suiza, lo que conlleva una menor cantidad de relojes disponible a un mayor precio. Por otro, existen ganancias potenciales de bienestar para Tailandia, ya que, al no ser capaz de competir, no puede desarrollar su industria relojera, con el consecuente desaprovechamiento de puestos de trabajo y de participación en la renta mundial. La protección temporal (figura B8.5) busca recuperar esas ganancias a cambio de la pérdida transitoria de la etapa del arancel.

Extensiones: como se expone en el tema B07, apartado 1.2.2, el argumento solo es válido si la industria supera los criterios de Mill y de Bastable, es decir, si llega a ser competitiva sin apoyo y si las ganancias descontadas superan los costes del arancel. De ahí que lo prioritario sea escoger industrias cuya demanda tenga amplio margen de crecimiento. Si además la industria tiene abundantes eslabonamientos hacia delante y hacia atrás, genera efectos desbordamiento o posee un valor estratégico que trasciende lo económico (p. ej. la aeronáutica), puede reportar beneficios añadidos.

Conclusiones:

▪ *Aspectos positivos*: el argumento de la industria naciente ha persuadido a muchos Gobiernos. Así, Estados Unidos mantuvo elevados aranceles sobre las manufacturas durante buena parte del s. XIX y Alemania adoptó una protección más moderada a partir del giro proteccionista de Bismarck (1879), lo que favoreció su industrialización. Tras la Segunda Guerra Mundial, Japón mantuvo amplios controles de importaciones hasta los años setenta. También suele citarse el arancel Smoot-Hawley (1930) de Estados Unidos, aunque es más bien un ejemplo de proteccionismo defensivo generalizado (agrícola e industrial), no de protección de industrias nacientes. Las represalias que provocó contribuyeron al colapso del comercio durante la Gran Depresión.

▪ *Limitaciones*: existen diversas críticas al argumento, tanto teóricas como empíricas:

- *Instrumentos alternativos*: incluso si se cumplen los criterios de Mill y de Bastable, podría conseguirse el desarrollo de la industria naciente a través de una política no arancelaria (p. ej. una subvención que disminuyese los costes de la industria sin gravar al consumidor).

- *Descuento intertemporal*: el criterio de Bastable exige la utilización de un factor de descuento intertemporal social, ya que es necesario descontar las ganancias del libre comercio posteriores al desarrollo de la industria. Determinar dicho factor puede exigir, en la práctica, elevados costes informacionales.

- *Fallo de mercado*: James Meade (1955b) y, posteriormente, Robert Baldwin (1969) muestran que las economías de escala internas o dinámicas[11] no son condición suficiente para la protección. Hace falta algún fallo de mercado adicional (p. ej. imperfecciones en el mercado de capitales que dificulten el endeudamiento inicial, imposibilidad de apropiarse de los desbordamientos de conocimiento o imperfecciones en el mercado de trabajo). De hecho, si se cumple el criterio de Bastable, los inversores privados racionales descontarán las ganancias futuras e invertirán en la industria naciente aunque no esté protegida y aunque incurran inicialmente en pérdidas:

  - *Economías de escala internas*: los productores, sabiendo que en el futuro obtendrán beneficios, están dispuestos a endeudarse para lograr el tamaño de planta óptimo.
  - *Economías de escala dinámicas*: están dispuestos a soportar pérdidas hasta alcanzar la experiencia necesaria.

- *Ámbito empírico*: no siempre es buena idea intentar desplazarse hacia industrias que tendrán ventaja comparativa en el futuro, ya que es condición necesaria que la protección del sector propicie efectivamente su competitividad. Si el país aún no dispone de una masa crítica de capital humano y productivo cualificado, proteger la industria resulta contraproducente, pues la protección no crea por sí sola esa capacidad y solo prolonga la supervivencia de empresas ineficientes.

[11] Este argumento no opera, sin embargo, con economías de escala externas, donde los costes medios no descienden porque una empresa aumente su tamaño o gane experiencia con el tiempo, sino por la entrada de más empresas. El descenso de costes es, por tanto, un efecto externo del que ninguna empresa puede apropiarse, por bien que funcionen los mercados de capitales.

2.3.3. Argumento centro-periferia y deterioro de la relación real de intercambio (Prebisch, 1949; Singer, 1950)

Origen: Raúl Prebisch (1949), en la Comisión Económica para América Latina (CEPAL), y Hans Singer (1950), en la sede de las Naciones Unidas, formularon casi a la vez el *argumento centro-periferia*.

Idea: la relación real de intercambio de la periferia, exportadora de productos primarios, tiende a deteriorarse frente al centro industrial.

Razonamiento: el deterioro se explica por la baja elasticidad-renta de la demanda de esos productos (ley de Engel) y por la asimetría en el reparto de las ganancias de productividad (tema B28, apartado 2.1.3). La recomendación defensiva derivada, industrializarse mediante protección en lugar de especializarse en productos primarios según la ventaja comparativa estática, fundamentó la industrialización por sustitución de importaciones.

Conclusiones:

- *Aspectos positivos*: la intuición de Prebisch y Singer de que la composición de la especialización condiciona el desarrollo reaparece en la ley de Thirlwall (tema B28, apartado 2.1.3) y en el modelo de Grossman y Helpman (tema B09, apartado 3.1.3).

- *Limitaciones*: su apoyo empírico es mixto, pues las series largas muestran un deterioro débil de los precios relativos de las materias primas, interrumpido por rupturas estructurales y superciclos alcistas, lo que debilita el argumento como justificación permanente de la protección.

2.3.4. Protección de las importaciones como promoción de las exportaciones (Krugman, 1984)

Origen: Paul Krugman (1984) formalizó un argumento frecuente en el debate sobre la industria japonesa, según el cual proteger el mercado interior puede hacer más competitiva a la empresa nacional también fuera de él.

Idea: en oligopolio, con mercados segmentados y coste marginal decreciente en la producción total, reservar el mercado interior a la empresa nacional le permite producir más, bajar su coste marginal y ganar cuota en todos los mercados, incluidos los de exportación.

Supuestos:

- *Entorno*: una empresa nacional y otra extranjera compiten a la Cournot en varios mercados segmentados, entre ellos el nacional.

- *Tecnología*: el coste marginal de cada empresa decrece con su producción total, por economías de escala estáticas, por aprendizaje o porque el gasto en I+D se reparte entre más unidades.

Razonamiento: un arancel o una cuota en el mercado nacional encarece o limita las ventas de la empresa extranjera en ese mercado, cuya función de reacción se desplaza hacia el origen, de modo que la empresa nacional vende más en casa. Al vender más en casa, produce más en total, su coste marginal baja y se vuelve más agresiva en los demás mercados. A la extranjera le ocurre lo contrario, pues pierde ventas, su coste marginal sube y la empresa extranjera retrocede en todos ellos. La protección de las importaciones acaba así elevando las exportaciones.

Conclusiones:

- *Aspectos positivos*: el modelo da a la vertiente defensiva un mecanismo propio, el vínculo entre mercados a través del coste marginal decreciente, y explica por qué un mercado interior grande y protegido puede convertirse en una ventaja competitiva internacional.

- *Limitaciones*: el resultado depende de la forma de competencia y de que el coste marginal sea decreciente. Además, la ganancia para el país no está garantizada, porque los consumidores pagan precios más altos en el mercado protegido, y la represalia del otro país la anula.

2.3.5. Conclusiones

Aspectos positivos: entre 1950 y 1975, principalmente, los países en desarrollo llevaron a la práctica el argumento de la industria naciente con la *industrialización por sustitución de importaciones* (ISI), que tuvo gran difusión en Latinoamérica. Mediante la protección del mercado interior, la ISI perseguía dos fines, el de dar a ciertas industrias margen para lograr economías de escala y de aprendizaje y el de desviar la demanda nacional de manufacturas de los productos extranjeros hacia la producción propia.

Limitaciones: la ISI presentó resultados desiguales. En muchos países latinoamericanos, la protección sostenida generó ineficiencias, comportamientos de búsqueda de rentas y dependencia del sector público, sin lograr la competitividad internacional esperada. El contraste con la orientación exportadora de los tigres asiáticos se convirtió en uno de los argumentos empíricos clásicos contra la protección prolongada.

2.4. Vertiente de la I+D

Por último, dada la importancia de los efectos dinámicos para defender la política comercial estratégica, este apartado trata la vertiente que busca aprovechar la I+D. Su modelo estratégico es el de Spencer y Brander (1983), mientras que el de Romer (1990) fundamenta esa vertiente en una externalidad y no en el desplazamiento de rentas.

2.4.1. La I+D como compromiso estratégico (Spencer y Brander, 1983)

Origen: Barbara Spencer y James Brander (1983) trasladaron a la I+D el argumento del desplazamiento de rentas.

Idea: si la I+D se decide antes que la producción y reduce el coste marginal, sirve como compromiso para ganar cuota frente a la rival, y una subvención pública puede reforzar ese compromiso.

Supuestos: una empresa nacional y otra extranjera compiten a la Cournot en un tercer mercado, en un juego en dos etapas. Primero eligen su gasto en I+D, que reduce su coste marginal, y después sus cantidades.

Razonamiento: el juego se resuelve hacia atrás. Como la I+D de cada empresa desplaza su función de reacción en la segunda etapa y reduce la producción de la rival, las empresas invierten en I+D más de lo que minimizaría el coste de producir la cantidad elegida. De ahí se siguen dos resultados de política:

- *Solo subvención a la I+D*: si el Gobierno no puede subvencionar las exportaciones, una subvención a la I+D desplaza rentas hacia la empresa nacional y eleva el bienestar del país, como la subvención a la exportación de Brander y Spencer (1985).

- *Con subvención a la exportación*: si dispone de ella, lo óptimo es subvencionar la exportación y gravar la I+D, porque la subvención a la exportación logra el compromiso directamente y la sobreinversión estratégica en I+D despilfarra recursos.

Conclusiones:

- *Aspectos positivos*: el modelo muestra que la I+D tiene un valor estratégico, además del productivo, y explica por qué los Gobiernos compiten en subvenciones a la investigación en sectores como el aeronáutico.

- *Limitaciones*: comparte las debilidades del desplazamiento de rentas, pues depende de la competencia a la Cournot, exige información que el Gobierno no suele tener y provoca represalias que anulan la ganancia.

2.4.2. Externalidades positivas: modelo de crecimiento con variedades de *inputs* (Romer, 1990)

> El modelo de aumento del número de variedades de *inputs* (Romer, 1990) y su extensión a la economía abierta (Rivera-Batiz y Romer, 1991) se desarrollan en el tema B09, apartado 3.1.2. Aquí interesa únicamente su implicación para la política comercial.

Origen: Paul Romer (1990) explica como resultado deliberado de la I+D, movida por incentivos, el progreso técnico que el modelo de Solow toma como exógeno.

Idea: el modelo de Romer establece que la investigación solo se produce si existe competencia imperfecta que permita apropiarse de sus frutos, y que aun así el equilibrio descentralizado innova por debajo del óptimo social. La razón es que el inventor captura los beneficios de su patente, pero no el excedente que su descubrimiento genera para los consumidores ni el efecto que produce sobre la productividad de la investigación ajena. Existe, por tanto, una brecha entre rendimiento privado y social que justifica la intervención pública.

Razonamiento (aplicación a la política comercial): si la política comercial estratégica logra que empresas de alta tecnología se asienten y operen en el país, el beneficio no se agota en ellas. Los desbordamientos de conocimiento alcanzan al conjunto del tejido productivo, y alrededor de esas empresas tienden a localizarse clústeres industriales, con proveedores y actividades complementarias que refuerzan el efecto mediante encadenamientos.

Conclusiones:

- *Aspectos positivos*: la brecha entre el rendimiento privado y el social de la investigación da al apoyo público a la I+D un fundamento de eficiencia que no depende del desplazamiento de rentas.

- *Limitaciones*: la externalidad justifica subvencionar la investigación, no proteger el sector, que son cosas distintas. Además, el razonamiento solo sostiene una política comercial si el desbordamiento es específico de que la actividad se localice en el territorio nacional. En cambio, si el conocimiento se difunde internacionalmente, la subvención genera un beneficio global financiado por el contribuyente propio, lo que constituye la objeción principal que la literatura opone a la política comercial estratégica de la I+D.

2.4.3. Costes fijos de I+D y oligopolio natural (Shaked y Sutton, 1982; 1983)

Los oligopolios naturales pueden estudiarse con el modelo de diferenciación vertical de Shaked y Sutton. Cuando la mejora de la calidad descansa sobre costes fijos de I+D, el número de empresas que sobrevive en el mercado está acotado por un máximo independiente del tamaño de este, lo que concentra la estructura de mercado y genera beneficios extraordinarios persistentes. La política estratégica de I+D puede, en este contexto, tratar de influir sobre qué país acoge a las empresas supervivientes del duopolio natural internacional.

> El modelo de Shaked-Sutton y la propiedad de finitud se desarrollan en el tema B06.

2.4.4. Regulación

Al igual que ocurre con el monopolio natural, en la medida en que el duopolio natural sea sostenible, conviene regularlo, con instrumentos análogos a los empleados en la regulación de los monopolios naturales (control de precios, obligaciones de acceso o supervisión de la conducta).

> Sobre la regulación de los mercados no competitivos, véase el tema A20.

2.5. Conclusiones

Pese a sus numerosas virtudes teóricas, contra la política comercial estratégica se han formulado varias críticas:

- *Sector público*: hay que tener en cuenta la financiación de muchas de las políticas y, además, todos los posibles fallos del sector público (p. ej. búsqueda de rentas o influencia de los grupos de presión).

- *Eficiencia económica*: proteger a la industria naciente no garantiza que el país llegue a desarrollar su ventaja comparativa, los instrumentos considerados son distorsionantes y muchas de las políticas dan lugar a un empobrecimiento del vecino.

- *Sostenibilidad*: las políticas de este tipo generan diferencias artificiales de productividad, difícilmente sostenibles a largo plazo.

- *Discrecionalidad*: puede dar lugar a fallos de coordinación en el ámbito internacional. Entre otras normas, el comercio internacional se rige por los acuerdos multilaterales que se pactan en las rondas negociadoras de la OMC, los cuales tienden a limitar las actuaciones de política comercial estratégica.

Justamente a causa de estas críticas, los países (en particular, los desarrollados) optan preferentemente por políticas de promoción exterior.

3. La política de promoción exterior: objetivos, instrumentos y efectos

La política de promoción exterior se estudia en sus objetivos, instrumentos e instituciones y en sus efectos.

3.1. Objetivos, instrumentos e instituciones

La **política de promoción exterior** pone al alcance de las empresas herramientas para internacionalizarse, sin condicionar directamente la ayuda a lo que cada una exporte.

3.1.1. Objetivos

Objetivo microeconómico: en cada empresa, la política de promoción exterior intenta reducir los costes de la internacionalización y aumentar sus beneficios, de modo que aumente el valor actual neto (VAN) de la inversión o de la actividad exportadora hacia el extranjero. Con información imperfecta sobre el mercado de destino, ese VAN es

$$\text{VAN} = -(A_0 + C_I) + \sum_{t=1}^{T} \frac{\mathbb{E}[Q_t]}{(1+k+\rho)^t}$$

donde A_0 es la inversión inicial, C_I el coste de reunir la información necesaria para exportar (estudios de mercado, prospección, requisitos legales del destino), $\mathbb{E}[Q_t]$ el valor esperado del flujo de caja que genera el proyecto en el año t y T su horizonte. La tasa de descuento o coste del capital es k, y ρ, la prima de riesgo que añade la incertidumbre sobre el mercado exterior (tema B02, apartado 2.3).

La promoción reduce el coste de la información (C_I) y la prima de riesgo (ρ), porque atenúa las barreras informacionales del apartado 1.2.2, y eleva los flujos de caja esperados ($\mathbb{E}[Q_t]$). Con los instrumentos financieros rebaja, además, el desembolso inicial (A_0) y el coste del capital (k).

Objetivo macroeconómico: la política busca la *internacionalización*, entendida aquí en sentido amplio, del conjunto de la economía. En lugar de perseguir éxitos puntuales, la ayuda se concentra en los primeros pasos de cada empresa en el exterior, con la meta de que la empresa arraigue de forma duradera en esos mercados y pueda después prescindir del respaldo público.

Por el lado del gasto, el PIB es $Y = C + I + G + (X - M)$, donde C es el consumo privado, I la inversión, G el consumo público y X y M las exportaciones y las importaciones de bienes y servicios. La promoción actúa sobre X, y su impulso eleva, *ceteris paribus*, la demanda agregada y el saldo exterior, $X - M$, cuya relación con la cuenta corriente y con el ahorro y la inversión nacionales se estudia en el tema B11, apartado 2.1.1.

3.1.2. Instrumentos

Las administraciones de los distintos países han traducido esa justificación teórica en instrumentos concretos, que aquí se agrupan en cuatro tipos.

Instrumentos que facilitan el acceso a la información: comprenden el suministro público de información útil para las empresas que quieren acceder a mercados exteriores (p. ej. estudios

de mercado, servicios de alertas de oportunidades de negocio y canales de distribución de información). Destaca, además, el "acceso a redes", puesto que el sector público está en una posición ventajosa para facilitar contactos en los mercados extranjeros gracias a sus redes de contactos en los países de destino.

Instrumentos de formación: el sector público puede ofrecer también recursos de formación y experiencia que facilitan la internacionalización (p. ej. los másteres y becas de ICEX España Exportación e Inversiones, seminarios o cursos de formación). Su finalidad es que la empresa conozca su potencial en el exterior y gane las competencias que le permitan medirse con rivales internacionales.

Instrumentos financieros: el sector público puede reducir la barrera financiera a través de líneas de crédito a la internacionalización (p. ej. las líneas del Instituto de Crédito Oficial).

> Sobre el crédito oficial a la exportación y los instrumentos españoles de apoyo financiero a la internacionalización, véase el tema B34, apartado 3.4 (Consenso OCDE).

Instrumentos de promoción de la marca o imagen de los productos nacionales: buscan incrementar el valor añadido de los productos exportados haciendo que sean más valorados por los consumidores extranjeros. Entre ellos destacan:

- *Ferias comerciales*: han adquirido una importancia creciente.

- *Planes integrales de desarrollo de mercados*: persiguen estrechar los lazos comerciales con los países que se consideran prioritarios (p. ej. el ICEX ha mantenido planes de este tipo con Brasil, China, India, Rusia, EE. UU. o los países del Golfo).

- *Marca país*: el objetivo es mejorar la imagen del país para que no quede sometida a círculos viciosos. Si la marca país no es favorable, las empresas tratarán de desmarcarse de ella (que no se les asocie), porque no actúa como una buena carta de presentación. Como resultado, los éxitos de las empresas en el extranjero no mejoran la imagen del país, porque no se asocian a la marca, y se genera un círculo vicioso que el sector público puede tratar de romper. En España, esta función correspondió al Alto Comisionado del Gobierno para la Marca España, sustituido en 2018 por la Secretaría de Estado de la España Global, que fue suprimida en 2021 (Real Decreto 585/2021). A ello se suman las acciones del ICEX para mejorar la imagen de los bienes y servicios españoles, como la marca Foods and Wines from Spain en mercados como EE. UU.

- *Misiones comerciales*: tanto directas como inversas (misiones de inversores, en las que se trae a inversores extranjeros para que conozcan las empresas nacionales).

- *Organismos autonómicos de promoción*: tratan de impulsar la imagen de su comunidad autónoma mediante campañas específicas que ponen el énfasis en sus rasgos distintivos.

La política de promoción de exportaciones en España: la Estrategia de Internacionalización de la Economía Española 2017-2027: además de estos instrumentos, los países diseñan estrategias de internacionalización. En el caso español, la *Estrategia de Internacionalización de la Economía Española 2017-2027* fija las principales líneas de acción para la internacionalización de la economía en su conjunto con un horizonte temporal de largo plazo, y se despliega mediante planes de acción bienales. Sus objetivos son asegurar que la contribución positiva del sector exterior al crecimiento económico se convierta en un fenómeno estructural y revierta en creación de empleo, y apoyar a los exportadores y atraer inversión hacia España. La Estrategia actúa en colaboración con otras instituciones y con el sector privado, con un enfoque principal hacia la pyme.

En su versión original, la Estrategia se articulaba en torno a seis grandes ejes de actuación. Los dos primeros eran el apoyo a la internacionalización adaptado al perfil de la empresa y la incorporación de la innovación, la tecnología, la marca y la digitalización. Les seguían el desarrollo del capital humano para la internacionalización y el aprovechamiento de las oportunidades de la política comercial común y de las instituciones financieras multilaterales. Los dos últimos eran la captación y consolidación de inversión extranjera de alto valor añadido y el refuerzo de la coordinación y complementariedad de todos los actores.

Tras su evaluación intermedia, la Estrategia fue actualizada y reagrupó su actuación en cinco líneas. Las dos primeras son la adaptación constante e innovadora a las necesidades de las empresas y el talento, la marca y la innovación como factores de competitividad. Las otras tres atienden a los foros internacionales y la política comercial europea, a la atracción de inversión extranjera de alto valor añadido y a la inteligencia económica. Las cinco líneas se guían por cinco principios rectores, la coordinación, la sostenibilidad, la digitalización, la igualdad de género y la seguridad económica.

3.1.3. Instituciones dedicadas a la promoción exterior

Expuestos los objetivos y los instrumentos de la política de promoción de exportaciones, queda por ver qué instituciones la aplican en la práctica. El edificio institucional para la promoción de la exportación cambia en función del país:

- *España*: la principal institución es ICEX España Exportación e Inversiones, entidad pública empresarial cuyo diseño original se inspiró en el Istituto nazionale per il Commercio Estero (ICE) italiano.

- *Reino Unido, Alemania y Estados Unidos*: en el Reino Unido, estas funciones correspondieron a UK Trade and Investment, integrada desde 2016 en el departamento ministerial de comercio internacional (el actual Department for Business and Trade). En Alemania desempeñan un papel central las Cámaras de Comercio en el exterior (AHK). En Estados Unidos, finalmente, existe una estructura descentralizada de una veintena de organizaciones encargadas de la promoción, directa o indirecta, de las actividades de exportación de las empresas estadounidenses.

3.2. Efectos de la promoción de exportaciones: el modelo de Melitz (2003)

Los modelos de oligopolio explican los efectos de la política comercial estratégica (apartado 2.2), mientras que, para los de la promoción exterior, el marco de referencia es el modelo de heterogeneidad de empresas de Melitz (2003). En él, la actividad exportadora exige asumir unos costes fijos de entrada en los mercados exteriores (prospección, adaptación del producto, redes de distribución), de modo que solo las empresas con una productividad superior a un determinado umbral exportan. Los instrumentos de la promoción exterior (información, formación, redes de contactos, financiación) actúan precisamente reduciendo esos costes fijos y, con ello, el umbral de productividad exportadora. Así, empresas que antes quedaban fuera de los mercados exteriores pasan a poder exportar, y se amplía el margen extensivo del comercio.

> El modelo de heterogeneidad de empresas se desarrolla por completo en el tema B06.

Conclusiones:

- *Aspectos positivos*: el modelo de Melitz es el marco teórico moderno para analizar los efectos del apoyo horizontal a la internacionalización. Predice qué empresas responden a la promoción, las situadas cerca del umbral de exportación (con frecuencia, pymes), y que su efecto se concentra en el margen extensivo del comercio.

- *Limitaciones*: la rebaja de los costes fijos de exportar eleva también el umbral de supervivencia en el mercado interior, porque los nuevos exportadores intensifican la competencia. Por ello, la promoción expulsa a las empresas menos productivas que solo venden en el país, y la ganancia agregada tiene perdedores identificables. Además, el modelo trata esos costes como un parámetro exógeno, sin explicar qué parte de ellos puede reducir el sector público, de modo que la eficacia de cada instrumento es una cuestión empírica.

Conclusiones

La política comercial moderna se aparta del *laissez faire* tanto por la vía estratégica como por la de la promoción exterior. Por un lado, la política comercial estratégica se justifica en oligopolios internacionales con interdependencia estratégica y beneficios extraordinarios. En su vertiente ofensiva, una subvención a la exportación desplaza rentas hacia la empresa nacional si compite a la Cournot (Brander y Spencer, 1985), mientras que el instrumento óptimo pasa a ser un impuesto si compite a la Bertrand (Eaton y Grossman, 1986). Con todo, la represalia y la libre entrada debilitan el argumento. En su vertiente defensiva, las economías de escala externas y la industria naciente solo justifican aranceles si se cumplen los criterios de Mill y Bastable. Por otro lado, la promoción exterior responde a fallos de mercado que encarecen la internacionalización, y en el modelo de Melitz actúa rebajando el umbral de productividad para exportar. En ambos casos, sin embargo, el fallo de mercado es condición necesaria pero no suficiente, porque el sector público también falla, y por eso los países desarrollados prefieren la promoción.

La política comercial estratégica aplica al comercio los modelos de oligopolio, de modo que el signo del instrumento óptimo depende de si se compite en cantidades o en precios (tema A19). Asimismo, amplía el *dumping* recíproco de la nueva teoría (tema B06), de la que procede también la heterogeneidad de empresas en que se apoya la promoción exterior. Prolonga, además, el análisis de los instrumentos clásicos de protección, entre cuyos argumentos figura la industria naciente (tema B07). En todo caso, su justificación exige comparar fallos de mercado y fallos del sector público (tema A23). Por último, los rendimientos crecientes que sustentan la industria naciente reaparecen en el modelo de Romer (tema B09) y en la elección entre la sustitución de importaciones y la orientación exportadora (tema B28).

La evidencia reciente sobre la industria naciente ha renovado el debate sobre la política industrial. Juhász (2018) muestra que las regiones francesas más aisladas de la competencia británica durante el bloqueo napoleónico ampliaron más su capacidad de hilatura mecánica del algodón y la mantuvieron tras el fin del bloqueo. A partir de esta evidencia, Juhász, Lane y Rodrik (2024) sostienen que la comparación entre fallos de mercado y fallos del sector público vuelve a estar abierta. Sigue en discusión, además, la eficacia de la promoción exterior. Lederman, Olarreaga y Payton (2010), con datos de 103 países, encuentran que las agencias de promoción de exportaciones elevan en promedio las exportaciones, aunque con rendimientos fuertemente decrecientes.

B9. Comercio internacional y crecimiento

Comercio internacional y crecimiento económico. Especial referencia a los efectos del comercio sobre el crecimiento.

Introducción

> **Introducción común.** Los temas B05, B06 y B09 comparten una introducción, que se desarrolla al comienzo del tema **B05**. Plantea las tres preguntas de la teoría del comercio internacional (por qué comercian los países, qué intercambian y cómo les afecta) y sus tres etapas, de la teoría tradicional a la nueva teoría y la novísima, que se pregunta por qué comercian las empresas.

La relación entre el **comercio internacional** y el **crecimiento económico** funciona en los dos sentidos. Crecer cambia lo que un país produce y demanda, y con ello su comercio y su relación real de intercambio. Comerciar, a su vez, puede elevar la renta y su ritmo de crecimiento.

En el primer sentido, el efecto depende del **sesgo del crecimiento**, procomercial si amplía sobre todo la producción del bien exportable y anticomercial si favorece al que compite con las importaciones, y de si el país es lo bastante grande para alterar los precios mundiales. Bhagwati (1958) llevó el argumento a su extremo con el **crecimiento empobrecedor**, el caso en que un país grande pierde bienestar porque el abaratamiento de sus exportaciones pesa más que el aumento de su producción.

En el segundo sentido, la pregunta es si el comercio solo eleva el nivel de la renta o también su ritmo de crecimiento. La teoría neoclásica apenas ofrecía efectos de nivel, ligados a la especialización, mientras que los **modelos de crecimiento endógeno en economía abierta** (Rivera-Batiz y Romer, 1991; Grossman y Helpman, 1990) muestran que el comercio puede elevar la tasa de crecimiento, porque amplía el mercado de las innovaciones y permite aprovechar el conocimiento generado en otros países. La evidencia empírica, desde Frankel y Romer (1999), respalda una relación positiva entre apertura y renta, aunque la identificación de la causalidad sigue siendo difícil, porque los países más ricos también comercian más.

La cuestión ha recobrado actualidad con la fragmentación geoeconómica, es decir, con la reorganización del comercio en torno a bloques geopolíticos. Según el FMI (2023), una fragmentación grave del comercio mundial podría costar hasta un 7 % del PIB mundial a largo plazo, y las pérdidas recaerían sobre todo en las economías emergentes y de renta baja.

Este tema analiza, en los dos sentidos, cómo se relacionan el comercio exterior y el crecimiento:

- *Del crecimiento al comercio*: análisis estático de sus efectos sobre el volumen de comercio, los precios relativos internacionales y el bienestar, según su origen y el tamaño del país.

- *Del comercio al crecimiento*: las principales aportaciones del pensamiento económico, de los clásicos a la novísima teoría.

- *Análisis dinámico*: los efectos del comercio internacional sobre el crecimiento económico, con especial atención a los modelos de crecimiento endógeno en economía abierta y a la evidencia empírica.

1. Efectos del crecimiento económico sobre el comercio internacional

El **crecimiento económico**, entendido como el incremento de la renta nacional de un país, tiene un efecto total sobre el volumen de comercio que se puede descomponer en dos efectos, el efecto sobre el consumo y el efecto sobre la producción. El crecimiento, además, puede tener repercusiones sobre la relación real de intercambio (RRI) con el resto de países en función del tamaño del país. Por ello se considera primero un país pequeño y después un país grande. En ambos casos, el análisis es primero positivo (los efectos sobre el volumen de comercio) y después normativo (los efectos sobre el bienestar). El origen del crecimiento económico, es decir, si se debe a la acumulación de factores o al progreso tecnológico, se deja inicialmente de lado, y al final se examina si influye en las predicciones.

1.1. Supuestos

En cualquiera de los dos casos, a la hora de entender los efectos del crecimiento económico sobre el comercio internacional se parte de un modelo base simplificado de equilibrio general competitivo con un país, el cual, independientemente de su tamaño, cumple los siguientes supuestos.

Entorno: la economía es abierta, con competencia perfecta en todos los mercados y libre comercio sin restricciones ni costes de transporte.

Agentes: los individuos son propietarios de los factores de producción y de la tecnología (economía de tipo Robinson Crusoe). Sus preferencias vienen dadas por una función de utilidad continua, creciente, cuasicóncava y diferenciable.

Tecnología: el país produce dos bienes (X e Y) con dos factores de producción (K y L). El bien X es el bien exportable y el bien Y el bien importable, y ambos son bienes normales. Las funciones de producción presentan productividades marginales positivas pero decrecientes y rendimientos constantes a escala. Los factores son homogéneos y perfectamente divisibles, y su cantidad total está dada.

1.2. País pequeño

En el contexto de un país pequeño, los efectos se caracterizan porque la relación real de intercambio no cambia. Se distinguen el análisis positivo, de los efectos sobre el volumen de comercio, y el normativo, de los efectos sobre el bienestar.

1.2.1. Análisis positivo: efectos sobre el volumen de comercio

Los efectos del crecimiento económico sobre el volumen de comercio pueden representarse gráficamente en el plano de bienes (figura B9.1). Se parte de una situación de equilibrio en esta economía abierta en la cual las decisiones de consumo y producción están separadas. El crecimiento económico supone un desplazamiento hacia afuera de la frontera de posibilidades de producción (FPP) y un desplazamiento paralelo de la restricción de recursos de la economía, al aumentar la renta nacional[1]. Pueden trazarse rayos-vectores desde el origen que pasen por el equilibrio inicial de consumo, C^0, y por el equilibrio inicial de producción, P^0. Si se supone que los nuevos equilibrios se sitúan sobre esos rayos, los triángulos de comercio son semejantes (es

[1] Analíticamente, esto supone pasar de una restricción de recursos inicial, $\text{RN}_0 = p_X^0 X_0 + p_Y^0 Y_0$, a otra mayor, $\text{RN}_1 = p_X^0 X_1 + p_Y^0 Y_1 > \text{RN}_0$, con los mismos precios relativos (de ahí que el desplazamiento sea paralelo).

decir, las exportaciones y las importaciones han crecido proporcionalmente). Los triángulos que describen el conjunto presupuestario de la economía también lo son.

Esto significa que el incremento relativo del comercio es idéntico al incremento relativo de la renta nacional, y se dice que el crecimiento económico es **neutral** con el comercio internacional. La neutralidad se puede comprobar observando que la ratio de crecimiento de las exportaciones y la ratio de crecimiento de la renta nacional medida en unidades del bien exportable son iguales, y análogamente para las importaciones:

$$\frac{\Delta \text{Exp}}{\text{Exp}} = \frac{\Delta \text{RN}_X}{\text{RN}_X} \qquad \text{y} \qquad \frac{\Delta \text{Imp}}{\text{Imp}} = \frac{\Delta \text{RN}_Y}{\text{RN}_Y}$$

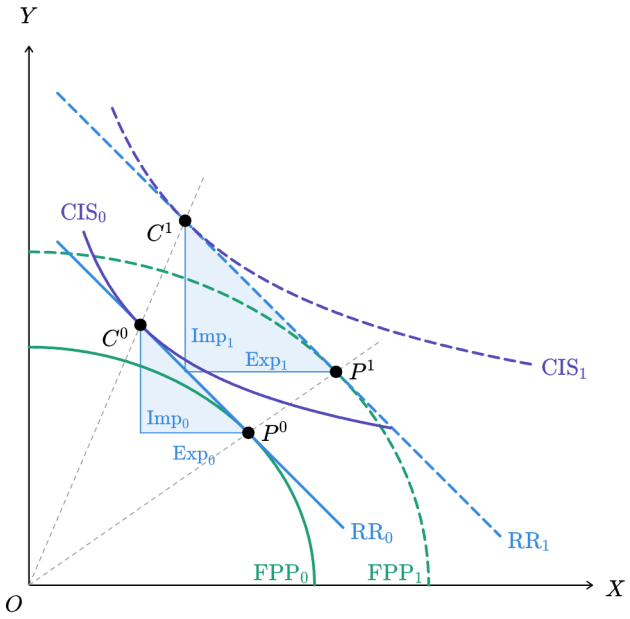

Figura B9.1: Crecimiento neutral en un país pequeño. Con precios constantes, la frontera se desplaza de FPP_0 a FPP_1 y la restricción de recursos, en paralelo, de RR_0 a RR_1, y los nuevos equilibrios de producción, P^1, y de consumo, C^1, quedan sobre los rayos que pasan por P^0 y C^0. Los triángulos de comercio (exportaciones de X e importaciones de Y) son semejantes, de modo que el comercio crece en la misma proporción que la renta nacional.

Sin embargo, hasta aquí se ha supuesto que tanto el consumo como la producción se desplazaban sobre una línea recta. El desplazamiento en línea recta no es condición necesaria para la neutralidad, ya que bastan sendas curvas cuyos puntos finales den lugar igualmente a triángulos de comercio y conjuntos presupuestarios semejantes. Cuando esto no ocurre, el crecimiento económico no es neutral con el comercio internacional y se dice que tiene un **sesgo**, cuyo tipo depende del efecto del crecimiento sobre el consumo y sobre la producción. Concretamente, se pueden dar cuatro casos:

1. *Sesgo anticomercial*: el incremento relativo del comercio internacional es menor que el incremento relativo de la renta nacional, es decir, $(\Delta\text{Com}/\text{Com}) < (\Delta\text{RN}/\text{RN})$.

2. *Sesgo ultra-anticomercial*: el crecimiento económico produce una caída absoluta del comercio internacional ($\Delta\text{Com} < 0$).

3. *Sesgo procomercial*: el incremento relativo del comercio internacional es mayor que el incremento relativo de la renta nacional, es decir, $(\Delta\text{Com}/\text{Com}) > (\Delta\text{RN}/\text{RN})$.

4. *Sesgo ultra-procomercial*: el crecimiento económico produce un incremento absoluto del comercio internacional mayor que el incremento absoluto de la renta nacional ($\Delta\text{Com} > \Delta\text{RN}$).

Efectos sobre el consumo: la relación entre el incremento de la renta nacional y el incremento de la demanda del bien importable Y determina el grado de autosuficiencia en el consumo de la economía. Se pueden dar cinco situaciones:

1. *Sesgo procomercial*: con el crecimiento, el consumo del bien importable aumenta más deprisa que la renta nacional, y, *ceteris paribus*, las importaciones absorben una fracción mayor de la renta.

2. *Sesgo ultra-procomercial*: el aumento del consumo del bien importable supera incluso el incremento absoluto de la renta, y el bien exportable pasa a consumirse menos[2].

3. *Sesgo anticomercial*: caso opuesto al primero, con un consumo del importable que crece a menor ritmo que la renta.

4. *Sesgo ultra-anticomercial*: el consumo del bien importable Y llega a reducirse en términos absolutos[3].

5. *Sesgo neutral*: en el caso de que la decisión de consumo se sitúe sobre el rayo-vector que pasa por el consumo inicial.

Gráficamente, en el plano bien X-bien Y (figura B9.2), esto se puede ilustrar observando dónde se sitúa la decisión de consumo en relación con la restricción de recursos:

1. *Restricción de recursos inicial*, RR_0: recoge las cestas al alcance de la economía dados la renta nacional y los precios relativos iniciales, y es una recta con pendiente $-p_X^0/p_Y^0$, el precio relativo inicial.

2. *Curva de indiferencia social*, CIS_0: reúne las cestas con un mismo bienestar social y es convexa, porque la función de utilidad es cuasicóncava. La dibujada es la más alta compatible con esa restricción.

3. *Equilibrio inicial*: la decisión de consumo viene determinada por el punto de tangencia entre la curva de indiferencia social y la restricción de recursos (punto 0), en el que se consumen X_0 e Y_0.

El crecimiento económico ($\Delta\text{RN} > 0$) provoca un desplazamiento en paralelo de la restricción de recursos, de RR_0 a RR_1, y, en función de dónde se sitúe la nueva tangencia con la curva de indiferencia social, se dará una situación u otra:

- *A la izquierda de la línea de neutralidad* (el rayo que pasa por 0 y N): el crecimiento habrá tenido un sesgo procomercial sobre el consumo.

- *A la derecha*: el sesgo será anticomercial.

- *Fuera del cono que describen los ejes trazados por el equilibrio inicial*: se estará en alguno de los casos extremos (ultra).

[2] Ello exige que X sea un bien inferior, lo que relaja el supuesto de que ambos bienes son normales.
[3] En este caso, el bien inferior es el importable Y (de nuevo, fuera del supuesto de normalidad).

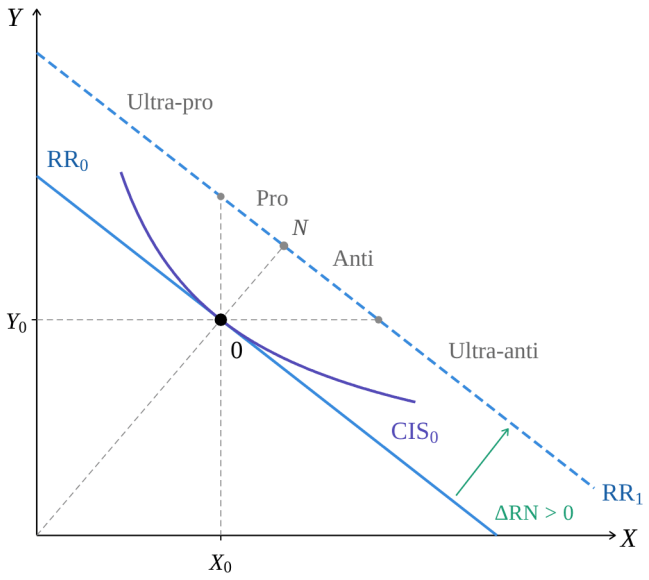

Figura B9.2: Sesgos del crecimiento en el consumo en un país pequeño. El crecimiento puede ser procomercial, neutral o anticomercial.

Sin embargo, el efecto que el incremento de la renta tenga sobre el consumo no es suficiente para determinar el sesgo comercial del crecimiento, puesto que es necesario saber qué efecto tiene también sobre la producción.

Efectos sobre la producción: el grado de autosuficiencia en la producción del bien importable viene dado, en este caso, por la relación entre el incremento de la renta nacional y el incremento de la producción del bien importable.

Gráficamente, los efectos sobre la producción se representan en el plano bien X-bien Y (figura B9.3):

1. *Frontera de posibilidades de producción inicial*, FPP_0: recoge las combinaciones máximas de X e Y que la economía puede producir con sus factores, y es cóncava, porque los costes de oportunidad son crecientes (tema B05, apartado 2.2).

2. *Recta de la RRI*, RRI: los precios internacionales, que el país pequeño toma como dados. Su pendiente es el precio relativo internacional, $-p_X/p_Y$.

3. *Equilibrio inicial*: la producción se sitúa en la tangencia entre ambas (punto 0), con X_0 e Y_0.

El tipo de efecto que el crecimiento económico tiene sobre la producción depende de la zona de la nueva frontera, FPP_1, en la que se sitúe el nuevo punto de producción, con dos diferencias respecto al caso del consumo. Primera, los nombres se invierten. Un crecimiento más que proporcional de la producción del importable es anticomercial, porque el país necesita importar menos. Segunda, el crecimiento desplaza la FPP y no la restricción de recursos, dado que el punto de producción está siempre sobre la frontera.

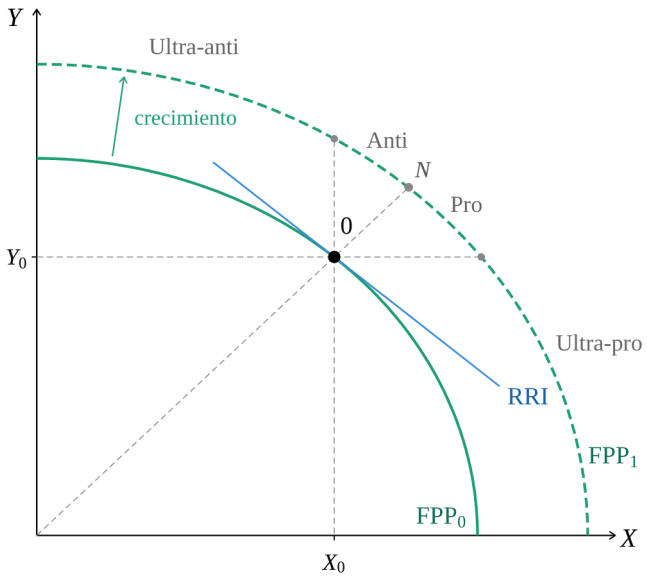

Figura B9.3: Sesgos del crecimiento en la producción en un país pequeño. El crecimiento puede ser procomercial, neutral o anticomercial.

Por último, si existe especialización completa (p. ej. con costes de oportunidad constantes, como en el modelo ricardiano), el país no produce nada del bien importable. En este caso, el aumento de la renta no tiene efecto sobre la composición de la producción y los efectos sobre el volumen de comercio vienen determinados exclusivamente por los efectos sobre el consumo. El sesgo del crecimiento sobre el comercio coincide así con el sesgo sobre el consumo, de modo que un crecimiento procomercial en el consumo es también procomercial en el comercio.

Los sesgos de la producción y del consumo no se combinan de forma sencilla en un sesgo del comercio internacional. Si ambos coinciden en su signo, se deduce de inmediato qué ocurre, en términos netos, con el volumen de comercio (p. ej. dos sesgos procomerciales arrojan un resultado global también procomercial). Cuando los signos se oponen, en cambio, el efecto global solo puede conocerse estimando las elasticidades-renta de demanda y de producción del importable.

1.2.2. Análisis normativo: efectos sobre el bienestar

Para un país pequeño, la variación de sus importaciones o exportaciones no altera el precio relativo internacional de los bienes (p_X/p_Y), que es la relación real de intercambio. Cualquiera que sea su tipo, el crecimiento económico aumenta entonces el bienestar social, porque eleva la curva de indiferencia social que el país alcanza.

1.3. País grande

Para un país grande se distinguen, como en el caso del país pequeño, el análisis positivo, ahora de los efectos sobre la relación real de intercambio, y el normativo, de los efectos sobre el bienestar, que pueden llegar a ser negativos.

1.3.1. Análisis positivo: efectos sobre la RRI

Si el país es grande, tendrá capacidad para alterar la relación real de intercambio. Así, además del efecto del crecimiento económico, se produce un efecto derivado del cambio en los precios relativos internacionales que dependerá del sesgo del crecimiento.

Gráficamente, en el plano bien X-bien Y (figura B9.4), el análisis se apoya en las curvas de oferta neta o demandas recíprocas. Las curvas muestran, para cada relación real de intercambio, la cantidad de exportaciones que cada país está dispuesto a ofrecer a cambio de las importaciones que desea obtener:

1. *Curva de oferta neta del país A*, CON_0^A: pasa por el origen y tiene dos tramos (la figura recoge solo el primero, en el que se producen los cortes).

- *Tramo inicial elástico*: al aumentar el precio relativo del bien exportado, el país ofrece más X para importar más Y.

- *Tramo final inelástico*: si el precio relativo de X es ya muy alto, el país no necesita exportar tanta cantidad de X para obtener las importaciones de Y que desea.

2. *Curva de oferta neta del país B*, CON^B: análoga a la de A, si bien para B las cantidades de X representan sus importaciones y las de Y sus exportaciones.

3. *Equilibrio*: el corte de ambas curvas (punto 0), en el que, para ambos países, lo que se ofrece en exportaciones coincide con lo que se demanda en importaciones. La RRI_0 es la pendiente de la recta que une el origen con ese punto.

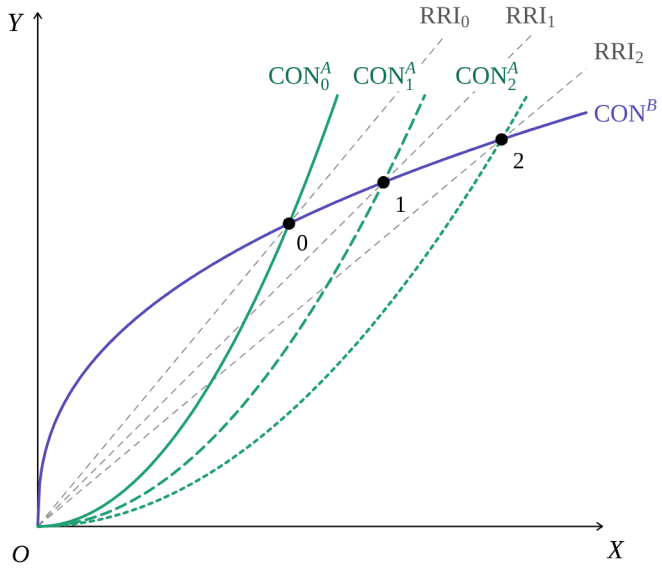

Figura B9.4: Crecimiento en un país grande. La curva de oferta neta se desplaza y la relación real de intercambio se deteriora.

Crecimiento económico: salvo en el caso ultra-anticomercial, eleva en términos absolutos la demanda de importaciones, sea cual sea su tipo. Por eso, cuando la curva de oferta neta de A pasa a CON_1^A o CON_2^A, los nuevos cruces con la de B (puntos 1 y 2) quedan a la derecha del

equilibrio inicial. Cuanto más procomercial es el crecimiento económico, más empeora la RRI, de RRI_0 a RRI_1 y RRI_2, puesto que su pendiente disminuye y se obtienen menos importaciones por cada unidad exportada.

Intuitivamente, un crecimiento más procomercial se traduce en más demanda de importaciones por parte de A. Como A es grande, el bien importable registra en el mundo un aumento apreciable de demanda y su precio internacional sube. Por la ley de Walras, A ofrece a la vez una mayor cantidad del bien exportable, cuyo precio cae. El encarecimiento del importable y el abaratamiento del exportable deterioran conjuntamente la relación real de intercambio.

Desde el punto de vista de la política económica se derivan dos implicaciones:

- *Orientación exportadora*: si un país grande orienta su política comercial a exportar, la oferta mundial del bien exportable supera lo que absorbe la demanda a los precios iniciales, lo que deteriora su relación real de intercambio.

- *Sustitución de importaciones*: si la política busca, en cambio, reemplazar importaciones con producción propia, el crecimiento adquiere un sesgo anticomercial que suaviza la caída de la RRI. Cuando ese sesgo es ultra-anticomercial, la RRI llega incluso a mejorar.

1.3.2. Análisis normativo: efectos sobre el bienestar (crecimiento neutral, aumentador o empobrecedor)

Hasta aquí se ha analizado en términos positivos el comportamiento de la relación real de intercambio. Ese comportamiento tiene, por extensión, efectos sobre el bienestar, que dependen de dos fuerzas contrapuestas, la ganancia de bienestar producida por el crecimiento económico y la pérdida derivada del deterioro de la relación real de intercambio. Se distinguen así dos casos, el primero de los cuales agrupa el crecimiento aumentador y el neutral en términos de bienestar:

1. *Crecimiento que aumenta o mantiene el bienestar*: la caída de la RRI es relativamente pequeña, de manera que el efecto del desplazamiento de la FPP domina y el bienestar del país aumenta. En el caso límite, ambos efectos se compensan y el bienestar se mantiene.

2. *Crecimiento empobrecedor*[4]: las pérdidas causadas por el deterioro de la RRI superan a las ganancias derivadas del crecimiento económico. Por lo tanto, la curva de indiferencia social final se sitúa más próxima al origen que la inicial.

El deterioro de la RRI es imprescindible para que haya crecimiento empobrecedor, pero no basta. La probabilidad de que la pérdida por ese deterioro pese más que las ganancias del crecimiento aumenta cuando se dan, además, otras condiciones:

- *Crecimiento débil*: la FPP se desplaza poco, de modo que el crecimiento, por sí solo, no permite acceder a una cesta de consumo mucho mejor que la de partida.

- *Elevada dependencia del comercio*: cuanto mayor es el peso del comercio en la economía, mayor es el impacto de un deterioro de la RRI.

- *Demanda exterior inelástica* del bien exportable (p. ej. alimentos o materias primas), de forma que el aumento de la oferta exportada provoca una caída más que proporcional de su precio relativo y reduce el valor de las exportaciones del país.

[4] La posibilidad de que crecer empobreciera a un país ya la había señalado Edgeworth (1894). Los primeros modelos formales sobre cómo el crecimiento altera la relación real de intercambio y el bienestar se deben a Johnson (1955) y a Bhagwati (1958), autor del término. Johnson (1967) mostró después que, en presencia de aranceles, el crecimiento puede empobrecer incluso a un país pequeño.

▪ *Ausencia de arancel óptimo*: el país no explota su poder de mercado mediante el arancel óptimo, que le permitiría contrarrestar el deterioro.

▪ *Sesgo muy procomercial* del crecimiento económico, pues cuanto más procomercial es el crecimiento, más se deteriora la RRI.

Gráficamente, en el plano bien X-bien Y (figura B9.5), se puede ilustrar un caso de crecimiento empobrecedor:

1. *Frontera de posibilidades de producción inicial*, FPP_0: recoge las combinaciones máximas de X e Y que el país puede producir y es cóncava, como en la figura B9.3.

2. *Recta de la RRI inicial*, RRI_0: la relación real de intercambio de partida, una recta cuya pendiente es el precio relativo internacional inicial.

3. *Curva de indiferencia social inicial*, CIS_0: reúne las cestas con el bienestar social alcanzado antes del crecimiento.

4. *Equilibrio inicial*: con la FPP y la RRI iniciales, la producción se sitúa en P^0 y el consumo en C^0.

Un crecimiento procomercial amplía la FPP con un sesgo hacia el sector exportador, de FPP_0 a FPP_1[5]. Como el país es grande, la RRI empeora y pasa de RRI_0 a RRI_1, con una pendiente menor en valor absoluto en el plano (X, Y). La producción se traslada a P^1 y el consumo a C^1, y la curva de indiferencia alcanzada, CIS_1, queda más cerca del origen que la inicial, con lo que el crecimiento resulta empobrecedor.

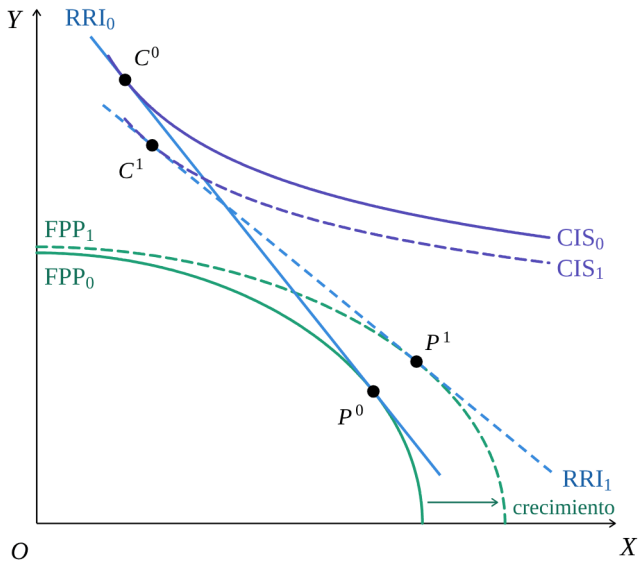

Figura B9.5: Crecimiento empobrecedor (Bhagwati, 1958). El deterioro de la relación real de intercambio más que compensa la expansión productiva.

[5]Puede deberse a un aumento de la dotación del factor en el que es intensivo el bien exportable o a progreso técnico neutral, o ahorrador de ese mismo factor, en el sector exportador (véase el apartado 1.4).

El sesgo del crecimiento se define, no obstante, a los precios relativos iniciales. Una vez incorporada la variación de la RRI, la producción final se desplaza a lo largo de la nueva FPP de vuelta hacia el bien importable, por lo que la asignación finalmente observada puede diferir del sesgo definido a precios constantes.

1.4. Efectos en función de la causa del crecimiento económico

Hasta ahora no se ha tenido en cuenta el origen del crecimiento económico al evaluar sus efectos sobre el comercio internacional. Se distinguen dos orígenes, la acumulación de factores y el progreso tecnológico.

1.4.1. Crecimiento económico generado por la acumulación de factores

El análisis del crecimiento económico debido a la acumulación de factores parte del teorema de Rybczynski (1955). Según este teorema, si con precios relativos de bienes y factores constantes crece la dotación de un solo factor, aumenta la producción del bien que lo usa con más intensidad y disminuye la del otro (véase la figura del tema B05, apartado 2.4.3). Se mantiene el modelo base, con el supuesto adicional de que el bien X (exportable) es relativamente intensivo en capital y el bien Y (importable) es relativamente intensivo en trabajo. De nuevo, los efectos varían según se trate de un país pequeño o de uno grande.

País pequeño (aumento del capital): si, desde el equilibrio inicial, aumenta la dotación de capital, la FPP se expande sobre todo en la dirección del bien X, cuya producción máxima posible crece más que la de Y. El efecto, por el teorema de Rybczynski, es un aumento de la producción del bien exportable y una disminución absoluta de la producción del bien importable, lo que supone un efecto ultra-procomercial sobre la producción (figura B9.3).

País pequeño (aumento del trabajo): si aumenta la dotación del factor que se utiliza intensivamente en el bien importable, la producción máxima posible de Y aumenta más que la de X. Por el mismo teorema, la producción de X cae en términos absolutos, con un efecto ultra-anticomercial sobre la producción.

En resumen, el crecimiento provocado por el aumento de la dotación de un factor es o bien ultra-procomercial o bien ultra-anticomercial en la producción, en función de si dicho factor se utiliza intensivamente en la industria exportadora o en la importadora.

En términos de distribución de la renta, dado que la RRI permanece constante, los precios relativos de los factores no se mueven y, por tanto, tampoco lo hacen las ratios capital-trabajo de cada sector. Puesto que existen rendimientos constantes y competencia perfecta, los productos marginales no cambian ni, con ellos, los rendimientos reales del capital y del trabajo. Los únicos cambios que produce el crecimiento son, pues, los relativos a la producción de ambos bienes y a la distribución de los factores entre las dos industrias.

País grande (aumento del capital): el país, al ofrecer más exportaciones, empuja p_X a la baja, y al demandar más importaciones, empuja p_Y al alza y, en consecuencia, la RRI empeora ($\downarrow p_X / \uparrow p_Y$). Es más, si se admite el desplazamiento interno de factores de un sector a otro, el sector del bien exportable se desarrolla mientras que el sector del bien importable se ve perjudicado. Esta relación entre la expansión del sector favorecido por la dotación y el declive del resto de sectores se conoce como *enfermedad holandesa*. *The Economist* acuñó el término en 1977

para referirse a la desindustrialización experimentada por los Países Bajos tras el descubrimiento de los yacimientos de gas natural de Groninga en 1959[6].

País grande (aumento del trabajo): la disminución de la demanda de importaciones hace bajar p_Y y la disminución de la oferta de exportaciones hace subir p_X, por lo que la RRI mejora ($\uparrow p_X / \downarrow p_Y$).

Crecimiento de ambos factores de producción: si se produce un crecimiento equilibrado, es decir, si crecen ambos factores en la misma proporción, la FPP se expande de forma homotética y el efecto sobre la producción es neutral.

1.4.2. Crecimiento económico generado por el progreso tecnológico

En el muy largo plazo la tecnología no tiene por qué ser constante y, por ello, se puede considerar que la función de producción incluye un factor que hace que esta no sea estable en el tiempo. Por **progreso técnico** se puede entender la posibilidad de producir en tres sentidos:

1. *Innovación de producto*: nuevas variedades de bienes ya existentes o bienes completamente nuevos.

2. *Innovación de proceso*: más producción con los mismos factores o la misma producción con menos factores.

3. *Mejora de la calidad*: versiones mejores de los productos existentes.

Entendido en la segunda acepción, el progreso técnico admite la misma distinción que el crecimiento económico. El progreso técnico es neutral cuando reduce en igual proporción el uso de ambos factores productivos y sesgado cuando economiza relativamente el de uno de ellos. Aquí se adopta el **criterio de Hicks**, según el cual, con una ratio capital-trabajo dada, una innovación tecnológica es neutral respecto al capital y al trabajo si y solo si deja constante el cociente de productividades marginales. En cambio, en el caso de que la productividad marginal de uno de los factores aumente en mayor proporción que la del otro, se habla de progreso técnico *sesgado*. En él, la relación marginal de sustitución técnica a lo largo del rayo-vector cambia y la isocuanta se desplaza hacia el origen de forma sesgada. Concretamente, el progreso técnico sesgado puede ser:

- *Intensificador del capital o ahorrador de trabajo*: en este caso, la productividad marginal del capital crece en mayor proporción que la del trabajo. Así, la $\text{RMST}_{KL} = \text{Pmg}_L/\text{Pmg}_K$ disminuye a lo largo del rayo-vector.

- *Intensificador del trabajo o ahorrador de capital*: ahora la productividad marginal del trabajo crece en mayor proporción que la del capital. Así, la RMST aumenta a lo largo del rayo-vector.

Para estudiar cómo afecta el progreso técnico al comercio internacional, se parte de nuevo del modelo reducido de equilibrio general con una economía abierta y dos sectores. Se distinguen dos casos, según el progreso técnico afecte a los dos sectores o solo a uno de ellos.

Progreso técnico en ambos sectores: si el progreso técnico es neutral y se produce al mismo ritmo en ambos sectores, el efecto es análogo al de un crecimiento equilibrado de las dotaciones y, por tanto, el efecto sobre la producción también es neutral.

[6] Además del canal real, operó un canal monetario, pues la entrada de divisas asociada al gas apreció el florín en términos reales y restó competitividad a las exportaciones del resto de sectores.

Progreso técnico neutral en un solo sector: sus efectos sobre la producción coinciden cualitativamente con los que tendría una mayor dotación del factor que el sector usa intensivamente. En primer lugar, la producción crece en el sector innovador y cae en el otro, como en el teorema de Rybczynski. Si el sector innovador es el exportador, el sesgo sobre la producción es ultraprocomercial y el volumen de comercio internacional se amplía. En segundo lugar, cuando el país es grande, el producto del sector innovador se abarata y el otro bien se encarece, por lo que la RRI empeora cuando la innovación tiene lugar en el sector exportador.

Progreso técnico sesgado en un solo sector: se distinguen dos situaciones:

- *Ahorrador de capital en el sector intensivo en capital*: tiene los mismos efectos cualitativos que el progreso técnico neutral en dicho sector.

- *Ahorrador de trabajo en el sector intensivo en capital*: su efecto es ambiguo, pues, a precios constantes, la producción del otro sector puede aumentar o disminuir según la intensidad de los efectos contrapuestos sobre la oferta y la demanda de factores.

1.5. Conclusiones

La sección analiza cómo afecta el crecimiento al volumen de comercio, a la relación real de intercambio y al bienestar. Su aportación es descomponer el efecto total en un efecto consumo y un efecto producción, y ligar el sesgo de este último a la causa del crecimiento, el factor que se acumula o el sector en el que se innova. El mismo esquema explica la enfermedad holandesa y las condiciones del crecimiento empobrecedor.

Su límite es que el crecimiento entra como un dato exógeno, de modo que la sección no explica sus causas ni la senda entre el equilibrio inicial y el final. El sentido inverso, cómo el comercio altera el crecimiento, es el objeto de las secciones 2 y 3.

2. Efectos del comercio sobre el crecimiento (I): de los clásicos a la novísima teoría

Hasta este punto se han analizado las consecuencias del crecimiento económico para el comercio internacional. Este apartado invierte la perspectiva y examina, a partir de las principales contribuciones del pensamiento económico, cómo influye el comercio internacional en el crecimiento.

2.1. Pensamiento clásico y neoclásico

Se repasan primero las aportaciones del pensamiento clásico, centradas en la especialización y la extensión del mercado, y después las del neoclásico, cuyo análisis de estática comparativa reduce las ganancias de la apertura a un efecto de nivel.

> Sobre la teoría tradicional del comercio internacional, véase el tema B05.

2.1.1. Pensamiento clásico

Las aportaciones clásicas giran en torno a dos conceptos, la **ventaja absoluta** (Adam Smith, 1776) y la **ventaja comparativa** (David Ricardo, 1817). La especialización completa en los bienes en que se posee ventaja tecnológica, absoluta o relativa, y el posterior intercambio elevan la producción global. Por la vía de la división del trabajo, el comercio incrementa así la producción de ambos países y tiene un efecto positivo sobre el crecimiento.

John Stuart Mill añadió dos matizaciones:

- *Demanda recíproca*: esta teoría completa a Ricardo, porque explica en qué punto del intervalo de intercambio mutuamente ventajoso se sitúa la relación real de intercambio. Si los costes comparativos delimitan ese intervalo, la intensidad relativa de la demanda de cada país determina el punto y, con él, cómo se reparte la ganancia.

- *Industria naciente*: este argumento, principal excepción admitida por la escuela clásica al libre comercio, justifica la protección temporal de una industria mientras adquiere la escala y experiencia necesarias. Esa protección queda condicionada a su carácter transitorio y a que la industria alcance la viabilidad sin ella.

De estas aportaciones, la que más interesa a este tema es la de Smith, que contiene una intuición genuinamente dinámica. Su tesis es que la división del trabajo está limitada por la extensión del mercado. De ello se infiere que especializarse solo compensa si existe mercado (demanda) suficiente para absorber la producción resultante, de modo que en un mercado reducido la especialización extrema no resulta rentable. De ahí se sigue una secuencia que se realimenta:

1. *Punto de partida*: la apertura amplía el mercado al que puede venderse.

2. *División del trabajo*: ese mercado mayor hace rentable una división del trabajo más fina.

3. *Productividad*: la especialización eleva la productividad, por la mayor destreza del trabajador, el ahorro de tiempo en los cambios de tarea y el estímulo a la mecanización.

4. *Realimentación*: la mayor productividad genera más renta, que amplía de nuevo el mercado.

Esa circularidad convierte el planteamiento en una teoría del crecimiento y no meramente de la asignación, pues la productividad deja de ser un dato previo al comercio para ser un resultado de él.

Parte de estas intuiciones fueron recuperadas por Allyn Young (1928) y por Kaldor bajo la noción de causación acumulativa, y constituyen el antecedente directo de los modelos de crecimiento endógeno que se examinan en el apartado 3. El argumento de la industria naciente de Mill añade una segunda intuición dinámica, la de que la ventaja comparativa no es un dato, sino algo que puede adquirirse mediante el aprendizaje, premisa sobre la que descansará buena parte de la literatura posterior del desarrollo.

2.1.2. Pensamiento neoclásico

El pensamiento neoclásico mantuvo la defensa del libre comercio con varias matizaciones derivadas de la teoría marginalista del valor. Por ejemplo, los costes de oportunidad crecientes hacen que la especialización deje de ser necesariamente completa, y la ventaja comparativa puede residir en dotaciones relativas de factores.

El crecimiento no ocupa, sin embargo, un lugar central en este cuerpo teórico, y la razón es de método. El análisis neoclásico del comercio es de **estática comparativa**, es decir, compara dos equilibrios con dotaciones y tecnología dadas, de modo que las ganancias de la apertura son de nivel y no de tasa de crecimiento. Por tanto, el comercio permite un consumo superior al de autarquía, pero agotado ese salto la economía no crece más deprisa por estar abierta. El interés se desplaza así hacia la asignación eficiente y los efectos redistributivos entre factores. Habrá que esperar a la nueva teoría del comercio y, sobre todo, a los modelos de crecimiento endógeno en economía abierta para que la apertura altere la tasa de crecimiento y no solo el nivel de renta.

> El modelo de Heckscher-Ohlin-Samuelson, los teoremas de Stolper-Samuelson y de Rybczynski y la paradoja de Leontief se desarrollan en el tema B05.

2.2. Nueva teoría

Dentro de la nueva teoría, la relación entre comercio y crecimiento se aborda primero con dos precursores, los desfases tecnológicos y el ciclo del producto, y después con la diferenciación vertical y la I+D.

> Sobre la nueva teoría del comercio internacional, véase el tema B06.

De la nueva teoría interesan aquí las aportaciones relacionadas con los factores que afectan a la relación entre comercio y crecimiento, como el progreso tecnológico, la I+D y la productividad.

2.2.1. Precursores de la nueva teoría: desfases tecnológicos (Posner, 1961; Hufbauer, 1966)

Origen: Michael Posner (1961) y Gary Hufbauer (1966) estudian los desfases tecnológicos (*technology gaps*) como origen del comercio internacional.

Idea: la innovación da al país que la genera una ventaja comparativa temporal en el nuevo producto, mientras los demás países no aprenden a imitarlo.

Razonamiento: concretamente, estos autores distinguen dos desfases y el desfase global neto que resulta de compararlos:

1. *Desfase de imitación* (*imitation lag*)[7]: la I+D y la iniciativa empresarial hacen que los países innovadores produzcan nuevos bienes y disfruten de un monopolio sobre ellos hasta que otros países aprenden a producirlos.

2. *Desfase de demanda*: mide cuánto tarda en aparecer demanda del producto en el país importador desde que se inventa en el innovador. Puede deberse, entre otras causas, a que la información circula despacio o a que las preferencias tardan en adaptarse a la novedad.

3. *Desfase global neto*: comparar ambos desfases indica si la innovación dará lugar a intercambios internacionales. Cuando imitar (copiar la técnica) lleva más tiempo que empezar a demandar, los países seguidores quieren el bien antes de saber fabricarlo, de modo que tienen que comprarlo fuera y surge el comercio. En el caso contrario, la copia llega antes que la demanda exterior, así que los seguidores pueden abastecerse solos y no hay comercio.

En resumen, aunque los países tengan preferencias y dotaciones relativas de factores idénticas, el progreso técnico que los innovadores generan de forma endógena les da una ventaja comparativa y, con ella, una base para comerciar.

Si existe una asimetría importante entre países, los países poco dinámicos, es decir, aquellos cuyos flujos de incorporación de innovaciones por unidad de tiempo son más lentos, se ven obligados a exportar bienes tradicionales a cambio de bienes innovadores. Esto puede dar lugar a un círculo vicioso, pues, al ser los bienes tradicionales de bajo valor añadido, la renta que esos países obtienen del comercio no es suficiente para invertir en I+D y mejorar su dinamismo. En cambio, cuando el dinamismo es similar en todos los países, el comercio internacional estimula el crecimiento económico, ya que las innovaciones se difunden rápidamente entre ellos.

Desarrollo gráfico: en el plano tiempo-saldo comercial del país A en el nuevo producto (figura B9.6), el saldo, BC^A, recorre tres fases:

1. *Desfase de demanda*: el país A innova en t_0, pero sus exportaciones no aumentan inmediatamente, puesto que la demanda exterior solo nace en t_1.

2. *Desfase de imitación*: a partir de t_1, y mientras este desfase siga corriendo, las exportaciones del país A crecen, es decir, existe comercio internacional precisamente porque imitar lleva más tiempo que demandar. Las exportaciones alcanzan un máximo y se mantienen en meseta mientras los costes de aprendizaje y otras barreras (informacionales, licencias) retrasan la transferencia tecnológica y la demanda exterior está ya cubierta.

3. *Estandarización*: cuando el país B completa la imitación y comienza a fabricar localmente (t_2), se produce la estandarización del producto y las exportaciones del país A disminuyen. En función de los costes relativos, la situación puede incluso revertirse y el país A pasar a ser importador del bien.

[7] Este, a su vez, se puede dividir en tres subcomponentes: el *foreign reaction lag* (tiempo que pasa entre el uso exitoso de la innovación en el país innovador y el momento en que algunas empresas del país importador advierten que los nuevos productos constituyen competencia potencial), el *domestic reaction lag* (tiempo que tardan todas las empresas del país importador en percibir que los nuevos productos constituyen competencia efectiva) y el *learning period* (tiempo que les lleva a las empresas del país importador aprender a producir y vender el producto).

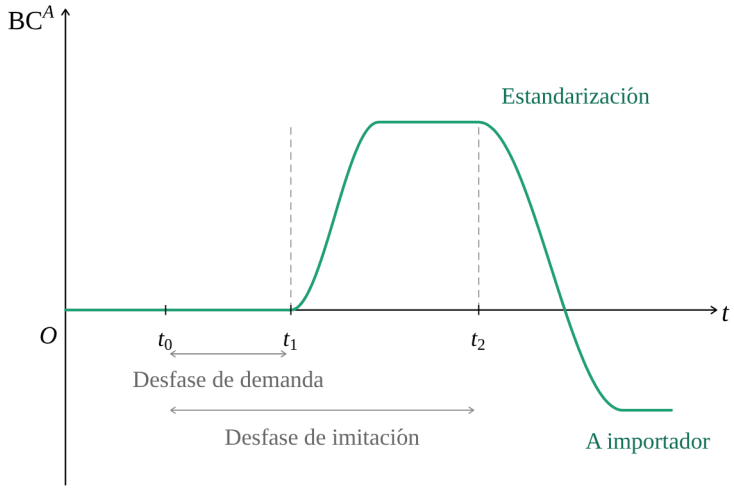

Figura B9.6: Desfases tecnológicos y ciclo del producto (Posner, 1961; Vernon, 1966). El país innovador exporta hasta que la imitación traslada la producción al exterior.

Conclusiones:

▪ *Aspectos positivos*: desde el punto de vista teórico, el modelo de desfases tecnológicos fue relevante para el desarrollo posterior de las teorías del crecimiento endógeno en un contexto internacional. Desde el punto de vista empírico, se ha usado para explicar la época dorada de crecimiento europeo de 1950 a 1973, basada en la absorción del desfase tecnológico con EE. UU., aunque existen otras hipótesis complementarias sobre aquel episodio.

▪ *Limitaciones*: los desfases (de imitación y de demanda) son exógenos al modelo, que no explica su duración ni los mecanismos de difusión de la tecnología. Se trata, pues, de un marco descriptivo más que de un modelo microfundamentado, carencia que abordarían más adelante los modelos norte-sur de ciclo del producto (Krugman, 1979) y la literatura de crecimiento endógeno.

2.2.2. Precursores de la nueva teoría: ciclo del producto (Vernon, 1966; Hirsch, 1967)

Origen: Raymond Vernon (1966) y Seev Hirsch (1967) desarrollaron una teoría con la que explicar por qué las innovaciones suelen concentrarse en los países desarrollados y acaban filtrándose después a economías menos avanzadas.

Idea: las proporciones de factores que exige un producto no son iguales en cada fase de su ciclo de vida, y en esa intensidad factorial cambiante está, según estos autores, la explicación.

Razonamiento:

1. *Etapa inicial*: solo las empresas de los países más avanzados reúnen lo que pide un producto recién inventado (un fuerte gasto en I+D, trabajadores cualificados, información que circule deprisa y capacidad para gestionar el riesgo). Los primeros compradores, además, suelen ser los hogares con más ingresos. Por eso, lo normal es que el producto se fabrique en un país desarrollado y que este lo exporte al resto del mundo.

2. *Etapa de madurez*: la tecnología se difunde y empiezan a fabricar el bien los países que pueden hacerlo más barato[8], con lo que las ventas exteriores del innovador comienzan a retroceder.

3. *Etapa de estandarización*[9]: la mano de obra cualificada deja de ser necesaria y la fabricación se desplaza a los países menos desarrollados, donde abunda el trabajo no cualificado. Los papeles se invierten, pues ahora estos países exportan el bien y los desarrollados lo importan.

Conclusiones:

- *Aspectos positivos*: la teoría del ciclo del producto analiza también, de manera implícita, la localización de la producción en los distintos países y, por tanto, se considera precursora de los modelos de la economía geográfica.

- *Limitaciones*: se trata de una teoría descriptiva, sin un modelo formal subyacente en su formulación original. Además, el propio Vernon (1979) reconoció que su vigencia se había debilitado, pues las empresas multinacionales, con redes globales de producción, lanzan hoy los productos de forma casi simultánea en múltiples mercados, lo que acorta o elimina las fases del ciclo.

2.2.3. Diferenciación vertical e I+D (Shaked y Sutton, 1982; 1983)

> El modelo de diferenciación vertical de Shaked-Sutton se desarrolla en el tema B06.

En un marco de diferenciación vertical con costes fijos de I+D, la competencia da lugar a un oligopolio natural internacional. Desde la óptica de este tema, el oligopolio natural puede redundar en crecimiento económico, debido a las inversiones en I+D que realizan las empresas para mantener su posición de calidad en el mercado.

2.3. Novísima teoría

La novísima teoría se ilustra con el modelo de heterogeneidad de empresas (Melitz, 2003), que añade la reasignación de recursos hacia las empresas más productivas, y con sus extensiones a la elección entre exportar e invertir y a la organización de la producción.

2.3.1. Modelo de heterogeneidad de empresas (Melitz, 2003)

> El modelo de heterogeneidad de empresas se desarrolla por completo en el tema B06.

Desde la perspectiva de la relación entre comercio y crecimiento, el modelo de Melitz ofrece dos predicciones fundamentales, a las que la literatura posterior añade un canal complementario:

1. *Efecto selección*: la apertura comercial separa a las empresas en tres grupos según su productividad (véase la figura del tema B06, apartado 3.1.3).

- *Menos productivas*: aunque producían en autarquía, abandonan el mercado.

[8] Es posible que la empresa innovadora empiece a establecer subsidiarias en otros países desarrollados, de modo que la exportación del país innovador hacia estos se aproxime a cero mientras el país innovador continúa exportando a los países en desarrollo. También sería posible recurrir a la concesión de licencias o a la adquisición de participaciones en plantas extranjeras. Sin embargo, Vernon señala que la alternativa de las licencias es inferior debido a las imperfecciones de los mercados internacionales de tecnología.

[9] La deslocalización puede compensar incluso en bienes intensivos en capital, pues el coste del capital puede pesar menos que otros costes o quedar contrarrestado por el menor coste de la mano de obra.

- *Productividad intermedia*: más productivas, pero no lo suficiente, venden solo en el mercado nacional.

- *Más productivas*: obtienen beneficios en los mercados de exportación, con lo que crecen y se convierten en exportadoras.

2. *Efecto productividad*: el hecho de que las empresas menos productivas abandonen el mercado y las más productivas ganen cuota hace que la productividad agregada del país sea mayor con comercio internacional. La razón es que los recursos que antes empleaban las empresas menos productivas pasan a las más productivas, que los utilizan más eficientemente.

3. *Efecto aprendizaje*: aunque en el modelo original la productividad de cada empresa es fija, el efecto productividad se puede ver reforzado por el aprendizaje asociado a la exportación (*learning by exporting*). Con ese aprendizaje, las empresas exportadoras adquieren conocimientos y habilidades que les permiten aumentar aún más su productividad.

En consecuencia, las ganancias de productividad agregada derivadas del comercio internacional, reforzadas en su caso por el efecto aprendizaje, elevan el nivel de la renta. En Melitz (2003) se trata de un efecto de nivel y no de tasa de crecimiento, pues la apertura eleva una sola vez la productividad agregada sin alterar su ritmo de crecimiento a largo plazo. El resultado, además, es general, ya que las mismas intuiciones se obtienen si se analiza un aumento del número de países a los que se puede exportar o una caída de los costes fijos o variables de exportación.

2.3.2. Extensiones: exportar o invertir (Helpman, Melitz y Yeaple, 2004) y organización óptima (Antràs y Helpman, 2004)

> Los modelos de Helpman, Melitz y Yeaple (2004) y de Antràs y Helpman (2004) se desarrollan en el tema B06, apartado 3.2.

Exportar o invertir: la principal predicción del modelo de Helpman, Melitz y Yeaple (2004) es que las empresas más productivas no solo son capaces de exportar, sino que las más eficientes de entre ellas pueden llevar a cabo inversión extranjera directa (IED). La sustitución de exportaciones por IED resulta tanto más probable cuanto mayores son los costes de transporte (dilema proximidad-concentración).

Organización óptima: Pol Antràs y Elhanan Helpman (2004) se preguntan dónde realizar el proceso productivo en un marco internacional y cómo estructurarlo (qué procesos realizar internamente y cuáles externalizar). Los autores estudian los determinantes de la internalización de procesos (la IED vertical) frente a su externalización o subcontratación. Entre esos determinantes destacan los conflictos de interés entre las partes, las diferencias en los costes de producción y el saber hacer de cada agente.

2.3.3. Conclusiones

Aspectos positivos: la novísima teoría identifica un canal específico por el que el comercio eleva la productividad agregada, la reasignación intraindustrial de recursos hacia las empresas más productivas. El canal resulta invisible tanto para la teoría tradicional (centrada en la reasignación entre sectores) como para la nueva teoría con empresas idénticas.

Limitaciones: en Melitz (2003), la productividad de cada empresa es fija y la reasignación eleva una sola vez el nivel de la productividad agregada, sin alterar su tasa de crecimiento.

Además, en una amplia clase de modelos que incluye el de Melitz con distribución de Pareto, la heterogeneidad de las empresas no añade ganancias de bienestar a las de los modelos con empresas idénticas (Arkolakis, Costinot y Rodríguez-Clare, 2012).

2.4. Conclusiones

No todo canal del comercio eleva la tasa de crecimiento además del nivel de renta. Para los clásicos, la especialización aumenta la producción, y Smith le da un alcance dinámico, pues un mercado mayor permite dividir más el trabajo y eleva una productividad que vuelve a ampliar el mercado. La industria naciente de Mill añade que la ventaja comparativa puede adquirirse mediante el aprendizaje. El análisis neoclásico, de estática comparativa, reduce en cambio las ganancias de la apertura a un efecto de nivel.

En la nueva teoría, los desfases tecnológicos dan al innovador una ventaja temporal. Si los países tienen un dinamismo parecido, el comercio difunde deprisa las innovaciones y estimula el crecimiento, y, si no, puede encerrar a los rezagados en un círculo vicioso. La I+D que sostiene la calidad en la diferenciación vertical puede redundar también en crecimiento. La novísima teoría suma la reasignación hacia las empresas más productivas, cuyo efecto en Melitz (2003) es, sin embargo, de nivel. Los efectos sobre la tasa los formalizan los modelos de crecimiento endógeno del apartado 3.1 y los contrasta la evidencia del apartado 3.2.

3. Efectos del comercio sobre el crecimiento (II): crecimiento endógeno y evidencia empírica

Los modelos de crecimiento endógeno formalizan esos canales en economía abierta, y la evidencia empírica permite contrastar si la apertura se asocia, de hecho, a un mayor crecimiento.

3.1. Modelos de crecimiento endógeno en economía abierta

Se exponen tres modelos: el de externalidades del capital, el de aumento del número de variedades de *inputs* y el de comercio, ventaja comparativa y crecimiento.

> Sobre los modelos de crecimiento endógeno, véase el tema A44.

3.1.1. Modelo de externalidades del capital con comercio internacional (Arrow, 1962; Romer, 1986; Lucas, 1988)

Origen: el modelo procede de los modelos de crecimiento endógeno de primera generación de Arrow (1962), Romer (1986) y Lucas (1988), extendidos a la economía abierta.

Idea: en esta primera generación de modelos de crecimiento endógeno nadie se propone innovar. La inversión en capital genera progreso tecnológico como efecto secundario, sin que medie una actividad deliberada. Así, la empresa invierte para producir, y al hacerlo acumula experiencia y conocimiento que no puede retener en exclusiva, de modo que el progreso técnico se modela como una externalidad positiva. La diferencia con la segunda generación es de mecanismo, porque allí la innovación se busca y se remunera mediante patentes, mientras que aquí sencillamente ocurre.

Supuestos:

- *Entorno*: economía abierta al exterior, con una función de producción agregada que incorpora, junto a los factores convencionales, un término de externalidad.

- *Agentes*: la tasa de ahorro s es exógena y constante, al igual que las tasas de crecimiento de la población n y de depreciación δ.

- *Tecnología*: cuando una empresa acumula capital per cápita, genera externalidades positivas para el resto de la economía, porque las demás pueden aprovechar la experiencia, los conocimientos y la tecnología que ese capital incorpora[10]. Cada empresa toma el nivel agregado de la externalidad como dado, de modo que no lo internaliza en su decisión de inversión.

- *Sector exterior*: el comercio internacional opera como catalizador de esas externalidades, al ampliar los canales de difusión del conocimiento mediante la importación de bienes de capital, la imitación y el contacto con la frontera tecnológica. Se representa por el parámetro β, que mide el grado de apertura comercial.

[10] Las externalidades del capital en la función de producción pueden surgir por el aprendizaje por la práctica (*learning by doing*, en la línea de Arrow, 1962) o por los desbordamientos de conocimiento (*knowledge spillovers*).

Planteamiento: la función de producción incorpora la externalidad E, proporcional al capital per cápita y amplificada por el grado de apertura.

$$Y = AK^{\alpha}L^{1-\alpha}E^{\eta}, \qquad E = \beta k, \qquad A > 0,\ \alpha \in (0,1)$$

Es decir, $Y = AK^{\alpha}L^{1-\alpha}(\beta k)^{\eta}$, donde A es la tecnología, K el *stock* de capital, L el de trabajo, k el capital per cápita, α la elasticidad-producto del capital y η la de la externalidad[11]. Expresada en términos per cápita, la función de producción es:

$$y = Ak^{\alpha+\eta}\beta^{\eta}, \qquad A > 0,\ \alpha \in (0,1)$$

La elasticidad relevante del capital deja de ser α y pasa a ser $\alpha + \eta$. La empresa percibe rendimientos decrecientes, pero la economía en su conjunto los experimenta mayores por efecto de la externalidad. De esa divergencia entre lo privado y lo social depende el resultado del modelo. El capital per cápita se acumula según la ecuación:

$$\dot{k} = sAk^{\alpha+\eta}\beta^{\eta} - (n+\delta)k$$

donde s es la tasa de ahorro, n la de crecimiento de la población y δ la de depreciación. Dividiendo por k, la tasa de crecimiento del capital per cápita es:

$$\frac{\dot{k}}{k} = sAk^{\alpha+\eta-1}\beta^{\eta} - (n+\delta)$$

Dado que β^{η} multiplica al término de ahorro, cuanto mayor es el grado de apertura comercial, mayor es la tasa de crecimiento del capital per cápita en cada momento.

Equilibrio: en el caso límite, si $\alpha + \eta = 1$, es decir, cuando el peso de la externalidad en el producto iguala a la participación del trabajo, el capital en sentido amplio presenta rendimientos marginales constantes y el modelo se reduce al modelo AK. La tasa de crecimiento de largo plazo resulta entonces

$$g_k = sA\beta^{\eta} - (n+\delta)$$

Desarrollo gráfico: el equilibrio del modelo se representa en el plano capital per cápita-tasa de crecimiento (figura B9.7):

1. *Curva de ahorro*, $sAk^{\alpha+\eta-1}\beta^{\eta}$: recoge la inversión por unidad de capital financiada con el ahorro, y es decreciente en general, porque con $\alpha+\eta < 1$ el capital tiene rendimientos decrecientes. Cuando $\alpha + \eta = 1$, la curva deja de serlo y se convierte en una horizontal, $sA\beta_0^{\eta}$.

2. *Recta de depreciación y crecimiento demográfico*, $n + \delta$: recoge la inversión por unidad de capital necesaria para mantener constante el capital per cápita y es horizontal, porque no depende de k.

3. *Equilibrio*: con $\alpha + \eta = 1$ y $sA\beta^{\eta} > n + \delta$, la distancia entre ambas, que es la tasa de crecimiento, $g_{k,0}$, permanece constante y no se anula, de modo que la economía no converge a un estado estacionario, sino que crece de forma sostenida.

[11] Con $\eta = 0$, la función se reduce a la Cobb-Douglas del modelo de Solow.

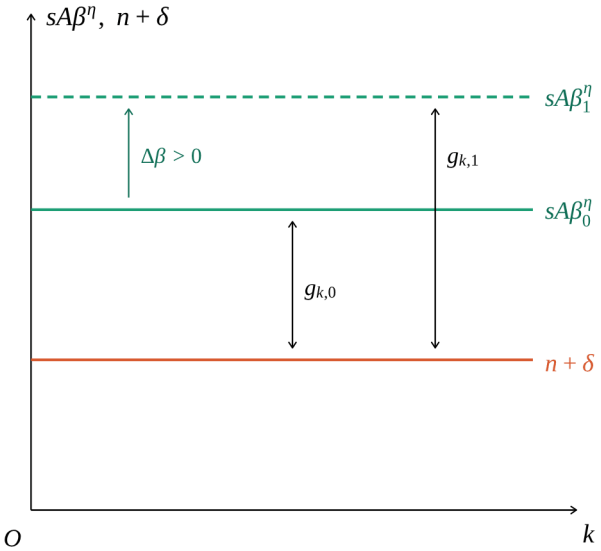

Figura B9.7: Modelo AK con apertura comercial. Un mayor grado de apertura, β, desplaza hacia arriba la recta de ahorro y eleva de forma permanente la tasa de crecimiento.

Estática comparativa: un aumento del grado de apertura ($\Delta\beta > 0$) desplaza la horizontal de ahorro hacia arriba, de $sA\beta_0^{\eta}$ a $sA\beta_1^{\eta}$, y eleva la tasa de crecimiento de forma permanente, de $g_{k,0}$ a $g_{k,1}$. El modelo aboga, así, por la apertura comercial como estímulo del crecimiento económico sostenido.

Conclusiones:

- *Aspectos positivos*: el modelo genera crecimiento sostenido a partir de la sola acumulación de capital, sin necesidad de postular un residuo exógeno, y proporciona un fundamento analítico directo a la apertura comercial como estímulo del crecimiento. Su implicación normativa es asimismo relevante. En la versión con ahorro optimizador (Romer, 1986), al no internalizar las empresas la externalidad que generan, la inversión de equilibrio es inferior a la socialmente óptima. Esa brecha justifica el apoyo público a la formación de capital y, en particular, a la apertura que amplifica sus efectos.

- *Limitaciones*: el mecanismo es mecánico, pues el progreso técnico aparece como subproducto automático de la inversión y no como resultado de una decisión, lo que impide analizar los incentivos a innovar, carencia que subsanarán los modelos de segunda generación. El resultado depende además críticamente del valor de η. Si $\alpha + \eta < 1$, la economía converge como en el modelo de Solow y el crecimiento endógeno desaparece, mientras que si $\alpha + \eta > 1$ la senda se vuelve explosiva. Por ello, el caso interesante exige una igualdad exacta difícil de justificar. Por último, el parámetro β recoge la apertura de forma reducida, sin distinguir entre sus distintos canales ni considerar que sus efectos pueden depender de la capacidad de absorción tecnológica del país receptor.

3.1.2. Modelo de aumento del número de variedades de *inputs* (Romer, 1990)

Origen: lo propuso Paul Romer (1990).

Idea: el progreso técnico, lejos de ser exógeno, resulta de la decisión deliberada de invertir en investigación, y esa decisión responde a incentivos económicos. La innovación adopta aquí la forma de un aumento del número de variedades de bienes de capital disponibles, y no de la mejora en la calidad de las existentes. El modelo demuestra que la I+D solo se produce si existe competencia imperfecta, pues nadie invertiría en investigar si el resultado fuese inmediatamente apropiable por todos.

Supuestos:

▪ *Entorno*: equilibrio general con horizonte infinito y tiempo continuo. El bien final actúa como numerario, con precio unitario. La población y la fuerza de trabajo permanecen constantes, lo que permite aislar el crecimiento de la renta per cápita del derivado del aumento demográfico.

▪ *Agentes*: tres sectores con estructuras de mercado distintas, y esa heterogeneidad es el núcleo del modelo.

- *Sector de investigación*: dedica recursos a obtener diseños de nuevas variedades de bienes de capital, con un coste F por invención. Opera en libre entrada, de forma que en equilibrio el valor de la patente iguala al coste de obtenerla.

- *Sector de bienes intermedios*: cada variedad es producida en régimen de monopolio por el titular de su patente, de duración indefinida[12]. El sector presenta rendimientos crecientes a escala, pues al coste fijo de la invención se añade un coste marginal constante.

- *Sector de bienes finales*: competitivo, combina trabajo y el conjunto de variedades de bienes de capital disponibles.

▪ *Tecnología*: el conocimiento presenta rasgos de bien público. Es *no rival*, pues su uso por una empresa no impide el de otra, y solo *parcialmente excluible*, ya que la patente protege el diseño concreto pero no las ideas generales que incorpora. De la no rivalidad se sigue el rasgo decisivo del modelo, la existencia de rendimientos crecientes a escala agregada. Se supone que las variedades entran de forma simétrica en la función de producción y que su elasticidad de sustitución es constante e igual a $\sigma = 1/(1-\alpha)$, con $\alpha \in (0,1)$.

Sector de bienes finales: competitivo, maximiza beneficios tomando los precios de los *inputs* como dados.

▪ *Planteamiento*: el productor elige el trabajo y la cantidad de cada variedad que maximizan su beneficio, $Y_t - w_t L_t - \sum_j p_{jt} x_{jt}$, donde w_t es el salario y p_{jt} el precio de la variedad j en unidades del bien final. Su función de producción emplea trabajo y un agregado de las N variedades de bienes de capital existentes.

$$Y_t = A\, K_t^{\alpha} L_t^{1-\alpha}, \qquad K_t = \left(\sum_{j=1}^{N_t} x_{jt}^{\alpha} \right)^{1/\alpha}$$

Sustituyendo, la expresión operativa resulta $Y_t = A\, L_t^{1-\alpha} \sum_{j=1}^{N_t} x_{jt}^{\alpha}$. Con variedades simétricas, $x_{jt} = x_t$, se obtiene $Y_t = A\, L_t^{1-\alpha} N_t x_t^{\alpha}$, de modo que la producción crece con el número de variedades aunque la cantidad total de bienes de capital, $N_t x_t$, permanezca constante. La razón de este resultado central es la concavidad de la función respecto de cada variedad, que hace productiva en sí misma la especialización.

[12] La duración indefinida de la patente es un supuesto simplificador. En la práctica, su expiración transformaría el mercado de la variedad en competitivo, eliminando el beneficio que remunera la invención.

▪ *CPO*: el productor contrata trabajo y cada variedad hasta que su productividad marginal iguala a su precio, $\partial Y_t/\partial L_t = w_t$ y $\alpha A L_t^{1-\alpha} x_{jt}^{\alpha-1} = p_{jt}$. De la segunda resulta la demanda de cada variedad, $x_{jt} = L_t\,(\alpha A/p_{jt})^{1/(1-\alpha)}$, decreciente en su precio y con elasticidad-precio constante $1/(1-\alpha)$.

▪ *CSO*: con $\alpha < 1$ la producción es estrictamente cóncava en el trabajo y en cada variedad, de manera que las CPO definen las demandas de factores.

Sector de bienes intermedios: el titular de cada patente produce su variedad en monopolio.

▪ *Planteamiento*: el monopolista elige el precio que maximiza su beneficio, $(p_j - 1)\,x_j(p_j)$, frente a la demanda isoelástica del sector final, con el coste marginal normalizado a la unidad.

▪ *CPO*: el ingreso marginal, αp_j, iguala al coste marginal, de donde $p_j = 1/\alpha > 1$, un margen constante sobre el coste marginal. El margen $(1-\alpha)/\alpha$ es lo que permite recuperar el coste de la invención, y es tanto mayor cuanto menos sustituibles sean las variedades entre sí.

▪ *CSO*: con elasticidad $1/(1-\alpha) > 1$, el beneficio crece con el precio por debajo de $1/\alpha$ y decrece por encima, por lo que el margen constante da el máximo.

Sector de investigación: la libre entrada iguala el valor presente de los beneficios monopolísticos futuros al coste de inventar, $\Pi/r = F$, donde Π es el beneficio por periodo de cada monopolista. De aquí se despeja el tipo de interés de equilibrio, $r = \Pi/F$.

Hogares: ahorran como en el modelo de Ramsey-Cass-Koopmans (tema A43, apartado 3.2).

▪ *Planteamiento*: maximizan una utilidad intertemporal con elasticidad de sustitución constante, $\int_0^\infty e^{-\rho t}(c^{1-\theta}-1)/(1-\theta)\,dt$, sujetos a $c + \dot{a} = w + ra$, donde c es el consumo per cápita[13], a los activos per cápita, r el tipo de interés, w el salario, ρ la tasa de preferencia temporal y $1/\theta$ la elasticidad de sustitución intertemporal.

▪ *CPO*: resulta la ecuación de Euler, $g_c = \dot{c}/c = (1/\theta)(r-\rho)$, según la cual el consumo crece mientras el tipo de interés supere a la tasa de preferencia temporal, y tanto más cuanto mayor es la elasticidad de sustitución intertemporal.

▪ *CSO*: el hamiltoniano es estrictamente cóncavo en c y lineal en a, y con la condición de transversalidad las CPO son suficientes.

Equilibrio: combinando la libre entrada en la investigación con la ecuación de Euler se obtiene la tasa de crecimiento de equilibrio, que es constante y depende exclusivamente de parámetros:

$$g_Y = f(\theta^-,\, F^-,\, A^+,\, L^+,\, \rho^-)$$

El crecimiento es endógeno, pues no depende de un progreso técnico exógeno, sino de la rentabilidad de investigar, que a su vez responde a la política económica.

Estática comparativa: la tasa de crecimiento aumenta con la productividad, A, y con el tamaño del mercado, L, que elevan el beneficio de cada monopolista y, con él, el tipo de interés de equilibrio. En cambio, disminuye con el coste de inventar, F, que reduce ese tipo de interés, y con θ y ρ, que reducen el ahorro destinado a investigar (si $r > \rho$). El efecto de L es el efecto escala que se discute en las limitaciones.

[13] En tiempo continuo se omite el argumento temporal: $c \equiv c(t)$ y $a \equiv a(t)$; se escribe solo en las condiciones inicial y terminal.

Extensiones (aplicación al comercio internacional): Rivera-Batiz y Romer (1991) trasladaron el modelo a la economía abierta y obtuvieron un resultado matizado, según el cual el efecto de la integración depende de cómo se produzcan las ideas:

- *Especificación de "equipo de laboratorio"*: la I+D emplea el bien final (como en la versión aquí expuesta, con coste F), y basta la integración de los mercados de bienes para elevar la tasa de crecimiento.

- *Especificación impulsada por el conocimiento*: la investigación se apoya en el *stock* de ideas, de modo que el comercio de bienes por sí solo no altera la tasa de crecimiento y solo la integración del conocimiento la eleva.

La apertura opera a través de tres canales:

1. *Ampliación del mercado*: eleva los beneficios esperados de cada innovación y, con ellos, el incentivo a investigar, efecto que L recoge en la solución.

2. *Acceso a variedades extranjeras*: el productor nacional dispone de un conjunto mayor de bienes de capital, lo que eleva su productividad por la vía de la especialización, con independencia de que innove o no.

3. *Difusión del conocimiento*: la integración permite que los investigadores de cada país aprovechen los descubrimientos ajenos, evitando la duplicación de esfuerzos que se produce cuando dos economías aisladas investigan lo mismo por separado. En la especificación impulsada por el conocimiento, este es el canal decisivo y el único que altera la tasa de crecimiento de forma permanente.

De los tres canales se sigue una implicación de política económica de primer orden. Si los efectos duraderos del comercio sobre el crecimiento operan por la vía del conocimiento y no por la del intercambio de bienes, la apertura comercial por sí sola resulta insuficiente y se requiere además capacidad de absorción tecnológica (capital humano e instituciones de investigación). La cuestión enlaza con la evidencia empírica del apartado 3.2.2.

Conclusiones:

- *Aspectos positivos*:

  - *Endogenización del progreso técnico*: a diferencia del modelo de Solow, donde el crecimiento a largo plazo depende de un residuo inexplicado, aquí resulta de decisiones de inversión sujetas a incentivos, y la no rivalidad de las ideas evita que los rendimientos decrecientes lo agoten.

  - *Fundamento de la política de I+D*: el equilibrio descentralizado crece por debajo del óptimo social, por dos distorsiones de naturaleza distinta. La estática es el margen monopolístico, que reduce la cantidad producida de cada variedad por debajo de la eficiente. La dinámica es que el inventor captura los beneficios de su patente pero no todo el excedente social que genera la nueva variedad, ni tampoco, en la versión original impulsada por el conocimiento, el efecto de su descubrimiento sobre la productividad de la investigación ajena. De ambas se sigue la justificación de la intervención pública. Una subvención a la investigación puede acercar la tasa de crecimiento a la óptima, pero corregir también la distorsión estática exige subvencionar la compra de bienes intermedios (o, de forma equivalente, la producción del bien final).

- *Limitaciones*:

- *Efecto escala*: la tasa de crecimiento es creciente con el tamaño de la población, de modo que el modelo predice que una economía mayor debería crecer más deprisa de forma permanente. Sin embargo, Charles Jones (1995) mostró que la evidencia no lo respalda, pues el número de investigadores en las economías avanzadas se multiplicó durante el s. XX sin que la tasa de crecimiento aumentara, lo que dio lugar a los modelos de crecimiento semiendógeno.

- *Simplificación de la innovación*: tratarla como mera adición de variedades ignora la destrucción creativa (las innovaciones desplazan a las anteriores), aspecto que abordarían Aghion y Howitt (1992) con los modelos de escalera de calidad.

- *Ausencia de fricciones*: no incorpora imperfecciones financieras ni institucionales, que son precisamente las que limitan la innovación en los países en desarrollo.

3.1.3. Comercio, ventaja comparativa y crecimiento (Grossman y Helpman, 1990; 1991)

Origen: Gene Grossman y Elhanan Helpman presentan el modelo en 1990 y lo desarrollan en su libro de 1991.

Idea: el modelo comparte con el de Romer la tesis de que el crecimiento sostenido descansa, en último término, en la I+D. De ahí que el comercio internacional pueda influir sobre ese crecimiento a través del coste relativo de investigar. Toda medida que abarate la I+D respecto de las actividades alternativas eleva el incentivo a innovar y, con él, la tasa de crecimiento a largo plazo.

Razonamiento: el efecto de la protección depende, en consecuencia, del patrón de ventaja comparativa del país, y su signo se invierte según cuál sea este:

1. *País que exporta bienes intensivos en I+D e importa bienes intensivos en trabajo* (p. ej. EE. UU.): gravar con un arancel el bien importado encarece en términos relativos el trabajo no cualificado y abarata el capital humano[14], factor en el que es intensiva la I+D, con lo que investigar resulta relativamente más barato. Los productores nacionales tienen así más incentivos para innovar y el crecimiento a largo plazo se acelera.

2. *País que importa bienes intensivos en I+D y exporta bienes intensivos en trabajo* (como los países en desarrollo): el mismo arancel tiene el efecto opuesto. Al encarecer en términos relativos la I+D, reduce los incentivos a innovar y frena el crecimiento a largo plazo.

La implicación de política económica depende del punto de partida, y al importador de bienes intensivos en I+D puede convenirle, en algunas especificaciones, incluso subvencionar esas importaciones.

Conclusiones:

▪ *Aspectos positivos*: el modelo proporciona el vínculo analítico entre el patrón de especialización y la tasa de crecimiento, y explica por qué una misma medida de política comercial puede resultar expansiva en una economía y contractiva en otra. Ofrece, además, un fundamento riguroso a la intuición estructuralista de que la composición de las exportaciones condiciona el desarrollo, si bien por una vía distinta a la del deterioro de la relación real de intercambio.

[14] Según el teorema de Stolper-Samuelson, cuando un bien se encarece en términos relativos, la retribución real del factor que su producción usa con más intensidad sube y la del otro baja.

- *Limitaciones*: se trata de resultados de *second best*, muy sensibles a los supuestos adoptados sobre la difusión internacional del conocimiento. Así, si el conocimiento fluye libremente entre países, la localización geográfica de la I+D pierde relevancia y con ella el argumento proteccionista, conclusión coherente con el resultado de Rivera-Batiz y Romer ya examinado. A ello se añade que el modelo prescinde de la economía política de la medida, en particular de la probabilidad de que la protección, una vez concedida, se perpetúe más allá de lo que justificaría el argumento tecnológico.

3.2. Evidencia empírica sobre los efectos del comercio internacional en el crecimiento

El repaso de la evidencia empírica empieza por los problemas de medición, la definición de la apertura y la causalidad inversa, y continúa con los principales estudios sobre crecimiento y comercio.

3.2.1. Medición

El contraste empírico de la relación entre apertura y crecimiento se enfrenta a dos problemas de medición fundamentales:

- *Variable apertura*: el primer problema es su definición. Los indicadores de resultado (la ratio comercio/PIB) dependen del tamaño y la geografía del país, mientras que los indicadores de política (aranceles medios, cobertura de barreras no arancelarias o índices compuestos como el de Sachs-Warner) son parciales y discutibles en su construcción.

- *Causalidad inversa o endogeneidad*: los países más ricos comercian más, por lo que una correlación positiva entre apertura y renta no demuestra que la primera cause la segunda. Frankel y Romer (1999) abordaron este problema instrumentando el comercio con sus determinantes geográficos (a partir de una ecuación de gravedad) y encontraron un efecto positivo y cuantitativamente elevado, aunque solo moderadamente significativo, del comercio sobre la renta per cápita. Rodríguez y Rodrik (2000) mostraron, no obstante, que buena parte de los resultados de esta literatura es frágil ante cambios en los indicadores de apertura y en las variables de control.

3.2.2. Estudios sobre crecimiento y comercio

Panagariya (2004): Arvind Panagariya, en su artículo de 2004, utiliza datos del Banco Mundial de las cuatro décadas anteriores y distingue tres grupos de países:

- *Milagros del crecimiento* (*miracles*): países cuya renta per cápita creció a una tasa media anual del 3 % o superior.

- *Fracasos del crecimiento* (*debacles*): países con tasas negativas.

- *Resto del mundo*: los demás países.

Con esta clasificación, Panagariya observó que el comercio internacional había crecido con fuerza en los países del "milagro" y poco en los del "fracaso", resultado que respalda la asociación entre apertura y crecimiento.

Sachs y Warner (1995): su índice clasifica una economía como cerrada si cumple al menos uno de cinco criterios, a saber, aranceles medios superiores al 40 %, barreras no arancelarias sobre

más del 40 % de las importaciones, una prima del mercado negro de divisas superior al 20 %, un monopolio estatal de las principales exportaciones o un sistema económico socialista. Entre 1970 y 1989, las economías en desarrollo abiertas crecieron en torno al 4,5 % anual per cápita, frente al 0,7 % de las cerradas. Rodríguez y Rodrik (2000) mostraron, sin embargo, que el resultado descansa sobre todo en la prima del mercado negro y en el monopolio exportador, que recogen más la inestabilidad macroeconómica y la calidad institucional que la política comercial.

Wacziarg y Welch (2008): en una línea complementaria, actualizaron el índice de Sachs-Warner con episodios de liberalización comercial. Estimaron que los países que liberalizan crecen, en promedio, en torno a 1,5 puntos porcentuales más al año que antes de la liberalización, con aumentos asociados de la inversión y de la ratio comercio/PIB.

Canales: la evidencia con datos de países y de empresas contrasta los mecanismos de los apartados 2.3 y 3.1:

- *Difusión de la I+D*: Coe y Helpman (1995) encuentran que la productividad total de los factores de un país depende también del capital de I+D de sus socios comerciales, ponderado por las importaciones, y tanto más cuanto más abierto es el país.

- *Insumos importados*: Amiti y Konings (2007) muestran, con datos de Indonesia, que la rebaja de los aranceles sobre los insumos eleva la productividad de las empresas que los importan al menos el doble que una rebaja equivalente de los aranceles sobre los productos finales. Por su parte, Goldberg, Khandelwal, Pavcnik y Topalova (2010) hallan en la India que la rebaja de los aranceles sobre los insumos explica, en promedio, el 31 % de los productos nuevos que lanzan las empresas, sobre todo porque les da acceso a variedades de insumos antes no disponibles.

- *Reasignación*: Pavcnik (2002) muestra que, tras la liberalización chilena, la productividad agregada subió en buena parte porque los recursos pasaron de las plantas menos productivas a las más productivas, como predice el modelo de Melitz (apartado 2.3.1).

Conclusiones

Los efectos del crecimiento sobre el comercio dependen del sesgo del crecimiento, que refleja su origen factorial o tecnológico, y del tamaño del país. En una economía pequeña con libre comercio, el crecimiento eleva siempre el bienestar. En una economía grande, en cambio, el deterioro de la relación real de intercambio puede superar las ganancias y dar lugar al crecimiento empobrecedor, un argumento que se invocó históricamente a favor de la sustitución de importaciones y del arancel óptimo. Por su parte, los efectos del comercio sobre el crecimiento operan a través de canales que la literatura ha precisado, desde la especialización hasta la difusión tecnológica y la reasignación hacia las empresas más productivas, y pueden elevar el nivel de renta o su tasa de crecimiento. Los modelos de crecimiento endógeno en economía abierta formalizan esos canales, y la evidencia respalda, con cautelas, una asociación positiva entre apertura y crecimiento.

El análisis estático se apoya en el modelo de Heckscher-Ohlin-Samuelson y en el teorema de Rybczynski, que relaciona la acumulación de un factor con la producción de cada bien (tema B05). Además, el crecimiento empobrecedor comparte su lógica con el arancel óptimo, porque ambos operan a través de la relación real de intercambio (tema B07). A su vez, los efectos del comercio sobre el crecimiento remiten a la reasignación hacia las empresas más productivas del modelo de Melitz (tema B06) y a la difusión de ideas de los modelos de crecimiento endógeno (tema A44). Esos mecanismos explican también los efectos dinámicos de la integración (tema B10). Por último, el pesimismo exportador que inspiró la sustitución de importaciones reaparece en la economía del desarrollo (tema B28).

Arkolakis, Costinot y Rodríguez-Clare (2012) muestran que, en una amplia clase de modelos, las ganancias del comercio dependen solo de la cuota del gasto interior y de la elasticidad del comercio. La clase incluye el modelo de Krugman y el de Melitz con distribución de Pareto, y en ella la heterogeneidad de las empresas no añade ganancias. Melitz y Redding (2015) replican que, fuera de esos supuestos, la selección de empresas abre un canal de bienestar adicional. Sigue abierta, además, la solidez de la relación entre apertura y crecimiento cuando cambian la medida de la apertura y las variables de control (apartado 3.2).

B10. Teoría de la integración económica

Teoría de la integración económica.

Introducción

Se habla de **integración económica** cuando dos o más países acuerdan ir rebajando los obstáculos que entorpecen su comercio recíproco, con la vista puesta en ciertos fines económicos y, a menudo, también políticos. Así ocurrió con el Zollverein o con la Comunidad Europea del Carbón y del Acero (CECA) y, después, la Comunidad Económica Europea (CEE)[1].

Desde la teoría del ***second best***, la integración puede considerarse un sustitutivo imperfecto del libre comercio, entendido como una situación sin barreras al comercio de ningún tipo. El libre comercio ofrece ganancias potenciales de eficiencia. Sin embargo, la apertura total e incondicional de un país al mercado internacional puede producir efectos adversos, sobre todo cuando se extiende precipitadamente a los movimientos de capital, como puso de manifiesto la crisis del sudeste asiático de 1997. Por ello, en el mundo real los países optan a menudo por eliminar sus barreras de manera gradual, mediante acuerdos más o menos comprometidos, en una aproximación de *second best* al libre comercio.

El **proceso de integración** adopta distintas formas en función del grado de compromiso entre los países. Ampliando la clasificación clásica de Béla Balassa (1961), cabe distinguir los siguientes estadios:

1. *Acuerdo de comercio preferencial*: la forma más básica de integración, que consiste habitualmente en una reducción de aranceles, aunque no necesariamente en su eliminación. Ejemplo: la Developing 8 Organization[2].

2. *Área de libre comercio*: *dentro del área* no se cobran aranceles, pero frente al resto del mundo cada miembro conserva los suyos[3]. Ejemplo: el Tratado entre México, Estados Unidos y Canadá (T-MEC), sucesor del Tratado de Libre Comercio de América del Norte (TLCAN).

3. *Unión aduanera*: al libre comercio interior se suma un arancel exterior *común*, el mismo en todos los miembros. Ejemplos: la unión aduanera de la CEE, completada en 1968, y la Unión Aduanera de África Austral.

[1] La creación del Zollverein (1834), la unión aduanera entre Estados alemanes liderada por Prusia y sin Austria, allanó el camino para la unificación alemana de 1871. La "preferencia imperial" fue un intento promovido por Joseph Chamberlain a finales del siglo XIX de reforzar la unidad del Imperio británico. Por su parte, el Tratado de París (1951) establecía objetivos económicos pragmáticos (la reducción de las barreras al comercio de carbón y acero) que en última instancia perseguían un fin político, la paz.

[2] Bangladés, Egipto, Indonesia, Irán, Malasia, Nigeria, Pakistán y Turquía, a los que se sumó Azerbaiyán en 2024.

[3] Este tipo de integración requiere elaboradas normas de origen que garanticen que las importaciones no puedan aprovechar la diferencia de estructuras arancelarias, entrando por el país con menores aranceles y desplazándose dentro del área hacia los países con aranceles más elevados.

4. *Mercado común*: unión aduanera en la que, además, el trabajo y el capital circulan libremente. Ejemplo: Mercosur en su diseño, aunque en la práctica es una unión aduanera imperfecta con movilidad de factores limitada.

5. *Mercado único*: mercado común del que desaparecen *todas* las trabas técnicas, jurídicas o burocráticas que frenan *de facto* los intercambios de bienes y el traslado de trabajo y capital. Ejemplo: el mercado único europeo, en vigor desde 1993.

6. *Unión económica*: mercado único en el que los países coordinan, además, sus políticas económicas. Ejemplo: la coordinación de políticas de la UE en el marco del Semestre Europeo o el Pacto por el Euro Plus (2011)[4].

7. *Unión monetaria*: unión económica en la que los países adoptan, además, una moneda común. Ejemplo: la zona del euro, desde 1999.

8. *Unión fiscal y unión política*: los estadios últimos del proceso, en los que se centralizan la política fiscal y, en último término, el poder político.

La teoría de la integración surge al terminar la Segunda Guerra Mundial. El orden económico que se reconstruye entonces deja atrás los acuerdos bilaterales de entreguerras y combina dos lógicas. La multilateral da lugar al FMI, al Banco Mundial y al GATT, y la regional, a bloques comerciales que adoptan sobre todo la forma de uniones aduaneras. La multiplicación de acuerdos regionales desde los años noventa y el bloqueo de las negociaciones multilaterales sitúan hoy la disyuntiva entre **regionalismo** y **multilateralismo** en el primer plano de la política comercial.

Este tema aborda la teoría de la integración económica como un análisis coste-beneficio en términos de bienestar, en tres partes:

- *Efectos estáticos*: se examinan primero la creación y la desviación del comercio que genera la integración, cuyo efecto neto sobre el bienestar es ambiguo.

- *Efectos dinámicos*: el tema estudia después sus beneficios dinámicos.

- *Proceso de integración*: la última parte trata los costes del proceso, el paso al mercado común y el debate entre regionalismo y multilateralismo.

Un anexo recoge la evidencia empírica disponible.

[4] El Pacto por el Euro Plus fue un compromiso político adoptado en marzo de 2011 por veintitrés Estados miembros de la UE, con el objetivo de reforzar la coordinación de las políticas económicas y contener la crisis de deuda soberana de la zona del euro.

1. Efectos estáticos de la integración económica

En la literatura económica, el debate sobre los beneficios de la integración se puede dividir en dos etapas:

1. *Primera etapa (años cincuenta y sesenta)*: se desarrolla la teoría tradicional, centrada en el estudio de las uniones aduaneras. A este mismo marco pertenecen aportaciones posteriores que lo prolongan, como la crítica de los Wonnacott (1981) y el teorema de Kemp y Wan (1976).

2. *Segunda etapa (desde los años setenta)*: a partir de Corden (1972) y en línea con los desarrollos de la nueva teoría del comercio de finales de esa década, se levanta parte de los supuestos neoclásicos. El énfasis pasa a las economías de escala, la competencia imperfecta y los aspectos geográficos.

La teoría tradicional se centra especialmente en las uniones aduaneras porque, tras el final de la Segunda Guerra Mundial, aparecieron bloques regionales como la Comunidad Económica Europea (1958), fruto de los Tratados de Roma (1957). En esta etapa destacan las aportaciones de Jacob Viner, James Meade, Richard Lipsey, Charles Cooper, Benton Massell, Robert Mundell y Murray Kemp.

1.1. Creación y desviación del comercio (Viner, 1950)

Origen: el punto de partida de este debate es la aportación de Jacob Viner en *The Customs Union Issue* (1950), que pone de manifiesto la ambigüedad que, en términos de bienestar, puede suponer la integración económica.

Idea: el autor sostiene que el efecto de una unión aduanera sobre el bienestar es, *a priori*, ambiguo, es decir, respecto a una situación de aranceles no discriminatorios, el bienestar puede tanto aumentar como disminuir.

1.1.1. Ideas previas

Hasta la aportación de Viner prevalecían dos ideas fundamentales:

1. *Unión aduanera y libre comercio*: el bienestar mundial es menor con la unión aduanera que con el libre comercio, dado que su arancel exterior común introduce una distorsión asignativa.

2. *Unión aduanera y reducción no discriminatoria*: se daba por hecho que la unión aduanera superaba siempre a una estructura arancelaria no discriminatoria (la que fija aranceles *erga omnes*), porque se la veía como un avance hacia el libre comercio. La unión suprime la distorsión entre los socios, aunque persista la que afecta al comercio con terceros países.

1.1.2. Razonamiento

La aportación de Viner a este debate tiene que ver con la identificación de dos efectos contrapuestos en la configuración de cualquier unión aduanera:

1. *Creación de comercio*: se refiere a la sustitución de la producción nacional de un país por importaciones procedentes de un país socio en el marco de la unión aduanera. Esta sustitución se debe al abaratamiento relativo de las importaciones del socio y da lugar a nuevo comercio entre los socios, con la consiguiente ganancia de eficiencia, al desplazarse la producción hacia el productor más eficiente de la unión.

2. *Desviación de comercio*: se refiere a la reorientación de las importaciones desde un tercer país con bajos costes de producción hacia un país socio con mayores costes de producción. Se trata de una derivación del comercio desde exportadores eficientes hacia exportadores más ineficientes, con la consiguiente pérdida de eficiencia.

En función de cuál de estos dos efectos prevalezca, se producirá una ganancia o una pérdida de bienestar global. La integración tampoco garantiza una mejora paretiana, es decir, del bienestar de todos los países respecto a la situación de partida.

Gráficamente, ambos efectos se representan en el plano cantidad-precio del bien, desde la óptica del país A (figura B10.1):

1. *Curva de oferta de A*, $S^A(y)$: recoge la producción de A a cada precio y es creciente, porque su coste marginal es creciente.

2. *Curva de demanda de A*, $D^A(y)$: recoge el consumo de A y es vertical en $\bar{y}$, porque Viner supone un consumo fijo.

3. *Precios de B y del resto del mundo*: horizontales en p^B y p^R, con $p^R < p^B$, porque ambos ofrecen a A cualquier cantidad a un precio constante, y, con el arancel específico t_e (por unidad), en $p^B + t_e$ y $p^R + t_e$.

4. *Equilibrio inicial*: con un arancel no discriminatorio, A importa del resto del mundo al precio $p^R + t_e$, produce y_0^s (punto 0) y consume $\bar{y}$ (punto 0′), de modo que importa M_0.

Con la unión aduanera, A pasa a importar de B al precio p^B, produce y_{UA}^s (punto 1) y consume $\bar{y}$ (punto 1′), con lo que importa M_{UA}. El área B mide la ganancia por creación de comercio y el área E, la pérdida por desviación de comercio.

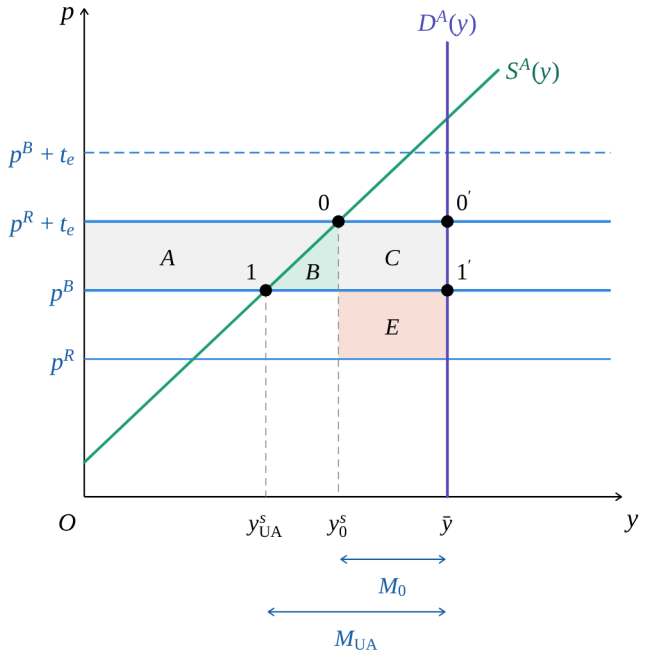

Figura B10.1: Creación y desviación de comercio (Viner, 1950) como efectos de una unión aduanera sobre el bienestar del país miembro.

Un ejemplo histórico es el de los productos agrícolas neozelandeses (mantequilla y cordero) en el Reino Unido. Tras su adhesión a la CEE en enero de 1973, el Reino Unido adoptó el arancel exterior común y la política agrícola común. Las dos medidas encarecieron las importaciones neozelandesas y desviaron parte de la demanda británica hacia proveedores comunitarios de mayor coste, que gozaban de libre acceso. Nueva Zelanda solo conservó contingentes específicos para algunos productos, como la mantequilla, y el Reino Unido aplicó el arancel exterior común y la PAC hasta el final del periodo transitorio del Brexit, el 31 de diciembre de 2020.

1.1.3. Conclusiones

Aspectos positivos: la aportación de Viner dio lugar a una amplia literatura sobre las uniones aduaneras, tanto teórica como empírica, como el estudio de Béla Balassa (1967) sobre los efectos creación y desviación de comercio en la Comunidad Económica Europea.

Limitaciones: por un lado, el estudio solo considera las barreras arancelarias y no analiza el papel de las barreras no arancelarias. Por otro, el análisis de Viner deja fuera la demanda, y esa es su carencia más grave. En él, el consumo queda fijado en $\bar{y}$ (véase la figura) y no reacciona a los precios relativos, ni antes ni después de la integración, con lo que desaparece el efecto sustitución.

1.2. Modelo canónico neoclásico de integración económica

Los primeros trabajos posteriores a la obra de Viner, entre los que destacan los de James Meade (1955a) y de Richard Lipsey (1957), se centraron, precisamente, en los efectos de la integración sobre el consumo. La consideración del lado de la demanda da lugar a una serie de efectos adicionales:

1. *Creación de consumo*: si la integración permite acceder a bienes con menores precios, la demanda nacional de productos extranjeros se incrementa, y el efecto creación opera, por tanto, también por el lado del consumo.

2. *Reorientación del consumo* (*intercountry substitution*): también se produce un efecto desviación por el lado del consumo, consistente en la reorientación del consumo desde terceros países hacia los países de la unión.

3. *Sustitución del consumo* (*intercommodity substitution*): si existen varios productos sustituibles entre sí, hay que considerar además los efectos sustitución y los cambios en los patrones de consumo.

1.2.1. Enfoque de equilibrio parcial (Meade, 1955a)

Origen: estas ideas se plantean primero en un análisis gráfico de equilibrio parcial basado en la obra de James Meade (1955a).

Idea: al incorporar una demanda con pendiente negativa, la unión aduanera genera, junto a la creación y la desviación de comercio, un efecto creación de consumo, y su efecto neto sobre el bienestar sigue siendo ambiguo.

Supuestos:

- *Entorno*: modelo neoclásico sin imperfecciones con tres países, dos de los cuales (A y B) van a formar una unión aduanera y son pequeños en relación con el tercero (R), que puede entenderse como el resto del mundo.

- *Agentes*: se supone que la demanda no es totalmente inelástica y, por tanto, tiene pendiente negativa.

- *Tecnología*: se considera que el coste marginal de A (es decir, su curva de oferta) no es constante, sino creciente, mientras que B y R ofrecen el bien a A a precios constantes, p^B y p^R.

Planteamiento: se compara el bienestar de A en dos situaciones. En la primera, una estructura arancelaria no discriminatoria grava por igual a B y al resto del mundo con el arancel específico t_e. En la segunda, una unión aduanera con B suprime el arancel entre los socios y lo mantiene frente al resto del mundo como arancel exterior común.

Desarrollo gráfico: los efectos de la unión aduanera se representan en el plano cantidad-precio, desde la óptica del país A (figura B10.2):

1. *Curva de oferta de A*, $S^A(y)$: recoge la producción de A a cada precio y es creciente, porque su coste marginal es creciente.

2. *Curva de demanda de A*, $D^A(y)$: recoge el consumo de A a cada precio y es decreciente, porque la demanda no es totalmente inelástica (Supuestos).

3. *Precios de B y del resto del mundo*: horizontales en p^B y p^R, porque ambos ofrecen el bien a A a precios constantes (Supuestos), y, con el arancel específico t_e, en $p^B + t_e$ y $p^R + t_e$.

4. *Equilibrio inicial* (estructura arancelaria no discriminatoria): el país A grava por igual al país B y al resto del mundo y cubre su exceso de demanda importando solo de este último, puesto que $p^R + t_e < p^B + t_e$. Produce y_0^s (punto 0), consume y_0^d (punto 0′) e importa $M_0 = y_0^d - y_0^s$. La recaudación arancelaria equivale a las áreas C+E.

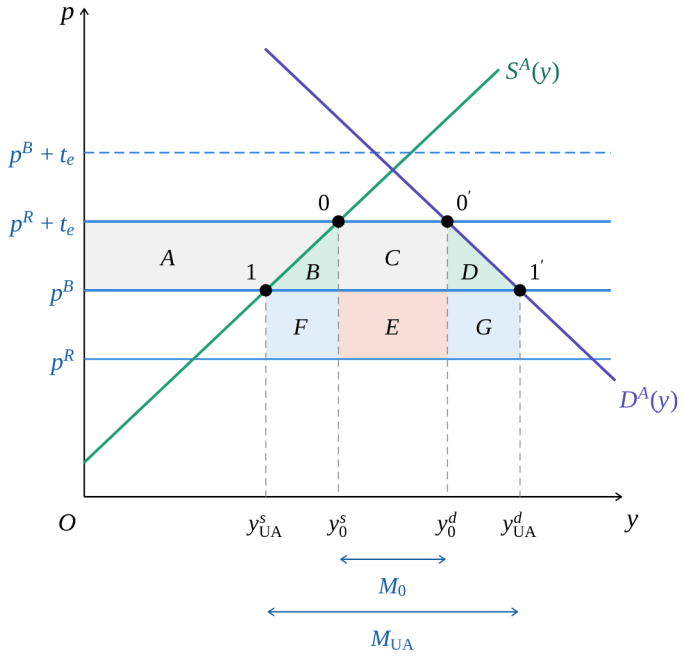

Figura B10.2: Modelo canónico de integración en equilibrio parcial (Meade, 1955a).

Estática comparativa: cuando A y B forman la unión aduanera, desaparecen los aranceles entre ellos y se fija un arancel exterior común frente al resto del mundo. Dado que el precio del bien importado de B sin arancel es inferior al del bien importado del resto del mundo con arancel ($p^B < p^R + t_e$), el país A pasa a importar solamente del país B. En consecuencia, los consumidores de A aumentan su consumo hasta y^d_{UA} (punto 1′) y los productores reducen su producción hasta y^s_{UA} (punto 1), de modo que las importaciones aumentan de M_0 a $M_{\text{UA}} = y^d_{\text{UA}} - y^s_{\text{UA}}$.

Bienestar: los efectos brutos de la integración sobre el bienestar de A son tres:

1. *Excedente del consumidor*: aumenta en las áreas A+B+C+D (efecto expansión del comercio).

2. *Excedente del productor*: disminuye en el área A, dado que los productores de A son menos competitivos que los de B.

3. *Excedente del Gobierno*: disminuye en C+E, puesto que ya no se importa del resto del mundo y, aunque el arancel exterior sigue vigente, su recaudación pasa a ser nula.

El efecto neto sobre la economía A es $(A + B + C + D) - A - (C + E) \Rightarrow \text{EN} = B + D - E$. En detalle, se distinguen tres efectos:

1. *Efecto creación de comercio* (área B): se corresponde con la ganancia de eficiencia productiva, pues la producción $y^s_0 - y^s_{\text{UA}}$ pasa de los productores de A a los del país B, que son más eficientes.

2. *Efecto creación de consumo* (área D): se corresponde con el excedente que ganan los consumidores de A que, al caer el precio, pasan a consumir cuando antes no lo hacían.

3. *Efecto desviación de comercio* (área E): se corresponde con la pérdida de eficiencia sobre la cantidad que antes abastecía el resto del mundo y que ahora produce el país B, menos eficiente que aquel.

De nuevo, se trata de una situación en la que no se puede hacer una predicción inequívoca sobre los efectos de la integración. Si la caída arancelaria es muy contundente y el precio del socio (p^B) no es mucho mayor que el del resto del mundo, el efecto desviación (área E) será pequeño en relación con el efecto creación (áreas B+D). En ese caso, la unión aduanera dará lugar a ganancias netas de bienestar. Sin embargo, bien podría darse la situación opuesta si el efecto desviación fuese dominante.

Conclusiones:

- *Aspectos positivos*: el análisis permite ilustrar gráficamente las intuiciones planteadas por Viner, incorporando el lado de la demanda.

- *Limitaciones*: se queda en el equilibrio parcial y deja fuera, por tanto, la posible sustitución cruzada entre mercados.

1.2.2. Enfoque de equilibrio general (Lipsey, 1957)

Origen: Richard Lipsey trasladó por primera vez el análisis de las uniones aduaneras al marco del equilibrio general.

Idea: en equilibrio general, la comparación entre una unión aduanera y un arancel no discriminatorio depende de la relación real de intercambio (RRI) con el socio, por lo que tampoco admite conclusiones inequívocas.

Supuestos: el país A, pequeño, produce con productividad marginal constante un bien exportable, X, y otro importable, Y. Su frontera de posibilidades de producción (FPP) es, en consecuencia, una recta, y A se especializa por completo.

Planteamiento: se compara el bienestar de A en tres situaciones, el libre comercio, un arancel no discriminatorio sobre las importaciones de Y procedentes del resto del mundo y una unión aduanera con B, cada una con los precios relativos que implica para el consumo.

Desarrollo gráfico: las tres situaciones se representan en el plano bien X-bien Y (figura B10.3):

1. *Recta de la RRI de libre comercio*, 1-2: recoge los precios internacionales y parte del punto de producción, 1. Su pendiente es el precio relativo internacional, y sobre ella están las cestas que A puede consumir con libre comercio.

2. *Curvas de indiferencia social*, CIS_0 y CIS_1: representan un bienestar tanto mayor cuanto más alejadas del origen.

3. *Equilibrio inicial*: la situación de partida es el libre comercio, con el que A produce en el punto 1, especializado por completo en el bien X, y alcanza su consumo óptimo en el punto LC, sobre la CIS_0.

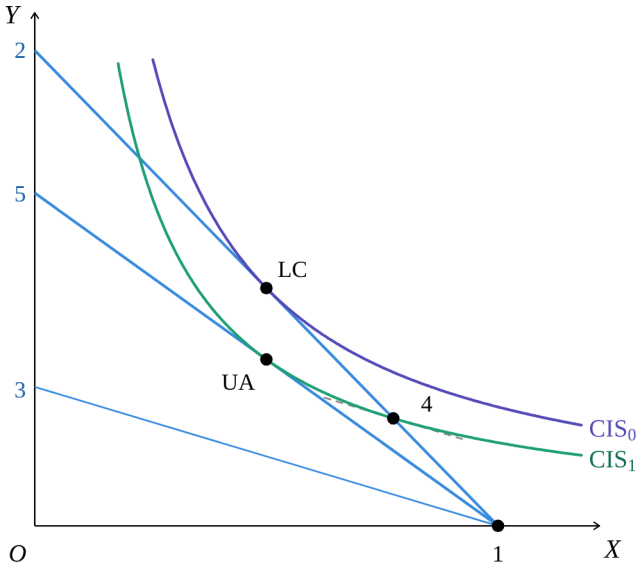

Figura B10.3: Modelo canónico de integración en equilibrio general (Lipsey, 1957).

Estática comparativa:

▪ *Arancel*: si A impone un arancel a las compras de Y al resto del mundo, Y se encarece en términos relativos y los precios interiores pasan a ser los de la línea 1-3, menos inclinada en el plano (X, Y). El consumo se desplaza entonces a un punto que cumple dos requisitos:

1. *Tangencia*: los consumidores deciden con el precio relativo interno, y el óptimo exige que la relación marginal de sustitución coincida con él (tangencia entre la línea de precios internos y la curva de indiferencia).

2. *Restricción externa*: la recaudación del arancel vuelve a los consumidores, así que el consumo ha de situarse en la RRI internacional que parte del punto de producción, el 1, que la especialización completa deja inalterado. Como A es pequeño, el arancel no modifica su RRI con el resto del mundo.

Ambas condiciones se cumplen en el punto 4, situado deliberadamente en la CIS_1, la curva de indiferencia por la que pasa también el punto UA de la unión aduanera.

▪ *Unión aduanera*: si A y B forman ahora una unión aduanera, A comercia con su socio según la línea 1-5, una RRI intermedia entre la de libre comercio, más favorable, y la relación de precios interiores con arancel (línea 1-3), menos favorable. Si con esa RRI A consume en UA, la unión no le aporta ni le resta bienestar frente al arancel no discriminatorio, dado que los dos puntos de consumo comparten curva de indiferencia. Esa igualdad es, sin embargo, solo un caso posible, porque el resultado depende de la RRI que A obtenga con el socio:

- *RRI con el socio peor que 1-5* (recta con menor pendiente): A no llega a la CIS_1 y su bienestar cae por debajo del que le daba el arancel no discriminatorio.
- *RRI con el socio mejor que 1-5* (recta con mayor pendiente): A alcanza una curva de indiferencia superior y queda mejor.

El equilibrio general, como antes el parcial, no permite decidir con carácter general qué es preferible, la unión aduanera o el arancel no discriminatorio. Lipsey lo atribuye a que ambas son situaciones de *second best* con respecto al libre comercio. Si se elimina solo una parte de las distorsiones y se mantienen otras, tanto en los países integrados como en los externos, nada garantiza que el bienestar mejore.

Conclusiones:

- *Aspectos positivos*: lleva el análisis de las uniones aduaneras al equilibrio general, con lo que recoge la sustitución entre bienes en el consumo que el enfoque parcial omite. También sitúa la teoría de la integración en el marco del *second best*, que desde entonces es su referencia.

- *Limitaciones*: descansa en supuestos muy restrictivos (dos bienes, productividad marginal constante con especialización completa y país pequeño). Además, es un análisis estático que deja fuera las economías de escala y la competencia imperfecta (apartado 2).

1.2.3. Determinantes del efecto neto

Del análisis anterior se desprende una serie de factores de los que depende el efecto neto de la integración económica (el signo entre paréntesis indica su relación con la ganancia de bienestar):

1. *Número de miembros de la unión aduanera* (+): cuantos más países participen, más probable resulta que la unión englobe a los productores más eficientes.

2. *Volumen de comercio previo entre los miembros* (+): si los socios ya comerciaban mucho entre sí, queda poco comercio por desviar.

3. *Volumen de comercio previo con el resto del mundo* (−): un comercio previo intenso con terceros países deja más compras que la unión puede desviar hacia los socios.

4. *Magnitud de los aranceles previos* (+): unos aranceles de partida altos dejan más margen para que la unión cree comercio.

5. *Magnitud del arancel exterior común* (−): un arancel común elevado frente a terceros amplía la desviación de comercio. El artículo XXIV del GATT exige que el arancel exterior común de una unión aduanera no sea, en su conjunto, más elevado ni más restrictivo que la incidencia general de los aranceles previos de sus miembros.

6. *Grado de sustituibilidad de los bienes de los miembros* (+): las economías parecidas, con mucho comercio intraindustrial potencial, ganan más por creación de comercio y pierden menos por desviación de comercio.

7. *Elasticidad-precio de la oferta y la demanda* (+): una oferta y una demanda más elásticas amplían la creación de comercio y la de consumo.

8. *Número de bloques comerciales* (ambiguo): Krugman (1991b) muestra que la relación con el bienestar mundial no es monótona. Un solo bloque mundial equivaldría al libre comercio y maximizaría el bienestar global. Al aumentar el número de bloques, el bienestar cae hasta alcanzar, en sus simulaciones, su mínimo con un número reducido de bloques (unos tres). A partir de ahí, la desviación de comercio disminuye y el bienestar vuelve a crecer con el número de bloques.

9. *Diferencia de costes entre el país (A) y su socio (B)* (+): cuanto menos eficiente sea A respecto a B, mayor la ganancia por creación de comercio.

10. *Diferencia de costes entre el socio (B) y el resto del mundo (R)* (−): cuanto menos eficiente sea B respecto a R, mayor la pérdida por desviación de comercio.

1.3. Extensiones al modelo canónico

A raíz de esta ambigüedad surge una serie de estudios posteriores que buscan determinar si existen fórmulas de integración preferibles a las uniones aduaneras, teniendo en cuenta elementos adicionales como el tamaño de los países o las políticas de compensación.

1.3.1. Tamaño de los países y relación real de intercambio (Mundell, 1964)

Origen: una de las primeras críticas se debe a Robert Mundell (1964). Señala que una de las principales limitaciones de la teoría tradicional es suponer que los países no pueden alterar la RRI, ni individualmente ni como unión.

Idea: el tamaño de los países influye en los efectos de la integración económica, puesto que determina su capacidad para alterar la RRI.

Razonamiento: las conclusiones cambian cuando los países tienen un tamaño relevante, tanto uno frente a otro como frente al resto del mundo. El análisis supone, además, que todos los bienes son sustitutivos brutos.

Con la unión, A compra a B una parte mayor de algunos bienes, lo que equivale a discriminar en su favor. B ve así crecer la demanda de sus exportaciones, y su RRI mejora tanto frente a A como frente al resto del mundo. La preferencia es recíproca, pues en otros bienes B favorece a su vez a A, cuya RRI mejora por la misma vía.

Entre los socios, el cambio en la RRI es de signo ambiguo, porque depende de cuál de las dos preferencias pese más. Frente a terceros, en cambio, no hay ambigüedad. Los miembros compran menos fuera y, dado el peso de la unión en el comercio mundial, esa retracción de la demanda abarata las exportaciones de terceros países, con lo que la RRI de la unión mejora.

Si a ello se añade la fijación de un arancel exterior común, el efecto puede ser aún mayor. La mejora de la RRI constituye, así, una ganancia adicional de la integración para los países grandes, invisible en el análisis de país pequeño.

Conclusiones:

- *Aspectos positivos*: incorpora el poder de mercado al análisis de la integración, anticipando la literatura sobre aranceles óptimos de las uniones.

- *Limitaciones*: la ganancia es, a escala global, un juego de suma cero (o negativa, por las distorsiones del arancel), obtenida a costa de terceros, por lo que invita a la retorsión. Además, resulta irrelevante para las uniones pequeñas, que carecen de peso en los mercados mundiales para alterar la RRI.

1.3.2. Reducciones arancelarias unilaterales (Cooper y Massell, 1965)

Origen: Charles Cooper y Benton Massell (1965) plantean que es preferible obtener el efecto creación de comercio mediante reducciones arancelarias unilaterales y no discriminatorias.

Idea: siempre que la reducción de aranceles sea no discriminatoria, esta produce las ganancias de la unión aduanera sin sus pérdidas.

Razonamiento: sobre el gráfico de equilibrio parcial de Meade (figura B10.2), el país A, en vez de formar una unión aduanera, reduce su arancel frente a todo el mundo hasta $t'_e = p^B - p^R$, es decir, aplica una política no preferencial tal que $p^R + t'_e = p^B$. Como el precio interior coincide exactamente con el que resultaría si se formara la unión aduanera, consumidores y productores disfrutan de los mismos excedentes que con ella.

La diferencia con la unión está en que A sigue comprando al resto del mundo, el proveedor más barato, y en que la recaudación arancelaria pasa de C+E a E+F+G. Para el país A el resultado neto es, sin ambigüedad, una mejora del bienestar (EN $= B + D + F + G$). Dicho de otro modo, una reducción arancelaria unilateral y no discriminatoria obtiene el efecto creación de comercio sin el efecto desviación y, por tanto, es preferible a una unión aduanera.

Conclusiones:

- *Aspectos positivos*: la aportación convierte la reducción arancelaria unilateral y no discriminatoria en la referencia con la que debe compararse toda unión aduanera.

- *Limitaciones*: omite la reciprocidad que ofrece la unión aduanera.

1.3.3. Efectos de la respuesta exterior (Wonnacott y Wonnacott, 1981)

Origen: Paul y Ronald Wonnacott (1981) critican precisamente este punto de la aportación de Cooper y Massell, pues el análisis de Cooper y Massell omite los aranceles del resto del mundo y del socio (los "aranceles extranjeros ausentes").

Idea: la comparación entre unión aduanera y liberalización unilateral cambia cuando se reconoce que los demás países también mantienen aranceles, pues solo la unión permite negociar su eliminación recíproca.

Razonamiento: si el socio B mantiene aranceles, la unión aduanera garantiza a los exportadores de A un acceso preferente al mercado de B que una reducción unilateral de A no puede conseguir. Esta reciprocidad mejora la RRI de A frente a su socio. Además, si los países de la unión actúan como un bloque grande frente al resto del mundo, pueden mejorar tanto su RRI como su bienestar conjunto.

Conclusiones:

- *Aspectos positivos*: rehabilita la unión aduanera frente a la crítica de Cooper y Massell al incorporar los aranceles ajenos, y ofrece la justificación teórica del valor de la reciprocidad, fundamento de la negociación comercial real.

- *Limitaciones*: las ganancias de acceso preferente dependen de que el socio mantenga aranceles significativos, por lo que el argumento pierde fuerza a medida que la liberalización multilateral reduce la protección, y la mejora de la RRI del bloque sigue obteniéndose a costa de terceros.

1.3.4. Uniones aduaneras con políticas de compensación: el teorema de Kemp y Wan (1976)

Origen: Murray Kemp y Henry Wan (1976) argumentan que las uniones aduaneras se pueden construir de tal forma que mantengan el bienestar de los países no participantes.

Idea: con políticas de compensación adecuadas, la unión aduanera es Pareto-superior, es decir, puede mejorar a sus miembros sin empeorar a nadie.

Razonamiento: el teorema establece que siempre existe un arancel exterior común que, combinado con un sistema de transferencias a tanto alzado entre los miembros, permite mejorar a los países que forman la unión sin empeorar al resto. Ese arancel exterior común mantiene inalterado el comercio (y los precios) con el resto del mundo, lo que, en la práctica, se aproxima a la exigencia del art. XXIV del GATT de no elevar la protección frente a terceros.

Conclusiones:

▪ *Aspectos positivos*: ofrece un fundamento teórico a la disciplina que el artículo XXIV del GATT (1947) ya imponía a los procesos de integración. De hecho, la creación de la unión aduanera de la CEE llevó al GATT a exigir compensaciones, y la Comunidad Andina desarrolló mecanismos para compensar a los países pequeños perjudicados.

▪ *Limitaciones*: es un teorema de existencia, que no describe cómo se forman las uniones reales, y es poco probable que las compensaciones internas a tanto alzado que exige se lleven a cabo plenamente en la práctica.

1.3.5. Áreas de libre comercio y normas de origen (Krueger, 1993)

Origen: Hirofumi Shibata (1967) extiende el análisis de las uniones aduaneras a las áreas de libre comercio (ALC), en las que cada miembro conserva su propio arancel frente a terceros, y Anne Krueger (1993) estudia las normas de origen que las sostienen.

Idea: sin arancel exterior común, las importaciones de terceros entrarían por el miembro de arancel más bajo y se reexportarían al resto (*deflexión de comercio*, *trade deflection*). Para evitarlo, el ALC exige *normas de origen*, que fijan qué bienes se consideran producidos en la zona y pueden circular en ella sin arancel.

Razonamiento: la diferencia con la unión aduanera tiene tres consecuencias:

1. *Normas de origen como protección*: para cumplirlas, el productor del bien final debe comprar en la zona una parte de sus insumos, aunque importarlos fuera más barato. La protección del bien final se extiende así a los insumos del socio, y la desviación del comercio, a los mercados de insumos (Krueger, 1993). A ello se suma el coste de acreditar el origen, que lleva a muchas empresas a pagar el arancel de nación más favorecida en lugar de pedir la preferencia.

2. *Competencia arancelaria*: como las importaciones de terceros tienden a entrar por el miembro de arancel más bajo, cada miembro tiene un incentivo a rebajar su arancel exterior para captar el comercio y los ingresos arancelarios (Richardson, 1995). Frente a la unión aduanera, cuyo arancel común se fija de forma conjunta y escapa a esa competencia, el ALC puede así generar menos desviación del comercio, aunque a costa de las distorsiones de las normas de origen.

3. *Versión del teorema de Kemp y Wan*: Panagariya y Krishna (2002) muestran que también existe un ALC que mejora a sus miembros sin perjudicar al resto del mundo, si cada miembro fija sus aranceles para mantener constante su comercio con terceros y las normas de origen se eligen de forma adecuada.

Conclusiones:

- *Aspectos positivos*: explica por qué la mayoría de los acuerdos regionales son ALC y no uniones aduaneras, ya que preservan la autonomía de cada miembro en su política comercial, y por qué las normas de origen son una pieza central de su negociación.

- *Limitaciones*: el balance de bienestar frente a la unión aduanera es ambiguo y depende de la estructura arancelaria de cada miembro y de lo restrictivas que sean las normas de origen, difíciles de medir.

1.4. Conclusiones

El análisis de Cooper y Massell padece dos carencias que comparte toda la teoría tradicional de la integración. La primera son sus supuestos neoclásicos, muy restrictivos, que ignoran fenómenos presentes en el comercio mundial, entre ellos las economías de escala, la competencia imperfecta o la diferenciación de productos. La segunda es su carácter estático, que le impide recoger los efectos dinámicos de la integración, objeto del apartado 2. Esas omisiones, junto con la de los aranceles ajenos que señalan los Wonnacott, resuelven la aparente paradoja de que se formen uniones aduaneras pese a que, según el propio análisis, la rebaja arancelaria unilateral sería preferible.

2. Efectos dinámicos de la integración económica

A raíz de las limitaciones del apartado 1.4 surgieron una serie de desarrollos que ofrecieron nuevas ideas a partir del levantamiento de algunos supuestos neoclásicos. El principal efecto de la integración económica es el aumento del tamaño de mercado, del que se derivan cuatro grandes familias de efectos dinámicos.

2.1. Efecto economías de escala: costes medios decrecientes (Corden, 1972)

Origen: la teoría tradicional de la integración contemplaba costes marginales constantes o crecientes, y el enfoque de equilibrio general de Lipsey, en concreto, supone productividad marginal constante. En muchas industrias, sin embargo, cada unidad sale más barata cuanto más se produce, porque las empresas disfrutan de economías de escala internas (costes medios decrecientes). Max Corden (1972) sostiene que con ellas la creación y la desviación de comercio conservan su validez y que a su lado surgen dos efectos nuevos.

Idea: la unión aduanera aumenta el tamaño de mercado y, por ello, ofrece ganancias potenciales derivadas de la explotación de las economías de escala. En este contexto aparecen dos efectos contrapuestos:

1. *Efecto reducción de costes*: un efecto positivo derivado del aprovechamiento de la escala de producción sobre un mercado mayor.

2. *Efecto supresión de comercio*: un efecto negativo, debido a que la protección de la unión permite nacer y sobrevivir a empresas que, sin ella, la oferta del resto del mundo habría desplazado.

Supuestos:

- *Entorno*: tres países (A, B y R) comercian un solo bien. A y B forman la unión aduanera y R es el resto del mundo. En el caso de referencia, los aranceles de partida son prohibitivos, de modo que cada país abastece por completo su mercado interno, mientras que en los otros dos casos algún país se abastece de R con arancel.

- *Agentes*: la demanda se considera lineal y existe simetría, de manera que las curvas de demanda son idénticas en A y en B.

- *Tecnología*: los costes unitarios se ordenan como en Viner, $p^R < p^B < p^A$, es decir, el resto del mundo produce más barato que el socio (B), y este, más barato que el propio país (A). En el primer caso, el arancel específico, t_e, es prohibitivo, $p^A \leq p^R + t_e$[5]. El coste medio decrece con la escala (economías de escala), y cada productor ofrece según su coste medio, que hace así de curva de oferta[6].

Razonamiento: los efectos sobre el bienestar de la unión aduanera dependen del patrón de producción previo a la unión. Por ello, el análisis distingue tres situaciones de partida, según haya un productor en cada país, en uno solo o en ninguno. No se considera el escenario de varios productores en un mismo mercado, ya que, al haber economías de escala, solo puede existir, a lo sumo, un productor en cada mercado.

Desarrollo gráfico: los efectos de la unión aduanera se representan en el plano cantidad-precio y coste (figura B10.4), para el primer caso:

[5] Formada la unión aduanera, $p^R < p^{\text{UA}} < p^B < p^A \leq p^R + t_e$, porque las economías de escala sitúan el precio de la unión por debajo del precio autárquico de B.

[6] Con economías de escala el coste marginal es inferior al medio, por lo que un precio igual al coste marginal generaría pérdidas. Por ello, el modelo supone que el precio se fija al nivel del coste medio.

1. *Curvas de demanda*, $D^A(y) = D^B(y)$: recogen el consumo de cada país a cada precio, son lineales y decrecientes e idénticas para ambos países (Supuestos).

2. *Curva de coste medio de A*, $\text{Cme}^A(y)$: decreciente, por las economías de escala internas.

3. *Curva de coste medio de B*, $\text{Cme}^B(y)$: también decreciente y por debajo de la de A, al ser B más eficiente.

4. *Oferta del resto del mundo*: infinitamente elástica al precio p^R, por debajo de ambas curvas de coste medio, y al precio $p^R + t_e$ con el arancel.

5. *Equilibrio inicial*: en la situación de partida, con aranceles prohibitivos, cada país abastece por completo su mercado interno, y se determinan dos equilibrios autárquicos (puntos 0^A y 0^B, a los precios p^A y p^B). El punto 0^R señala el consumo de referencia al precio mundial p^R, el que se alcanzaría con libre comercio.

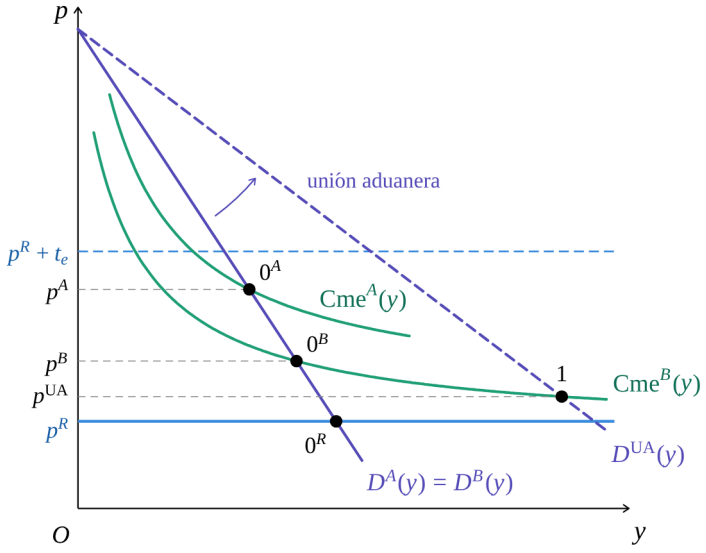

Figura B10.4: Efectos de la integración con economías de escala (Corden, 1972). Con un productor en cada país, la unión genera creación de comercio y un efecto reducción de costes.

Estática comparativa: con la unión aduanera, el productor de B, que tiene costes menores que el de A, pasa a abastecer también todo el mercado de A. La curva de demanda a la que se enfrenta el productor de B pivota entonces sobre el eje de ordenadas hacia afuera, pues ahora agrupa la demanda de ambos mercados ($D^{UA} = D^A + D^B$), y se alcanza un equilibrio en el punto 1, al precio p^{UA}. Esto implica dos efectos:

1. *Efecto creación de comercio*: el productor de A cierra y todo el consumo de A se cubre con importaciones procedentes de B, más baratas, lo que eleva el excedente del consumidor de A.

2. *Efecto reducción de costes*: la mayor escala de B, que ahora sirve a los dos mercados, abarata su coste medio por las economías de escala internas. Esa ganancia adicional beneficia a los consumidores de ambos países, cuyo excedente aumenta.

Faltan, en cambio, la desviación de comercio, porque A no compraba a R, y la supresión de comercio, porque ninguna empresa nueva surge al amparo de la unión. Con un productor en cada país antes de la unión, el balance es, por tanto, inequívocamente positivo. Los otros dos casos se analizan del mismo modo:

- *Un solo productor, en B*: de partida, B está en autarquía y A, sin productor, comercia con aranceles y se abastece por completo de R, pues $p^R < p^B < (p^R + t_e) < p^A$. Constituida la unión aduanera, A se abastece de B, dado que $p^B < p^R + t_e$. No hay creación de comercio, porque A carecía de producción propia, ni supresión de comercio, porque no nace ninguna empresa nueva. Sí hay desviación de comercio, porque las compras de A pasan de R, el oferente más eficiente, a B, y reducción de costes, porque la producción se concentra en B y aprovecha las economías de escala. En consecuencia, los efectos son ambiguos.

- *Ningún productor*: de partida, A y B comercian con aranceles y se abastecen por completo de R. Constituida la unión aduanera, surge en B un productor que atiende su propio mercado y el de A. No hay creación de comercio, porque ninguno de los dos países producía, ni reducción de costes, porque la producción no se concentra, sino que se traslada íntegra de R a B. Sí hay desviación de comercio, porque las importaciones pasan de R a B, y supresión de comercio, porque la unión permite que nazca y se mantenga una empresa que, sin ella, quedaría desplazada por la oferta del resto del mundo. Por ello, los efectos son negativos, salvo que la creación de consumo, al caer el precio de $p^R + t_e$ a p^{UA}, compense la desviación y la supresión de comercio.

El modelo de Corden lleva así a una regla según el número de productores tras la unión. Si se reduce, el bienestar mejora, aunque menos de lo posible en caso de especialización perversa (apartado 3.1.3). Si no varía, el resultado es ambiguo. Si aumenta, el bienestar empeora, salvo que la creación de consumo compense la desviación y la supresión de comercio.

Conclusiones:

- *Aspectos positivos*: incorpora las economías de escala internas a la teoría de las uniones aduaneras, con los efectos reducción de costes y supresión de comercio, y da así un fundamento formal a las ganancias de la integración ligadas al aumento del tamaño de mercado.

- *Limitaciones*: es un análisis de equilibrio parcial con un solo bien, demandas idénticas, a lo sumo un productor por mercado y precios iguales al coste medio, sin un tratamiento explícito de la competencia imperfecta, que incorporan los modelos posteriores. Además, supone que la producción se concentra en el socio más eficiente, algo que no está garantizado.

2.2. Efectos procompetitivos y de variedad: modelo de Krugman (1979b)

La escala no es el único canal dinámico de la integración, que afecta también a la variedad de bienes disponible y a la intensidad de la competencia.

> El modelo de competencia monopolística en economía abierta (Krugman, 1979b) se desarrolla en el tema B06, apartado 2.1.2.

En un contexto de competencia imperfecta y preferencia por la variedad, el aumento del tamaño de mercado que supone la integración produce dos familias de efectos, que se pueden ilustrar a partir del modelo de Krugman (1979b):

1. *Efecto variedad o innovación en bienes finales*: si se entiende la innovación como un aumento del número de variedades disponibles, el mayor tamaño de mercado conduce a un aumento del número de variedades al alcance de los consumidores, con la consiguiente ganancia de bienestar.

2. *Efectos procompetitivos*: la integración promueve la competencia entre empresas a escala internacional y, con ella, la "disciplina de mercado". Esta disciplina produce, por un lado, un *efecto precio*, por el que la mayor competencia eleva la elasticidad percibida de la demanda y reduce los márgenes (*mark-ups*) sobre el coste marginal, con la consiguiente caída de los precios (véase la figura del tema B06, apartado 2.1.2). Por otro, unos *efectos no precio*:

- *Ineficiencia-X*: según Leibenstein, se reduce gracias al esfuerzo adicional que aplican empresas y trabajadores para sobrevivir en un entorno más competitivo.

- *Búsqueda de rentas*: los grupos de presión nacionales pueden perder poder, lo que puede reducir los comportamientos de búsqueda de rentas (p. ej. un *lobby* sectorial pierde influencia cuando el país puede abastecerse del exterior).

- *Innovación*: el efecto es ambiguo, pues la competencia puede estimularla y, a la vez, desincentivarla al comprimir los márgenes con los que se financia.

2.3. Efectos sobre el crecimiento: variedades (Romer, 1990) y desbordamientos (Lucas, 1988)

Dos efectos vinculan la integración con el crecimiento a largo plazo, el de innovación en bienes intermedios y el de desbordamiento. Ambos se derivan de que la integración crea un mercado mayor que el nacional y se ilustran con los modelos de crecimiento endógeno.

> Los modelos de crecimiento endógeno se desarrollan en el tema A44, apartados 1.3 y 3.1, y su versión en economía abierta, en el tema B09, apartado 3.1.

Efecto innovación en bienes intermedios: la innovación se entiende aquí como un aumento del número de variedades de *inputs*. En un marco de competencia imperfecta en el sector de los bienes intermedios, la integración comercial amplía el mercado de cada variedad y eleva los beneficios de la innovación, lo que estimula la I+D y da lugar a crecimiento endógeno. Esto se ilustra a partir del modelo de I+D de Romer (1990).

Efecto desbordamiento: el aumento del tamaño de mercado incentiva la inversión y la acumulación de capital, lo que genera externalidades positivas (aprendizaje por la práctica, desbordamientos de conocimiento) sobre el resto de la economía y puede sostener el crecimiento endógeno. Esto se ilustra a través del modelo de externalidades del capital (Arrow, 1962; Romer, 1986; Lucas, 1988), expuesto en el tema B09.

2.4. Efecto selección: modelo de Melitz (2003)

Cuando las empresas difieren en productividad, la integración tiene un efecto adicional, el de selección. El efecto selección consiste en un traspaso de la producción desde las empresas menos productivas hacia las más productivas, que se ilustra a partir del modelo de Melitz (2003). Este efecto es especialmente importante desde el punto de vista de la política económica por dos motivos. Por un lado, el abandono de empresas puede dar lugar a desempleo, con los consiguientes

costes económicos y sociales. Por otro, es un mecanismo que genera cambios en la productividad. En particular, si las empresas que se ven obligadas a salir son las menos productivas, la productividad agregada de la industria crece tras la liberalización.

> El modelo de heterogeneidad de empresas se desarrolla en el tema B06, apartado 3.1.

2.5. Conclusiones

En conjunto, los efectos dinámicos (escala, competencia, variedad, crecimiento y selección) constituyen hoy el principal argumento económico en favor de la integración. Los efectos estáticos de creación y desviación son de una sola vez y de signo ambiguo, mientras que los dinámicos se despliegan en el tiempo. Los de escala, competencia, variedad y selección elevan el nivel de la renta y de la productividad, y solo los de crecimiento endógeno alteran su tasa de crecimiento a largo plazo. Las estimaciones del informe Cecchini (1988) atribuían a las economías de escala y a la mayor competencia en torno a la mitad de las ganancias del mercado único europeo.

3. El proceso de integración económica: retos y dificultades

Hasta aquí se ha examinado el debate sobre los beneficios de la integración, tanto estáticos como dinámicos. Sin embargo, en la realidad existe, además, una serie de costes que limitan los procesos de integración económica.

3.1. Costes

Esos costes son cuatro, la pérdida de soberanía económica, la transmisión de los ciclos, la especialización perversa y la concentración geográfica.

3.1.1. Pérdida de soberanía económica

Uno de los principales costes de la integración es la pérdida de soberanía económica nacional. Por ejemplo, la fijación de un arancel exterior común supone una limitación de la autonomía de la política comercial y exige el cumplimiento de los acuerdos de cooperación establecidos entre los miembros, lo cual puede dar lugar a fallos de coordinación.

Dani Rodrik (2011) plantea esta tensión en su "trilema político de la economía mundial", conocido como **trilema de Rodrik** (tema B21, Anexo A.1.1), según el cual no es posible tener a la vez integración económica profunda, Estado-nación soberano y política democrática. A lo sumo pueden conjugarse dos de los tres elementos, de modo que profundizar en la integración exige ceder soberanía nacional (como en el proceso integrador de la UE) o subordinar las demandas democráticas internas a las exigencias de la economía global.

3.1.2. Transmisión de los ciclos

La integración acelera la transmisión internacional del ciclo económico y puede acentuar sus picos y valles, porque cuanto más integrados están los mercados de bienes y las cadenas de producción, más rápidamente se propagan las perturbaciones entre países. Dos episodios ilustran esta idea. El primero es el gran colapso del comercio de 2009, cuando la crisis financiera global se transmitió a través de las cadenas de valor y el comercio mundial cayó mucho más que el PIB. El segundo, la pandemia de COVID-19 en 2020, en la que las disrupciones de las cadenas de suministro sincronizaron la caída de la producción entre economías muy alejadas del foco inicial.

3.1.3. Especialización perversa (Grubel, 1967; Robson, 1984)

Origen: autores como Herbert Grubel (1967) y Peter Robson (1984) argumentan que el efecto reducción de costes derivado de una unión aduanera no está garantizado, porque una unión aduanera puede dar lugar a la llamada *especialización perversa*.

Idea: con los supuestos de Corden, cabría esperar que la unión concentrase toda la producción en B. Sin embargo, puede suceder que A, pese a que su curva de coste medio queda más alta que la de B, llegue a la unión con un coste medio menor que el de su socio, porque algún "accidente histórico" le permitió alcanzar antes una escala mayor. La noción de "accidente histórico" introduce un elemento dinámico.

Razonamiento: basta con que el precio de entrada al mercado esté por debajo del coste medio de las primeras unidades de B para impedir su entrada.

En ese caso, la unión deja ambos mercados en manos del productor de A. El resultado es una *especialización perversa*, ineficiente porque el mercado queda para quien no es el productor potencialmente más eficiente. Esta consideración ofrece un argumento plausible sobre la dicotomía norte-sur. Así, en una integración entre socios de desarrollo desigual, el socio avanzado que alcanzó antes una gran escala puede retener la producción aunque el menos desarrollado sea potencialmente más eficiente.

Desarrollo gráfico: la especialización perversa se representa en el plano cantidad-precio y coste (figura B10.5):

1. *Curvas de demanda*: la de cada país, $D^A(y) = D^B(y)$, y la de la unión, $D^{\text{UA}}(y)$, que suma ambas.

2. *Curva de coste medio de A*, $\text{Cme}^A(y)$: decreciente, por las economías de escala.

3. *Curva de coste medio de B*, $\text{Cme}^B(y)$: decreciente y por debajo de la de A para cualquier volumen, pues B es potencialmente más eficiente.

4. *Equilibrio*: antes de la unión, A opera con un coste c_0^A (punto 0^A) inferior al de B, c_0^B (punto 0^B), porque alcanzó antes una mayor escala. Con la unión, A abastece todo el mercado ampliado al precio p_{UA}^A (punto 1), por encima del que permitiría B, p_{UA}^B (círculo hueco).

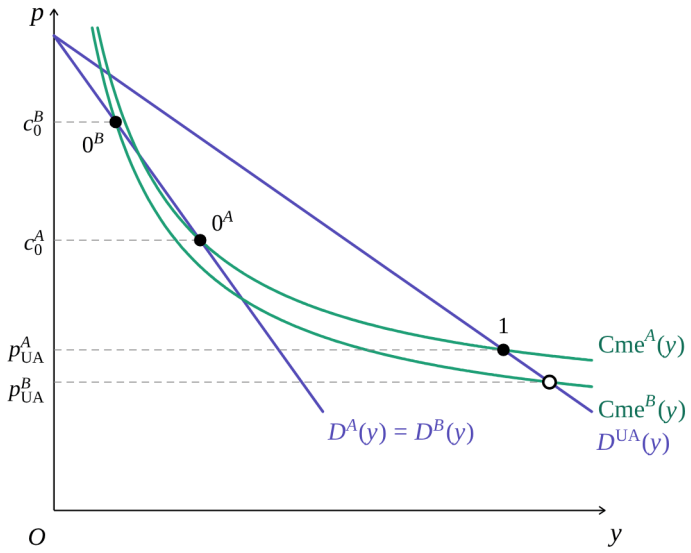

Figura B10.5: Especialización perversa. Por un accidente histórico, el productor potencialmente menos eficiente de la unión (A) abastece todo el mercado ampliado (punto 1). El círculo hueco señala el resultado eficiente, con B como productor.

Conclusiones:

▪ *Aspectos positivos*: matiza el resultado de Corden, según el cual la unión entre países con un productor cada uno es inequívocamente beneficiosa. También introduce la dependencia de la

historia en la localización de la producción, una idea que la nueva economía geográfica desarrolla con sus equilibrios múltiples.

- *Limitaciones*: es un argumento de posibilidad más que una predicción, pues no indica cuándo ni con qué frecuencia se produce el accidente histórico, y comparte los supuestos restrictivos del modelo de Corden (equilibrio parcial, un solo bien y un productor por mercado).

3.1.4. Concentración geográfica: la nueva economía geográfica (Krugman, 1991a)

> El modelo centro-periferia se desarrolla en el tema B06, apartado 2.3.3.

Las aportaciones de la llamada **nueva economía geográfica** ponen el énfasis en el hecho observado de que la riqueza, la población y el empleo están distribuidos espacialmente de forma desigual, tanto en el ámbito internacional como en el regional.

Estos desequilibrios espaciales pueden surgir por dos tipos de determinantes:

- *Primera naturaleza*: tienen que ver con la distribución desigual de los recursos naturales y se derivan de factores geográficos como la presencia de materias primas o la proximidad a vías naturales de comunicación (ríos, océanos).

- *Segunda naturaleza*: derivan de la acción humana y, principalmente, de factores económicos como la estructura de mercado o las reglas de fijación de precios.

A su vez, existen dos tipos de fuerzas, las de **aglomeración o centrípetas**, que concentran la actividad en los mercados de mayor tamaño y productividad, y las de **dispersión o centrífugas**, que la esparcen por el espacio (costes de transporte, congestión, demanda inmóvil).

Estas ideas de la nueva economía geográfica se han formalizado en el **modelo centro-periferia** (Krugman, 1991a). El modelo consta de dos regiones, un bien homogéneo producido por trabajadores inmóviles y un bien diferenciado, sujeto a costes de transporte de tipo *iceberg*, producido en competencia monopolística por trabajadores móviles que migran hacia la región con mayor salario real.

La interacción entre las fuerzas centrípetas y centrífugas genera dos tipos de equilibrio, el simétrico, con la actividad repartida entre ambas regiones, y el de aglomeración, con toda la industria concentrada en el centro y la otra región reducida a periferia.

El resultado central (ilustrado por el diagrama de *tomahawk*, figura B10.6) es que el equilibrio depende del grado de libertad comercial, φ. Se distinguen tres tramos:

- *Equilibrio con simetría*: se da con costes de comercio altos.

- *Histéresis comercial*: a partir del *punto de sostenimiento*, φ_S, la aglomeración pasa a ser sostenible y coexisten equilibrios múltiples, con histéresis y dependencia de los accidentes históricos.

- *Aglomeración completa*: superado el *punto de ruptura*, φ_B, la simetría se vuelve inestable y cualquier perturbación concentra toda la industria en una región.

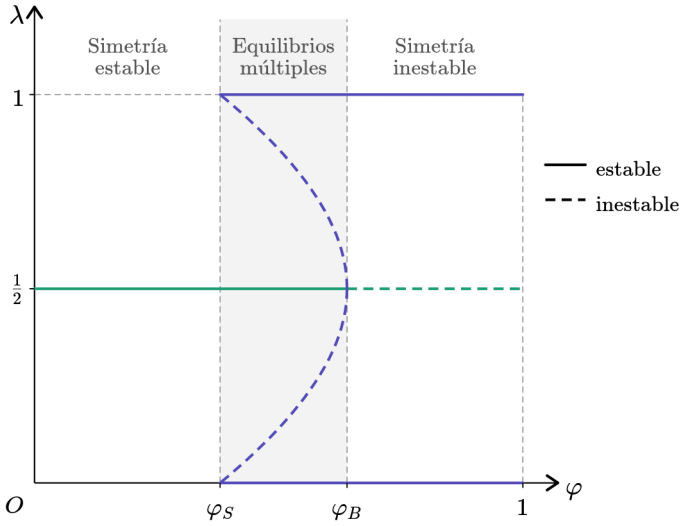

Figura B10.6: Diagrama de *tomahawk* del modelo centro-periferia (Krugman, 1991a). Con la cuota de una región en la industria, λ, frente a la libertad comercial, φ, la simetría ($\lambda = 1/2$) es estable hasta el punto de ruptura, φ_B, y la aglomeración ($\lambda = 0$ o 1) es sostenible desde el punto de sostenimiento, φ_S. Entre ambos hay equilibrios múltiples, y los inestables van en trazo discontinuo.

La implicación directa para la teoría de la integración es que la reducción de los costes de comercio puede convertir de forma abrupta pequeñas asimetrías espaciales en procesos de aglomeración que concentren la industria en el centro y vacíen la periferia. Por tanto, la integración no garantiza la convergencia regional, lo que ofrece un fundamento teórico a las políticas de cohesión territorial como las de la UE.

3.2. Profundización de la integración: el mercado común (MacDougall, 1960)

Como se indica en la introducción, el mercado común es un estadio posterior a la unión aduanera, en la medida en que además permite la libre circulación de los factores productivos. La configuración de este tipo de mercado añade a las ganancias de la unión aduanera las de la libre movilidad de los factores, aunque con ganadores y perdedores propios.

Origen: el modelo de MacDougall (1960), también conocido como *modelo de MacDougall-Kemp*, permite ilustrar en un solo gráfico los movimientos del factor trabajo[7] y su impacto sobre el bienestar.

Idea: la libre movilidad del trabajo iguala su retribución entre países y elimina la pérdida de eficiencia de la asignación autárquica, con una ganancia inequívoca de bienestar global pero, de nuevo, con ganadores y perdedores.

[7] La versión original del modelo, debida a MacDougall (1960), se aplicaba a los movimientos internacionales del capital, y autores posteriores la aplicaron a la movilidad del factor trabajo.

3.2.1. Supuestos

Entorno: dos países, A y B (por concreción, EE. UU. y México, respectivamente), con competencia perfecta en los mercados de bienes y de factores, de modo que los factores se retribuyen según el valor de su producto marginal. Se considera un entorno de corto plazo en el que el capital permanece fijo en cada país (el modelo puede estudiarse análogamente para la movilidad del capital). Las dotaciones factoriales no son necesariamente iguales en ambos países, por lo que las productividades marginales tampoco lo son, y de ellas depende la posición inicial de las curvas. La dotación total de trabajo se considera constante y exógena (no hay crecimiento poblacional), por lo que el análisis se centra en la demanda de trabajo de cada país.

Tecnología: la tecnología es neoclásica de buen comportamiento y no lineal en cada país, del tipo $X^A = f(\bar{K}^A, L^A)$ y $X^B = f(\bar{K}^B, L^B)$, con productividad marginal del trabajo positiva y decreciente ($f'_L > 0$, $f''_L < 0$). Cada país produce un único bien con dos factores homogéneos. De estos supuestos se deriva que los resultados del modelo dependen exclusivamente de las diferencias entre países en el valor del producto marginal del trabajo.

3.2.2. Modelo

Planteamiento: la demanda de trabajo de cada país coincide con el valor de su productividad marginal del trabajo, ya que la CPO de la maximización del beneficio exige que el salario iguale dicho valor. La curva es también decreciente en el empleo de cada país, medido desde su propio origen, por la productividad marginal decreciente, y, con las tecnologías habituales (p. ej. Cobb-Douglas), convexa.

Equilibrio: en autarquía, EE. UU. remunera a sus trabajadores con un salario w_0^A superior al salario mexicano w_0^B. En un mercado común, esa diferencia incentiva la migración de trabajadores mexicanos hacia EE. UU. Con plena flexibilidad, la migración continúa hasta que el salario se iguala en ambos países, $\text{VPmg}_L^A = \text{VPmg}_L^B = w_1$.

Desarrollo gráfico: los efectos del mercado común se representan en el plano cantidad de trabajo-salario (figura B10.7), con la dotación mundial de trabajo en el eje horizontal y un origen para cada país, O^A y O^B:

1. *Demanda de trabajo de A*, $\text{VPmg}_L^A(L^A)$: recoge el salario al que A emplea cada cantidad de trabajo, igual al valor de su productividad marginal. Es decreciente en el empleo de A, medido desde O^A, porque esa productividad marginal es decreciente (Planteamiento).

2. *Demanda de trabajo de B*, $\text{VPmg}_L^B(L^B)$: recoge lo mismo para B y es decreciente en el empleo de B, medido desde O^B, por la misma razón.

3. *Equilibrio*: en autarquía, con el reparto del trabajo L_0, los salarios son w_0^A en EE. UU. y w_0^B en México. En el mercado común, la migración de B a A desplaza el reparto hasta L_1, donde ambas curvas se cortan (punto 1) y el salario se iguala en w_1. El área sombreada entre ambas curvas es la pérdida de eficiencia que se elimina.

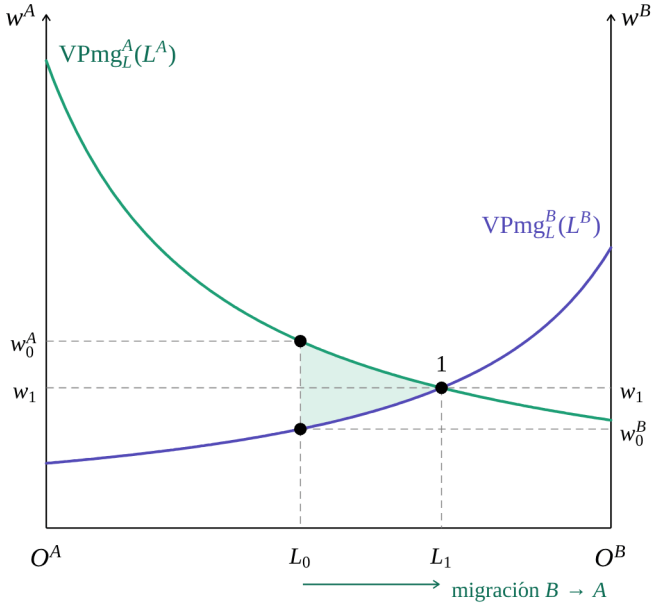

Figura B10.7: Mercado común (MacDougall, 1960). La migración de B a A iguala los salarios y elimina la pérdida de eficiencia de la asignación autárquica (área sombreada).

Bienestar: constituido un mercado común, la migración produce los siguientes efectos:

- *Propietarios del capital del país receptor*: ganan, pues la caída del salario eleva su excedente.

- *Trabajadores residentes del país receptor*: pierden, debido a esa misma bajada salarial. Esta pérdida está en la base de los argumentos de los movimientos antiinmigración.

- *Trabajadores migrantes*: ganan, pues pasan a percibir un salario superior al de su país de origen.

- *País de origen*: los trabajadores que permanecen ganan (el salario sube al reducirse la oferta de trabajo), mientras que los propietarios del capital pierden. Además, la producción total del país de origen cae con la salida de los migrantes, y solo las remesas pueden compensar esa pérdida, de ahí su importancia para muchos países en desarrollo.

En general, la magnitud de estos efectos depende de las dotaciones factoriales de ambos países, que determinan la posición de las curvas, y de la elasticidad de la demanda de trabajo, que determina su pendiente. Con todo, el efecto neto global es una ganancia de bienestar por la eliminación de la pérdida de eficiencia (el "triángulo" entre las curvas). Si no se consideran las remesas, esa ganancia global es el saldo de una ganancia neta para el país receptor y una pérdida neta, menor, para el país emisor.

3.2.3. Conclusiones

Aspectos positivos: muestra en un solo gráfico que la libre movilidad del factor trabajo eleva el bienestar global, al producir una mejor asignación de los recursos. Como en términos de equidad da lugar a ganadores y perdedores, explica también la aparición de políticas que restringen los movimientos internacionales de factores.

Limitaciones: supone trabajo homogéneo y que los migrantes gastan sus rentas en el país de destino. Sin embargo, los migrantes son a menudo jóvenes formados, con lo que la fuga de cerebros agrava el efecto negativo sobre el país de origen. Si se incorporan las remesas, en cambio, el país de origen podría salir beneficiado, pues la masa salarial que perciben los migrantes al salario w_1 supera el valor de la producción que habrían generado en origen. Por su parte, el país de destino no resultaría perjudicado, ya que sus residentes conservan la parte del triángulo de eficiencia situada por encima de w_1.

3.3. Regionalismo y multilateralismo

Fuera de los modelos, la integración se materializa en acuerdos concretos. Se estudian los acuerdos comerciales regionales y el debate sobre su encaje en el sistema multilateral de comercio.

> El artículo XXIV del GATT se desarrolla en el tema B33, apartado 1.2.2, y el debate entre regionalismo y multilateralismo, en ese mismo tema, apartado 3.5.4, y en el tema B35, apartado 1.2.

Los **acuerdos comerciales regionales** (*regional trade agreements*, RTA), si bien no constituyen en sí mismos una fase del proceso integrador, se han multiplicado en las últimas décadas[8]. Los RTA engloban tanto uniones aduaneras como tratados de libre comercio y, según han ido proliferando y sumando socios, han adquirido un carácter transversal que obliga a los países a cooperar en ámbitos cada vez más amplios. Esto ha llevado a hablar de un "**nuevo regionalismo**" (Ethier, 1998), orientado al mercado y a la integración profunda, en contraposición al "viejo regionalismo" de los años sesenta y setenta, basado en políticas de sustitución de importaciones.

La proliferación de acuerdos ha abierto, además, el debate sobre su relación con el sistema multilateral:

- *Piedras de construcción* (*building blocks*): para unos, los bloques regionales favorecen el libre comercio mundial, pues consolidan liberalizaciones y generan efectos dominó de adhesión (Baldwin, 1993).

- *Piedras de tropiezo* (*stumbling blocks*): para otros, encabezados por Bhagwati, los bloques desvían comercio, restan incentivos a la negociación multilateral y generan una maraña de normas de origen y aranceles superpuestos, el llamado *spaghetti bowl*, que eleva los costes de transacción del comercio mundial.

[8] Según la OMC, el GATT recibió 124 notificaciones de acuerdos comerciales regionales entre 1948 y 1994, y en 2026 hay casi 400 en vigor.

Conclusiones

En el plano estático, la integración económica es un ejercicio de *second best*. Por ello, el efecto de la integración sobre el bienestar depende del saldo entre la creación y la desviación de comercio y de rasgos como el número de miembros, los aranceles previos o la similitud de las economías. En el plano dinámico, las economías de escala, los efectos procompetitivos y de variedad, el estímulo a la innovación y la selección de las empresas más productivas inclinan la balanza a favor de la integración. Frente a ello pesan la cesión de soberanía, la transmisión de las perturbaciones, la especialización perversa y la concentración de la actividad en el centro. De ahí que los procesos maduros se acompañen de políticas de compensación y cohesión, en la lógica del teorema de Kemp-Wan, y de la libre movilidad de factores propia del mercado común.

Los efectos dinámicos de la integración proceden de la nueva teoría, con las economías de escala, la variedad y la selección de empresas (tema B06), y del crecimiento endógeno en economía abierta, en el que un mercado mayor acelera la innovación (tema B09). En la escala de integración, la unión monetaria añade un coste nuevo, la pérdida del tipo de cambio como instrumento de ajuste, que analiza la teoría de las áreas monetarias óptimas (tema B16). Por su parte, la experiencia muestra procesos de distinta profundidad, del mercado único europeo (tema B39) a los acuerdos no comunitarios (tema B35). Por último, la compatibilidad del regionalismo con el sistema multilateral depende de la disciplina del artículo XXIV del GATT (tema B33) y del problema de las normas de origen superpuestas (tema B07).

Baldwin (1993) sostiene que el regionalismo es una piedra de construcción del multilateralismo, y Bhagwati (2008), que es una piedra de tropiezo (apartado 3.3). Limão (2006) halla que EE. UU. redujo menos sus aranceles multilaterales en los productos que importaba con preferencias, en línea con la segunda tesis. Sigue abierto también el debate sobre el contenido de los acuerdos recientes, que regulan además la propiedad intelectual, la inversión, las normas sanitarias y laborales o el arbitraje entre inversores y Estados. Rodrik (2018) advierte de que estas disposiciones encajan mal en la teoría del comercio y pueden redistribuir renta en favor de la banca internacional, la industria farmacéutica o las multinacionales más que ganar eficiencia.

Anexo. Evidencia empírica y medición de la integración

La medición de los efectos de la integración y la evidencia empírica disponible complementan el análisis teórico del tema.

A.1. Medición

La medición de los efectos de la integración económica resulta uno de los ejercicios más difíciles de la literatura del comercio internacional, debido a la dificultad para delimitar los efectos y las variables que actúan en el proceso integrador. Entre las metodologías empleadas destacan cuatro:

Modelos residuales: en los inicios de la CEE era muy importante conocer los efectos *ex post* de la integración. Los modelos residuales utilizaban como indicador la proporción de exportaciones intrarregionales entre los países de la unión, de modo que un aumento del indicador suponía que los países de la unión comerciaban cada vez más entre sí. De acuerdo con este indicador, entre 1970 y comienzos de la década de 2000 la proporción de comercio intrarregional aumentó con claridad en el TLCAN (NAFTA, por sus siglas en inglés) y el Mercosur. En cambio, apenas varió en la UE, que partía ya de niveles muy elevados (en torno al 60 %), y se mantuvo estable en la Asociación de Naciones del Sudeste Asiático (ASEAN). La limitación de estos modelos es que esta tasa no distingue entre los efectos creación y desviación de comercio ni contempla si dichas áreas habrían mejorado sus exportaciones respecto a terceros países.

Modelos de intensidad del comercio: se basan en comparar el comercio entre dos socios con un nivel "natural", el que cabría esperar dado el peso de cada uno en el comercio mundial. Si el comercio observado es mayor que ese nivel natural, está sesgado a favor del socio por factores naturales (lengua común, frontera común) o institucionales (la pertenencia a un acuerdo de integración). De acuerdo con este indicador, en el mismo periodo el sesgo regional del comercio se intensificó en el TLCAN y, sobre todo, en el Mercosur, apenas varió en la UE y se redujo en la ASEAN.

Modelos gravitatorios: constituyen hoy el instrumento estándar de la evaluación *ex post*. Estiman el efecto de los acuerdos incluyendo variables indicativas de pertenencia en una ecuación de gravedad que controla por el tamaño de los socios, la distancia y la resistencia multilateral. Los desarrollos recientes tratan, además, la endogeneidad de la firma de los acuerdos (la decisión de firmarlos depende de factores no observados que también afectan al comercio).

Modelos de equilibrio general computable: posteriores a los gravitatorios, permiten la evaluación *ex ante* de los acuerdos mediante la simulación de economías calibradas. Organismos como la Comisión Europea los han utilizado profusamente para estimar los efectos del mercado único o de los acuerdos comerciales en negociación. Su limitación es la sensibilidad de los resultados a la calibración y a los supuestos de comportamiento. Una referencia histórica de la evaluación *ex ante* es el informe Cecchini (1988) sobre "el coste de la no Europa". Combinó estudios microeconómicos de costes con simulaciones macroeconométricas (no un modelo de equilibrio general computable propiamente dicho) y estimó ganancias del orden del 4,25 % al 6,5 % del PIB comunitario derivadas de la culminación del mercado único.

A.2. Resultados

La evidencia disponible apunta, en general, a efectos positivos de la integración:

Creación de comercio: los acuerdos aumentan de forma sustancial el comercio bilateral entre sus miembros (Baier y Bergstrand, 2007).

Comercio de servicios: los acuerdos con disciplinas sobre servicios aumentan el comercio de servicios entre sus miembros, aunque la evidencia sobre la magnitud del efecto y sobre qué tipo de acuerdo lo produce es menos concluyente que para las mercancías.

Crecimiento y productividad: la evidencia es menos concluyente. Campos, Coricelli y Moretti (2019) estiman que, sin la integración, la renta per cápita de los países que se adhirieron a la UE habría sido en torno a un 10 % menor en sus diez primeros años de pertenencia, con Grecia como excepción. En cambio, Vamvakidis (1999) encuentra que las economías crecieron más tras una liberalización comercial amplia, pero menos tras incorporarse a un acuerdo regional.

Nivel de empresa: Bustos (2011) encuentra que el Mercosur indujo la modernización tecnológica de las empresas argentinas. La reducción de los aranceles brasileños aumentó los ingresos esperados de exportación de esas empresas, y las de las industrias con mayores rebajas elevaron más deprisa su inversión en tecnología.

Inversión extranjera directa: en los países en desarrollo, la pertenencia a la OMC y a más acuerdos comerciales preferenciales se asocia a mayores entradas de inversión extranjera directa (Büthe y Milner, 2008).

Inmigración: Ortega y Peri (2014) encuentran que la apertura a la inmigración eleva a largo plazo la renta per cápita del país receptor. El efecto opera a través de la productividad total de los factores, por la mayor diversidad de habilidades productivas y, en parte, por una innovación más intensa.

B11. Balanza de pagos

Balanza de pagos: concepto, medición e interpretación.

Introducción

Las **cuentas internacionales** de una economía resumen las relaciones económicas entre sus residentes y los no residentes y se componen de tres estados:

1. *Balanza de pagos*: estado estadístico que resume las *transacciones* económicas entre un país y el resto del mundo durante un periodo de tiempo determinado.

2. *Posición de inversión internacional* (PII): estado estadístico que recoge, referido a una fecha concreta, el valor de los activos y pasivos *financieros* de la economía frente al exterior[1]. Constituye un subconjunto de los activos y pasivos que figuran en el balance nacional[2].

3. *Cuenta de otras variaciones de activos y pasivos financieros*: recoge las variaciones de la posición debidas a fenómenos distintos de las transacciones entre residentes y no residentes, como las revalorizaciones.

Todas ellas se elaboran conforme a las directrices del FMI, recogidas en el sexto *Manual de Balanza de Pagos y Posición de Inversión Internacional* (MBP6), aplicado en la UE desde 2014. El manual está armonizado con el Sistema de Cuentas Nacionales (SCN) y con el Sistema Europeo de Cuentas Nacionales y Regionales (SEC 2010), lo que permite comparar y agregar las cuentas internacionales con el resto de estadísticas macroeconómicas. En 2025 se aprobó su sucesor, el séptimo manual (MBP7), junto con el nuevo SCN 2025, que se aplicarán en los próximos años.

El interés por las cuentas internacionales se debe a que reflejan los **desequilibrios externos**, uno de los grandes debates de la ciencia económica desde que David Hume (1752) describió el mecanismo automático de ajuste por flujos de metales preciosos. El debate se ha intensificado con la liberalización comercial y financiera posterior a la Segunda Guerra Mundial, con los llamados desequilibrios globales de los años dos mil y con las perturbaciones recientes. De hecho, según el FMI, los saldos por cuenta corriente mundiales se ampliaron en 2024 en 0,6 puntos del PIB mundial, y en torno a dos tercios de ese aumento correspondían a saldos excesivos, concentrados

[1] Su saldo (activos menos pasivos) es la posición neta, acreedora o deudora. La PII se ordena por categorías funcionales: inversión directa, inversión de cartera, derivados financieros (distintos de reservas) y opciones de compra de acciones de empleados, otra inversión y activos de reserva. El SCN, en cambio, clasifica solo por instrumento.

[2] El balance nacional abarca, junto a la PII, los activos no financieros, así como los activos financieros y los pasivos que mantienen entre sí los residentes (FMI, 2009, p. 8).

en China, EE. UU. y la zona del euro[3]. Estos desequilibrios importan, en primer lugar, porque las economías que se endeudan mucho y muy deprisa con el exterior son vulnerables a las **paradas súbitas** de la entrada de capitales (*sudden stops*). Importan también porque pueden reflejar una asignación ineficiente del ahorro y porque, cuando las reservas se agotan o la deuda externa se revela insostenible, desembocan en crisis de balanza de pagos. El principal instrumento para analizarlos es la balanza de pagos, que, aunque siempre está equilibrada en sentido contable, puede reflejar desequilibrios en sentido *económico*.

De las tres cuentas internacionales, este tema se centra en la balanza de pagos, que permite entender los desequilibrios que se producen en las relaciones económicas internacionales. Para ello, el tema se organiza en tres partes:

- *Fundamentos*: se estudian los fundamentos teóricos y contables de la balanza de pagos, su concepto y su medición.

- *Estructura contable*: el tema describe su estructura, la de las demás cuentas internacionales y su relación con las cuentas nacionales.

- *Interpretación económica*: la última parte va de los factores que afectan al conjunto de las cuentas a la lectura de cada una de ellas y de la posición de inversión internacional, y se cierra con otras presentaciones de la balanza de pagos y la evidencia sobre los desequilibrios.

[3] FMI, *External Sector Report* (2025), con datos de 2024. En ediciones anteriores (datos de 2019), el FMI ya calificaba de excesivos los superávits de Alemania y los Países Bajos y los déficits de EE. UU., Canadá y el Reino Unido.

1. Fundamentos de la balanza de pagos: concepto y medición

Se exponen el fundamento teórico de la balanza de pagos, su concepto y base normativa, sus principios contables y su medición.

1.1. Fundamento teórico: el sector exterior en el flujo circular de la renta

> Este apartado amplía con el resto del mundo el flujo circular de la renta del tema A30, apartado 1.1. No hace falta exponerlo con detalle en el ejercicio oral, pero conviene entenderlo, porque explica por qué la balanza de pagos está siempre en equilibrio contable.

Origen: la idea de una "balanza" de las transacciones de un país con el exterior es mercantilista (balanza comercial, s. XVI-XVII). Hume (1752), anticipado por Gervaise (1720), le dio su análisis clásico con el mecanismo de flujo de especie y precios, que muestra que el saldo exterior no puede ser permanente. Después, Meade (1951) construyó el marco moderno que integra la balanza en el análisis macroeconómico de la renta.

Idea: se parte del flujo circular de la renta (tema A30, apartado 1.1) y se añade un cuarto agente, el *resto del mundo*, que engloba a todas las unidades no residentes que operan con los residentes.

Razonamiento: el resto del mundo abre el circuito en tres puntos:

- *Mercado de bienes*: las exportaciones son una inyección de gasto y las importaciones, una fuga.
- *Distribución de la renta*: los residentes reciben y pagan rentas primarias y transferencias del exterior.
- *Financiación*: cada una de esas operaciones tiene una contrapartida financiera (cobro, pago, crédito), de modo que la nación adquiere activos o emite pasivos frente al exterior.

De ello se siguen dos intuiciones. Primera, como toda operación con no residentes tiene una contrapartida, el registro de las relaciones exteriores es necesariamente de partida doble y la balanza de pagos está siempre en equilibrio contable, aunque no económico. Segunda, el saldo de las operaciones no financieras con el exterior es, a la vez, la brecha entre el ahorro y la inversión de la nación y su préstamo o endeudamiento neto frente al resto del mundo (la variación de su posición financiera, salvo revalorizaciones y otras variaciones de volumen). Esta lectura se desarrolla en el apartado 3.

Conclusiones:

- *Aspectos positivos*: ofrece un marco común a la balanza de pagos y a las cuentas nacionales, que relaciona el saldo exterior con el ahorro, la inversión y la posición financiera de la nación. Sirve, además, de base a la interpretación de los saldos y a las teorías del ajuste (tema B12).
- *Limitaciones*: las relaciones que se obtienen son identidades contables, que se cumplen siempre y no indican causalidad. No explican, por tanto, qué determina el saldo exterior ni cómo se corrige un desequilibrio, cuestiones que corresponden a las teorías del ajuste de la balanza de pagos.

1.2. Concepto, base normativa y elementos a contabilizar

Se presentan el concepto y la base normativa de la balanza de pagos, cuya metodología fija el FMI en sus manuales, y los elementos que registra, los flujos y las posiciones frente al exterior.

1.2.1. Concepto y base normativa

Concepto: se denomina *balanza de pagos* al documento contable en el que se registran, para un periodo de tiempo dado, las operaciones económicas llevadas a cabo por los residentes de un país con quienes no lo son. Su estructura en cuentas se expone en el apartado 2.1.

Objetivo: ofrecer un marco integrado para el análisis de los desequilibrios exteriores de una economía, principal canal por el que las políticas macroeconómicas de los países se afectan unas a otras. El marco sirve también para la formulación de la política económica (política cambiaria, gestión de reservas, financiación exterior).

Base normativa: la metodología es la del FMI, que ha publicado siete ediciones del Manual de Balanza de Pagos, de las que destacan cuatro:

1. *MBP1* (1948): continuación de la labor de la Sociedad de Naciones, consistía en cuadros normalizados para la declaración periódica de datos al FMI, sin análisis conceptual.

2. *MBP5* (1993): armonizado con el SCN-93, sacó de la cuenta corriente los activos no producidos y las transferencias de capital, que pasaron a una nueva cuenta de capital. De ahí que la antigua cuenta de capital pasara a llamarse cuenta financiera. Además, separó los servicios de las rentas (las actuales rentas primarias, antes "servicios factoriales") e incorporó la posición de inversión internacional.

3. *MBP6* (2009): armonizado con el SCN-2008, mantiene la estructura del MBP5 y la actualiza en tres frentes. El primero es la globalización (residencia de individuos y empresas con conexiones con varias economías, remesas, entidades con fines específicos, procesos de producción globalizados). El segundo son las cuestiones de balance (mayor detalle de la PII, de las revalorizaciones y de las otras variaciones de volumen, para el análisis de vulnerabilidad y sostenibilidad). El tercero es la innovación financiera (derivados, titulización, cadenas de propiedad complejas en la inversión directa, reestructuraciones de deuda y garantías).

4. *MBP7* (2025): publicado en marzo de 2025 como *Manual integrado de balanza de pagos y posición de inversión internacional*, armonizado con el SCN-2025, actualiza el MBP6 para recoger la creciente interconexión económica, la digitalización y la innovación en los mercados financieros. Su implantación en las estadísticas nacionales será gradual, por lo que las estadísticas publicadas, y este tema, siguen todavía el MBP6.

La concordancia entre el MBP6 y el SCN-2008 es completa en la valoración, la delimitación de las unidades residentes, el momento de registro, los activos y pasivos financieros y las transferencias de capital. Por ello, en la mayoría de los países, la balanza de pagos y la PII se compilan primero y se incorporan después a la cuenta del resto del mundo de las cuentas nacionales (apartado 2.3).

1.2.2. Elementos a contabilizar: flujos y posiciones

El capítulo 3 del MBP6 distingue entre dos tipos de partidas:

1. *Flujos*: reflejan la creación, transformación, intercambio, transferencia o extinción de valor económico y suponen, para una unidad institucional, una alteración en la composición, el volumen o el valor de sus activos y pasivos dentro de un periodo contable. Se recogen como transacciones o como otros flujos:

- *Transacciones*: interacciones entre dos unidades institucionales que ocurren por mutuo acuerdo o en virtud de la ley y que implican un intercambio de valor o una transferencia. Pueden ser monetarias o no monetarias (remuneración en especie, trueque) y se registran en la balanza de pagos.

- *Otros flujos*: cambios de volumen, valor o clasificación de un activo o pasivo que no tienen su origen en una transacción entre residentes y no residentes. Su registro se efectúa en una cuenta específica, la de otras variaciones de los activos y pasivos financieros. En las cuentas internacionales solo se registran otros flujos para el oro en lingotes y para los activos y pasivos financieros que constituyen derechos u obligaciones frente a no residentes.

2. *Posiciones*: nivel de los activos o pasivos financieros frente al exterior en un momento determinado. La posición de inversión internacional las recoge al comienzo y al final del periodo, y las posiciones de dos momentos se conectan por los flujos ocurridos entre ambos.

1.3. Principios contables

Los cinco principios que rigen el registro son los mismos que los de las cuentas nacionales (tema A30, apartado 1.4).

1.3.1. Delimitación de la economía nacional: criterio de residencia

Cada unidad institucional es residente del territorio económico con el que guarda un vínculo más estrecho, que se concreta en su **centro de interés económico predominante**. Una unidad es residente de un territorio cuando mantiene en él una instalación en la que realiza, o tiene previsto realizar, transacciones y actividades económicas a una escala significativa. Esa actividad ha de ser indefinida o durar un periodo finito pero prolongado (en la práctica, un año o más). Por consiguiente, la residencia se determina así para cada tipo de unidad:

- *Hogares*: son residentes del territorio en el que mantienen su vivienda principal. La presencia durante un año o más, o la intención de permanecer, basta para determinar la residencia. El lugar de trabajo, en cambio, no determina la residencia, y todos los miembros de un hogar comparten la de este aunque pasen tiempo en el extranjero, siempre que no lleguen a tener allí su centro de interés económico predominante.

- *Empresas*: son residentes del territorio desde el que producen bienes y servicios en un volumen significativo. Por ello, una filial en el extranjero es residente del país de acogida, y su relación con la matriz se registra como inversión directa. Asimismo, cuando una misma unidad jurídica realiza operaciones sustanciales en dos o más territorios, se reconoce en cada uno una unidad institucional separada (sucursal).

- *Administraciones públicas*: los organismos públicos de cualquier nivel (central, regional y local) son residentes, igual que el personal destinado en sus embajadas y bases militares en el extranjero.

Hay dos casos especiales. Los organismos internacionales se consideran residentes de un territorio económico propio, distinto de aquel en que tienen su sede, y las instituciones sin fines de lucro al servicio de los hogares (ISFLSH), de la economía en que están constituidas y registradas jurídicamente.

1.3.2. Sistema de registro: partida doble vertical

El criterio básico es la **partida doble**, según la cual cada transacción da lugar al menos a dos asientos en las cuentas del mismo país (partida doble *vertical*). Uno es un **crédito** (exportaciones, rentas recibidas, disminuciones de activos financieros, aumentos de pasivos) y el otro, un **débito** (importaciones, rentas pagadas, aumentos de activos, disminuciones de pasivos). Ello garantiza que la suma de los créditos sea igual a la de los débitos, es decir, el equilibrio contable. En términos de las cuentas nacionales, un crédito del residente es un recurso y un débito, un empleo. En cambio, en la cuenta del resto del mundo del SEC, el mismo flujo se registra desde la óptica del no residente, por lo que los signos se invierten.

Dos ejemplos ilustran este criterio. Primero, una importación de tubos metálicos para construir un barco, financiada con un préstamo del exterior, genera un débito en la cuenta corriente y un crédito en la cuenta financiera (aumento de pasivos). Segundo, cuando los fondos estructurales de la UE financian una carretera, la ayuda se anota como crédito en la cuenta de capital y, como contrapartida, el depósito recibido figura como débito en la cuenta financiera (aumento de activos).

1.3.3. Momento de registro: devengo

El registro de los flujos tiene lugar cuando el valor económico se crea, se transforma, se intercambia, se transfiere o se extingue, momento que no coincide necesariamente con el cobro o el pago. Así, el momento de registro viene determinado por el traspaso de la **propiedad económica**, es decir, de los riesgos y beneficios del activo. Los bienes, los activos no financieros no producidos y los activos financieros se registran cuando se transfiere la propiedad económica, y los servicios, cuando se prestan. Las rentas primarias y secundarias se registran cuando surge el derecho a cobrar o la obligación de pagar, y las revalorizaciones y los cambios de volumen, en el periodo comprendido entre las fechas en que se valoran las posiciones.

El MBP6 recomienda el devengo porque brinda la información más completa. El principio de caja (registro cuando se recibe o desembolsa efectivo) no es un principio alternativo, sino una limitación de las fuentes. Por eso, los sistemas de declaración basados en cobros y pagos deben ajustarse al devengo, y solo excepcionalmente se admiten los criterios de compromiso o de vencimiento.

1.3.4. Criterio de valoración: precios de mercado

Los **precios de mercado** son la base de valoración de las transacciones, de los otros flujos y de las posiciones. El precio de mercado es el valor de cambio corriente, es decir, aquel al que se intercambian efectivamente por dinero, o podrían intercambiarse, los bienes, los servicios, la mano de obra y los activos. Se trata del precio del intercambio concreto, no de un precio cotizado, mundial o medio de una clase de intercambios. Tampoco es necesariamente un precio de mercado libre, pues la transacción puede darse en monopolio o monopsonio, y las que implican *dumping* o descuentos también se valoran a precios de mercado. Cuando el precio de mercado no es observable, se recurre a valoraciones equivalentes a las de mercado. Así, los activos y pasivos

financieros negociados con regularidad se valoran a sus cotizaciones, y los que no se negocian o se negocian con muy poca frecuencia, a un precio equivalente al de mercado (p. ej. el valor presente de sus flujos de caja).

En cuanto a las exportaciones e importaciones, el FMI recomienda que ambas se valoren **FOB** (*free on board*, sin flete ni seguro) en la frontera aduanera de la economía exportadora. Las estadísticas aduaneras de comercio de mercancías, en cambio, siguiendo las directrices de Naciones Unidas, valoran las exportaciones FOB y las importaciones CIF (*cost, insurance and freight*, con flete y seguro). Ello obliga a un ajuste y explica parte de las asimetrías bilaterales entre fuentes.

1.3.5. Unidad de cuenta

Las cuentas internacionales se compilan generalmente en moneda nacional, si bien la conversión a una unidad de cuenta de referencia es un requisito para elaborar cuentas comparables, y un país cuya moneda sufre fuertes tensiones puede elaborarlas en otra, como el euro o el dólar. Los datos en moneda extranjera son útiles, además, para la gestión de la liquidez internacional y ante una inflación elevada, fuertes oscilaciones cambiarias o tipos de cambio múltiples. Cada transacción se anota en la moneda en que se realizó y se convierte al tipo de cambio del momento del flujo, y las posiciones, al de la fecha de balance.

1.4. Medición

La medición comprende el sistema de recogida y sus límites, las fuentes de información y la práctica de la medición.

1.4.1. Sistema de recogida y límites al registro

La *Guía de compilación del MBP6* distingue dos grandes fuentes, los **sistemas de registro de transacciones internacionales** (ITRS), basados en las liquidaciones que comunican los bancos, y la declaración directa y las encuestas a empresas. En los ITRS es habitual fijar umbrales para limitar la carga de declaración:

- *Umbral de exención*: por debajo de él, las operaciones no se declaran.
- *Umbral de simplificación*: por debajo de él, las operaciones se declaran de forma agregada o sin clasificar por concepto, de modo que el compilador debe asignarlas a partidas mediante encuestas o a partir de la estructura de las operaciones de mayor importe.

La Guía no fija cuantías, pero exige que se elijan de forma que la calidad de los datos siga siendo aceptable (FMI, *Guía de compilación del MBP6*, 2014, §4.12-4.14).

Con la adopción del MBP6, muchos compiladores han sustituido los ITRS por la declaración directa de los residentes con transacciones o posiciones exteriores significativas, complementada con encuestas y registros administrativos. Ello mejora la calidad del registro por devengo, pero exige combinar fuentes heterogéneas y deja fuera las operaciones de los declarantes por debajo de los umbrales, que se estiman.

1.4.2. Fuentes de información

Los compiladores combinan, en general, cuatro tipos de fuentes:

- *Estadísticas aduaneras y de comercio*: la información sobre mercancías procede de las declaraciones de aduana para el comercio con terceros países y, dentro de la UE, donde no existen formalidades aduaneras, de la declaración estadística INTRASTAT.

- *Registros de inversión exterior*: recogen la inversión extranjera directa en el país y la del país en el exterior.

- *Encuestas*: sobre el comercio internacional de servicios y el gasto de los visitantes no residentes.

- *Datos de las administraciones públicas*: flujos con organismos internacionales y, en los Estados miembros, con la Unión Europea (fondos agrícolas y estructurales, recursos propios).

1.4.3. La medición en la práctica

Compiladores y base jurídica: en España, el Banco de España elabora la balanza de pagos y la PII, operaciones del Plan Estadístico Nacional. Por su parte, el Banco Central Europeo (BCE) agrega las de la zona del euro (Orientación BCE/2011/23) y Eurostat, las de la UE (Reglamento (CE) 184/2005).

Desde 2013, la fuente principal española es la declaración directa que regula la Circular 4/2012 del Banco de España. El residente que opera con no residentes o mantiene activos o pasivos frente a ellos declara con periodicidad mensual, trimestral o anual según su importe (umbrales de 300 y 100 M€), y por debajo de 1 M€ solo si se le requiere. En 2014 se extinguió el antiguo sistema de declaración a través de los bancos.

Límites en una economía globalizada: la medición choca con tres problemas:

- *Discrepancia mundial*: la suma de los saldos corrientes de todos los países, que debería ser nula, no lo es, y las asimetrías bilaterales revelan errores de cobertura, valoración y residencia.

- *Entidades con fines especiales*: sociedades sin actividad real, domiciliadas por razones fiscales, por las que la inversión solo pasa de camino a su destino. Damgaard, Elkjaer y Johannesen (2019) estiman que esta inversión extranjera directa (IED) "fantasma" se acerca al 40 % de la IED mundial en 2017 (unos 15 billones de dólares de 40) y que Luxemburgo y los Países Bajos albergan casi la mitad.

- *El caso irlandés*: el traslado a Irlanda de propiedad intelectual y otros activos de grandes multinacionales elevó su PIB en torno a un 25 % en 2015, sin un aumento parecido de la renta de los residentes. Por ello, la oficina estadística irlandesa (CSO) publicó por primera vez en 2017 la renta nacional bruta modificada (RNB*), que descuenta la depreciación de esos activos y los beneficios de las sociedades redomiciliadas y refleja mejor la economía interior (apartado 3.2.1).

1.5. Conclusiones

Con el resto del mundo incorporado al flujo circular de la renta, toda operación con no residentes tiene una contrapartida, de ahí que la balanza de pagos esté siempre en equilibrio contable, aunque no económico. El saldo no financiero mide, además, la brecha entre ahorro e inversión y el préstamo o endeudamiento neto frente al exterior. El MBP6 registra flujos y posiciones con los cinco principios de las cuentas nacionales (residencia, partida doble, devengo, precios de mercado y unidad de cuenta). Esa concordancia con el SCN-2008 permite compilar primero la balanza de pagos e incorporarla después a la cuenta del resto del mundo.

La medición, en cambio, descansa en fuentes propias, cuya calidad determina el tamaño de los errores y omisiones, y en una economía globalizada choca con la discrepancia mundial, la inversión directa "fantasma" y casos como el irlandés. La sección 2 ordena esos registros en cuentas, las de la balanza de pagos y la posición de inversión internacional con la cuenta que la concilia con los flujos.

2. Estructura contable de las cuentas internacionales

Se describen la estructura de la balanza de pagos, las demás cuentas internacionales y su relación con las cuentas nacionales.

2.1. Estructura de la balanza de pagos

La balanza de pagos agrupa sus cuentas en función de la naturaleza de los recursos económicos suministrados y recibidos:

Tabla B11.1: Estructura de la balanza de pagos según el MBP6.

Cuenta	Qué registra	Créditos (+)	Débitos (−)	Saldo
Cuenta corriente (CC)	Transacciones no financieras: bienes, servicios, rentas primarias y rentas secundarias	Exportaciones, rentas y transferencias recibidas	Importaciones, rentas y transferencias pagadas	SB_{CC}
Cuenta de capital (CK)	Transferencias de capital y activos no financieros no producidos (ANFNP)	Transferencias recibidas, cesiones de ANFNP	Transferencias pagadas, adquisiciones de ANFNP	SB_{CK}
Cuenta financiera (CF)	Variaciones netas de los activos y pasivos financieros exteriores	Aumentos de pasivos y disminuciones de activos	Aumentos de activos y disminuciones de pasivos	$\text{SB}_{\text{CF}} = \Delta\text{AF} - \Delta\text{PF}$
Errores y omisiones (EO)	Discrepancia estadística			SB_{EO}

Dado que cualquier transacción da lugar a dos asientos que se compensan, las tres cuentas, junto con los errores y omisiones, dan lugar a una identidad contable[4]:

$$\text{SB}_{\text{CC}} + \text{SB}_{\text{CK}} + \text{SB}_{\text{EO}} = \text{SB}_{\text{CF}}$$

donde SB indica el saldo de la cuenta correspondiente. Es decir, el saldo de las transacciones no financieras (exportaciones netas, rentas, transferencias) es conceptualmente igual al saldo de la cuenta financiera, que recoge la entrada o salida neta de financiación del país. En la práctica, sin embargo, ambos difieren en el saldo de errores y omisiones.

Convención de signos: la cuenta corriente y la de capital se presentan en términos brutos, como créditos y débitos. La cuenta financiera, a diferencia de ellas, se presenta en términos netos, como adquisición neta de activos financieros y como pasivos netos incurridos, ambos con signo positivo cuando aumentan. Las columnas de créditos y débitos de la tabla anterior solo reflejan la lógica de la partida doble.

Registro de los flujos: dos ejemplos ilustran la mecánica. Primero, la importación de un bien de equipo de Estados Unidos genera un débito en la cuenta corriente. El pago con cargo a un

[4]FMI (2009), §2.18 y §2.24. Los activos de reserva forman parte de la cuenta financiera.

depósito en un banco estadounidense reduce los activos financieros del residente, lo que en la cuenta financiera se registra como una adquisición neta de activos negativa, que compensa el débito corriente. Segundo, la compra de acciones de una empresa estadounidense pagada con ese mismo depósito es un intercambio de activos que se registra íntegramente dentro de la cuenta financiera (aumento de inversión de cartera, disminución de otra inversión), sin alterar su saldo.

El anexo registra cinco operaciones internacionales habituales y obtiene los saldos de las cuentas.

2.1.1. Cuenta corriente: bienes, servicios, rentas primarias y rentas secundarias

La **cuenta corriente** muestra los flujos entre residentes y no residentes de bienes y servicios y de rentas primarias y secundarias, registrados por sus valores brutos.

Tabla B11.2: Cuenta corriente.

Débitos (−)		Créditos (+)	
Importaciones de bienes (M_B)	xxx	Exportaciones de bienes (X_B)	xxx
Importaciones de servicios (M_S)	xxx	Exportaciones de servicios (X_S)	xxx
Rentas primarias pagadas (R_P)	xxx	Rentas primarias recibidas (R_R)	xxx
Transferencias corrientes pagadas (Z_P)	xxx	Transferencias corrientes recibidas (Z_R)	xxx
Saldo: $\mathbf{SB_{CC}} = \mathbf{SB}_B + \mathbf{SB}_S + \mathbf{SBR}_1 + \mathbf{SBR}_2$	**xxx**		

Cuenta de bienes o balanza comercial: recoge los bienes que cambian de propietario económico de un país a otro. Quedan fuera, por no ser comercio, los que cruzan la frontera de forma temporal sin ese traspaso, como los enviados a transformar o a reparar. Las exportaciones se anotan como créditos y las importaciones, como débitos, y su saldo es el saldo comercial, $\mathrm{SB}_B \equiv \mathrm{BC} = X_B - M_B$.

Cuenta de servicios: registra la prestación de servicios entre residentes y no residentes, con saldo $\mathrm{SB}_S = X_S - M_S$. El MBP6 distingue doce componentes, de los que destacan:

1. *Servicios de manufactura sobre insumos propiedad de otros*: valor de la transformación realizada por empresas residentes sobre bienes propiedad de no residentes, y viceversa (refino de petróleo, ensamblaje electrónico, empaquetado), sin que el bien conste como exportación ni importación.

2. *Mantenimiento y reparaciones*: misma lógica para la reparación de bienes (embarcaciones, aeronaves), excluidas las construcciones.

3. *Transporte*: traslado de personas u objetos, incluidos los servicios postales y de mensajería, pero no las telecomunicaciones.

4. *Viajes*: compras de bienes y servicios de los visitantes no residentes en viajes personales o de negocios de menos de un año, y viceversa.

5. *Construcción*: obras realizadas por empresas residentes en otros países y viceversa. Sin embargo, si la obra supera el año, se constituye una sucursal residente y la relación pasa a ser de inversión directa.

6. *Seguros y pensiones* y *servicios financieros*: registran el valor del servicio. En los seguros es el cargo implícito por el servicio y no las primas brutas, y en los servicios financieros, las comisiones explícitas y los márgenes implícitos, como los servicios de intermediación financiera medidos indirectamente (SIFMI). No incluyen los flujos de principal (cuenta financiera) ni los rendimientos (rentas primarias).

7. *Propiedad intelectual*: cargos por el uso (licencias) de patentes, marcas, derechos de autor, procesos y diseños industriales y franquicias, y licencias de reproducción y distribución. En cambio, la venta definitiva de los resultados de la I+D no figura aquí sino en otros servicios empresariales (servicios de I+D), y la de activos de comercialización (marcas, logotipos, nombres de dominio), en la cuenta de capital.

8. *Telecomunicaciones, informática e información* y *otros servicios empresariales* (I+D, consultoría, servicios técnicos y relacionados con el comercio).

Cuenta de rentas primarias: registra las rentas vinculadas a la producción y a la propiedad de activos entre residentes y no residentes, con saldo $\text{SBR}_1 = R_R - R_P$, donde R_R son las rentas recibidas de no residentes y R_P las pagadas. Comprende tres subcuentas:

1. *Rentas del trabajo* (remuneración de asalariados): créditos por los servicios laborales que los residentes prestan a empleadores no residentes (temporeros, trabajadores fronterizos) y débitos por los que los no residentes prestan a empleadores residentes.

2. *Rentas de la inversión*: créditos por los intereses, dividendos y beneficios reinvertidos que los residentes reciben por sus activos frente al exterior, y débitos por los que pagan a no residentes por los pasivos emitidos. Se clasifican según las categorías funcionales de la cuenta financiera (rentas de la inversión directa, de cartera, de otra inversión y de reservas).

3. *Otras rentas primarias*: rentas de recursos naturales, impuestos sobre la producción y las importaciones y subvenciones. En la UE se incluyen aquí los impuestos sobre la producción y las importaciones pagados a las instituciones de la Unión y las subvenciones recibidas de ellas, como las del Fondo Europeo Agrícola de Garantía (FEAGA).

Cuenta de rentas secundarias: registra las transferencias corrientes entre residentes y no residentes, con saldo $\text{SBR}_2 = Z_R - Z_P$, donde Z_R son las transferencias corrientes recibidas y Z_P las pagadas. El MBP6 define las transferencias de capital de forma positiva y las corrientes por exclusión, de modo que son corrientes todas las que no son de capital. Se desagregan en los sectores privado (remesas de emigrantes) y público (ayuda oficial al desarrollo, recursos propios IVA y RNB de la UE, cotizaciones y prestaciones sociales transfronterizas).

Saldo: el de la cuenta corriente resulta de restar a la suma de las exportaciones y las rentas y transferencias recibidas la suma de las importaciones y las rentas y transferencias pagadas:

$$\text{SB}_{\text{CC}} = \underbrace{(X_B - M_B) + (X_S - M_S)}_{X-M} + \text{SBR}_1 + \text{SBR}_2$$

donde $X - M$ es el saldo de bienes y servicios. El saldo de la cuenta corriente equivale a la brecha entre el ahorro y la inversión nacionales, $\text{SB}_{\text{CC}} = S - I$ (apartado 3.2.2), y puede mostrar superávit o déficit.

2.1.2. Cuenta de capital: transferencias de capital y ANFNP

Tabla B11.3: Cuenta de capital.

Débitos (−)		Créditos (+)	
Transferencias de capital pagadas	xxx	Transferencias de capital recibidas	xxx
Adquisiciones de ANFNP	xxx	Cesiones de ANFNP	xxx
Saldo: $\mathbf{SB_{CK}}$	**xxx**		

Transferencias de capital: son la principal partida. En ellas se cede la propiedad de un activo que no sea efectivo ni existencias, se obliga a una o ambas partes a adquirir o disponer de un activo, o el acreedor condona una deuda.

Activos no financieros no producidos: comprenden tres categorías:

- *Activos de comercialización*: marcas, cabeceras de publicaciones, logotipos y nombres de dominio.

- *Recursos naturales*: la tierra, los derechos sobre minerales, bosques y pesca y el espectro electromagnético. En cambio, si solo se cede su uso, sin traspaso de la propiedad, la operación se registra como renta primaria.

- *Contratos, arrendamientos y licencias*: activos creados por el ordenamiento jurídico, entre ellos los arrendamientos operativos, las licencias públicas y los derechos de compra.

Saldo: el saldo conjunto de la cuenta corriente y la de capital mide la *capacidad o necesidad de financiación* del país frente al exterior, $\text{CIN} = \text{SB}_{\text{CC}} + \text{SB}_{\text{CK}}$.

2.1.3. Cuenta financiera: ID, IC, DF, OI y RI

La **cuenta financiera** muestra cómo se cubre la necesidad de financiación del país o cómo el país financia al exterior si tiene exceso de ahorro. Esta cuenta, cuyos flujos se registran en términos netos, se organiza en cinco **categorías funcionales**, que agrupan los instrumentos según la motivación económica de la relación:

Tabla B11.4: Cuenta financiera.

Adquisición neta de activos financieros (ΔAF)		Pasivos netos incurridos (ΔPF)	
Inversión directa en el exterior	xxx	Inversión directa del exterior	xxx
Inversión de cartera: activos	xxx	Inversión de cartera: pasivos	xxx
Derivados financieros y OCAPE: activos	xxx	Derivados financieros y OCAPE: pasivos	xxx
Otra inversión: activos	xxx	Otra inversión: pasivos	xxx
Activos de reserva (ΔRI)	xxx		
Saldo: $\mathbf{SB_{CF} = \Delta AF - \Delta PF}$ (préstamo neto (+) / endeudamiento neto (−))	**xxx**		

1. *Inversión directa* (ID): inversión transfronteriza realizada con el objetivo de establecer un interés duradero y cierto control o influencia sobre una empresa no residente. El MBP6 exige

una participación de al menos el 10 % del capital con derecho a voto. Esta categoría incluye las participaciones en el capital, los beneficios reinvertidos y los préstamos entre empresas relacionadas.

2. *Inversión de cartera* (IC): inversión transfronteriza en valores negociables (acciones y participaciones en fondos, incluidos los derechos de suscripción, e instrumentos de deuda) distinta de la directa y de los activos de reserva. Busca optimizar el binomio rentabilidad-riesgo y es más volátil.

3. *Derivados financieros y opciones de compra de acciones por empleados* (DF; OCAPE para las opciones de compra de acciones por empleados): se separan porque su función es transferir riesgo y no financiar.

4. *Otra inversión* (OI): categoría residual que incluye préstamos y depósitos, créditos comerciales y anticipos, arrendamientos financieros, cesiones y adquisiciones temporales de valores, los préstamos de ayuda oficial al desarrollo, las cuotas en organismos internacionales y la estimación de la entrada neta de billetes por turismo.

5. *Activos de reserva* (RI): activos exteriores a disposición inmediata de las autoridades monetarias para financiar desequilibrios de pagos e intervenir en el mercado de cambios: oro monetario, derechos especiales de giro (DEG), posición de reserva en el FMI y otros activos (divisas y depósitos, valores). Un aumento de reservas se registra como adquisición neta de activos, con signo positivo. En regímenes flexibles sus variaciones deberían ser moderadas, pues el ajuste se produce vía tipo de cambio. Los criterios de suficiencia de las reservas se exponen en el apartado 3.2.4.

Las categorías funcionales se cruzan con la clasificación por instrumentos, común al SCN-2008 y al SEC-2010 (tema A30, apartado 2.3):

Tabla B11.5: Categorías funcionales e instrumentos financieros en el MBP6.

Instrumento	ID	IC	DF	OI	RI
Oro monetario y DEG				asignaciones (pasivo)	tenencias
Efectivo y depósitos				✓	✓
Valores representativos de deuda	entre empresas relacionadas	✓			✓
Préstamos	entre empresas relacionadas			✓	posición FMI
Participaciones en el capital y en fondos de inversión	≥ 10 %	< 10 %		otras participaciones	✓
Seguros, pensiones y garantías estandarizadas				✓	
Derivados financieros y OCAPE			✓		✓
Otras cuentas pendientes de cobro o pago				✓	

Saldo: el de la cuenta financiera es el *préstamo neto* (+), cuando la economía adquiere más activos de los pasivos en que incurre y proporciona así financiación neta al resto del mundo, o el *endeudamiento neto* (−), en caso contrario:

$$\mathrm{SB_{CF}} = \Delta\mathrm{ID} + \Delta\mathrm{IC} + \Delta\mathrm{DF} + \Delta\mathrm{OI} + \Delta\mathrm{RI}$$

donde cada término es la adquisición neta de activos financieros menos los pasivos netos incurridos de la categoría correspondiente. Conceptualmente, el saldo coincide con $\mathrm{SB_{CC}} + \mathrm{SB_{CK}}$.

2.1.4. Errores y omisiones netos

En principio, la balanza de pagos está equilibrada por la partida doble, pero en la práctica las imperfecciones de los datos generan una discrepancia. La partida de **errores y omisiones netos** se obtiene restando de la cuenta financiera el agregado de las cuentas corriente y de capital, $\mathrm{SB_{EO}} = \mathrm{SB_{CF}} - (\mathrm{SB_{CC}} + \mathrm{SB_{CK}})$. Un valor positivo indica que los créditos de las cuentas corriente y de capital están infravalorados, que sus débitos están sobrevalorados, que el aumento neto de activos financieros está sobrevalorado o que el aumento neto de pasivos está infravalorado. Un valor persistente y del mismo signo suele indicar flujos financieros no registrados.

2.2. Otras cuentas internacionales

Las demás cuentas internacionales son la posición de inversión internacional y la cuenta de otras variaciones de activos y pasivos financieros.

2.2.1. Posición de inversión internacional

La posición de inversión internacional (PII), definida en la introducción, incluye entre los activos el oro monetario en lingotes, único activo financiero sin pasivo de contrapartida. Se organiza con las mismas categorías funcionales e instrumentos que la cuenta financiera. La diferencia entre activos y pasivos es la **PII neta**:

$$\mathrm{PIIN}_t = \mathrm{AFE}_t - \mathrm{PFE}_t$$

donde AFE_t son los activos financieros frente al exterior y PFE_t los pasivos. Si $\mathrm{PIIN}_t > 0$, el país es acreedor neto, y si $\mathrm{PIIN}_t < 0$, deudor neto. No debe confundirse la posición acreedora o deudora, que es un *stock*, con la capacidad o necesidad de financiación, que es un flujo. La PIIN varía con las transacciones que recoge la cuenta financiera y con las otras variaciones:

$$\Delta\mathrm{PIIN}_t = \mathrm{SB}_{\mathrm{CF},t} + \mathrm{COVAPF}_t = \mathrm{SB}_{\mathrm{CF},t} + \mathrm{EV}_{F,t} + \mathrm{OVV}_t$$

donde $\mathrm{EV}_{F,t}$ es el efecto valoración y OVV_t las otras variaciones de volumen. En ausencia de otras variaciones, un superávit por cuenta corriente y de capital (préstamo neto) aumenta la posición acreedora o reduce la deudora, y viceversa.

2.2.2. Cuenta de otras variaciones de activos y pasivos financieros

La cuenta de otras variaciones de activos y pasivos financieros (COVAPF), ajena a la balanza de pagos, recoge los flujos que no nacen de transacciones entre residentes y no residentes. Comprende el **efecto valoración** (variaciones de los precios de los activos y de los tipos de cambio) y las **otras variaciones de volumen** (cancelaciones unilaterales de deuda, reclasificaciones, cambios de residencia), y concilia así la balanza de pagos con la PII. En economías con grandes posiciones brutas, el efecto valoración puede superar en un año al saldo de la cuenta financiera.

2.3. Relación con las cuentas nacionales

La cuenta del resto del mundo del SEC-2010 registra las mismas operaciones que la balanza de pagos, desde la óptica de los no residentes:

Tabla B11.6: Correspondencia entre las cuentas internacionales y la cuenta del resto del mundo.

Balanza de pagos (MBP6)	Cuenta del resto del mundo (SEC-2010)	Saldo
Bienes y servicios	Cuenta exterior de bienes y servicios (V.I)	Saldo exterior de bienes y servicios
Rentas primarias y secundarias	Cuenta exterior de rentas primarias y transferencias corrientes (V.II)	Saldo exterior corriente
Cuenta de capital	Cuenta de capital del resto del mundo (V.III.1)	Capacidad/necesidad de financiación
Cuenta financiera	Cuenta financiera del resto del mundo (V.III.2)	Adquisición neta de activos menos pasivos
COVAPF	Cuenta de otras variaciones de activos (V.III.3)	—
PII	Balances del resto del mundo (V.IV)	Patrimonio financiero neto

Puesto que la cuenta del resto del mundo se elabora desde la óptica de los no residentes, sus saldos tienen el signo contrario a los de la balanza de pagos, de modo que un superávit corriente nacional es un déficit del resto del mundo. Salvo ese cambio de signo, los saldos coinciden, de forma que la capacidad o necesidad de financiación de la nación en las cuentas nacionales es igual a $\mathrm{SB}_{\mathrm{CC}} + \mathrm{SB}_{\mathrm{CK}}$. Asimismo, el saldo de la cuenta financiera de la economía total, en el que los activos y pasivos entre residentes se cancelan, coincide con el de la balanza de pagos, que difiere de $\mathrm{SB}_{\mathrm{CC}} + \mathrm{SB}_{\mathrm{CK}}$ en los errores y omisiones:

$$\mathrm{CIN} = S + \mathrm{SB}_{\mathrm{CK}} - I = \mathrm{SB}_{\mathrm{CC}} + \mathrm{SB}_{\mathrm{CK}} \simeq \mathrm{SB}_{\mathrm{CF}}^{\mathrm{BP}}$$

donde S e I son el ahorro y la inversión nacionales. Esta correspondencia permite pasar de la contabilidad exterior a la interpretación económica.

> La secuencia de cuentas nacionales y la derivación de la capacidad o necesidad de financiación se desarrollan en el tema A30, apartados 2.1 a 2.4.

2.4. Conclusiones

Las cuentas internacionales forman un sistema cerrado y coherente. En la balanza de pagos, la partida doble iguala el saldo de las cuentas corriente y de capital con el de la cuenta financiera, salvo por los errores y omisiones. La suma de las dos primeras es la capacidad o necesidad de financiación de la nación, y la cuenta financiera muestra, por categorías funcionales, cómo se presta o se obtiene esa financiación. La posición de inversión internacional recoge el *stock* resultante, que no debe confundirse con aquel flujo, y la cuenta de otras variaciones concilia ambos mediante los efectos de valoración y de volumen.

La armonización con el SEC-2010, que solo invierte los signos, permite además leer cada saldo exterior como reflejo de uno interno, de modo que el saldo corriente equivale a la brecha entre ahorro e inversión nacionales. Sobre esa doble lectura descansa la interpretación económica de la sección 3, que va de los factores comunes a todas las cuentas al análisis de cada una y de la PII.

3. Interpretación económica de las cuentas internacionales

Entendida la estructura contable, queda por ver su utilidad económica como punto de partida del análisis de los desequilibrios exteriores (apartado 1.2.1), que, cuando son persistentes, dan lugar a problemas de coordinación y conflictos de interés entre países. Se estudian, por este orden, los factores que afectan a la vez a todas las cuentas, el análisis de cada una de ellas, la lectura de la posición de inversión internacional y de la riqueza nacional y, por último, otras presentaciones de la balanza de pagos y la evidencia empírica.

3.1. Factores que afectan de forma global a las cuentas internacionales

Algunos factores influyen a la vez en todas las cuentas internacionales, porque condicionan tanto el comercio de bienes y servicios como las rentas y los flujos de capital. Se agrupan en dos bloques, el de la competitividad de la economía y el de los rasgos estructurales de su inserción exterior.

Competitividad y otros determinantes: el primer bloque reúne cuatro grupos de factores:

1. *Competitividad-precio y no precio*: la competitividad-precio depende de los precios y costes del país en relación con los de sus competidores, expresados en la misma moneda. La competitividad no precio depende de la calidad, la diferenciación y la imagen de marca de los productos, del servicio posventa y del cumplimiento de los plazos de entrega.

2. *Contenido tecnológico de los flujos*: los bienes y servicios de mayor intensidad tecnológica compiten más por calidad y diferenciación que por precio, y la inversión directa que incorpora tecnología la transfiere al país receptor y refuerza su capacidad exportadora (tema B18, apartado 3.2.3).

3. *Cuestiones institucionales y relaciones exteriores*: los acuerdos comerciales y de inversión, la pertenencia a uniones aduaneras o monetarias y las sanciones o represalias comerciales condicionan el acceso a los mercados y a la financiación exterior.

4. *Otros factores*: los naturales, como la dotación de recursos, el clima o las catástrofes, que alteran la oferta exportable y las necesidades de importación (p. ej. de energía o de alimentos), y las expectativas, que mueven sobre todo los flujos financieros y el tipo de cambio.

Indicadores de competitividad-precio: el tipo de cambio efectivo real (TCER) deflactado por los costes laborales unitarios (CLU) compara los costes del país con los de sus socios comerciales, expresados en la misma moneda. Con el tipo de cambio directo del libro (tema B13, apartado 1.1.2) y en tasas de variación:

$$g_e = g_E + g_{\text{CLU}^*} - g_{\text{CLU}}$$

donde e es el TCER, E el tipo de cambio efectivo nominal (unidades de moneda nacional por unidad de una cesta de monedas extranjeras) y CLU y CLU^* los costes laborales unitarios del país y la media ponderada de los de sus socios. Un aumento de e es una depreciación real, es decir, una ganancia de competitividad-precio. Los índices que publican el BCE y el FMI usan, sin embargo, la cotización inversa, de modo que en ellos un aumento indica una apreciación (tema A30, apartado 3.4.3).

Los costes laborales unitarios crecen con el coste laboral por asalariado (CLA), que incluye las cotizaciones sociales a cargo del empleador, y disminuyen con la productividad aparente del trabajo (PAT):

$$g_{\text{CLU}} = g_{\text{CLA}} - g_{\text{PAT}}$$

de modo que un país gana competitividad-precio si sus CLU crecen menos que los de sus socios, porque sus costes laborales crecen menos o su productividad más, o si su moneda se deprecia en términos nominales.

La relación real de intercambio (RRI), definida en el tema B05, apartado 1.1, mide el poder de compra de las exportaciones en términos de importaciones:

$$\text{RRI} = \frac{P_X}{P_M}$$

donde P_X y P_M son los índices de precios (o de valores unitarios) de las exportaciones y de las importaciones, ambos en moneda nacional. Una mejora de la RRI eleva la renta real del país, pues el mismo volumen de exportaciones paga más importaciones, pero, si se debe a una subida de los precios de exportación, puede ir acompañada de una pérdida de competitividad-precio y de cuota de mercado.

Inflación: la comparación entre el índice de precios de consumo (IPC) y el deflactor del PIB completa la lectura exterior con la de los precios (tema A30, apartado 3.3.2). El IPC incluye bienes importados y el deflactor no. Por ello, un IPC que crece más que el deflactor suele reflejar inflación importada, debida al encarecimiento de las importaciones en moneda nacional, por la subida de sus precios o por la depreciación de la moneda. El deflactor, por su parte, se descompone por el lado de las rentas en tres componentes por unidad de producto, los costes laborales unitarios, el excedente bruto de explotación (y renta mixta) y los impuestos netos sobre la producción y las importaciones. Ello permite distinguir si la inflación interna procede de los salarios, de los márgenes o de los impuestos.

Rasgos estructurales: el segundo bloque recoge seis rasgos de la inserción exterior de la economía:

1. *Grado de apertura*: cuanto mayor es el peso del comercio exterior en el PIB, más sensibles son las cuentas a la demanda mundial y a los precios relativos (sus indicadores se exponen en el apartado 3.2.1).

2. *Número de sectores competidores*: una base exportadora diversificada, con muchos sectores capaces de competir en el exterior, amortigua las perturbaciones que afectan a un producto concreto.

3. *Concentración de los flujos*: la dependencia de pocos productos, mercados o acreedores aumenta la vulnerabilidad ante perturbaciones específicas, como la caída del precio de la materia prima que se exporta o la crisis del principal socio comercial.

4. *Régimen cambiario*: determina si el ajuste exterior se produce a través del tipo de cambio o de las reservas, los precios y la renta internos (tema B15, apartado 1.4).

5. *Grado de integración internacional*: la integración comercial, a través de las cadenas globales de valor, y la financiera amplían las fuentes de demanda y de financiación, pero también la exposición a las perturbaciones externas y al contagio (tema B18).

6. *Estabilidad y seguridad jurídica*: la estabilidad macroeconómica y política y la protección de los derechos de propiedad atraen capital estable, sobre todo inversión directa, y reducen la prima de riesgo que exigen los acreedores.

3.2. Análisis de cada cuenta

Cada cuenta responde, además, a factores propios y admite una lectura económica distinta. La cuenta corriente se analiza por subcuentas y por su saldo, la cuenta de capital aporta ahorro adicional y la cuenta financiera muestra cómo se financia la economía.

3.2.1. Cuenta corriente: análisis de las subcuentas

Balanza de bienes y servicios: su estudio permite analizar la especialización del país, detectar la dependencia importadora de ciertos productos (energía, bienes de equipo) y medir el *grado de apertura* con tres indicadores (tema A30, apartado 3.4.3). En ellos, X y M son las exportaciones e importaciones de bienes y servicios:

- *Tasa de apertura*: $\text{TA} = (X + M)/\text{PIB}$, peso del comercio exterior en la economía.
- *Tasa de cobertura*: $\text{TC} = X/M$, proporción de las importaciones que se pagan con exportaciones.
- *Tasa de penetración de las importaciones*: $\text{TP} = M/\text{PIB}$, dependencia de la oferta exterior, que a menudo se mide sobre la demanda interna en lugar de sobre el PIB.

Estos indicadores miden el comercio en términos brutos, no de valor añadido. Así, un bien importado que apenas se transforma antes de reexportarse cuenta entero como exportación, y, con cadenas globales de valor extendidas, eso supone contar dos veces el mismo valor (de ahí las medidas de comercio en valor añadido, TiVA). La especialización se mide con el índice de Balassa (ventaja comparativa revelada) y el peso del comercio intraindustrial, con el índice de Grubel y Lloyd, que se desarrollan en el tema B06, apartado 1.1.1.

La variación del saldo de bienes y servicios mide, además, la aportación de la demanda externa al crecimiento. Como $\text{PIB} = A + (X - M)$, donde $A = C + I + G$ es la demanda interna o absorción, la tasa de crecimiento del PIB en volumen se descompone en dos aportaciones:

$$g_{\text{PIB}} = \frac{\Delta A_t}{\text{PIB}_{t-1}} + \frac{\Delta (X - M)_t}{\text{PIB}_{t-1}}$$

donde el primer término, el crecimiento de la demanda interna ponderado por su peso en el PIB, es la aportación de la demanda interna y el segundo, la de la demanda externa neta, ambas en puntos porcentuales. Una aportación negativa de la demanda externa no indica por sí sola una pérdida de competitividad, porque puede deberse a las importaciones que arrastra una demanda interna pujante.

La apertura reduce también la eficacia de la política fiscal, porque parte del aumento de la demanda se filtra hacia las importaciones. El multiplicador de la economía abierta es, por ello, menor que el de la cerrada, y tanto más cuanto mayor es la propensión marginal a importar (tema A32, apartado 1.2.1, y tema B12, apartado 1.3.1).

Balanza de rentas primarias: es el vínculo entre el producto interior y la renta nacional, es decir, entre el valor añadido generado dentro de las fronteras y el que perciben los residentes:

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

donde RNB es la renta nacional bruta y SBR_1 el saldo de rentas primarias. En las economías con una PII deudora, SBR_1 tiende a ser negativo ($\text{RNB} < \text{PIB}$) por los intereses y dividendos

pagados sobre sus pasivos exteriores. La cuantía de esos pagos depende del tipo de interés y de la moneda en que está denominada la deuda externa. En particular, cuando un país no puede endeudarse en el exterior en su propia moneda (el llamado *pecado original*, tema B17, apartado 2.4.2), una depreciación encarece en moneda nacional los intereses que paga. En Irlanda, la RNB es muy inferior al PIB por los beneficios, repatriados o reinvertidos, de las multinacionales allí radicadas.

Balanza de rentas secundarias: añade a la RNB las transferencias corrientes netas para obtener la renta nacional bruta disponible, RNBD = RNB+SBR_2. Entre ellas, las transferencias oficiales que financian gasto corriente (apartado 2.1.1), como el apoyo presupuestario, elevan la renta disponible del receptor sin generar deuda. Las ayudas a la inversión, en cambio, son transferencias de capital y no forman parte de la renta disponible.

Las rentas secundarias han ganado atención con la globalización por las *remesas*, ingresos de los hogares transferidos entre países, vinculados sobre todo a la migración, en efectivo o en especie y por canales formales o informales. Las remesas incluyen renta primaria (principalmente, la remuneración de asalariados) y renta secundaria (las transferencias personales entre hogares). En la balanza de pagos se definen de forma algo más amplia que en las estadísticas del movimiento de personas, ya que su definición no se apoya en las nociones de migración, empleo y relaciones familiares. En muchas economías superan a la ayuda oficial o a la inversión directa y son una fuente estable de financiación exterior y de reducción de la pobreza.

3.2.2. Cuenta corriente: análisis del saldo

El saldo de la cuenta corriente equivale a la brecha ahorro-inversión de la nación, y existen cuatro enfoques para interpretarlo:

Identidad ahorro-inversión nacional: $\text{SB}_{\text{CC}} = S - I$, donde S e I son el ahorro y la inversión nacionales. Un déficit puede entenderse como insuficiente ahorro nacional, lo cual suele ser negativo, o como elevada inversión, lo cual no necesariamente lo es.

Identidad por sectores institucionales: desagregando el ahorro y la inversión entre el sector privado y el público:

$$\text{SB}_{\text{CC}} = (S_P - I_P) + (S_G - I_G)$$

donde S_P e I_P son el ahorro y la inversión privados y $S_G - I_G$ el saldo público, que, sumadas sus transferencias netas de capital, da el saldo presupuestario (capacidad o necesidad de financiación de las Administraciones públicas). El ahorro privado debe financiar la inversión privada, el déficit público y el saldo frente al exterior, de modo que un exceso de inversión privada sobre el ahorro privado se compensa con superávit público o con déficit corriente.

La coincidencia de déficit corriente y déficit público se conoce como *déficits gemelos* (*twin deficits*), término acuñado en los años ochenta a raíz de la experiencia de la Administración Reagan. Según la hipótesis de los déficits gemelos, la causalidad opera por dos canales. Por una parte, el aumento del déficit público estimula la renta y las importaciones ($\uparrow (G - T) \rightarrow \uparrow$ PIB $\rightarrow \uparrow M$). Por otra, eleva el tipo de interés, lo que, con movilidad de capitales y tipo de cambio flexible, atrae capitales, aprecia la moneda y reduce las exportaciones.

Enfoque de la absorción: sumando y restando la demanda interna:

$$\text{SB}_{\text{CC}} = (X - M) + \text{SBR}_1 + \text{SBR}_2 = (\text{PIB} - A) + \text{SBR}_1 + \text{SBR}_2 = \text{RNBD} - A$$

donde A es la absorción. Si se prescinde de las rentas secundarias, la identidad se reduce a $\mathrm{SB}_{\mathrm{CC}} = \mathrm{RNB} - A$, su forma habitual con el producto nacional bruto (PNB), equivalente a la RNB. Así, un déficit significa que el país gasta más de lo que produce y percibe, y se corrige reduciendo el gasto o, vía medidas estructurales, produciendo más.

Enfoque intertemporal: un déficit no tiene por qué ser negativo ni un superávit preferible. El saldo exterior es el instrumento con el que la economía suaviza su consumo ante perturbaciones temporales o adelanta la inversión cuando su rentabilidad esperada es alta, por lo que el juicio depende de las causas del saldo y no solo de su signo. A la inversa, un superávit puede reflejar un consumo débil o la incapacidad del sistema financiero para canalizar el ahorro hacia la inversión interna.

> Los enfoques de ajuste de la balanza de pagos (elasticidades, absorción, monetario e intertemporal) se desarrollan en el tema B12, y el modelo intertemporal en que se apoya la última lectura, en su apartado 3.1.

3.2.3. Cuenta de capital: ahorro adicional

Una transferencia de capital del resto del mundo no incrementa los pasivos exteriores del país receptor ni los activos del que la realiza, puesto que es una transacción unilateral, sin contraprestación. Sin embargo, sí supone una aportación de ahorro al receptor, que aumenta sus disponibilidades para financiar su formación de capital. Por eso la capacidad o necesidad de financiación de la nación suma al ahorro el saldo de la cuenta de capital, $\mathrm{CIN} = S + \mathrm{SB}_{\mathrm{CK}} - I$ (apartado 2.3), formado por las transferencias netas de capital recibidas y las cesiones netas de ANFNP. Si $S + \mathrm{SB}_{\mathrm{CK}} < I$, existe necesidad de financiación exterior, y si $S + \mathrm{SB}_{\mathrm{CK}} > I$, capacidad de financiación, que financia la inversión de otros países y mejora la posición de inversión internacional. Los flujos de capital redistribuyen esta capacidad entre países y permiten a unos gastar por encima de su renta si otros gastan por debajo.

En los Estados miembros de la UE, una parte importante de las transferencias de capital recibidas procede de los fondos europeos destinados a inversión, como las ayudas estructurales a infraestructuras (apartado 1.3.2). Las cesiones de ANFNP, como la venta de derechos sobre recursos naturales, de licencias o de marcas, aportan recursos del mismo modo.

3.2.4. Cuenta financiera: composición de la financiación

La cuenta financiera mide cómo se financia el préstamo o endeudamiento neto. Si la economía ha gastado más de lo que su renta permite, tendrá que emitir pasivos frente al exterior, liquidar activos exteriores o disponer de reservas, mientras que, si ha gastado menos, acumula activos o cancela pasivos. La composición también importa, porque la financiación vía inversión directa es más estable que la de cartera o los depósitos a corto plazo, y la denominación en moneda extranjera de los pasivos determina la exposición a una depreciación.

En el caso del endeudamiento, la financiación puede llegar como un flujo autónomo de fondos hacia el país o tener que generarse mediante medidas de política económica, uso de reservas o ajuste de la balanza de pagos. La financiación autónoma depende de la estabilidad económica y política, de la existencia de lazos económicos duraderos y de las aptitudes tecnológicas y de gestión del país, especialmente relevantes para la inversión directa.

Cada categoría funcional (apartado 2.1.3) informa de un aspecto distinto de la financiación:

▪ *Inversión directa*: la de nueva planta (*greenfield*) crea capacidad productiva desde cero, mientras que la que se dirige a instalaciones existentes (*brownfield*), normalmente mediante fusiones y adquisiciones, transfiere la propiedad y amplía menos el capital instalado (tema B18, apartado 1.2.2).

▪ *Inversión de cartera*: es propensa a los comportamientos de rebaño (tema B17, apartado 2.3). Incluye el capital especulativo a corto plazo (*hot money*), que entra y sale según los diferenciales de rentabilidad y se retira hacia activos seguros en las crisis (huida hacia la calidad, *flight to quality* o *flight to safety*). Una financiación muy dependiente de la inversión de cartera es, por tanto, una señal de vulnerabilidad.

▪ *Otra inversión*: además de los préstamos y depósitos bancarios, recoge la recepción de préstamos oficiales, bilaterales o multilaterales (p. ej. del FMI), que en las crisis sustituyen a la financiación de mercado.

▪ *Derivados financieros*: no financian, sino que reflejan la cobertura de riesgos (cambiario, de tipo de interés o de precio de las materias primas) y la participación de los residentes en los mercados de materias primas.

▪ *Activos de reserva*: sus variaciones muestran la intervención de las autoridades. Las reservas caen bruscamente en las paradas súbitas de la entrada de capitales (*sudden stops*), cuando el banco central vende divisas para sostener el tipo de cambio o cubrir el desequilibrio. El trilema de Mundell-Fleming, o trinidad imposible (tema A32, apartado 1.3.4), acota su uso, pues con libre movilidad del capital defender un tipo de cambio fijo exige renunciar a la autonomía monetaria. Su suficiencia se juzga con tres criterios habituales (FMI, *Assessing Reserve Adequacy*, 2011, y tema B15, apartado 3.1.1):

- *Cobertura de importaciones*: reservas que cubran al menos tres meses de importaciones.
- *Regla de Guidotti-Greenspan*: reservas de al menos el 100 % de la deuda externa a corto plazo por vencimiento residual.
- *Cobertura del dinero amplio*: reservas en torno al 20 % de M2, frente al riesgo de fuga de depósitos.

3.3. Posición de inversión internacional y hoja de balance

La cuenta financiera y la COVAPF explican el cambio de la PII entre el comienzo y el final del periodo (apartado 2.2), de modo que la lectura de la financiación conduce a la de la posición. El **enfoque de la hoja de balance** amplía el análisis de la balanza de pagos a los *stocks*. Su justificación es que la estructura financiera de las economías (composición de activos y pasivos por instrumento, plazo, moneda y sector) ha sido una fuente central de vulnerabilidad en las crisis recientes. La PII y el balance nacional se leen desde cuatro puntos de vista.

Situación acreedora o deudora: el signo de la PII neta (apartado 2.2.1) indica si el país es acreedor o deudor frente al resto del mundo, y su cociente sobre el PIB, la magnitud de esa posición. Dos ejemplos ilustran su dispersión. Hong Kong era acreedor neto por casi seis veces su PIB a finales de 2025, y Grecia tenía una PII neta del −137,6 % del PIB en el tercer trimestre de ese año[5]. Una posición muy deudora eleva los pagos de rentas primarias y expone al país a un cambio de actitud de sus acreedores.

[5] Departamento de Censos y Estadística de Hong Kong, estadísticas de balanza de pagos y PII del cuarto trimestre de 2025 (20 de marzo de 2026), con una PII neta de 19,5 billones de dólares de Hong Kong (5,9 veces el PIB), y Eurostat, serie *tipsii40* (PII neta en porcentaje del PIB).

Composición de la riqueza nacional: la riqueza de una economía es la suma de sus activos no financieros y su PII neta (tema A30, apartado 3.2.2):

$$\text{AN} = K + \text{PIIN}$$

donde AN es la riqueza nacional y K el *stock* de activos no financieros situados en el país. Se supone que la inversión y el ahorro se miden netos del consumo de capital fijo y que no hay otras variaciones de volumen, operaciones con ANFNP ni errores y omisiones. En ese caso, la variación de la riqueza nacional es la suma de la inversión, del saldo de la cuenta financiera y de los efectos de valoración del capital y de la posición exterior:

$$\Delta\text{AN} = I + \text{SB}_{\text{CF}} + \text{EV}_K + \text{EV}_F \;\rightarrow\; \Delta\text{AN} = S + \text{SB}_{\text{CK}} + \text{EV}_K + \text{EV}_F$$

donde EV_K y EV_F son las revalorizaciones de los activos no financieros y de la posición exterior. De ahí que una economía solo pueda aumentar su riqueza ahorrando, recibiendo transferencias de capital u obteniendo ganancias de capital, y que de esas tres vías solo controle directamente la primera.

El Banco Mundial (*The Changing Wealth of Nations 2024*) amplía esta medida a la riqueza integral, que suma al capital producido el capital natural (bosques, pesquerías, reservas de combustibles fósiles), el capital humano (valor presente de las rentas futuras del trabajo) y los activos exteriores netos. Así puede valorarse si el desarrollo es sostenible, es decir, si el país repone con capital producido, capital humano o activos exteriores el capital natural que agota.

Madurez económica del país: la hipótesis de las fases de la balanza de pagos asocia la posición exterior al grado de desarrollo. Un país joven es deudor, pues importa capital y registra déficits corrientes. Al madurar, obtiene superávits comerciales con los que atiende las rentas de su deuda y, cuando su cuenta corriente pasa a superávit, reduce su posición deudora hasta convertirse en acreedor. El acreedor maduro puede financiar déficits comerciales con las rentas de sus activos exteriores. Sin embargo, la evidencia empírica respalda poco una secuencia tan regular, porque la relación entre la fase de la balanza de pagos y el nivel de desarrollo no es monótona (Ono, 2014).

Sostenibilidad y vulnerabilidad exterior: una posición deudora es sostenible si el país puede atender su servicio de forma continuada, sin financiación excepcional ni un ajuste brusco de la absorción. La condición formal, que relaciona el superávit comercial necesario para estabilizar la deuda exterior sobre el PIB con la diferencia entre el tipo de interés real y la tasa de crecimiento, se desarrolla en el tema B12, apartado 3.3. La vulnerabilidad depende, además, de la composición de los pasivos (deuda o inversión directa, plazo, moneda y sector deudor), que el enfoque de la hoja de balance pone en primer plano. Un descalce de monedas, por ejemplo, convierte una depreciación en una pérdida de patrimonio de los deudores, el efecto de hoja de balance de las crisis gemelas (tema B17, apartado 3.2).

3.4. Otras presentaciones y evidencia empírica

Se exponen las presentaciones alternativas de la balanza de pagos (analítica, monetaria y sectorial) y la evidencia empírica sobre los desequilibrios, de los globales de 2000-2008 a los posteriores a 2020.

3.4.1. Presentaciones alternativas de la balanza de pagos

Presentación analítica: parte de la distinción de Meade (1951) entre transacciones *autónomas*, que responden a sus propios motivos (comerciar, invertir, prestar), y *acomodantes* o compensatorias, que las autoridades realizan para cubrir el desequilibrio que dejan las primeras. Una "línea" separa ambas y deja por encima las autónomas, cuyo saldo mide el desequilibrio, y por debajo, las que lo financian. Según dónde se trace se obtienen los saldos clásicos, de menor a mayor amplitud:

- *Saldo comercial*: recoge solo los bienes y aproxima la competitividad, pero ignora servicios, rentas y capitales.
- *Saldo por cuenta corriente*: añade servicios y rentas y equivale a la brecha ahorro-inversión.
- *Capacidad o necesidad de financiación*: añade la cuenta de capital (apartado 2.1.2).
- *Saldo básico*: suma a la cuenta corriente los flujos de capital a largo plazo, sobre todo la inversión directa. Indica si el déficit corriente se financia con recursos estables o con capital a corto plazo, reversible.
- *Saldo global*: abarca todas las transacciones salvo las reservas y las partidas vinculadas a ellas (crédito del FMI y financiación excepcional), de modo que su signo indica si las autoridades han tenido que financiar el desequilibrio. La presentación analítica del FMI se basa en este saldo.

La *financiación excepcional* (MBP6, apéndice 1) agrupa los acuerdos con que las autoridades atienden necesidades de balanza de pagos como alternativa o complemento a las reservas y al crédito del FMI: condonaciones y donaciones intergubernamentales, canjes de deuda por capital, préstamos de apoyo a la balanza de pagos, reprogramaciones y refinanciaciones, pagos anticipados y recompras de deuda, y acumulación o pago de atrasos.

La línea entre transacciones autónomas y acomodantes es nítida con tipo de cambio fijo y controles de capital, el mundo de Meade. Con libre movilidad de capitales, en cambio, los flujos privados reaccionan a las expectativas, y una entrada de capital a corto plazo puede ser tan acomodante como una venta de reservas. Por eso el MBP6 no propone un saldo único y advierte de que la presentación analítica no sirve para todos los fines (§14.16), ya que cada saldo responde a una pregunta distinta.

Presentación monetaria: vincula la balanza de pagos con las estadísticas monetarias y muestra cómo los flujos con el exterior afectan a los balances del sistema bancario y a la creación de liquidez interna a través de la variación de los activos exteriores netos.

Presentación sectorial: desagrega los flujos y posiciones exteriores por sectores institucionales (Administraciones públicas, banco central, instituciones de depósito, otras instituciones financieras y sector privado no financiero), lo que permite identificar qué sector origina el desequilibrio exterior y cuál soporta la exposición. En la Unión Económica y Monetaria (UEM), los saldos TARGET del banco central, derivados de los pagos transfronterizos liquidados en T2 (el sistema que sustituyó a TARGET2 en marzo de 2023), son la partida que puede reflejar la financiación oficial de los desequilibrios privados. No obstante, también responden a otros factores, como las compras de activos del Eurosistema.

Análisis comparativo: el contraste bilateral de las cuentas entre economías socias permite detectar focos de vulnerabilidad, dependencia de un mercado o acreedor concreto y exposición a efectos contagio, además de servir de control de calidad estadístico mediante las asimetrías bilaterales.

3.4.2. Evidencia empírica sobre los desequilibrios

Desequilibrios globales (2000-2008): la liberalización financiera de los años noventa culminó en un periodo de desequilibrios externos elevados y persistentes (*global imbalances*), que fueron una de las principales amenazas para la estabilidad de la economía mundial antes de la crisis financiera global:

1. *EE. UU. y Reino Unido*: presentaban déficits por cuenta corriente elevados y persistentes (en EE. UU., en torno al 6 % del PIB en 2006), que financiaban consumo más que inversión. Su corrección exigía mayor ahorro nacional, privado y público, y, por parte de los acreedores, dejar de acumular pasivos americanos. Sin embargo, esto supondría una caída del dólar que reduciría el valor real de las tenencias de los propios acreedores, un problema de acción colectiva que contribuyó a perpetuar el desequilibrio.

2. *Superávits contrapartida*: Alemania, China, Japón y los exportadores de petróleo.

3. *Periferia del euro*: Grecia, Portugal y España registraron déficits por cuenta corriente en torno al 10 % del PIB o superiores en 2007-2008, financiados en gran medida con deuda bancaria a corto plazo y acompañados de posiciones de inversión internacional muy deudoras. Tras la crisis, los corrigieron mediante la contracción de la absorción y la devaluación interna.

Crisis financiera global: provocó un colapso súbito de los flujos comerciales y financieros, transmitido casi de inmediato por las cadenas globales de valor. En este *great trade collapse*, el volumen del comercio mundial cayó en torno a un 12 % en 2009 (más de un 20 % en valor), unas cinco veces más que el PIB. La *sudden stop* de la financiación exterior de los países deficitarios de la periferia del euro forzó un ajuste de sus cuentas corrientes vía contracción de la absorción, mientras que la corrección de los superávits fue mucho menor. La asimetría del ajuste entre deudores y acreedores es la lección central del episodio.

Desequilibrios tras 2020: la guerra comercial entre Estados Unidos y China (desde 2018, reabierta a mayor escala en 2025), la pandemia y la invasión de Ucrania (2022) han alterado la composición de los flujos más que el patrón de los saldos. Entre esos cambios de composición figuran la relocalización de cadenas de valor, el *shock* energético y las sanciones. En cambio, el déficit corriente de EE. UU. y los superávits de China y de la zona del euro persisten, y la atención se desplaza cada vez más a las posiciones brutas y a la denominación de los pasivos.

> La globalización y la coordinación internacional de políticas se desarrollan en los temas B18 y B19; el papel del FMI en los desequilibrios de balanza de pagos, en el tema B22; las crisis de balanza de pagos, en el tema B17.

3.5. Conclusiones

La lectura de las cuentas va de los flujos a los *stocks*. La competitividad y los rasgos estructurales de la inserción exterior afectan a todas ellas a la vez, pero cada una responde a factores propios. Las subcuentas corrientes informan de la especialización, de la aportación de la demanda externa al crecimiento y del paso del producto interior a la renta nacional. El saldo corriente, brecha entre ahorro e inversión, se juzga por sus causas y no solo por su signo. La cuenta de capital aporta ahorro adicional, y la cuenta financiera indica cuánto se financia el país y con qué estabilidad.

Solo la PII, leída con el enfoque de hoja de balance, muestra si esos flujos acumulan una posición vulnerable, por su signo, por la composición de los pasivos y por su efecto sobre la riqueza nacional. Con libre movilidad de capitales, ningún saldo mide por sí solo el desequilibrio. La

crisis financiera global mostró, además, un ajuste asimétrico, que recayó sobre la absorción de los deudores de la periferia del euro más que sobre los superávits.

Conclusiones

La balanza de pagos comparte con las cuentas nacionales su marco conceptual y se integra en ellas a través de la cuenta del resto del mundo. Por ello, la capacidad o necesidad de financiación de la nación es a la vez un saldo interno (ahorro más transferencias netas de capital recibidas, menos inversión) y un saldo exterior (cuenta corriente más cuenta de capital). Todas las cuentas dependen de la competitividad y de los rasgos estructurales de la economía. La cuenta corriente mide la brecha entre ahorro e inversión, la cuenta financiera muestra la composición de la financiación que la cubre y la posición de inversión internacional, leída con el enfoque de hoja de balance, revela la acumulación de vulnerabilidades. Los desequilibrios globales muestran, además, que el ajuste recae sobre todo en los deudores.

La balanza de pagos es la versión exterior de las identidades de la contabilidad nacional, en las que el saldo frente al resto del mundo refleja la brecha entre ahorro e inversión (tema A30). Las teorías del ajuste dan contenido de comportamiento a esas identidades, al explicar qué variables cierran la brecha (tema B12). A su vez, la lectura de la PII como hoja de balance permite anticipar las crisis en que se combinan los descalces de moneda y de vencimientos (tema B17) y justifica una red de seguridad financiera global que provea liquidez a los deudores (tema B22). Por último, la fragmentación de la producción en cadenas globales de valor hace que los saldos bilaterales midan mal la interdependencia comercial (tema B18).

La medición de las cuentas se complica en una economía globalizada, en la que las entidades con fines especiales canalizan inversión directa "fantasma" (apartado 1.4.3). Johnson y Noguera (2012) muestran, además, que el comercio medido en valor añadido altera de forma apreciable los saldos bilaterales. Sigue abierto el debate sobre la llamada *materia oscura* (*dark matter*). La PII de EE. UU. se volvió cada vez más deudora mientras sus rentas netas de inversión seguían siendo positivas. Hausmann y Sturzenegger (2005) lo atribuyen a activos no registrados, como la tecnología y los conocimientos técnicos que incorporan sus inversiones en el exterior, ligados al privilegio exorbitante del dólar. Curcuru, Dvorak y Warnock (2008) replican que, con datos coherentes, el diferencial de rentabilidad de la cartera de valores es casi nulo, de modo que la paradoja obedecería más a problemas de medición que a activos ocultos.

Anexo. Registro de operaciones en la balanza de pagos

Un país registra en un año las operaciones con el exterior que siguen, en millones de euros, sin otras transacciones. Cada una se anota por partida doble (apartado 1.3.2), en la cuenta corriente como crédito o débito y en la cuenta financiera como variación neta de activos (ΔAF) o de pasivos (ΔPF), con signo positivo cuando aumentan (apartado 2).

1. *Exportación de bienes*: una empresa residente vende a un no residente bienes por 100, que cobra en un depósito en un banco no residente. Se anota un crédito de 100 en bienes y un aumento de los activos de otra inversión de 100.

2. *Importación de bienes*: una empresa residente compra a un no residente bienes por 60 y los paga con ese depósito. Se anota un débito de 60 en bienes y una disminución de los activos de otra inversión de 60.

3. *Préstamo del exterior*: un banco no residente concede a una empresa residente un préstamo de 50, que abona en una cuenta de la empresa en el propio banco. Aumentan en 50 los pasivos de otra inversión (el préstamo) y en 50 los activos de otra inversión (el depósito). La operación es un intercambio de activos financieros por pasivos financieros, de modo que no altera el saldo de la cuenta financiera.

4. *Venta de divisas del banco central*: el banco central interviene en el mercado de divisas y vende a un banco residente divisas por 30, que el banco deposita en un banco no residente. Los activos de reserva disminuyen en 30 ($\Delta\text{RI} = -30$) y los activos de otra inversión aumentan en 30. La venta cambia la composición de los activos exteriores del país, pero no su saldo, y la caída de las reservas es la huella de la intervención.

5. *Inversión directa*: una empresa no residente compra por 70 el 60 % de una empresa residente y paga en un depósito del vendedor en un banco no residente. A la vez, una empresa residente invierte 40 en una filial en el extranjero, con cargo a sus depósitos en el exterior. Los pasivos de inversión directa aumentan en 70 y los activos, en 40, mientras que los activos de otra inversión aumentan en $70 - 40 = 30$. La inversión directa neta, $\Delta\text{AF}_{\text{ID}} - \Delta\text{PF}_{\text{ID}} = 40 - 70 = -30$, es una entrada neta de financiación del exterior.

Tabla B11.7: Registro de las operaciones en la balanza de pagos, en millones de euros.

Operación	Cuenta corriente	ΔAF	ΔPF
1. Exportación	Bienes: crédito 100	Otra inversión: 100	
2. Importación	Bienes: débito 60	Otra inversión: -60	
3. Préstamo del exterior		Otra inversión: 50	Otra inversión: 50
4. Venta de divisas		Reservas: -30; otra inversión: 30	
5. Inversión directa		Inversión directa: 40; otra inversión: 30	Inversión directa: 70
Total	$\text{SB}_{\text{CC}} = 40$	$\Delta\text{AF} = 160$	$\Delta\text{PF} = 120$

Los saldos son $\text{SB}_{\text{CC}} = 100 - 60 = 40$, $\text{SB}_{\text{CK}} = 0$ y $\text{SB}_{\text{CF}} = \Delta\text{AF} - \Delta\text{PF} = 160 - 120 = 40$ (tabla B11.7). Sin errores ni omisiones, se cumple la identidad $\text{SB}_{\text{CC}} + \text{SB}_{\text{CK}} = \text{SB}_{\text{CF}}$, y el país presta 40 al resto del mundo, su capacidad de financiación. Las operaciones 3 a 5 no alteran ese saldo, porque son puramente financieras, pero sí su composición, con una entrada neta de inversión directa de 30 y una pérdida de reservas de 30.

B12. Ajuste de la balanza de pagos

Mecanismos de ajuste de la balanza de pagos. Especial referencia al enfoque intertemporal de balanza de pagos. Análisis de sostenibilidad del déficit y de la deuda exterior.

Introducción

Introducción común. Los temas B11 y B12 comparten una introducción, que se desarrolla al comienzo del tema **B11**. Presenta las tres cuentas internacionales de una economía (balanza de pagos, posición de inversión internacional y cuenta de otras variaciones), elaboradas según el MBP6, y explica su interés, que nace de los desequilibrios externos que reflejan y que pueden desembocar en crisis cuando las reservas se agotan o la deuda externa se revela insostenible.

De esos desequilibrios nace una amplia literatura sobre el **ajuste de la balanza de pagos**, es decir, sobre los mecanismos que devuelven a una economía a una posición exterior sostenible. Cada enfoque sitúa el origen del desequilibrio en un lugar distinto y propone, por ello, una vía de corrección distinta.

Los enfoques de la cuenta corriente lo buscan en los precios relativos o en la renta. Por la vía de los precios, una devaluación mejora la balanza comercial si se cumple la **condición de Marshall-Lerner**, que exige que las exportaciones y las importaciones sean suficientemente sensibles a los precios, aunque a corto plazo puede empeorarla antes de mejorarla, como describe la **curva J**. Por la vía de la renta, el **enfoque de la absorción** de Alexander (1952) recuerda que el saldo exterior es la diferencia entre la producción y el gasto interior, de modo que solo mejora si el gasto cae en relación con la renta. Por su parte, el **enfoque monetario** (Johnson, 1958, 1972) ve en el desequilibrio la contrapartida de un desequilibrio del mercado de dinero, y el **enfoque intertemporal** trata la cuenta corriente como el resultado de decisiones de ahorro e inversión, lo que traslada la pregunta a la sostenibilidad del déficit.

En la periferia del euro, España llegó a registrar en 2007 un déficit por cuenta corriente cercano al 10 % del PIB y, sin tipo de cambio propio, lo corrigió mediante la contracción de la demanda interna y la moderación de los costes. En 2013 obtuvo capacidad de financiación frente al exterior por primera vez desde 1997 (Banco de España, 2014).

Existe consenso en que un tipo de cambio flexible no garantiza el ajuste externo, por su volatilidad, sus efectos inflacionistas, la incertidumbre sobre los precios extranjeros y sus posibles desalineamientos respecto de los fundamentos. Por ello, el objetivo de este tema es el estudio de las teorías sobre el ajuste de la balanza de pagos en un régimen cambiario fijo, que se organizan en tres bloques:

- *Ajuste de la cuenta corriente*: vía tipo de cambio real, en línea con los autores clásicos y neoclásicos, y vía renta, en línea con la teoría keynesiana (sección 1).

- *Enfoque activos*: las teorías basadas en la cuenta financiera, desarrolladas en los años sesenta y setenta (sección 2).

- *Enfoque intertemporal*: el modelo intertemporal de la balanza de pagos y el análisis de la sostenibilidad del déficit exterior (sección 3).

1. Teorías sobre el ajuste de la cuenta corriente

Los fundamentos de las teorías modernas se encuentran en las aportaciones de los autores clásicos, neoclásicos y keynesianos.

1.1. Enfoques de tipo de cambio real

En esta categoría se encuadran dos aportaciones de los autores clásicos y neoclásicos, el mecanismo flujo-especie de Hume y el enfoque de elasticidades de Marshall y Lerner. En ambos, el saldo exterior se corrige a través del tipo de cambio real.

1.1.1. Enfoque de flujo-especie (Hume, 1752)

Origen: el *mecanismo flujo-especie* se debe a la obra del filósofo y economista británico David Hume en *Of the Balance of Trade* (1752).

Idea: la balanza comercial se ajusta automáticamente en respuesta a *shocks* sin necesidad de que el tipo de cambio nominal varíe.

Supuestos:

- *Entorno*: mundo neoclásico sin imperfecciones en el que se cumple la teoría cuantitativa del dinero[1].

- *Sector exterior*: la balanza de pagos se reduce a la balanza comercial y los desequilibrios se saldan mediante movimientos de metal (especie), sin movimientos autónomos de capital. El sistema de tipos de cambio se considera fijo con convertibilidad estricta[2] (la oferta monetaria tiene que estar respaldada totalmente por reservas, p. ej. oro). La convertibilidad estricta se garantiza a través de intervenciones no esterilizadas, es decir, se admite que la oferta monetaria varíe con las variaciones de reservas internacionales.

Razonamiento: el mecanismo flujo-especie opera en cuatro etapas:

1. *Déficit comercial y drenaje de reservas*: si un país se enfrenta a un déficit de balanza comercial persistente, las reservas internacionales (p. ej. oro) tienden a disminuir, puesto que salen más de las que entran.

2. *Caída de la oferta monetaria*: como se quiere mantener el tipo de cambio nominal, la oferta monetaria del país cae al caer las reservas.

3. *Deflación*: por la teoría cuantitativa del dinero, la caída de la oferta monetaria provoca una caída proporcional en los precios nacionales.

[1] Los supuestos son los habituales del pensamiento clásico. Hay competencia perfecta y plena flexibilidad de precios y salarios en los mercados de bienes y de trabajo, y la igualdad entre ahorro e inversión asegura, junto con lo anterior, el pleno empleo. La velocidad de circulación es estable, de modo que se cumple la teoría cuantitativa del dinero.

[2] La convertibilidad en sentido estricto implica que debe existir un compromiso ilimitado e irrevocable por parte de la autoridad monetaria de atender las desviaciones de reservas, por pequeñas que sean. Esto implica que la autoridad monetaria no debe crear dinero sin respaldo en reservas exteriores y el sistema bancario privado tampoco, lo cual se podría garantizar mediante la prohibición de sistemas de reserva fraccionaria.

4. *Ganancia de competitividad-precio exterior*: dado que el tipo de cambio nominal permanece fijo, esta caída del nivel general de precios provoca una depreciación del tipo de cambio real que hará al país más competitivo. Por ello, se producirá un aumento de sus exportaciones netas y se corregirá el déficit.

Corolario I: si un país presenta superávit comercial, se dará la situación contraria. Por lo tanto, la balanza de pagos se ajusta de modo automático a través de ajustes reales del tipo de cambio ante cualquier desequilibrio externo.

Corolario II: la única variable (nominal) de ajuste es, pues, el nivel general de precios. No se necesita el tipo de cambio nominal para equilibrar la balanza de pagos.

Conclusiones:

▪ *Aspectos positivos*: la teoría del mecanismo flujo-especie fue fundamental como base teórica para la defensa del sistema de tipo de cambio fijo del patrón oro. Este sistema, nacido en Inglaterra a comienzos del siglo XIX, se extendió internacionalmente a partir de la década de 1870.

▪ *Limitaciones*: la teoría presenta dos:

- *Asimetría del ajuste*: un país deficitario debe restringir su base monetaria para mantener su oferta monetaria respaldada, pero un país superavitario no tiene por qué expandir su base monetaria, por lo que podría acumular reservas de forma indefinida.

- *Evidencia histórica*: durante la Gran Depresión (1929-1933), los problemas exteriores de los países ligados al patrón oro ponen en duda que el ajuste sea automático. Gana terreno entonces la idea de que corregir el saldo exterior puede requerir decisiones deliberadas de política económica, en particular devaluaciones. La oleada de devaluaciones posterior no mejoró, sin embargo, las balanzas de pagos en su conjunto, porque las devaluaciones de unos países tendían a anular las de otros. Con todo, quienes devaluaron o salieron antes del patrón oro también se recuperaron antes (Eichengreen y Sachs, 1985).

1.1.2. Enfoque de elasticidades (Marshall, 1923; Lerner, 1944)

Origen: a raíz de estas limitaciones surgen las aportaciones de Abba Lerner (1944), inspiradas en la obra de Alfred Marshall (1923).

Idea: los dos autores estudian en qué condiciones una devaluación mejora la balanza de pagos. La respuesta de la balanza comercial depende de cuán elásticas sean exportaciones e importaciones.

Ambos autores identifican dos efectos contrapuestos de la devaluación, que mejora la balanza comercial cuando el primero domina al segundo:

▪ *Efecto volumen* (+): aumenta la cantidad de las exportaciones nacionales y disminuye la cantidad de importaciones.

▪ *Efecto precio o valor* (−): las importaciones se encarecen en moneda nacional, de modo que, para un mismo volumen importado, aumenta el gasto en importaciones.

Supuestos:

▪ *Entorno*:

- *Renta extranjera dada*: economía abierta cuya evolución no influye en la renta extranjera, que se mantiene constante (no hay efectos de repercusión).

- *Precios fijos*: el supuesto se aplica tanto al nivel general de precios nacional como al extranjero. Las fluctuaciones del tipo de cambio nominal se trasladan así al tipo de cambio real y, con él, a la competitividad-precio.

▪ *Sector exterior*:

- *Balanza comercial*: la balanza de pagos solo recoge intercambios comerciales[3] y está inicialmente en equilibrio.

- *Demanda de exportaciones*: no es infinitamente elástica ($\partial P_X^*/\partial e < 0$, donde P_X^* es el precio de las exportaciones expresado en divisa).

- *Ofertas infinitamente elásticas*: lo son la oferta de exportaciones y la de importaciones, respecto a su precio. Esto implica que el precio de las exportaciones (en moneda nacional) y el precio de las importaciones (en divisa) son constantes: $\partial P_X/\partial e = 0$, $\partial P_M^*/\partial e = 0$. Aquí, P_X y P_M^* son índices de precios de los bienes exportados e importados, no niveles generales de precios.

- *Variaciones cambiarias*: las del tipo de cambio real son infinitesimales.

Razonamiento: la condición se obtiene en cinco pasos:

1. *Equilibrio inicial*: se parte de una balanza comercial en equilibrio, $\text{BC} = X - eM^* = 0$, donde X son los ingresos por exportaciones en moneda nacional, e es el tipo de cambio real directo y M^* es el gasto en importaciones en divisa.

2. *Diferenciación*: el efecto de una devaluación se obtiene derivando el saldo respecto al tipo de cambio real directo:

$$\frac{\partial \text{BC}}{\partial e} = \frac{\partial X}{\partial e} - M^* - e\frac{\partial M^*}{\partial e}$$

3. *Normalización*: se divide por el gasto en importaciones:

$$\frac{1}{M^*}\frac{\partial \text{BC}}{\partial e} = \frac{1}{M^*}\frac{\partial X}{\partial e} - 1 - \frac{e}{M^*}\frac{\partial M^*}{\partial e}$$

4. *Elasticidades*: en el equilibrio se cumple que $M^* = X/e$. Sustituyendo y expresando en términos de las elasticidades (definidas en valor absoluto):

$$\frac{1}{M^*}\frac{\partial \text{BC}}{\partial e} = \frac{e}{X}\frac{\partial X}{\partial e} - 1 - \frac{e}{M^*}\frac{\partial M^*}{\partial e} = \varepsilon_{Xe} + \varepsilon_{Me} - 1$$

5. *Condición de Marshall-Lerner*: de la expresión anterior se deduce cuándo $\partial \text{BC}/\partial e > 0$, resultado conocido como la *condición de Marshall-Lerner*:

$$\frac{\partial \text{BC}}{\partial e} > 0 \leftrightarrow \varepsilon_{Xe} + \varepsilon_{Me} > 1$$

[3] Esto implica que se prescinde de las rentas, las transferencias y los flujos de capital.

Es decir, una devaluación o una depreciación mejora el saldo comercial siempre que la suma de las elasticidades-tipo de cambio real de las exportaciones y las importaciones sea superior a la unidad. Si ambas demandas son suficientemente sensibles al tipo de cambio, el efecto volumen domina al efecto precio.

La sensibilidad de la balanza comercial a ambos efectos depende de la magnitud de esas elasticidades, que a su vez depende, entre otros factores, de los siguientes:

- *Estructura competitiva de mercados internacionales*: cuanta más competencia, mayor elasticidad.
- *Grado de diferenciación*: cuanto menor diferenciación, mayor elasticidad.
- *Grado de sustituibilidad*: cuanto mayor sustituibilidad, mayor elasticidad.

Corolario (política económica): si no se cumple la condición de Marshall-Lerner, el mercado de divisas puede ser inestable (anexo) y una devaluación empeoraría el saldo comercial, lo que justificaría la intervención en los mercados de divisas para garantizar la estabilidad del sistema.

Conclusiones:

- *Aspectos positivos*: ofrece una mayor microfundamentación del vínculo entre la balanza comercial y el tipo de cambio real.
- *Limitaciones*: la experiencia de la Gran Depresión puso en duda su validez, pues las devaluaciones aplicadas resultaron ineficaces, lo que llevó a cuestionar que las elasticidades fuesen lo suficientemente altas como para cumplir la condición. De hecho, los primeros estudios estadísticos, durante las décadas de 1930 y 1940, mostraban elasticidades entre 0 y 0,5. Esto provocó lo que Fritz Machlup llamó *pesimismo de las elasticidades*, la creencia en la ineficacia de las devaluaciones.

1.2. Extensiones al análisis de elasticidades

Surgen además una serie de aportaciones, algunas anteriores a Lerner, que tratan de explicar estas limitaciones y ampliar el enfoque, en concreto la generalización de Joan Robinson (1937) y los trabajos que profundizan en los efectos de una devaluación.

1.2.1. Contribución de Joan Robinson (1937)

Origen: Joan Robinson (1937) generaliza el análisis. Para ello levanta dos supuestos del enfoque de elasticidades, la elasticidad infinita de las curvas de oferta de exportaciones e importaciones y el equilibrio inicial de la balanza.

Idea: el efecto de una devaluación no depende solo de las dos elasticidades de demanda, sino de cuatro elasticidades, las de demanda y las de oferta de exportaciones e importaciones. La condición resultante es la de Bickerdike-Robinson-Metzler, de la que Marshall-Lerner es el caso particular con ofertas infinitamente elásticas y comercio inicialmente equilibrado.

Razonamiento: con ofertas de elasticidad finita, parte de la devaluación se compensa con cambios de precios (suben los de exportación en moneda nacional y bajan los de importación en divisa), lo que amortigua la ganancia de competitividad. Si, además, se parte de un déficit inicial, aparece un efecto valoración, por el que el resultado difiere según la moneda en que se mida, de modo que el saldo puede mejorar en divisa y empeorar en moneda nacional.

Conclusiones:

- *Aspectos positivos*: gana realismo y generalidad, y muestra que la condición de Marshall-Lerner es un caso límite.

- *Limitaciones*: el precio de la generalidad es la ambigüedad, pues la condición resultante apenas ofrece guía operativa, y el enfoque sigue siendo de equilibrio parcial.

1.2.2. Profundización en los efectos de una devaluación

Efecto J (Magee, 1973): desde una perspectiva dinámica y empírica, explica que el efecto precio puede manifestarse de forma inmediata, mientras que el efecto volumen presenta retardos[4]. Al principio, los volúmenes importados y exportados, fijados en contratos multiperiodo, no varían, pero el precio en moneda nacional de las importaciones facturadas en moneda extranjera aumenta. Por ello, la mejora de la balanza comercial no solo depende de las elasticidades, sino también del horizonte temporal considerado.

Si la condición de Marshall-Lerner no se cumple a corto plazo pero sí a medio y largo plazo, el saldo de la balanza comercial empeora en el corto plazo para luego recuperarse y mejorar respecto al punto inicial[5]. La velocidad del ajuste dependerá de la capacidad para ajustarse a los nuevos precios. Sin embargo, la evidencia empírica es mixta. De hecho, la revisión de Bahmani, Harvey y Hegerty (2013) concluye que la condición de Marshall-Lerner no se cumple en cerca de la mitad de los casos estudiados, de modo que la recuperación que dibuja el efecto J no está garantizada.

Gráficamente, en el plano tiempo-balanza comercial (figura B12.1), la *curva J* recoge el saldo comercial antes y después de la devaluación en t_0. El efecto precio lo reduce de inmediato y el efecto volumen, con retardo, lo eleva después por encima del nivel inicial.

[4] Helen Junz y Rudolf Rhomberg (1973) distinguen cinco tipos de retardos. El de reconocimiento (*recognition lag*) es el tiempo que tardan compradores y vendedores en darse cuenta de que la situación competitiva ha cambiado, y es menor cuanto mayor sea la interconexión entre los mercados. El de decisión (*decision lag*) es el tiempo que tardan en darse nuevas órdenes de compra y venta, y depende de la flexibilidad de los agentes para tomar decisiones y reajustar sus contratos. El de entrega (*delivery lag*) va desde que se recibe la orden de compra hasta que el bien está listo para su entrega, y crece con el periodo de producción de los sectores exportadores (cuanto más larga la cadena de valor, mayor es). El de sustitución (*replacement lag*) es el tiempo que tardan los importadores extranjeros en agotar sus inventarios y amortizar sus equipos importados, y depende de la gestión de inventarios y de la vida económica de la maquinaria importada. Por último, el de producción (*production lag*) es el tiempo que tardan los exportadores en establecer nuevas líneas de negocio.

[5] Se distinguen tres fases. La primera va de la devaluación (t_0) al vencimiento de los contratos negociados antes de ella. En esta fase, el efecto de la devaluación no es necesariamente negativo (la curva J puede partir de un punto por encima o por debajo del eje), porque depende de la moneda en que estén denominados los contratos de exportación e importación. Cuanto mayor sea el porcentaje de contratos de exportación en moneda extranjera y menor el de contratos de importación en esa moneda, más probable es que el efecto inicial sea positivo. En la segunda fase se ajustan los precios internacionales de los contratos negociados después de la devaluación, y en ella se produce el *pass-through* limitado, de modo que el efecto, de nuevo ambiguo, depende del coeficiente de traslación. En la tercera se produce el ajuste en cantidades, sujetas a mayores retardos.

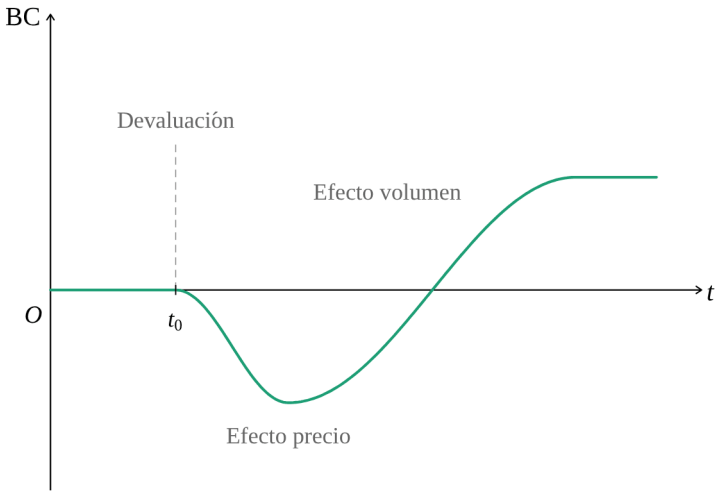

Figura B12.1: Curva J. Tras una devaluación la balanza comercial empeora a corto plazo antes de mejorar.

Ejemplo: Marshall-Lerner a corto y a largo plazo

Un país parte del equilibrio comercial, con exportaciones de 100 en moneda nacional e importaciones de 100 en divisa ($e = 1$), y devalúa un 10 %. Las elasticidades-tipo de cambio real de las exportaciones y las importaciones son de 0,2 y 0,3 en el primer año y de 0,8 y 0,7 a largo plazo.

- *Corto plazo*: $\varepsilon_{Xe} + \varepsilon_{Me} = 0{,}5 < 1$. Las exportaciones suben a 102 y las importaciones bajan a 97 en divisa, que cuestan ahora $1{,}1 \cdot 97 = 106{,}7$ en moneda nacional, así que BC $= -4{,}7$. Domina el efecto precio.
- *Largo plazo*: $\varepsilon_{Xe} + \varepsilon_{Me} = 1{,}5 > 1$. Las exportaciones suben a 108 y las importaciones bajan a 93, que cuestan 102,3, y BC $= 5{,}7$. Domina el efecto volumen.
- *Lectura*: la misma devaluación empeora primero el saldo y lo mejora después. La curva J aparece porque las elasticidades crecen con el tiempo, a medida que vencen los contratos y los compradores cambian de proveedor.

Efecto *pass-through* limitado (Krugman, 1987b): tras una devaluación, los exportadores tienden a resistirse a que disminuya el precio de sus productos expresado en moneda extranjera. El porcentaje de devaluación que se traslada a los precios de exportación en divisa (el coeficiente *pass-through*) no tiene entonces por qué ser del 100 %. Por ello, el efecto sobre la balanza comercial es ambiguo, ya que los exportadores amplían sus márgenes en moneda nacional, pero el aumento del volumen exportado es menor que con una traslación completa.

Este porcentaje dependerá de varios factores:

- *Poder de mercado*: el de los exportadores.
- *Predictibilidad*: la de las variaciones en el tipo de cambio (cuanto más predecible, más se trasladará a precios).

- *Magnitud*: la de la variación del tipo de cambio.

- *Persistencia*: la permanencia o transitoriedad del *shock* (cuanto menos dure, menos se trasladará a precios).

- *Apertura*: la de la economía.

Efecto inflacionista de la devaluación: el enfoque de equilibrio parcial considera los precios fijos, y así la devaluación nominal se traduce en una devaluación real. Sin embargo, una devaluación con frecuencia provoca inflación[6] y, por lo tanto, una depreciación nominal no se traslada plenamente al tipo de cambio real, que podría incluso llegar a apreciarse.

Conclusiones:

- *Aspectos positivos*: los tres efectos explican por qué la respuesta del saldo comercial a una devaluación es lenta e incompleta, y concilian el pesimismo de las elasticidades con los casos en que la condición de Marshall-Lerner se cumple a medio plazo.

- *Limitaciones*: en el plano metodológico, el enfoque de elasticidades es, en su conjunto, de equilibrio parcial. Deja fuera cómo repercute la mejora comercial en las demás variables de la economía, empezando por la renta, laguna que da origen a los enfoques siguientes.

1.3. Enfoques de la renta

El enfoque contrapuesto al enfoque elasticidades pone énfasis en la renta como mecanismo de ajuste, en línea con la teoría de ciclos. El multiplicador del comercio exterior se remonta a Roy Harrod (1933).

1.3.1. Enfoque del multiplicador keynesiano del comercio exterior (Machlup, 1943)

Origen: Fritz Machlup (1943) aplica el principio del multiplicador al comercio exterior, lo que incorpora el mercado de bienes y servicios al análisis del ajuste de la balanza de pagos.

Idea: la mejora de la balanza comercial depende no solo de la sensibilidad de las importaciones y de las exportaciones al tipo de cambio, sino también de las propensiones marginales a consumir y a importar de los individuos.

Supuestos: son muy similares a los del enfoque elasticidades (corto plazo con precios rígidos, balanza de pagos reducida a la balanza comercial…), con dos precisiones:

- *Metodología*: este enfoque puede formularse con el mercado de bienes del modelo IS-LM-BP (desarrollado con posterioridad, en los años sesenta), si bien adopta dos supuestos simplificadores para obtener la ecuación fundamental del modelo:

  - *Imposición*: los impuestos son de suma fija (T_0).

  - *Importaciones*: dependen exclusivamente de la renta nacional. Esto supone un alejamiento fundamental del modelo IS-LM-BP canónico, simplificación que permite aislar el efecto renta.

[6] Una devaluación genera inflación por tres canales. El primero es el aumento del precio nacional de las importaciones. Si son bienes finales, pesa tanto más cuanto mayor sea su peso en la cesta de consumo, y, si son bienes intermedios, eleva los costes de producción, que acaban trasladándose a los precios. El segundo es la menor competencia internacional, que da mayor poder de negociación a los productores nacionales y les permite ampliar sus márgenes. El tercero es el traslado de la subida de precios a reivindicaciones salariales más elevadas, que puede desencadenar una espiral inflacionista (efectos de segunda ronda).

▪ *Sector exterior*: se cumple la condición de Marshall-Lerner, de modo que una devaluación mejora el saldo comercial. Como en el modelo las importaciones no dependen del tipo de cambio, ello equivale a suponer $\partial X/\partial e > 0$.

Planteamiento: el modelo parte de la condición de equilibrio en el mercado de bienes:

$$Y = C_0 + c(Y - T_0) + I_0 - vr + G_0 + X(e) - mY, \qquad c \in (0,1)$$

donde C_0 es el componente autónomo de la función de consumo agregada, $c \in (0,1)$ es la propensión marginal a consumir de la economía, Y es la renta, T_0 los impuestos de suma fija, I_0 el componente autónomo de la función de inversión agregada, $v > 0$ la sensibilidad de la inversión agregada al tipo de interés, r el tipo de interés, G_0 el gasto público, $X(e)$ las exportaciones de la economía (que sí dependen del tipo de cambio e) y $m > 0$ la propensión marginal a importar de la economía.

Estática comparativa (*efecto renta-exportación*): reordenando, se obtiene que el efecto de una mejora de las exportaciones sobre la renta de la economía es:

$$Y = \frac{1}{1-c+m}(A_0 + X) \to \frac{\partial Y}{\partial X} = \frac{1}{1-c+m} > 0$$

donde $A_0 = C_0 - cT_0 + I_0 - vr + G_0$ recoge todos los componentes autónomos de las funciones agregadas. El *multiplicador del comercio exterior* es, típicamente, superior a la unidad, si bien menor que el multiplicador de una economía cerrada, pues la propensión a importar constituye una filtración adicional.

Efecto balanza comercial-exportación (efecto total): considerando exclusivamente el término correspondiente a la balanza comercial, $\text{BC} \equiv X - mY$, se tiene que el efecto de una mejora de las exportaciones sobre la balanza comercial de la economía es:

$$\text{BC} = X - mY \to \frac{\partial \text{BC}}{\partial X} = 1 - m\frac{\partial Y}{\partial X} = \frac{1-c}{1-c+m} = \frac{s}{s+m} \in (0,1)$$

donde $s = 1 - c \in (0,1)$ es la propensión marginal al ahorro de los hogares de la economía. Es decir, un aumento de las exportaciones (p. ej. inducido por una devaluación) mejora el saldo comercial, pero menos que uno a uno, ya que la renta adicional se reparte entre consumo, ahorro e importaciones, y la parte que se filtra hacia importaciones minora la mejora. El efecto será mayor cuanto mayor sea la propensión marginal a ahorrar, que reduce el multiplicador y, con él, las importaciones inducidas, o menor la propensión marginal a importar.

Conclusiones:

▪ *Aspectos positivos*:

- *Propensiones marginales*: amplía el enfoque elasticidades al considerar las propensiones marginales de los individuos.
- *Recursos ociosos*: en la medida en que los precios son rígidos, supone que existen recursos ociosos y que, por tanto, la economía puede atender la demanda externa mediante el aumento de su producción sin que aumenten los precios.

▪ *Limitaciones*: el auge económico vivido tras la Segunda Guerra Mundial, conocido como la "edad dorada del capitalismo" (1945-1973), invitaba a abandonar el supuesto de recursos ociosos, sobre el que descansa todo el resultado. Además, el modelo ignora los efectos monetarios y de precios de la devaluación, supone cumplida la condición de Marshall-Lerner en lugar de derivarla y omite las repercusiones sobre la renta del resto del mundo, relevantes para países grandes.

1.3.2. Enfoque absorción (Alexander, 1952)

Origen: en este contexto, Sidney Alexander (1952) desarrolla el enfoque absorción de la balanza de pagos.

Idea: la mejora de la balanza de pagos depende de la situación de empleo de recursos de la economía.

Supuestos: son similares a los anteriores, salvo en los precios, que se consideran rígidos si existen recursos ociosos y flexibles en el pleno empleo.

Planteamiento: el modelo parte de la ecuación básica de la demanda agregada: $Y = C + I + G + (X - M) \rightarrow Y = A + \text{BC}$[7], donde A es la *absorción interna* y refleja la demanda total de bienes finales de una economía, independientemente de su origen.

La absorción interna, a su vez, tiene un componente autónomo y un componente que depende de la renta: $A = A_0 + cY$, donde $c \in (0, 1)$ es un parámetro constante y exógeno que refleja la propensión marginal a consumir[8] al igual que antes.

Balanza comercial y absorción: reordenando, se obtiene el saldo de la balanza comercial en términos de la absorción interna:

$$\text{BC} = Y - A \rightarrow \text{BC} = Y - (A_0 + cY) \rightarrow \text{BC} = (1 - c)Y - A_0$$

Efecto balanza comercial-tipo de cambio: derivando con respecto al tipo de cambio real, se obtiene el efecto de su variación sobre la balanza comercial:

$$\frac{\partial \text{BC}}{\partial e} = (1 - c)\frac{\partial Y}{\partial e} - \frac{\partial A_0}{\partial e}$$

La expresión anterior muestra que una variación del tipo de cambio produce dos efectos:

▪ *Efecto renta* (+): recogido en el primer término del lado derecho de la ecuación. Refleja cómo las devaluaciones (o depreciaciones) permiten mejorar el saldo de la balanza comercial al aumentar las exportaciones y, con ellas, la renta, teniendo en cuenta que parte de ese aumento de renta se filtrará en forma de consumo.

▪ *Efecto absorción* (signo ambiguo): recogido en el segundo término del lado derecho (incluido el signo negativo). Refleja cómo una devaluación (o depreciación) reduce el saldo de la balanza comercial si eleva directamente la absorción autónoma (A_0) y lo mejora si la reduce, como ocurre en pleno empleo a través de los canales monetario y redistributivo.

[7] Aquí se abstrae de las rentas primarias y secundarias.

[8] Si se considerase que la demanda de inversión y las compras del Gobierno dependen de la renta, se hablaría de propensión marginal a absorber.

En consecuencia, el efecto total de una devaluación sobre la balanza comercial dependerá de cuál predomine, lo que varía con la situación de empleo de la economía:

Desempleo: si existen recursos ociosos, el efecto será positivo siempre que la propensión marginal a consumir sea inferior a la unidad (y tanto mayor cuanto más alejada esté de ella), porque el efecto renta será mayor que el efecto absorción[9]. La devaluación permite entonces que la producción nacional cubra la demanda interna y la externa sin que aumenten los precios y, por tanto, sin afectar a la absorción autónoma. Ello podría explicar por qué muchos países recurrieron a devaluaciones competitivas en los años treinta.

Pleno empleo: si no existe desempleo, el efecto renta será nulo[10], puesto que no quedan recursos disponibles para atender la demanda total de bienes nacionales, tanto interna como externa. Por tanto, el efecto total sobre la balanza comercial dependerá del efecto absorción. Alexander contempla dos canales por los que la devaluación reduce directamente la absorción y puede, así, mejorar el saldo comercial:

- *Canal monetario*: recoge dos efectos:
  - *Efecto Keynes*: para una oferta monetaria dada, la inflación reduce la oferta real de dinero, lo que eleva el tipo de interés y disminuye la absorción: $\uparrow e \rightarrow\uparrow P \rightarrow\downarrow (M^s/P) \rightarrow\uparrow i \rightarrow\downarrow C, \downarrow I \rightarrow\downarrow A$.
  - *Efecto Pigou o efecto saldos reales*: para una oferta monetaria dada, la inflación deteriora la riqueza real de los individuos. De esta forma, decidirán ahorrar más y consumir menos, lo cual reducirá la absorción: $\uparrow e \rightarrow\uparrow P \rightarrow\downarrow C \rightarrow\uparrow S \rightarrow\downarrow A$.
- *Canal redistributivo*: recoge otros dos:
  - *Rémora fiscal, progresividad en frío o progresión fría* (*fiscal drag*): si existe un sistema de impuestos progresivos y este no se actualiza, la inflación y los aumentos salariales nominales harán que una mayor proporción de gente se encuentre en los tramos altos del impuesto. Mientras el Estado no amplíe su gasto, el resultado es un trasvase de renta de los agentes privados a las arcas públicas: $\uparrow e \rightarrow\uparrow P \rightarrow\uparrow w \rightarrow\uparrow \tau w \rightarrow\downarrow C \rightarrow\downarrow A$.
  - *No indexación de los salarios*: cuando los salarios nominales suben menos que los precios, porque su indexación es imperfecta, la renta real pasa de los asalariados a los perceptores de beneficios. Como estos ahorran, por hipótesis, una fracción mayor de su renta, la absorción interna cae.

Desarrollo gráfico: el enfoque absorción se representa en el plano renta-precios (figura B12.2):

1. *Oferta agregada* (OA): recoge la producción que la economía ofrece a cada nivel de precios y presenta dos tramos. Es horizontal mientras existen recursos ociosos, con precios rígidos, y vertical en la renta de pleno empleo, Y_{PE}, donde los precios son flexibles y no quedan recursos para producir más.

[9] En la teoría de ciclos se contempla que la propensión marginal a consumir supere la unidad (p. ej. por endeudamiento). En ese caso, la balanza comercial no mejoraría con el aumento de la producción.

[10] No obstante, este podría ser positivo en dos casos. El primero se da si el país es grande y puede afectar a la relación real de intercambio (RRI). Al devaluar, cae su demanda de importaciones y, con ella, su precio, y si esta caída es mayor que la del precio de las exportaciones provocada por la devaluación, mejora la RRI, lo que eleva la renta real aunque no aumente la producción. El segundo se da si el tipo de cambio está desalineado (sobreapreciado), porque la devaluación corrige la distorsión.

2. *Demanda agregada* (DA): recoge las combinaciones de renta y precios con las que la producción iguala la demanda de bienes nacionales, $A+\mathrm{BC}$. Tiene pendiente negativa porque una subida de los precios reduce la absorción por los efectos Keynes y Pigou y, con el tipo de cambio nominal fijo, aprecia el tipo de cambio real y reduce las exportaciones netas. Su posición depende de la absorción autónoma y de las exportaciones.

3. *Equilibrio*: en el corte de ambas curvas. Con desempleo, la demanda agregada (DA_0) corta el tramo horizontal en el punto 0. En pleno empleo, la demanda agregada ($\mathrm{DA}_{\mathrm{PE},0}$) corta el tramo vertical en el punto 0_{PE}.

Estática comparativa: una devaluación ($\Delta e > 0$) eleva las exportaciones y desplaza la demanda agregada a la derecha. Su efecto depende del tramo de la oferta agregada:

- *Desempleo*: la demanda pasa de DA_0 a DA_1 y la economía, del punto 0 al 1. La producción y, por ende, la renta nacional aumentan sin que suban los precios.

- *Pleno empleo*: la demanda pasa de $\mathrm{DA}_{\mathrm{PE},0}$ a $\mathrm{DA}_{\mathrm{PE},1}$ y la economía, del punto 0_{PE} al 1_{PE}. La devaluación solo produce inflación, porque en el tramo vertical de la oferta la producción no puede aumentar.

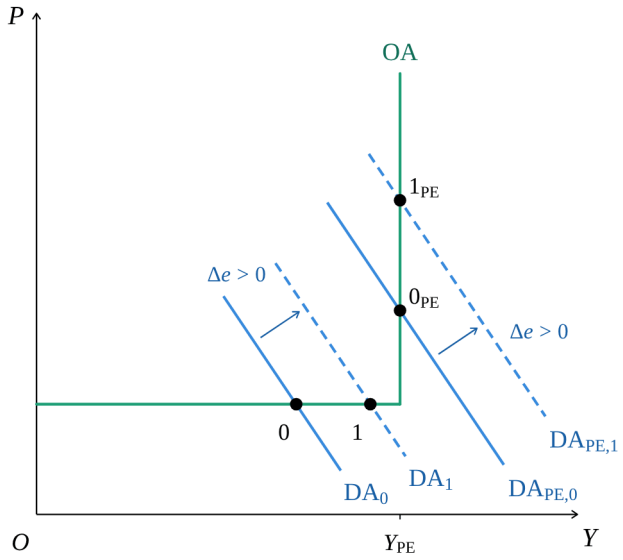

Figura B12.2: Enfoque absorción (Alexander, 1952). La devaluación mejora la balanza solo si la producción crece más que la absorción.

Extensiones: en términos de política económica, una devaluación podría ser insuficiente e incluso ineficaz para corregir la balanza comercial, por lo que debería combinarse con programas de contención o reducción del gasto interno. La combinación óptima depende de las características del país (p. ej. la propensión a absorber o la rigidez del mercado de trabajo), y el diagrama de Swan la formaliza.

Conclusiones:

- *Aspectos positivos*: integra la situación de empleo de la economía en el análisis y desplaza el foco del equilibrio parcial al equilibrio de la demanda agregada. De este modo muestra que una

devaluación solo funciona si libera recursos reales (si la producción puede crecer o la absorción caer).

▪ *Limitaciones*: los canales que reducen la absorción en pleno empleo (Keynes, Pigou, redistributivos) son empíricamente débiles y lentos, y la cuenta financiera sigue ausente del análisis. Además, el enfoque tiende a presentarse como sustituto del de elasticidades cuando, en realidad, ambos son complementarios.

1.3.3. Diagrama de Swan (Swan, 1955; Salter, 1959)

Origen: Trevor Swan expone el diagrama en su trabajo escrito en 1955 y publicado en 1963. Por su parte, Wilfred Salter (1959) desarrolla el mismo análisis, con bienes comerciables y no comerciables.

Idea: en un régimen de cambio fijo pero ajustable, la política económica persigue dos objetivos, el equilibrio interno (pleno empleo sin inflación) y el equilibrio externo (un saldo por cuenta corriente objetivo). Para ello necesita dos instrumentos, la absorción interna y el tipo de cambio real, ninguno de los cuales basta por separado.

Planteamiento: el tipo de cambio real es $e = EP^*/P$, de modo que un aumento de e es una depreciación. Sin rentas ni transferencias, el saldo por cuenta corriente coincide con la balanza comercial, que mejora con e (se supone cumplida la condición de Marshall-Lerner) y empeora con la absorción, A, por las importaciones. Cada objetivo define una curva.

Desarrollo gráfico: los equilibrios interno y externo se representan en el plano absorción-tipo de cambio real (figura B12.3):

1. *Equilibrio interno* (EI): combinaciones con las que la demanda de producción nacional, A+BC, iguala la producción de pleno empleo, $\bar{Y}$. Su pendiente es negativa, porque una absorción mayor solo es compatible con el pleno empleo sin inflación si una apreciación desvía demanda hacia bienes extranjeros. Por encima de la curva hay inflación y, por debajo, desempleo.

2. *Equilibrio externo* (EE): combinaciones con las que la balanza comercial alcanza su objetivo, $\mathrm{BC} = \mathrm{BC}^T$. Su pendiente es positiva, porque una absorción mayor eleva las importaciones y exige una depreciación que restablezca el saldo. Por encima de la curva hay superávit y, por debajo, déficit.

3. *Equilibrio*: las dos curvas se cortan en el punto 0, el único que cumple ambos objetivos, y dividen el plano en cuatro zonas de desequilibrio. Esas zonas son la I, de inflación y déficit, la II, de inflación y superávit, la III, de desempleo y superávit, y la IV, de desempleo y déficit.

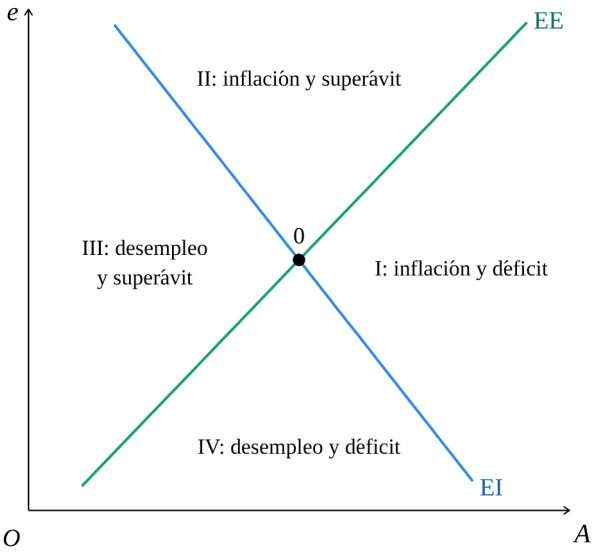

Figura B12.3: Diagrama de Swan. Las curvas de equilibrio interno EI y de equilibrio externo EE se cortan en el punto 0 y delimitan cuatro zonas de desequilibrio, que combinan inflación o desempleo con déficit o superávit.

Estática comparativa: con un solo instrumento, el ajuste no alcanza en general los dos objetivos.

- *Solo absorción*: la economía se desplaza en horizontal. En las zonas I y III la absorción corrige los dos desequilibrios a la vez (en I, reducir el gasto modera la inflación y el déficit). En las zonas II y IV, en cambio, cada objetivo pide un movimiento opuesto (en IV, el desempleo exige más gasto y el déficit, menos).

- *Solo tipo de cambio*: la economía se desplaza en vertical. En las zonas II y IV el tipo de cambio corrige ambos desequilibrios, pues en IV una devaluación reduce el desempleo y el déficit y en II una revaluación modera la inflación y el superávit, y en las zonas I y III se enfrenta al mismo dilema.

Incluso en las zonas favorables, el instrumento único alcanza una curva antes que la otra y solo lleva al punto 0 si la economía parte de la misma altura que el punto 0 (con la absorción como instrumento) o de la misma abscisa (con el tipo de cambio). El resultado ilustra la *regla de Tinbergen* (1952), según la cual alcanzar dos objetivos independientes exige al menos dos instrumentos independientes.

Extensiones: si cada instrumento se encomienda a un objetivo, el ajuste converge cuando cada uno se asigna al objetivo sobre el que tiene mayor efecto relativo (principio de clasificación efectiva de los mercados, Mundell, 1962). Por ejemplo, con una propensión a importar reducida, la absorción se asigna al equilibrio interno y el tipo de cambio, al externo. La misma lógica, aplicada a las políticas monetaria y fiscal, se trata en el tema A40, apartado 1.3, y los dos objetivos como criterio de elección del régimen cambiario, en el tema B15, apartado 1.1.

Conclusiones:

- *Aspectos positivos*: integra los enfoques de elasticidades y de absorción al distinguir las políticas que modifican el nivel del gasto de las que lo desvían entre bienes nacionales y extranjeros, y muestra con sencillez por qué el ajuste exterior exige combinar instrumentos.

- *Limitaciones*: es un análisis estático que ignora la cuenta financiera y trata el tipo de cambio real como instrumento, cuando la autoridad solo controla el nominal, cuyo efecto real se erosiona si la devaluación eleva precios y salarios.

1.4. Conclusiones

Hasta los años setenta, el consenso atribuía, por lo general, a la cuenta corriente la causa de los desequilibrios exteriores. Al fin y al cabo, todos los enfoques de esta sección surgieron cuando los flujos de capitales apenas empezaban a expandirse.

Desde esa década, sin embargo, la apertura gradual de las cuentas de capital ha llevado a buscar en los flujos financieros la causa última de los desequilibrios. De ahí la objeción principal a los enfoques anteriores, que desatienden los flujos financieros que hacen de contrapartida de la balanza comercial y no contemplan, entre otros, los movimientos autónomos de capital.

2. Teorías basadas en el desequilibrio en la cuenta financiera: enfoque activos

Así nace el **enfoque de la balanza de pagos cabeza abajo** (o enfoque activos), cuyo nombre alude a que lee la balanza de pagos al revés. Los desequilibrios comerciales pasan a verse como efecto de los financieros y no como su causa, con lo que la causalidad se invierte.

Intuitivamente, con libre circulación de capitales, la oferta y la demanda de divisas responden no solo a su uso como medio de cambio internacional, sino también a su uso como depósito de valor (dinero u otros activos).

El caso de Tailandia lo ilustra. En 1995-1996, antes de que estallara la crisis asiática, su déficit por cuenta corriente rondaba el 8 % del PIB y respondía a la llegada masiva de *hot money*, es decir, de inversión especulativa a corto plazo.

Estas ideas se ilustran con dos modelos. El **modelo monetario** se centra en el dinero como activo financiero y el **modelo de cartera** considera un espectro más amplio de activos financieros.

2.1. Modelo monetario (Johnson, 1958, 1972)

Origen: el economista canadiense Harry Johnson formuló el enfoque monetario de la balanza de pagos en 1958 y lo sistematizó en 1972, revitalizando las ideas que David Hume había anticipado con el mecanismo flujo-especie.

Idea: los desequilibrios exteriores nacen en el mercado monetario, y los propios flujos monetarios devuelven la balanza de pagos al equilibrio de forma automática.

2.1.1. Supuestos

Entorno: país pequeño y abierto, régimen de cambio fijo, libre circulación de capitales y mundo neoclásico (comercio sin restricciones y pleno empleo), por lo que el nivel de renta viene determinado por factores reales exógenos.

Además, se cumple la ley del precio único entre países, tanto para activos financieros como para mercancías: $i = i^*$, $P = E \cdot P^*$. Existe plena sustituibilidad de bienes y activos nacionales y extranjeros.

Agentes: el multiplicador monetario es unitario, por lo que base monetaria y oferta monetaria se usan indistintamente.

2.1.2. Modelo

Planteamiento: el modelo consta de dos ecuaciones del mercado de dinero:

- *Demanda de dinero*: es proporcional a la renta nominal, $M^d = kPY$, con $P = E \cdot P^*$ por la ley del precio único. Como la renta la fijan factores reales y el tipo de interés es el internacional, ante un cambio del crédito interno la demanda no varía, $M^d = M_0^d$.

▪ *Oferta de dinero*: es endógena, pues, aunque la autoridad monetaria controla el crédito interno, las reservas se ajustan a través de la balanza de pagos. Se recoge en una identidad basada en el balance del banco central, cuyo activo está formado por el crédito interno y las reservas internacionales (activos netos frente al exterior)[11]: $M^s = \text{CI} + R$.

Equilibrio: el mercado de dinero se equilibra cuando $M^d = M^s \rightarrow M_0^d = M^s$. De esta condición y de la identidad de la oferta de dinero se derivan el mecanismo de ajuste del modelo y sus principales resultados:

$$M^d = \text{CI} + R \rightarrow \Delta R = \Delta M^d - \Delta \text{CI}$$

Con la demanda de dinero dada, $\Delta R = -\Delta \text{CI}$, y los cambios en las reservas internacionales se deben a cambios en el crédito interno (*shocks* monetarios). Así, el nivel de reservas ajusta la balanza de pagos ante los *shocks* que desequilibran el mercado monetario.

Estática comparativa (política monetaria expansiva): un aumento del crédito interno actúa en dos pasos:

1. *Mercado monetario interno*: partiendo del equilibrio, un aumento del crédito interno crea un exceso de oferta de saldos reales, pues la demanda de dinero y los precios no varían: $\uparrow M^s \leftrightarrow \uparrow \text{CI} \rightarrow M^s > M^d$.

2. *Mercado de bienes y sector exterior*: los agentes tratan de eliminar ese exceso adquiriendo bienes y activos. Con pleno empleo y precios fijos, el exceso de demanda solo puede cubrirse con bienes y activos extranjeros[12]. Esa compra provoca una salida de reservas de la misma cuantía que el aumento del crédito interno y un déficit de balanza de pagos de igual magnitud (por cuenta corriente y cuenta financiera, excluidos los activos de reserva). De ahí que se hable de enfoque de la BP "cabeza abajo": $\text{DA} > \text{OA} \rightarrow \uparrow \text{DA}^* \rightarrow \downarrow R$. Por tanto, los desequilibrios externos son un fenómeno monetario. Si la autoridad esteriliza la pérdida de reservas con nuevo crédito interno, el drenaje se prolonga.

Estática comparativa (crecimiento de la renta y devaluación): cuando lo que cambia es la demanda de dinero, el modelo da dos predicciones que lo distinguen de los enfoques de la sección 1 (Frenkel y Johnson, 1976):

▪ *Crecimiento de la renta*: eleva la demanda de saldos, $\Delta M^d = kP\,\Delta Y$, que, con el crédito interno dado, solo puede atenderse con una entrada de reservas. El crecimiento produce así superávit, al contrario que en el multiplicador del comercio exterior (apartado 1.3.1), en el que más renta significa más importaciones.

▪ *Devaluación*: con $P = E \cdot P^*$, eleva los precios internos y, con ellos, la demanda nominal de dinero, $\Delta M^d = kYP^*\,\Delta E$. Para recomponer sus saldos reales, los agentes gastan menos de lo que ingresan y la balanza registra un superávit con entrada de reservas. Sin embargo, el superávit es transitorio y se agota cuando la oferta monetaria ha crecido en la misma proporción que los precios, de modo que la devaluación no mejora la balanza de forma permanente.

Extensiones: las implicaciones de política económica son dos:

[11] En un régimen de cambio fijo, las variaciones de R son la contrapartida del saldo neto de todas las transacciones con el exterior, es decir, del saldo de la balanza de pagos: $\Delta R = X - M + F$, donde F son las entradas netas de capital.

[12] Con todo, la carga del ajuste tiende a recaer sobre los movimientos de capital, por la mayor facilidad de realizar transacciones internacionales de capital.

- *Política monetaria activa*: no tiene efectos sobre el equilibrio interno, puesto que la economía se sitúa en el pleno empleo, y solo afecta transitoriamente al equilibrio externo (déficit). Deja, no obstante, una pérdida permanente de reservas, ya que los cambios en el crédito interno se contrarrestan con los cambios en las reservas y la oferta monetaria vuelve a su nivel inicial[13]. Por lo tanto, una política monetaria activa no es recomendable, porque el ajuste es automático y una política expansiva puede agotar las reservas.

- *Programación del crédito interno*: si la autoridad monetaria conoce la demanda de dinero y las reservas de la economía, sí puede fijar un crecimiento del crédito interno compatible con el equilibrio exterior. De ahí que muchos programas de estabilización del FMI para los países en desarrollo (PED) se basen en límites al crecimiento del crédito interno para evitar salidas de reservas, según el modelo de Polak (1957) que se estudia en el tema B22, apartado 3.2.

2.1.3. Conclusiones

Aspectos positivos: ofrece una explicación teórica a situaciones en las que los desequilibrios externos se generan por desequilibrios en el mercado monetario.

Limitaciones: se dirigen, principalmente, a la adopción de la ley del precio único:

- *Bienes*: existen bienes no comerciables en los que no tiene por qué cumplirse la ley del precio único.

- *Activos*: el grado de sustituibilidad entre activos nacionales y extranjeros no es perfecto, relajación que da lugar al modelo de cartera.

2.2. Modelo de cartera (McKinnon, 1969)

Origen: Ronald McKinnon (1969) desarrolla un modelo basado en la aplicación de la teoría de la selección de carteras de Harry Markowitz al estudio de la balanza de pagos.

Idea: los desequilibrios en la balanza de pagos siguen siendo causados por los desequilibrios en el mercado monetario, si bien los desequilibrios monetarios también se verán influidos por las decisiones de cartera de los agentes económicos.

2.2.1. Supuestos

Entorno: los supuestos son similares (país pequeño y abierto, precios fijos, libertad de movimiento de capital), con la salvedad de que existe *sustituibilidad imperfecta* entre los activos nacionales y extranjeros[14]. Esto se debe a factores tales como información imperfecta, riesgo cambiario o riesgo político, entre otros. Se definen tres activos financieros:

- *Dinero* (M): activo que no ofrece rendimiento alguno.

[13] En este contexto aparece el trilema de Mundell-Fleming (trinidad imposible), según el cual no se puede garantizar simultáneamente tipo de cambio fijo, libre circulación de capitales y una política monetaria autónoma (tema A32, apartado 1.3.4).

[14] La sustituibilidad imperfecta se observa en la realidad incluso en países con libre circulación de capitales, como muestra el sesgo doméstico (*home bias*) de los bancos de la zona del euro respecto a los bonos soberanos.

▪ *Bonos nacionales* (B): activos que otorgan un rendimiento[15], r, no comerciables internacionalmente ni canjeables por bonos extranjeros. De ello se deriva, como consecuencia fundamental para el modelo, que el flujo de capital financiero entre el país y el resto del mundo se determina exclusivamente por el flujo de bonos extranjeros.

▪ *Bonos extranjeros* (B^*): estos también otorgan un rendimiento, r^*, y, a diferencia de los bonos nacionales, sí son comerciables internacionalmente.

2.2.2. Modelo

Planteamiento: la demanda de activos financieros se reparte entre el mercado de dinero y el de bonos, y su oferta determina la riqueza:

▪ *Demanda de dinero*: se demanda una proporción de la riqueza, tal que $M^d = \omega_M W$, con $\omega_M = f(r^-, r^{*-})$.

▪ *Demanda de bonos*: igualmente, se demandará una proporción de la riqueza en bonos nacionales, $B^d = \omega_B W$, con $\omega_B = f(r^+, r^{*-})$, y en bonos extranjeros, $B^{d*} = \omega_{B^*} W$, con $\omega_{B^*} = f(r^-, r^{*+})$, es decir, cada proporción aumenta con el rendimiento propio y disminuye con el del activo alternativo.

▪ *Oferta de activos o riqueza*: la riqueza de la economía vendrá dada por la oferta de activos financieros: $W = M^s + B^s + E \cdot B^{s*}$, donde, al ser fijo el tipo de cambio, E puede normalizarse a la unidad ($E = 1$), lo que reduce la expresión a $W = M^s + B^s + B^{s*}$.

Equilibrio: la restricción de riqueza implica (ley de Walras) que la suma de los excesos de demanda es nula:

$$(M^d - M^s) + (B^d - B^s) + (B^{d*} - B^{s*}) = 0$$

Así, si los mercados nacionales (dinero y bonos nacionales) presentan en conjunto un desequilibrio, el mercado de bonos extranjeros presenta un desequilibrio compensador. Si presentan en conjunto un exceso de demanda, existirá un exceso de oferta en el mercado de activos extranjeros y viceversa.

Desarrollo gráfico: el equilibrio del mercado de dinero se representa en el plano dinero-tipo de interés, M-r (figura B12.4):

1. *Demanda de dinero* ($M^d = \omega_M(r, r^*)\, W$): con W y r^* dados, tiene pendiente negativa, porque la proporción de la riqueza que se mantiene en dinero disminuye con el rendimiento de los bonos nacionales.

2. *Oferta monetaria* (M_0^s): es vertical, porque la fijan el crédito interno y las reservas ($M^s = \mathrm{CI} + R$), no el tipo de interés.

3. *Equilibrio*: en el punto 0, con el tipo de interés r_0, las carteras tienen su composición óptima y no hay flujos de reservas.

Estática comparativa (política monetaria expansiva): las implicaciones del modelo, sobre todo para la política económica, se aprecian al analizar una política monetaria expansiva:

[15] Los bonos ofrecen un tipo de interés nominal i, que, con precios fijos, coincide con su rendimiento r.

▪ *Caída del tipo de interés*: cuando aumenta la oferta monetaria por una expansión del crédito interno ($\Delta \text{CI} > 0$, de M_0^s a $M_{0'}^s$), el tipo de interés nacional desciende de r_0 a $r_{0'}$ (punto $0'$) y, así, sube el precio de los bonos nacionales.

▪ *Reajuste de carteras*: los individuos compran bonos extranjeros y se deshacen de dinero y de bonos nacionales. Por la ley de Walras, el exceso de oferta de dinero tiene como contrapartida un exceso de demanda de bonos extranjeros, que el país pequeño satisface adquiriéndolos en el mercado internacional. La compra provoca una salida de reservas y un déficit de balanza de pagos por la salida de capital financiero nacional.

▪ *Fin del ajuste*: la salida de reservas ($\Delta R < 0$) reduce la oferta monetaria, de $M_{0'}^s$ a M_1^s, hasta que las carteras alcanzan su nueva composición óptima en el punto 1. Al tratarse de un ajuste de *stocks*, la salida de capital se agota y desaparece el desequilibrio externo.

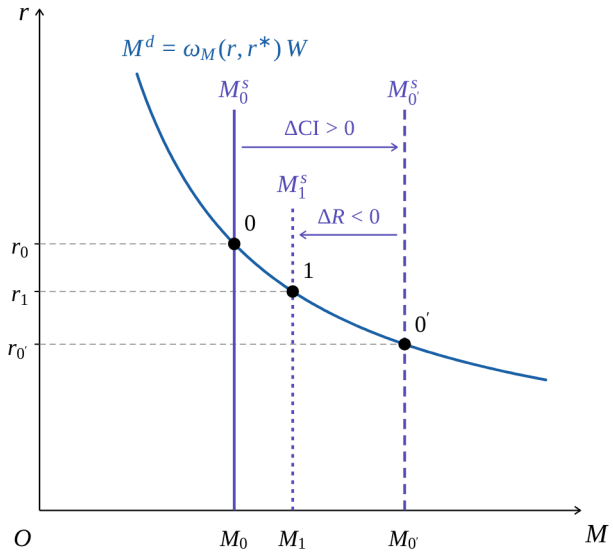

Figura B12.4: Modelo de cartera (McKinnon, 1969). Una expansión del crédito interno reduce el tipo de interés ($0 \to 0'$) y la salida de reservas reduce la oferta monetaria sin devolverla a su nivel inicial ($0' \to 1$), a diferencia del ajuste completo del modelo monetario.

En resumen, las implicaciones son similares a las del modelo monetario, dado que la idea fundamental es la misma. No obstante, existe una diferencia esencial, y es que un desequilibrio en el mercado monetario puede ser compensado, al menos parcialmente, por un desequilibrio en el mercado de bonos nacionales, al no ser sustitutivos perfectos de los bonos extranjeros.

En consecuencia, el desequilibrio en el mercado monetario puede tener mayor magnitud que el desequilibrio externo y, por lo tanto, el desequilibrio monetario no se corregirá exclusivamente vía reservas, sino también con variaciones en las cantidades demandadas de los distintos activos. La oferta monetaria final queda entre la inicial y la ampliada ($M_0^s < M_1^s < M_{0'}^s$), con $r_1 < r_0$, mientras que en el modelo monetario, con $\Delta R = -\Delta \text{CI}$, volvería a M_0^s. Esto hace que la política monetaria tenga un mayor grado de autonomía para corregir desequilibrios internos.

2.2.3. Conclusiones

Aspectos positivos: gana realismo al admitir la sustituibilidad imperfecta entre activos, restaura un margen de autonomía para la política monetaria y constituye el fundamento de los modelos de cartera del tipo de cambio (tema B14, apartado 2.3).

Limitaciones: es un modelo estático, sin expectativas ni valoración del riesgo explícita, cuyos resultados dependen del menú de activos considerado, y mantiene los supuestos de precios fijos y renta dada.

2.3. Conclusiones

El enfoque activos invierte la causalidad de la sección 1, pues, con libre circulación de capitales, el desequilibrio exterior es la contrapartida de un desequilibrio en el mercado de dinero o de activos. En el modelo monetario, con sustituibilidad perfecta, una expansión del crédito interno se traduce en una pérdida de reservas de igual cuantía. La oferta monetaria vuelve así a su nivel inicial, y la política monetaria carece de autonomía duradera (trilema de Mundell-Fleming). El modelo de cartera admite, en cambio, que los bonos nacionales y extranjeros no son sustitutivos perfectos. Por ello, parte del desequilibrio monetario se absorbe en el mercado de bonos nacionales y la política monetaria recupera un margen de autonomía.

Con tipo de cambio fijo, ambos modelos advierten además de que los desequilibrios persistentes anticipan crisis de balanza de pagos, intuición que desarrollan los modelos de crisis cambiarias (tema B17, apartado 1). No obstante, el límite común de ambos modelos es el carácter estático del análisis, que el enfoque intertemporal de la sección 3 viene a superar.

3. Enfoque intertemporal y sostenibilidad del déficit y de la deuda exterior

Los modelos de la balanza de pagos presentados hasta ahora son estáticos, es decir, no tienen en cuenta explícitamente el factor temporal. Sin embargo, la balanza de pagos recoge fenómenos dinámicos, lo que motivó el cambio de enfoque de los años ochenta.

Según los modelos neoclásicos de comercio internacional (p. ej. el modelo de Haberler), la liberalización comercial hace posible desvincular intratemporalmente las decisiones de producción y de consumo. Por su parte, los modelos de equilibrio intertemporal introducen la liberalización financiera, con la que esas decisiones pueden desvincularse también intertemporalmente, pues cabe prestar y endeudarse en el exterior[16].

De este modo se entiende que los movimientos de bienes y de capitales se produzcan de forma simultánea. Solo es posible sostener en el tiempo un saldo deudor en la cuenta corriente si la cuenta financiera registra un saldo de igual magnitud y signo, es decir, si entran capitales. Sin acceso al mercado de capitales, los déficits corrientes solo podrían cubrirse mediante transferencias de capital recibidas, la venta de activos no financieros no producidos (licencias, marcas…) o la de reservas internacionales.

Por ello, una acumulación continuada de déficits por cuenta corriente puede desembocar en problemas de sostenibilidad de la deuda, en la medida en que las obligaciones frente al exterior podrían ser tales que la economía no tenga recursos para satisfacerlas.

3.1. Modelo intertemporal de la balanza de pagos (Sachs, 1981; Obstfeld y Rogoff, 1995)

Origen: Jeffrey Sachs (1981) aplicó el análisis de la elección intertemporal a la cuenta corriente de una economía abierta, y Maurice Obstfeld y Kenneth Rogoff (1995) lo sistematizaron como enfoque intertemporal de la cuenta corriente.

Idea: con acceso al mercado internacional de capitales, el consumo y la inversión de un país dejan de estar restringidos por su producción corriente, de modo que el saldo por cuenta corriente emerge como resultado de la optimización intertemporal de los agentes.

> La ecuación de Euler, el consumidor-productor con el teorema de separación de Fisher y el problema de horizonte infinito con la condición de no-Ponzi, en los que se apoya el modelo, se desarrollan en el tema A29, apartados 1.4.1, 2.4.2 y 3.1. Los determinantes de los flujos internacionales de capital que se derivan del modelo se estudian en el tema B18, apartado 3.1.1, y su aplicación a la financiación de los países en desarrollo, en el tema B29, apartado 1.

3.1.1. Supuestos

Entorno: la economía es pequeña, con mercados perfectos y completos, información perfecta (plena certeza sobre la renta futura) y horizonte infinito en tiempo discreto.

[16] En los modelos neoclásicos de comercio internacional, la economía no se endeuda con el exterior. Puede consumir más de lo que produce de un bien a cambio de consumir menos de lo que produce del otro, de modo que el exceso de consumo de un bien se intercambia por el déficit de consumo del otro.

Agentes: economía de agente representativo que maximiza su utilidad intertemporal según el altruismo intergeneracional puro.

Tecnología:

▪ *Bien único*: se produce un único bien de consumo con una función de producción con un único factor productivo (capital) con productividad marginal positiva ($f'(K) > 0$) pero decreciente ($f''(K) < 0$): $Y_t = f(K_t)$.

▪ *Naturaleza dual*: el factor productivo es, a su vez, bien de consumo, de modo que se utiliza para producir y para consumir.

▪ *Sin depreciación*: no existe depreciación del capital, por lo que la ley de acumulación de capital de la economía es $K_{t+1} = K_t + I_t$, donde I_t es la inversión bruta.

Sector exterior:

▪ *Cuenta corriente*: constituye, junto con la función de producción y la ley de acumulación del capital, una de las restricciones de recursos de la economía. Todo el déficit por cuenta corriente se financia con deuda exclusivamente (otras formas de financiación, como las acciones, no se tienen en cuenta). Por tanto, el saldo de la cuenta corriente será $\text{SB}_{\text{CC},t} = B_{t+1} - B_t$, donde B_t son los activos netos frente al exterior en forma de deuda. Si el saldo de la cuenta corriente es positivo, la economía mostrará capacidad de financiación, lo que se traducirá en un cambio positivo en la posición neta respecto al resto del mundo y viceversa.

▪ *Transversalidad*: se impone una condición que evita que se dé un juego de Ponzi y se originen sendas explosivas de deuda. De lo contrario, el país podría endeudarse indefinidamente sin verse obligado en ningún momento a devolver la cuantía recibida.

3.1.2. Modelo

Planteamiento (economía cerrada): el análisis parte del modelo en autarquía y muestra después los efectos de la apertura exterior. El modelo de autarquía se puede resumir en el programa de maximización del planificador benevolente, que maximiza la utilidad social de acuerdo con una función continua, aditivamente separable, creciente y cóncava, sujeto a las restricciones de recursos de la economía:

$$\max_{\{C_t, I_t\}} U = \sum_{t=0}^{\infty} \beta^t u(C_t), \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} C_t + I_t = Y_t = f(K_t) \\ K_{t+1} = K_t + I_t \\ K_0 \text{ dado} \end{cases}$$

donde C_t, I_t e Y_t son el consumo, la inversión y la renta de cada periodo, K_t el factor de producción, con K_0 dado, $\beta = 1/(1+\rho) \in (0,1)$ el factor de descuento intertemporal y $\rho > 0$ la tasa de descuento subjetiva[17]. En economía cerrada no existe acceso al ahorro exterior, por lo que los bonos están en oferta neta nula y el tipo de interés, r, es endógeno (tipo de interés de autarquía).

[17] Las diferencias de ρ entre países dan distinta forma a sus curvas de indiferencia intertemporales, con mayor sesgo hacia el consumo presente en unos y hacia el futuro en otros.

La cuenta corriente no aparece como restricción separada. En economía cerrada es nula en todo periodo ($B_t = 0$) y, en economía abierta, queda consolidada en la restricción intertemporal, que resulta de iterar la restricción por periodos $B_{t+1} = (1+r^*)B_t + Y_t - C_t - I_t$, con $\mathrm{SB}_{\mathrm{CC},t} = B_{t+1} - B_t$. El saldo corriente de cada periodo emerge, así, como resultado endógeno de las decisiones óptimas, y la única exigencia externa genuina es la condición de solvencia.

Planteamiento (economía abierta): si la economía se abre comercial y financieramente al exterior, cambian dos cosas:

- *Financiación exterior*: la economía puede endeudarse al tipo de interés internacional, que le viene dado por tratarse de un país pequeño.

- *Contrapartida financiera*: un exceso del gasto interno (consumo más inversión) sobre la producción en un periodo ($C_t + I_t > Y_t$) da lugar a un déficit comercial en ese periodo, y un exceso del ahorro sobre la inversión ($C_t + I_t < Y_t$), a un superávit. En ambos casos, los movimientos comerciales tienen una contrapartida financiera directa.

Formalmente, la restricción de recursos por periodos se sustituye por la restricción presupuestaria intertemporal de la economía abierta, en la que B_t pasa a recoger los activos netos frente al exterior en forma de deuda:

$$\sum_{t=0}^{\infty} \frac{C_t + I_t}{(1+r^*)^t} = (1+r^*)B_0 + \sum_{t=0}^{\infty} \frac{Y_t}{(1+r^*)^t}, \qquad \lim_{t\to\infty} \frac{B_{t+1}}{(1+r^*)^t} = 0$$

donde r^* es el tipo de interés mundial, dado para el país pequeño, B_0 los activos netos iniciales frente al exterior, y la condición de no-Ponzi garantiza que no se producen sendas explosivas de deuda.

CPO: en economía cerrada y con solución interior, el método de los multiplicadores de Lagrange[18] da las condiciones de primer orden:

$$\frac{du}{dC_t} = \frac{du}{dC_{t+1}} \frac{1+r}{1+\rho} \;\Rightarrow\; \frac{du/dC_t}{du/dC_{t+1}} = \frac{1+r}{1+\rho} \tag{1}$$

$$\underbrace{\frac{df}{dK_{t+1}}}_{\mathrm{Pmg}_K} = r \tag{2}$$

$$(1 + \frac{df}{dK_{t+1}}) = \frac{1}{\beta} \cdot \frac{du/dC_t}{du/dC_{t+1}} \tag{3}$$

La condición (1), la ecuación de Euler, iguala el cociente de utilidades marginales, $(du/dC_t)/(du/dC_{t+1})$, al rendimiento bruto del ahorro, $1 + r$, corregido por el factor de descuento intertemporal, $\beta = 1/(1+\rho)$. Por tanto, el perfil temporal del consumo depende de cómo se sitúe el tipo de interés, r, frente a la tasa subjetiva de descuento, ρ. La senda es creciente si $r > \rho$, decreciente si $r < \rho$ y constante si $r = \rho$ (tema A29, apartado 1.4.1).

La condición (2), de inversión, iguala la productividad marginal del capital al tipo de interés. Dado que $r = df/dK$, la condición (3) iguala la relación marginal de transformación intertemporal

[18] Los programas de optimización dinámicos se resuelven, con carácter general, mediante ecuaciones de Bellman. No obstante, en el caso de problemas sencillos en certidumbre, como el presente (incluso con horizonte infinito, junto con la condición de transversalidad), basta con aplicar el método de Lagrange o el método de sustitución.

(la tasa de traslado de renta entre periodos vía inversión) a la relación marginal de sustitución intertemporal (la tasa de traslado vía ahorro), $(du/dC_t)/[\beta\, du/dC_{t+1}]$.

En economía abierta, las condiciones (1) a (3) se mantienen, pero con el tipo de interés mundial, r^*, en lugar del nacional. De ello se sigue el *teorema de separación de Fisher* (tema A29, apartado 2.4.2). Como la economía, pequeña y abierta, puede prestar y endeudarse a r^*, la inversión se fija donde $df/dK = r^*$ con independencia de las preferencias, y el consumo se elige después sobre la restricción intertemporal. En economía cerrada, en cambio, el tipo de interés es endógeno y debe garantizar $S = I$, de modo que consumo e inversión se determinan conjuntamente.

CSO: la concavidad estricta de u y la de f ($f''(K) < 0$) hacen cóncavo el objetivo y convexo el conjunto factible. Con la condición de transversalidad, las condiciones de primer orden son necesarias y suficientes. La apertura conserva esta propiedad, porque la restricción intertemporal es lineal en C_t e I_t.

Desarrollo gráfico: con dos periodos, el modelo intertemporal se representa en el plano consumo presente-consumo futuro, C_1-C_2 (figura B12.5):

1. *Frontera de posibilidades de producción intertemporal* (FPP): recoge las combinaciones de consumo presente y futuro alcanzables trasladando recursos mediante la inversión, y es cóncava por los rendimientos decrecientes de f.

2. *Curvas de indiferencia sociales* (CIS_0, CIS_1): recogen las combinaciones de consumo presente y futuro que dan a la economía el mismo bienestar. Estas curvas son decrecientes, porque renunciar a consumo presente exige más consumo futuro, y convexas, porque la relación marginal de sustitución intertemporal disminuye a medida que aumenta C_1.

3. *Recta presupuestaria*: refleja la restricción presupuestaria intertemporal de la economía. El valor presente del consumo no puede exceder el valor presente de la producción neta de inversión. La pendiente de la recta es $-(1+r)$ en autarquía y $-(1+r^*)$ con la apertura (recta mundial), porque cada unidad de consumo presente a la que se renuncia rinde $1+r$, o $1+r^*$, unidades de consumo futuro.

4. *Equilibrio*: en autarquía, el punto de triple tangencia entre las tres, en el que producción y consumo coinciden ($P^0 = C^0$)[19], sobre la curva CIS_0. La pendiente común, $-(1+r)$, define el tipo de interés de autarquía. El consumo presente está limitado por la producción corriente, y la inversión (la producción de mañana) está limitada por el ahorro nacional, porque la economía solo dispone de sus propios recursos.

Estática comparativa: a partir del equilibrio de autarquía, la apertura se analiza en tres etapas:

1. *Apertura*: la economía se abre comercial y financieramente en condiciones de escasez de capital, de modo que el tipo de interés nacional es superior al extranjero, $r = f'(K) > r^*$, y el país resulta, *a priori*, atractivo para los inversores extranjeros. Gráficamente, la recta mundial es más plana que la de autarquía, y el equilibrio se desdobla en dos tangencias separadas sobre ella (teorema de separación). En el punto de producción P^1, la frontera de posibilidades de producción es tangente a la recta mundial ($f'(K) = r^*$, con más inversión que en autarquía), y en el punto de consumo C^1 lo es la curva de indiferencia social más alta alcanzable.

[19] Como el modelo incluye inversión, incluso en autarquía el país produce más de lo que consume (la diferencia es la inversión). En puridad, los puntos P no representan la producción, sino la producción neta de inversión, es decir, los recursos nacionales disponibles para el consumo. En los modelos estáticos de comercio (p. ej. Haberler), sin inversión, los puntos P sí son la producción.

2. *Efectos a corto plazo* ($t = 1$): la brecha horizontal entre C^1 y P^1 en el primer periodo es el déficit por cuenta corriente ($\text{SB}_{\text{CC}} < 0$). Suben el consumo y la inversión y, con ello, la absorción interna. Como la renta se mantiene constante, se genera un déficit corriente[20] financiado con la entrada de capital, que aumenta los pasivos frente al exterior.

3. *Efectos a medio plazo* ($t = 2$): la inversión del primer periodo eleva la producción (la diferencia vertical entre P^1 y P^0), lo que permite el repago del servicio de la deuda (principal más intereses) y revierte el signo de los saldos. La ganancia de bienestar de la apertura se aprecia en que C^1 se sitúa sobre una curva de indiferencia, CIS_1, superior a la de autarquía, CIS_0.

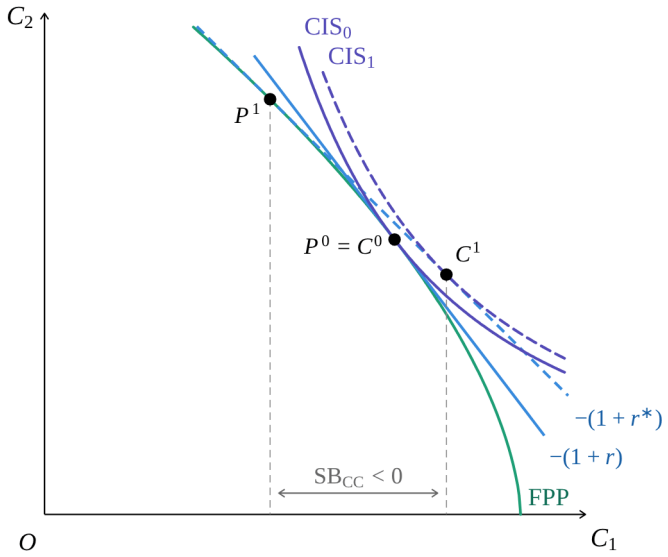

Figura B12.5: Modelo intertemporal de la balanza de pagos. La apertura financiera permite separar consumo y producción intertemporales y financiar un déficit corriente.

Extensiones: del análisis se infieren tres resultados principales:

▪ *Movimientos de capital*: la dirección de los flujos internacionales de capital, que en el modelo adoptan la forma de deuda, depende exclusivamente de la diferencia entre el tipo de interés de autarquía y el mundial. Así, los países con mayores tipos de interés de autarquía (por una mayor productividad marginal del capital o una mayor impaciencia) reciben capital, y viceversa (tema B18, apartado 3.1.1).

▪ *Balanza de pagos*: un déficit comercial no es necesariamente negativo desde el punto de vista del bienestar, pues puede resultar de la optimización intertemporal de los agentes, al financiar inversión productiva o suavizar el consumo ante *shocks* transitorios (tema B11, apartado 3.2.2).

▪ *Perturbaciones transitorias y permanentes*: si $r^* = \rho$, la ecuación de Euler hace constante el consumo, que se sitúa en $C_t = r^* B_t + Y_t^P - I_t^P$, donde $X_t^P = \frac{r^*}{1+r^*} \sum_{s \geq t} X_s/(1+r^*)^{s-t}$ es el valor permanente (la anualidad) de cada variable. Llevado a la restricción por periodos, ese consumo da la *ecuación fundamental de la cuenta corriente* (Obstfeld y Rogoff, 1995), $\text{SB}_{\text{CC},t} = (Y_t - Y_t^P) - (I_t - I_t^P)$, de modo que el país solo presta o se endeuda en la medida en que

[20] Gráficamente, con $r^* < r$ es imposible que el país registre un superávit corriente en el primer periodo (que salgan capitales), pues ello supondría consumir a la izquierda de P^1, y la curva CIS_1 cortaría a la CIS_0.

la producción o la inversión se apartan de su nivel permanente. Una caída transitoria de la producción genera así un déficit, porque el país se endeuda para suavizar el consumo. Una caída permanente, en cambio, reduce el consumo en la misma cuantía y deja inalterada la cuenta corriente. Por último, una mejora esperada de la productividad del capital genera un déficit, porque eleva hoy la inversión y la renta permanente.

3.1.3. Conclusiones

Aspectos positivos: el modelo dota a la balanza por cuenta corriente de microfundamentos. Con ello explica fenómenos que los enfoques tradicionales no captan, como que la valoración de un déficit corriente dependa de la naturaleza, transitoria o permanente, de las perturbaciones.

Limitaciones: sus predicciones chocan con dos regularidades empíricas. En primer lugar, el *puzzle* de Feldstein y Horioka (1980) muestra que el ahorro y la inversión nacionales están mucho más correlacionados de lo que predice un modelo con perfecta movilidad de capitales. En segundo lugar, la paradoja de Lucas (1990) consiste en que el capital fluye poco hacia los países pobres. Ello apunta a riesgo de impago, fricciones informativas y diferencias institucionales, que el modelo básico ignora al suponer un único tipo de interés mundial y ausencia de restricciones al endeudamiento.

3.2. Desarrollos posteriores del enfoque intertemporal

Tres desarrollos posteriores retoman las limitaciones empíricas del modelo intertemporal para explicar la dirección de los flujos internacionales de capital, a partir del capital humano, del ahorro precautorio y de las características demográficas.

3.2.1. La importancia del capital humano (Lucas, 1990)

A la luz del modelo anterior, se podría pensar que el capital tenderá a desplazarse de los países ricos hacia los países pobres. La razón es que los países pobres, con un menor *stock* de capital por trabajador, tendrían en principio una mayor productividad marginal del capital y, por tanto, ofrecerían una mayor rentabilidad.

Sin embargo, el capital no fluye hacia los PED tanto como cabría esperar. Es más, en conjunto, los países desarrollados han sido con frecuencia importadores netos de capital. Entre los propios PED se produce, además, lo que Pierre-Olivier Gourinchas y Olivier Jeanne denominan "**rompecabezas de la asignación**" (*allocation puzzle*), es decir, que el capital fluye menos hacia los países con mayor crecimiento de la productividad[21].

Uno de los factores explicativos puede encontrarse en el trabajo de Robert Lucas *Why Doesn't Capital Flow from Rich to Poor Countries?* (1990), donde concluye que el capital humano desempeña un papel fundamental.

Lucas demuestra[22] cómo una función neoclásica agregada predice una productividad marginal del capital en la India 58 veces superior a la productividad marginal del capital en Estados Unidos. Esa diferencia debería atraer hacia la India entradas de capital mucho más fuertes que hacia

[21] Gourinchas y Jeanne (2013) muestran que las importaciones netas de capital de los PED están negativamente correlacionadas con el crecimiento de su productividad, patrón que se explica en gran parte por los flujos de capital del sector público.

[22] Para ello, se apoya en el trabajo de Robert Summers y Alan Heston (1988), según el cual la producción por trabajador de EE. UU. es unas 15 veces la de la India, y Lucas supone una participación del capital en la renta de 0,4 ($\alpha = 0{,}4$).

EE. UU. Sin embargo, nada de esto concuerda con los datos, ni en términos de productividad, ni en términos de flujos internacionales de capital.

Lucas señala que la función de producción del modelo neoclásico es inadecuada si no tiene en cuenta el capital humano. En efecto, cuando se introduce el capital humano, la ratio de productividades se reduce de 58 a 5, y cuando se consideran las externalidades de dicho capital humano, la ratio se reduce a la unidad. Todo ello permite explicar por qué el capital no se traslada masivamente de los países ricos a los PED.

3.2.2. La importancia del ahorro precautorio (Bewley, 1986; Aiyagari, 1994)

Por otro lado, a partir de los modelos de Truman Bewley y S. Rao Aiyagari, Mendoza, Quadrini y Ríos-Rull (2009) destacan la incidencia del **ahorro precautorio** en los países en desarrollo. Con mercados financieros poco desarrollados, el riesgo idiosincrático (riesgo individual no asegurable por la incompletitud de los mercados) eleva las tasas de ahorro precautorio[23], pues los agentes tratan de "autoasegurarse".

La consideración del ahorro precautorio conlleva que el tipo de interés de autarquía se encuentra por debajo del que prevalecería con mercados completos (en estado estacionario, la tasa de descuento subjetiva, ρ). Los diferenciales en los tipos de interés de autarquía entre países vendrán determinados, así, por su distinto grado de desarrollo financiero, del que depende la capacidad de cada país para asegurar el riesgo idiosincrático.

En este contexto, el capital tenderá a fluir desde los PED, que ofrecen menores tipos de interés debido a su mayor propensión a autoasegurarse, hacia los países desarrollados, en los que existe mayor desarrollo financiero.

3.2.3. La importancia de las características demográficas (Bernanke, 2005)

Finalmente, también es relevante la estructura demográfica del país. Según la hipótesis del ciclo vital (tema A33, apartado 2.2), la cohorte de jubilados ahorra menos y la cercana a la jubilación ahorra en previsión de esta. Por ello, un mayor peso de la primera tiende a elevar el tipo de interés, y el de la segunda, a reducirlo.

Por su parte, Ben Bernanke (2005) atribuye el exceso global de ahorro (*global saving glut*) en parte al elevado ahorro de las economías avanzadas con poblaciones envejecidas, que ahorran en previsión de la jubilación. Sin embargo, la causa principal es, para él, el giro de las economías emergentes hacia superávits por cuenta corriente y acumulación de reservas. Ello contribuye a explicar el déficit por cuenta corriente de EE. UU. y el bajo nivel de los tipos de interés reales a largo plazo.

3.3. Análisis de la sostenibilidad del déficit y de la deuda exterior

Que el déficit exterior pueda ser eficiente no garantiza que sea sostenible, porque el exceso de endeudamiento exterior puede plantear problemas de solvencia y de liquidez. Un modelo sencillo basado en la balanza de pagos permite analizar la sostenibilidad de la deuda exterior.

[23] Estas tasas, a su vez, vendrán determinadas por la propensión a la prudencia de los hogares y la volatilidad de los *shocks* de riesgo idiosincrático.

3.3.1. Consideraciones

El enfoque intertemporal presenta el endeudamiento exterior como una decisión racional, fruto de una optimización entre periodos. No obstante, un comportamiento racional no asegura que cualquier volumen de deuda sea óptimo, y surgen dificultades, en particular, cuando:

- *Impaciencia*: el déficit exterior que se genera es abultado y financia sobre todo el consumo, no la inversión, porque los agentes valoran en exceso el consumo presente (su tasa de descuento subjetiva es muy alta).

- *Mala planificación*: afecta a los proyectos de inversión.

- *Incertidumbre*: es elevada, por las dudas sobre la viabilidad de los proyectos o la falta de credibilidad institucional.

Todas estas situaciones pueden derivar en un **exceso de endeudamiento exterior** que, a su vez, puede dar lugar a dos tipos de problemas:

Problemas de solvencia: se refieren a las dificultades de un país para hacer frente al servicio de la deuda en el largo plazo, según sus fundamentos. Se habla de *sostenibilidad de la deuda* cuando un país puede atender ese servicio de manera continuada, sin necesidad de financiación extraordinaria ni de un ajuste interno de gran magnitud, y el modelo que sigue lo concreta en una ratio deuda externa-PIB constante. En una situación de insolvencia, el país no podrá atender el servicio de la deuda sin reducir su *stock*.

Problemas de liquidez: se refieren a las dificultades para hacer frente a los vencimientos de la deuda en el periodo en curso, por la asincronía entre ingresos y pagos. En una situación de iliquidez, el país no puede atender el servicio de su deuda sin reprogramarla, es decir, sin aplazar parte de los vencimientos del periodo corriente.

Por todo ello, desde el punto de vista de política económica, el acceso a la financiación externa aporta ganancias de bienestar, pero debe ir acompañado de una política de endeudamiento prudente.

3.3.2. Modelo de sostenibilidad de la deuda exterior basado en la BP

Origen: la dinámica de la deuda exterior se analiza con la misma aritmética que Domar (1944) aplicó a la deuda pública (tema A39, apartado 2.1), a partir de la identidad contable de la balanza de pagos. Milesi-Ferretti y Razin (1996) sistematizaron ese análisis como criterio de sostenibilidad de la cuenta corriente.

Idea: como el crecimiento diluye la carga de la deuda exterior, basta para estabilizar su ratio sobre el PIB un superávit comercial por unidad de deuda que compense lo que el tipo de interés real excede a la tasa de crecimiento.

Supuestos:

- *Metodología*: análisis contable de la dinámica de la deuda, a partir de la identidad de la balanza de pagos.

- *Entorno*: país pequeño abierto al exterior ($i = i^*$), que crece a una tasa constante y exógena g_Y, con libre movimiento de capital.

▪ *Agentes*: los consumidores son propietarios del factor de producción y de la tecnología. Entre sus tenencias de riqueza se encuentra deuda pública nacional, deuda pública extranjera y participaciones en empresas.

▪ *Tecnología*: se produce exclusivamente a partir del capital productivo, el cual no se deprecia: $Y = f(K)$.

▪ *Sector exterior*: no existen transferencias corrientes redistributivas (p. ej. cotizaciones sociales o subsidios), de modo que no hay rentas secundarias. Se cumple la ecuación de Fisher, es decir, la relación entre los tipos de interés nominal y real y la inflación es tal que $1+r^* = (1+i^*)/(1+\pi^*)$. Además, se supone que la inflación nacional coincide con la extranjera ($\pi = \pi^*$), como ocurre con tipo de cambio fijo y paridad del poder adquisitivo (PPA). El déficit se cubre exclusivamente con deuda pública o privada, sin inversión directa ni en acciones.

Planteamiento: el punto de partida es la identidad contable de la balanza de pagos, $\mathrm{SB}_{\mathrm{CC},t} + \mathrm{SB}_{\mathrm{CK},t} = \mathrm{SB}_{\mathrm{CF},t}$, donde B_t denota ahora el *stock* de deuda externa neta al inicio del periodo. Desagregando el comportamiento de cada cuenta se obtiene:

▪ *Cuenta corriente*: su saldo comprende la balanza comercial, $\mathrm{BC}_t = X_t - M_t$, y las rentas primarias, que aquí son los intereses de la deuda externa acumulada: $\mathrm{SB}_{\mathrm{CC},t} = \mathrm{BC}_t - i^* B_t$.

▪ *Cuenta de capital*: su saldo es nulo, al no considerarse transferencias de capital: $\mathrm{SB}_{\mathrm{CK},t} = 0$.

▪ *Cuenta financiera*: su saldo viene dado por la variación del *stock* de deuda externa neta, es decir, por el nuevo endeudamiento: $\mathrm{SB}_{\mathrm{CF},t} = -(B_{t+1} - B_t)$, negativo cuando entran capitales.

Sustituyendo en la identidad, se obtiene la ecuación dinámica de la deuda en términos nominales:

$$\mathrm{BC}_t - i^* B_t = -(B_{t+1} - B_t) \to B_{t+1} = (1 + i^*)B_t - \mathrm{BC}_t$$

Es decir, la deuda del periodo siguiente es la deuda viva capitalizada al tipo de interés, minorada por el superávit comercial que se destina a su servicio.

Para llegar a la ecuación dinámica de la deuda en proporción del PIB basta con dividir por el PIB nominal del periodo siguiente, $P_{t+1}Y_{t+1} = P_t Y_t (1 + \pi^*)(1 + g_Y)$, utilizar la ecuación de Fisher y operar:

$$b_{t+1} = \frac{1 + i^*}{(1 + \pi^*)(1 + g_Y)} b_t - \mathrm{bc}_{t+1} = \frac{1 + r^*}{1 + g_Y} b_t - \mathrm{bc}_{t+1}$$

donde las minúsculas denotan ratios sobre el PIB ($b_t \equiv B_t/(P_t Y_t)$ y $\mathrm{bc}_{t+1} \equiv \mathrm{BC}_t/(P_{t+1}Y_{t+1})$) y r^* es el tipo de interés real mundial que da la ecuación de Fisher.

Equilibrio: la condición de sostenibilidad de la deuda establece que la ratio deuda exterior-PIB sea estable, es decir, $b_{t+1} = b_t$, de donde:

$$b_t = \frac{1 + r^*}{1 + g_Y} b_t - \mathrm{bc}_{t+1} \to \mathrm{bc}_{t+1} = \left(\frac{1 + r^*}{1 + g_Y} - 1 \right) b_t \to \frac{\mathrm{bc}_{t+1}}{b_t} = \frac{r^* - g_Y}{1 + g_Y}$$

La última igualdad traduce la sostenibilidad exterior en un objetivo de saldo comercial. En concreto, el superávit (bc_{t+1}) ha de igualar el coste efectivo de la deuda exterior viva ($r^* b_t$), una vez descontada la parte que diluye el crecimiento ($g_Y\, b_t$) y con el ajuste del factor ($1 + g_Y$).

Estática comparativa: con deuda positiva, el superávit comercial que estabiliza la ratio aumenta con el tipo de interés real mundial, r^*, y disminuye con la tasa de crecimiento, g_Y, porque el crecimiento del PIB diluye la deuda en proporción del PIB. El signo de $r^* - g_Y$ separa dos casos:

- $g_Y > r^*$: el país puede registrar un déficit comercial proporcional a la diferencia entre ambas tasas sin que su ratio de deuda aumente, de modo que el déficit comercial es sostenible.

- $g_Y < r^*$: el país debe registrar un superávit comercial proporcional a la diferencia entre ambas tasas para que su ratio de deuda no aumente, y ese superávit es mayor cuanto mayor sea la deuda heredada, b_t.

Un ejemplo ilustrativo de los problemas de sostenibilidad exterior es el caso de la periferia de la zona del euro. Varias economías periféricas (Grecia, Portugal o España) acumularon en la década de 2000 elevados déficits por cuenta corriente, financiados con entradas de capital procedentes del resto de la zona. Conforme al análisis expuesto, esos déficits no eran sostenibles. La corrección posterior, durante la crisis de deuda soberana (2010-2014), exigió un ajuste severo de la absorción interna y una reducción de costes y precios relativos (la llamada *devaluación interna*), que devolvió las cuentas corrientes de esas economías al equilibrio o al superávit.

Extensiones: más allá de la condición formal, la sostenibilidad del déficit depende de varios factores:

- *Volumen de deuda*: una deuda más abultada pone más en riesgo su propia sostenibilidad, por dos razones principales:

  - *Prima de riesgo*: el riesgo de impago aumenta con la deuda acumulada y, en consecuencia, los prestamistas exigen una prima más alta.

  - *Riesgo cambiario*: si la deuda exterior se ha contraído en divisa, cualquier depreciación de la moneda propia aumenta su importe en moneda nacional.

- *Tasa de crecimiento*: el destino del capital financiero exterior es decisivo. Si se invierte en actividades productivas, y no en consumo, impulsa el crecimiento, y un mayor crecimiento aligera la carga de la deuda y la hace más sostenible.

- *Evolución de los tipos de interés*: si el tipo de interés sube con fuerza, la sostenibilidad también se resiente, de modo que cuentan la capacidad del país para negociar en los mercados de capitales exteriores y su credibilidad.

- *Expectativas*: aunque no aparecen explícitamente en la fórmula, condicionan tanto la capacidad de endeudamiento del país, que se reduce si la economía no cubre las expectativas del mercado, como la llegada de inversión directa productiva y, con ella, el crecimiento[24].

- *Absorción interna*: si es elevada, desvía hacia el consumo o la inversión recursos reales que podrían destinarse al servicio de la deuda.

[24] Durante los años setenta, muchos países en desarrollo destinaron buena parte de los préstamos exteriores a grandes proyectos de inversión sin viabilidad económica. Además, un cambio de expectativas puede adelantar una crisis de deuda o incluso provocarla, en particular cuando los inversores muestran comportamientos de rebaño.

▪ *Controles de capitales*: cerrar al país el acceso a la financiación internacional, o encarecérsela hasta tipos muy altos, le deja menos recursos para pagar su deuda. El motivo es que el crédito exterior constituye una fuente de financiación que le permite liberar los medios con que saldar los compromisos adquiridos con sus acreedores[25].

Conclusiones:

▪ *Aspectos positivos*: el modelo ofrece una regla operativa y parsimoniosa (la relación entre r^*, g_Y y el superávit comercial) que constituye la base de los análisis de sostenibilidad de la deuda del FMI.

▪ *Limitaciones*:

- *Financiación exclusiva con deuda*: el modelo supone que el déficit se cubre por entero con aumentos de los pasivos financieros netos[26]. Deja fuera la financiación con capital (inversión directa y en acciones), que no obliga a pagar intereses fijos, y otras fuentes de recursos exteriores. Unos ingresos netos por rentas primarias del trabajo, los que perciben los trabajadores residentes por trabajar para empleadores no residentes, ayudarían a sostener el déficit. Asimismo, las rentas secundarias, transferencias como las remesas o la ayuda oficial al desarrollo, pueden pesar mucho, sobre todo en los PED. Por último, vender activos no financieros no producidos (derechos sobre recursos naturales, licencias o marcas) mejora la cuenta de capital y, por esa vía, la capacidad de financiación del país.

- *Nivel de la ratio*: el criterio habitual de sostenibilidad, que la ratio deuda externa-PIB no crezca, se queda corto, ya que nada dice del nivel en que esa ratio se estabiliza.

3.4. Conclusiones

La evidencia disponible confirma el papel de los determinantes examinados, si bien con capacidad explicativa desigual. Diversos estudios (p. ej. Chinn y Prasad, 2003) encuentran que la posición fiscal, el nivel de desarrollo, la posición acreedora o deudora neta y la estructura demográfica explican una parte sustancial de la variación de los saldos por cuenta corriente entre países. En cambio, el tipo de cambio real presenta una asociación más débil e inestable de lo que sugiere el análisis teórico.

Dos regularidades tienen especial relevancia para la política económica. La primera es que los desequilibrios exteriores presentan una elevada persistencia, lo que apunta a determinantes estructurales antes que cíclicos. La segunda regularidad es la paradoja de Lucas, que se atribuye a las diferencias de capital humano, a la calidad institucional, al riesgo de expropiación y a la acumulación preventiva de reservas.

[25] No obstante, estos préstamos también aumentan el volumen de deuda, de modo que pueden aliviar un problema de liquidez a corto plazo pero generar uno de solvencia a largo plazo, salvo que la confianza en el país se mantenga y el país siga recibiendo financiación.

[26] El "aumento de los pasivos financieros netos" recoge tanto aumentos de la deuda exterior como disminuciones de las reservas.

Conclusiones

Los enfoques tradicionales del ajuste, pensados para un régimen de cambio fijo con escasa movilidad de capitales, sitúan el origen del desequilibrio en la cuenta corriente. Lo corrigen mediante el tipo de cambio real (flujo-especie de Hume, elasticidades y condición de Marshall-Lerner) o mediante la renta (multiplicador del comercio exterior y absorción). Con la liberalización financiera de los años setenta, el enfoque activos invierte la causalidad y hace de los desequilibrios externos la contrapartida de desequilibrios en los mercados de dinero y de activos. Desde los años ochenta, el enfoque intertemporal dota a la cuenta corriente de microfundamentos y muestra que un déficit transitorio puede reflejar decisiones óptimas de ahorro e inversión. Por ello, la pregunta relevante deja de ser cómo evitarlo y pasa a ser en qué condiciones es sostenible, lo que, cuando $r^* > g_Y$, exige un superávit comercial que cubra el coste efectivo de la deuda exterior corregido por el crecimiento.

Los enfoques del ajuste dan contenido de comportamiento a los saldos de la balanza de pagos (tema B11). El enfoque monetario y el de cartera son las versiones con cambio fijo de los modelos de activos del tipo de cambio (tema B14), y la incompatibilidad que señalan entre una paridad fija y la expansión del crédito interno es el mecanismo de las crisis de primera generación (tema B17). Por su parte, el mecanismo de ajuste que deja cada régimen cambiario es uno de los criterios para elegirlo (tema B15). Por último, la absorción y el enfoque monetario inspiran el diseño de los programas del FMI, que combinan la reducción de la demanda interna con el cambio en los precios relativos (tema B22).

La facturación en dólares de buena parte del comercio mundial pone en duda la eficacia de la devaluación que supone el enfoque de elasticidades. Gopinath et al. (2020) muestran que, por ello, los precios y los volúmenes comerciales responden más al tipo de cambio del dólar que al bilateral. Sigue abierto, además, el debate sobre el origen de los desequilibrios globales. Frente al exceso global de ahorro de Bernanke (2005), expuesto en el apartado 3.2.3, Borio y Disyatat (2011) objetan que los saldos netos ocultan los flujos brutos que financiaron el auge del crédito. Atribuyen la crisis a un exceso de elasticidad del sistema monetario y financiero internacional, no de ahorro.

Anexo. Condición de Marshall-Lerner como condición de estabilidad

Además de indicar cuándo una devaluación mejora el saldo comercial (apartado 1.1.2), la condición de Marshall-Lerner es una **condición de estabilidad del mercado de divisas**. Su lectura depende del sistema de intercambio:

- *Sistema de intercambio comercial monetario*: si la oferta y la demanda de divisas derivan exclusivamente de las operaciones comerciales, una perturbación que incline al país hacia el déficit genera un exceso de demanda de divisas que aprecia la divisa (deprecia la moneda nacional). El equilibrio será estable si esa depreciación corrige el déficit, lo que ocurre precisamente cuando se cumple la condición de Marshall-Lerner. Si no se cumple, la depreciación amplía el déficit y el sistema se aleja del equilibrio, lo que, en un régimen fijo, puede conducir al agotamiento de las reservas.

- *Sistema de intercambio comercial puro (trueque)*: sin dinero ni mercado de divisas, el análisis se formula con las curvas de oferta recíproca de Marshall, y el precio relevante es la relación real de intercambio. Estas curvas pueden cortarse en varios puntos, con equilibrios alternativamente estables e inestables. Un equilibrio es estable si la suma de las elasticidades de las demandas recíprocas supera la unidad (condición análoga a la de Marshall-Lerner). En ese caso, ante un exceso de oferta de las exportaciones de un país, el deterioro de su relación real de intercambio restablece el equilibrio. En los equilibrios inestables, en cambio, la relación real de intercambio se aleja de ellos hacia el equilibrio estable más próximo.

Gráficamente, en el plano divisas-tipo de cambio (figura B12.6), el primer caso se representa con dos curvas:

1. *Demanda de divisas* (D_{div}): procede de las importaciones y tiene pendiente negativa, porque una depreciación las encarece.

2. *Oferta de divisas* (S_{div}): procede de las exportaciones. Con una demanda de exportaciones inelástica, a partir de cierto tipo de cambio la depreciación reduce los ingresos en divisas y la curva se vuelve hacia atrás.

3. *Equilibrio*: las curvas se cortan dos veces. En e_0 (Estable) se cumple la condición de Marshall-Lerner, y un exceso de demanda de divisas, al depreciar la moneda, se corrige. En e_1 (Inestable) la suma de las elasticidades no supera la unidad, y la depreciación amplía el exceso de demanda y aleja el tipo de cambio del equilibrio.

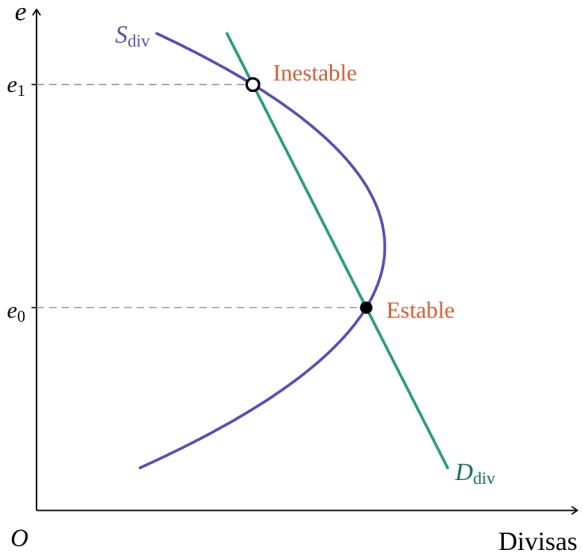

Figura B12.6: Condición de Marshall-Lerner y estabilidad del mercado de divisas. Con una demanda de exportaciones inelástica, la oferta de divisas se vuelve hacia atrás. El equilibrio superior, en el que la suma de las elasticidades no supera la unidad, es inestable.

B13. Mercado de divisas

Mercados de divisas: operaciones e instrumentos.

Introducción

Tal y como subrayó Paul Volcker, "el tipo de cambio es el precio más importante de una economía"[1]. En efecto, la evolución macroeconómica depende de forma decisiva de los tipos de cambio, que condicionan el comercio exterior, la inflación y el *valor* en moneda nacional de la deuda externa. Esa influencia hace del **mercado de divisas** uno de los mercados con más peso en el sistema económico internacional. Con ese nombre genérico se alude a la red de instituciones a través de la cual unas monedas nacionales se compran y se venden a cambio de otras. Por la misma razón, la forma en que cada país organiza la relación entre su moneda y las demás, su **régimen cambiario**, es una de sus decisiones de política económica más importantes.

Históricamente, la relevancia de los tipos de cambio ha crecido con la integración económica internacional. Durante el patrón oro clásico (c. 1870-1914) los tipos de cambio eran fijos respecto del oro; el periodo de entreguerras combinó la volatilidad con devaluaciones competitivas; y el sistema de Bretton Woods (1944) estableció paridades fijas pero ajustables frente al dólar. Tras la Segunda Guerra Mundial, la rápida expansión del comercio y de los movimientos de capital multiplicó las operaciones de cambio entre monedas. El fin de la convertibilidad del dólar en oro (1971) y la generalización de la flotación (1973) abrieron una etapa de tipos de cambio flexibles, caracterizada por una elevada volatilidad y por la libertad de cada país para elegir su régimen cambiario, reconocida en los Acuerdos de Jamaica (1976).

Hoy, el mercado de divisas es el mayor mercado financiero del mundo. Según la encuesta trienal del Banco de Pagos Internacionales (BIS)[2], el volumen medio de negociación *diario* alcanzó en abril de 2025 los 9,6 billones de dólares, mientras que el comercio mundial de bienes y servicios ronda los 35 billones de dólares *anuales*. El valor anual de las transacciones de divisas multiplica así por setenta, aproximadamente, el del comercio que, en principio, deberían financiar. El crecimiento de la negociación de divisas ha ido acompañado de una transformación constante del funcionamiento del mercado, impulsada por el avance tecnológico y la innovación financiera, y de un debate recurrente sobre qué determina los tipos de cambio y qué régimen cambiario es preferible.

El tema se ocupa del mercado de divisas como mercado, es decir, de quién opera en él, con qué contratos y con qué grado de eficiencia. Su negociación es descentralizada y extrabursátil, no se interrumpe durante los días laborables, se canaliza a través de bancos creadores de mercado

[1] "A nation's exchange rate is the single most important price in its economy; it will influence the entire range of individual prices, imports and exports, and even the level of economic activity. So it is hard for any government to ignore larger swings in its exchange rate." Volcker y Gyohten, *Changing Fortunes: The World's Money and the Threat to American Leadership* (1992).

[2] BIS (2025), *Triennial Central Bank Survey. OTC foreign exchange turnover in April 2025*, Basilea, Banco de Pagos Internacionales.

y plataformas electrónicas y está muy concentrada, pues el dólar interviene en el 89,2 % de las operaciones (BIS, 2025). Sus buenas prácticas descansan, sobre todo, en un código de adhesión voluntaria, el Código Global de Conducta del mercado de divisas (2017).

El tema estudia el mercado de divisas según el esquema siguiente:

- *Concepto, características y agentes*: concepto de mercado de divisas, sus características y los agentes que operan en él.

- *Instrumentos y operaciones*: los contratos que se negocian y las operaciones que se hacen con ellos.

- *Eficiencia y regulación*: la eficiencia del mercado de divisas y su regulación.

1. Mercado de divisas: concepto, características y agentes

El estudio del mercado parte de lo que en él se negocia, la divisa, y de su precio, el tipo de cambio, para describir después cómo funciona y quiénes operan en él.

1.1. Divisa y tipo de cambio

El **mercado de divisas** es el marco organizativo en el que agentes de todo el mundo intercambian moneda extranjera y en el que se forma su precio, el tipo de cambio. Conviene precisar antes ambos conceptos.

1.1.1. Concepto de divisas

Definición: las *divisas* son los medios de pago denominados en moneda extranjera que se aceptan con carácter general para saldar operaciones internacionales. En sentido amplio, además de los billetes y monedas extranjeros de curso legal, incluyen los depósitos bancarios denominados en moneda extranjera y los documentos que dan derecho a disponer de ellos (p. ej. cheques, talones o tarjetas de crédito)[3]. No incluyen, en cambio, los instrumentos de deuda en moneda extranjera.

Tipos: por su grado de convertibilidad, se distinguen tres clases de divisas:

1. *Plenamente convertibles*: se intercambian sin restricciones, al contado (*spot*) y a plazo (*forward*), tanto si el tipo de cambio lo fija el mercado como si lo fija la autoridad. Son ejemplos de esta clase:

- *Majors*[4]: las cinco principales, el dólar estadounidense, el euro, el yen, la libra esterlina y el franco suizo (USD, EUR, JPY, GBP y CHF).

- *Minors*: el dólar canadiense, el australiano y el neozelandés (CAD, AUD y NZD), a los que se suman las tres coronas escandinavas.

- *Exóticas*: algunas de ellas (p. ej. el peso mexicano o el *rand* sudafricano).

2. *Parcialmente convertibles*: la conversión está limitada a ciertas operaciones[5] o a ciertos agentes, con el renminbi chino[6] como ejemplo paradigmático. Según quién pueda convertir libremente la moneda, se habla de dos variantes:

- *Convertibilidad externa*: solo la tienen los no residentes.

- *Convertibilidad interna*: solo la tienen los residentes.

[3] Como incluyen depósitos, el tipo de interés ya no es solo un coste de oportunidad de la divisa, sino también una fuente de rentabilidad de esta.

[4] No deben confundirse con las monedas de "libre uso" del Convenio Constitutivo del FMI (art. XXX, f). Estas son hoy las cinco de la cesta del derecho especial de giro (DEG): USD, EUR, CNY, JPY y GBP, lista que incluye el renminbi y no el franco suizo.

[5] El Convenio Constitutivo del FMI prohíbe, salvo aprobación del Fondo o al amparo del régimen transitorio del art. XIV, las restricciones a los pagos y transferencias por transacciones internacionales corrientes (art. VIII, sección 2), pero admite los controles de capitales (art. VI, sección 3).

[6] Vinculado al dólar hasta 2005 (y de nuevo entre 2008 y 2010), flota desde entonces dentro de una banda estrecha con referencia a una cesta de divisas. Es convertible por cuenta corriente desde 1996, pero no plenamente por cuenta financiera, y forma parte de la cesta del DEG desde 2016.

3. *No convertibles* (*blocked currencies*): los controles de cambio impiden cambiarlas libremente, así que la moneda circula solo en las transacciones internas y no se negocia abiertamente en el mercado de divisas.

Algunos de los motivos por los cuales puede imponerse la no convertibilidad son las dificultades de balanza de pagos, el temor a que se produzcan fugas masivas de capitales hacia ubicaciones *offshore* o la elevada volatilidad debida a la especulación.

Sin llegar a ese extremo, el real brasileño, el peso argentino o la rupia india tienen restringida la plena convertibilidad por razones históricas, por lo que buena parte de su negociación exterior se articula mediante instrumentos *non-deliverable*.

1.1.2. Concepto de tipo de cambio

Definición: el *tipo de cambio* es el precio relativo de un par de monedas (p. ej. USD/EUR). La primera moneda es la moneda base (*base currency*) y la segunda moneda es la moneda de cotización (*quote currency*). Obsérvese que esta notación de mercado sigue el orden opuesto al de la notación algebraica (cociente) que emplea el libro, pues el par USD/EUR expresa euros por dólar, es decir, EUR/USD en forma de cociente.

Tipos: a partir de esta definición se distinguen cuatro:

1. *Tipo de cambio nominal* (E): es el precio relativo o tasa de cambio de dos monedas y, por tanto, no tiene en cuenta su poder adquisitivo. Se distinguen tres cotizaciones, dos según que la moneda base sea la extranjera o la nacional y una tercera entre dos monedas extranjeras:

- *Tipo de cambio directo o continental*: precio de una unidad de moneda extranjera expresado en moneda nacional, $E = M/M^*$ ($\uparrow E \rightarrow$ depreciación). Este tema y el resto del libro siguen esta convención.

- *Tipo de cambio inverso o americano*: precio de una unidad de moneda nacional expresado en moneda extranjera, $1/E = M^*/M$ ($\uparrow 1/E \rightarrow$ apreciación). El *Financial Times*, por ejemplo, utiliza esta cotización.

- *Tipo de cambio cruzado*: el tipo de cambio entre dos monedas en el que ninguna de ellas es la moneda del país, obtenido a partir de las cotizaciones de ambas frente a una tercera moneda vehicular (generalmente, el dólar).

2. *Tipo de cambio real* (e): corrige el tipo nominal por los niveles de precios de los dos países y mide cuántas unidades de bienes nacionales cuesta una unidad de bienes extranjeros, $e = E \cdot (P^*/P)$[7].

3. *Tipo de cambio efectivo nominal*: promedio geométrico de los tipos nominales bilaterales frente a los socios comerciales, ponderados por la importancia de cada uno en el comercio del país.

Por ejemplo, el BCE publica índices de tipos de cambio efectivos nominales del euro frente a tres grupos de socios comerciales, cuya composición se ajusta con cada ampliación de la zona del euro (la última, la entrada de Bulgaria en enero de 2026):

- *Grupo estrecho* (TCE-12): frente a los principales socios comerciales.

[7] Obsérvese que, si cada país está especializado en un único bien ($P = P_X$ y $EP^* = P_M$), el tipo de cambio real es la inversa de la relación real de intercambio (Bajo y Monés, 2000, p. 474).

▪ *Grupo ampliado* (TCE-17): añade al anterior otros socios relevantes, como China y varios Estados miembros de la UE ajenos a la zona del euro.

▪ *Grupo amplio* (TCE-40): añade un conjunto más extenso de países y es el índice de referencia del BCE.

4. *Tipo de cambio efectivo real*: resulta de multiplicar el tipo de cambio efectivo nominal, en cotización directa, por una media ponderada de precios o costes externos (p. ej. costes laborales unitarios) relativa a los precios o costes del propio país, como en $e = E \cdot (P^*/P)$. Los índices que publican el BCE y el FMI usan, sin embargo, la cotización inversa, de modo que en ellos un aumento indica apreciación (véase el tema A30).

1.2. Características del mercado de divisas

El mercado de divisas abarca tanto la infraestructura física como el conjunto de instituciones necesarias para negociar. Desde el punto de vista descriptivo, ha experimentado una gran transformación y hoy presenta los siguientes rasgos:

Dimensión: además de ser el mayor del mundo (véase la introducción), es un mercado profundo y líquido. Es también intertemporal, pues tiene un segmento de operaciones *spot* (entrega prácticamente inmediata) y otro de operaciones *forward* (entrega a plazo).

Composición y estructura:

▪ *Concentración*: pocas divisas y pocas plazas reúnen la mayor parte del volumen, según la encuesta trienal del BIS de 2025:

- *Por divisas*: el dólar, presente en el 89,2 % de las operaciones, representa el 44,6 % del volumen negociado[8]. A distancia se sitúan el euro, con el 14,5 %, y el yen, con el 8,4 %.

- *Por ubicación geográfica*: el 38 % de la negociación se realiza en el Reino Unido y el 19 %, en EE. UU.

▪ *Carácter interbancario*: la negociación se desarrolla sobre todo entre bancos y otras instituciones financieras, que concentran en torno al 95 % del volumen según la encuesta trienal del BIS de 2025. Las operaciones con clientes no financieros, más ligadas al comercio y la inversión real, son, por el contrario, una fracción mínima (el reparto por agentes se detalla en el apartado 1.3).

Regulación y funcionamiento:

▪ *Mercado OTC*: se trata de un mercado no organizado u *over-the-counter* (OTC), descentralizado, sin cámara de compensación y con contratos a medida[9].

▪ *Mercado continuo*: la negociación no se interrumpe en todo el día porque se desplaza con los husos horarios. Empieza en Asia (Tokio y Hong Kong), pasa por Oriente Medio (Baréin) a Europa (Fráncfort y Londres) y acaba en América (Nueva York y San Francisco), donde enlaza con la apertura asiática siguiente. Como las sesiones se solapan, el arbitraje entre plazas mantiene una única cotización.

[8] En el informe trienal del BIS las cuotas son el doble porque cada operación se computa una vez por cada una de sus dos monedas, de modo que suman el 200 %.

[9] Ser OTC no lo convierte en un mercado opaco en sus precios, porque gracias a la tecnología todos los participantes acceden a cotizaciones en tiempo real.

Infraestructura y tecnología: la innovación financiera ha automatizado casi todo el ciclo de una operación, que descansa en tres tipos de sistemas:

- *Sistemas de información*: difunden en tiempo real los precios y las noticias que mueven el mercado. Las plataformas más extendidas son LSEG Workspace (heredera de Refinitiv Eikon, antes Reuters) y Bloomberg.

- *Sistemas de casación automática*: emparejan electrónicamente las órdenes de compra y de venta de cualquier plaza sin intervención de un operador. La negociación electrónica interbancaria al contado se concentra en EBS (integrado hoy en CME Group) y en LSEG FX Matching (antiguo Reuters Matching).

- *Sistemas de compensación y liquidación*: en algunos mercados y para parte de las operaciones, los sistemas de pago y compensación, como CHIPS en EE. UU. o T2 (antes TARGET2) en la zona del euro, limitan el riesgo de contrapartida. Específicamente para las divisas, el sistema CLS (Continuous Linked Settlement, 2002) elimina el riesgo de liquidación (el llamado *riesgo Herstatt*) mediante el mecanismo de pago contra pago.

1.3. Agentes que participan en el mercado de divisas

Históricamente, los participantes en el mercado de divisas eran comerciantes de bienes y servicios. Sin embargo, la liberalización de los mercados de capitales del s. XX y la innovación financiera han dado lugar a la participación de un número mucho mayor de agentes.

1.3.1. Creadores de mercado (*market makers* o *dealers*)

Los **creadores de mercado** (*market makers* o *dealers*) son típicamente grandes bancos comerciales e instituciones financieras encargadas de proveer liquidez a los mercados[10]. Dicho de otro modo, garantizan un volumen de intercambios suficiente para que los demás participantes operen de manera continuada, y actúan por ello como mayoristas.

Su remuneración procede de las comisiones y del diferencial (*spread*) entre el precio al que el creador de mercado compra (*bid*) y el precio al que vende (*ask*).

La encuesta trienal del BIS recoge los datos de más de 1100 *dealers* declarantes de 52 jurisdicciones, pero la creación de mercado está en manos de unos pocos grandes bancos, entre ellos BNP Paribas, Deutsche Bank o UBS. Según la encuesta trienal del BIS de 2025, los *dealers* declarantes representan en torno al 46 % del volumen de negocio, algo menos que el 47 % de 2022 y más que el 38 % de 2019[11].

1.3.2. Corredores o intermediarios de mercado (*brokers*)

Los **intermediarios** o *brokers* ponen en contacto a los creadores de mercado con quienes quieren comprar o vender divisas. Facilitan la liquidez sin proveerla, porque no operan por cuenta propia ni mantienen nunca una **posición abierta**, es decir, una operación con ganancias o pérdidas todavía no realizadas (en una posición cerrada ya se han materializado).

Como no asumen riesgo, a diferencia de los *dealers*, solo ingresan comisiones o, en su lugar, un margen fijo de puntos sobre el *spread* natural del mercado (Anexo A.1). Reciben a la vez las

[10] Cotizan precios de compra y de venta a cualquier contraparte y en todo momento, aunque sin obligación formal de hacerlo, porque dejar de proveer liquidez tendría para ellos elevados costes reputacionales.

[11] Fuente: BIS (2025), *Triennial Central Bank Survey*, https://www.bis.org/statistics/rpfx25_fx.htm

cotizaciones de compra (*bid*) y de venta (*ask*) de muchos *dealers* y trasladan a cada cliente la más favorable, de modo que también los inversores pequeños operan a precios que por su cuenta no obtendrían.

1.3.3. Resto de agentes: instituciones financieras no creadoras y agentes no financieros

Tradicionalmente, los agentes privados que acudían al mercado de divisas en busca de liquidez eran las **instituciones financieras no creadoras de mercado** y las **sociedades no financieras**. No obstante, en los últimos años se ha incorporado un elevado número de operadores minoristas y de empresas de negociación algorítmica por cuenta propia (*proprietary trading firms*, PTF).

Instituciones financieras: las no creadoras de mercado concentran en torno al 50 % de las operaciones en 2025 (frente al 55 % de 2019). Su composición es diversa, pues en el lado privado figuran bancos comerciales y de inversión, fondos de pensiones, aseguradoras y *hedge funds*, y en el oficial, bancos centrales y fondos soberanos.

Las instituciones privadas operan en el mercado de divisas para invertir en el exterior y diversificar sus carteras, para cubrir el riesgo de cambio de esas inversiones y, en el caso de los *hedge funds*, para tomar posiciones especulativas.

Las instituciones oficiales apenas suman en torno al 1 % de las operaciones. Aun así, los bancos centrales influyen en el mercado por sus funciones de regulación, supervisión y señalización. Sus compras y ventas solo alcanzan un volumen relevante en episodios puntuales, como las intervenciones de Japón en 2010-2011, 2022, 2024 y 2026. Al estar ligada sobre todo a la política cambiaria, su actuación varía según el régimen cambiario.

Agentes no financieros: incluyen, sobre todo, a las sociedades no financieras y a los Gobiernos, así como a los inversores minoristas o particulares, y representan en torno al 5 % restante de las operaciones. Los inversores minoristas, históricamente irrelevantes, han ganado peso con las plataformas de agregación minorista (*retail aggregators*).

1.4. Conclusiones

La divisa sirve de medio de pago en las transacciones internacionales y su convertibilidad va de la plena a la nula. Su precio, el tipo de cambio, admite cotizaciones nominales, reales y efectivas. El mercado en que se negocia, el mayor del mundo, es OTC, continuo y muy informatizado, y se concentra en el dólar, en el Reino Unido y en las operaciones entre instituciones financieras. La concentración alcanza también a los agentes, entre los que unos pocos grandes bancos actúan como creadores de mercado, que proveen liquidez y asumen riesgo a cambio del *spread*, mientras que los *brokers* median sin tomar posiciones abiertas. Las demás instituciones financieras privadas acuden al mercado para invertir, cubrirse o especular. Los bancos centrales, en cambio, influyen por sus funciones de regulación y señalización, con un volumen de operaciones pequeño.

Entendido qué se negocia, dónde y quién lo hace, la sección 2 estudia primero los instrumentos, que son las herramientas de las operaciones, y después las operaciones mismas.

2. Instrumentos y operaciones

La sección presenta los instrumentos, del contado a las opciones, y las operaciones de obtención de liquidez, cobertura, arbitraje, especulación e intervención oficial. Los instrumentos son esencialmente contratos entre las partes. Desde el punto de vista histórico, se distinguen los instrumentos tradicionales (el contado y el plazo, tan antiguos como el propio comercio internacional) y los instrumentos fruto de la innovación financiera desarrollada desde la década de los setenta, que operó en dos vertientes:

Innovación de proceso: las telecomunicaciones y la informática abarataron y aceleraron las operaciones, y con ello integraron las distintas plazas en un mercado mundial.

Innovación de producto: la libertad de movimientos de capitales trajo necesidades nuevas de cobertura y de financiación, a las que respondieron instrumentos como los *swaps* y las opciones.

En todos los casos, los porcentajes de volumen se refieren a la cuota de cada instrumento sobre el volumen medio diario negociado en el mercado mundial de divisas[12].

2.1. Instrumentos

En el mercado se negocian cinco instrumentos, de la compraventa al contado a los derivados. Son los contratos *spot*, los de futuros y a plazo, los *swaps* de tipos de cambio, los *swaps* de divisas y las opciones.

> El concepto, el mecanismo, la valoración y las estrategias de futuros, *forwards*, opciones y *swaps* se desarrollan en el tema B25, secciones 1 a 3. Aquí se recoge solo lo propio del mercado de divisas: el peso de cada instrumento, su uso en las operaciones con divisas y los FX *swaps*, instrumento característico de este mercado que el tema B25 solo resume.

2.1.1. Contratos *spot* (31 % del volumen)

El cambio de monedas al contado es tan antiguo como el comercio entre comunidades monetarias distintas, de los cambistas medievales a la banca moderna. Sin embargo, el mercado *spot* contemporáneo, interbancario y continuo, se articuló con el telégrafo. Así, el primer cable trasatlántico duradero (1866) permitió al tipo de cambio GBP/USD cotizar simultáneamente en Londres y Nueva York, y de ahí que el par libra-dólar se siga denominando *cable*.

Definición: los contratos *spot* o al contado son acuerdos de compraventa entre dos partes para su ejecución inmediata. Pactada la operación en t, la liquidación (*settlement*), es decir, el pago y la entrega de las divisas, se produce en la fecha *spot*, por lo general en $t+2$ como máximo.

Tipos: el contado no tiene modalidades propias y solo se distingue por la fecha valor, que puede ser el mismo día de la contratación (t), el día hábil siguiente ($t+1$) o la fecha *spot* estándar ($t+2$). La encuesta del BIS computa las tres como contado[13].

[12] Salvo indicación en contrario, las cuotas son las de abril de 2025, según la encuesta trienal del BIS (*Triennial Central Bank Survey*, 2025), que no incluye los derivados negociados en mercados organizados. Fue un mes de volatilidad excepcional por el episodio arancelario estadounidense.

[13] Fuente: BIS (2022), *Triennial Central Bank Survey of Foreign Exchange and OTC Derivatives Markets. Reporting guidelines for turnover in April 2022*, apartado 5.1, https://www.bis.org/statistics/triennialrep/2022survey_guidelinesturnover.pdf. En cambio, las operaciones a muy corto plazo del tipo *tomorrow/next*, *overnight* o *spot/next* no se computan como contado, sino como FX *swaps* (apartado 2.1.3), y tampoco las patas *spot* de estos.

Funcionamiento: son los contratos de compraventa de divisas más sencillos y se negocian en el segmento *spot* del mercado. En el segmento interbancario, la negociación electrónica se concentra en los sistemas de casación automática (EBS y LSEG FX Matching), y la liquidación puede hacerse por pago contra pago en CLS (apartado 1.2). Los contratos *spot* representan en torno al 31 % del volumen y se utilizan sobre todo en operaciones comerciales y de obtención de liquidez.

Conclusiones:

- *Aspectos positivos*: su sencillez y la entrega inmediata los convierten en el instrumento natural para disponer de la divisa en el momento en que se necesita.

- *Limitaciones*: no protegen frente a la variación del tipo de cambio de los cobros y pagos aplazados, cuya cobertura exige el segmento *forward*. Si las dos monedas no se entregan a la vez, los contratos *spot* exponen además al riesgo de liquidación (riesgo Herstatt), que elimina CLS.

2.1.2. Contratos de futuros (*futures*) y contratos a plazo (*forward*) (19 % del volumen)

Los acuerdos de entrega futura a precio pactado son antiquísimos, y sus antecedentes como mercados organizados (Ámsterdam en el s. XVII, Dōjima en 1730) se recogen en el tema B25, apartado 1.1.1. Para el mercado de divisas, el hito es 1972, cuando el CME (Chicago Mercantile Exchange) creó el IMM (International Monetary Market), cuyos futuros sobre divisas fueron los primeros futuros financieros.

Definición: a diferencia de los contratos *spot*, los futuros y los *forwards* fijan hoy el precio y la obligación de una compraventa de divisas en una fecha futura.

Tipos: según el mercado en que se negocian, se distinguen dos, cuyo concepto, mecanismo y diferencias se desarrollan en el tema B25, apartados 1.1.1 a 1.1.3:

- *Futuros*: contratos estandarizados que se negocian en mercados organizados con cámara de compensación.

- *Forwards*: contratos bilaterales y a medida que se negocian en mercados OTC.

Funcionamiento: el tipo de cambio pactado, el tipo *forward*, refleja el diferencial de intereses entre las dos monedas (paridad cubierta de intereses). Dado el carácter OTC del mercado de divisas, en él domina el *forward*, con en torno al 19 % del volumen que mide la encuesta del BIS. Los futuros sobre divisas, negociados en mercados organizados (p. ej. el CME) y ajenos a esa encuesta, suponen, en cambio, una fracción marginal. Los *forwards* se utilizan sobre todo para cobertura y especulación.

Conclusiones:

- *Aspectos positivos*: fijan hoy el tipo de cambio de un cobro o pago futuro y eliminan el riesgo de cambio con más eficacia que las opciones, sin exigir prima.

- *Limitaciones*: obligan a liquidar al tipo pactado aunque el tipo de cambio evolucione favorablemente. Los *forwards*, bilaterales y sin cámara de compensación, conservan además el riesgo de contrapartida, y los futuros exigen márgenes y liquidación diaria de pérdidas.

2.1.3. Contratos de *swaps* de tipos de cambio (Forex o FX *swaps*) (42 % del volumen)

Los FX *swaps* se generalizaron como instrumento interbancario de gestión de la liquidez y la financiación en divisas tras el colapso de Bretton Woods y la generalización de la flotación (1973), y su uso se multiplicó con la globalización financiera de las décadas siguientes.

Definición: los *swaps de tipos de cambio* son contratos de permuta entre dos partes con dos patas de intercambio acordadas simultáneamente:

- *Spot* o *near leg*: una parte compra a la otra, o le vende, una cantidad de divisas al tipo *spot* del día.

- *Forward* o *far leg*: en la fecha pactada, las partes deshacen ese intercambio por la misma cantidad al tipo *forward* que fijaron al contratar.

Tipos: según la fecha de la pata cercana, la encuesta del BIS distingue los *swaps spot/forward*, con esa pata al contado, y los *forward/forward*, con las dos patas a plazo. Incluye también los *swaps* a muy corto plazo (*overnight*, *tomorrow/next*, *spot/next*), con los que se renuevan de un día para otro las posiciones de contado[14].

Funcionamiento: económicamente, un FX *swap* equivale a un préstamo garantizado, en el que cada parte presta temporalmente su moneda y recibe la otra como colateral. Con en torno al 42 % del volumen, es el instrumento más negociado del mercado y se utiliza, generalmente, en operaciones de cobertura, financiación y gestión de la liquidez en divisas. No debe confundirse con los *swaps* de divisas o *currency swaps*.

> **Ejemplo: un FX *swap* a seis meses**
>
> Scotiabank (un banco canadiense) necesita euros y un banco de la zona del euro necesita dólares canadienses. Ambos acuerdan usar un *swap* de tipo de cambio con vencimiento a seis meses. La pata *spot* se fija en 1 M€ por 1,25 M$CA, es decir, CAD/EUR = 0,8. La pata *forward* se fija en 0,79 CAD/EUR, tipo *forward* que refleja el diferencial de intereses entre ambas monedas (paridad cubierta de intereses).
> Así, hoy, Scotiabank recibe 1 M€ del banco de la zona del euro y le entrega 1,25 M$CA y, en seis meses, Scotiabank devuelve 1 M€ y recibe 1,27 M$CA (1 M€ · (1/0,79) = 1,2658 M$CA), lo cual le permite cubrir el riesgo de cambio.

Conclusiones:

- *Aspectos positivos*: como préstamo garantizado, apenas presenta riesgo de crédito, y combina en una sola operación la cobertura y la obtención de liquidez temporal en la otra divisa.

- *Limitaciones*: subsisten el riesgo de reposición, si la contraparte incumple antes de la pata *forward* y hay que rehacer la operación a los precios del momento, y el de liquidación, hoy mitigado por sistemas como CLS.

[14] Fuente: BIS (2022), *Triennial Central Bank Survey of Foreign Exchange and OTC Derivatives Markets. Reporting guidelines for turnover in April 2022*, apartado 5.1, https://www.bis.org/statistics/triennialrep/2022survey_guidelinesturnover.pdf.

2.1.4. Contratos de *swaps* de divisas (*currency swaps*) (2 % del volumen)

El antecedente de los *swaps* de divisas son los préstamos paralelos (*parallel loans*) con los que las empresas sorteaban los controles de cambios británicos en los años setenta. El primer *currency swap* moderno, entre IBM y el Banco Mundial en 1981, se describe en el tema B25, apartado 3.1.

Definición: los *swaps de divisas* son contratos de permuta en los que las partes se pagan intereses en monedas distintas durante el plazo pactado y, por lo general, intercambian los principales al vencimiento a un tipo de cambio fijado de antemano. A diferencia de los FX *swaps*, en los que el diferencial de intereses queda incorporado en el tipo de la pata *forward*, en ellos sí hay pagos periódicos de intereses.

Tipos: como el interés de cada moneda puede ser fijo o variable, caben tres combinaciones, fijo contra fijo, fijo contra variable y variable contra variable, que el tema B25 recoge en el apartado 3.1.1.

Funcionamiento: su uso típico es el de una empresa que se endeuda en su moneda y permuta los flujos para financiarse en otra. El tema B25 ilustra ese caso en el apartado 3.1 y desarrolla su valoración en el 3.2. En comparación con los FX *swaps*, los *swaps* de divisas, de mayor plazo, representan apenas en torno al 2 % del volumen negociado.

Conclusiones:

- *Aspectos positivos*: como los FX *swaps*, eliminan el riesgo de cambio, en su caso el de una financiación a más largo plazo en otra moneda, pues el tipo de cambio al que se devuelven los principales se fija de antemano.

- *Limitaciones*: a diferencia de los FX *swaps*, presentan un riesgo de impago apreciable.

2.1.5. Contratos de opciones (7 % del volumen)

Las primeras opciones sobre divisas cotizadas se listaron en la Bolsa de Filadelfia (PHLX) en 1982, y el mercado OTC se desarrolló durante esa misma década, apoyado en la adaptación del modelo de Black-Scholes a las divisas realizada por Garman y Kohlhagen (1983).

Definición: una opción sobre divisas da a quien la compra el derecho, sin obligación alguna, de comprar o de vender una divisa a un tipo de cambio pactado, el precio de ejercicio (*strike price*), en una fecha determinada o hasta ella (vencimiento).

Tipos: según el derecho que otorgan, se distinguen dos, a los que se suman modalidades más complejas (sus tipos, su mecanismo y su valoración se desarrollan en el tema B25, apartados 2.1 y 2.2):

- *Call* (opción de compra): da derecho a comprar la divisa al precio de ejercicio.

- *Put* (opción de venta): da derecho a venderla al precio de ejercicio.

- *Otras opciones*: las *average rate options*, las *barrier options*, los *swaptions* de divisas y los *currency warrants*.

Funcionamiento: a cambio del derecho, el comprador paga una prima y ejerce la opción solo si el tipo de cambio evoluciona a su favor. Las opciones representan en torno al 7 % del volumen (incluidas las modalidades más complejas), frente al 4 % de 2022. Su volumen se ha más que duplicado en el último trienio, impulsado por la volatilidad cambiaria. Como los *forwards*, se utilizan sobre todo para cobertura y especulación.

Conclusiones:

- *Aspectos positivos*: protegen frente a una evolución desfavorable del tipo de cambio sin renunciar a la favorable, pues no obligan a ejercer.

- *Limitaciones*: exigen el pago de una prima y cubren el riesgo con menos eficacia que los *forwards* y los futuros, porque generan ratios de cobertura volátiles.

2.2. Operaciones

Vistos los instrumentos, queda por estudiar qué operaciones realizan con ellos los agentes. Las operaciones se clasifican según dos criterios.

Según la fecha de entrega de la divisa:

- *Operaciones al contado o spot*: con entrega en un máximo de dos días hábiles (*t+2*), se realizan en el segmento *spot* del mercado apoyándose en los contratos al contado.

- *Operaciones a plazo o forward*: con entrega posterior a *t+2*, se realizan en el segmento *forward* mediante *forwards*, *swaps* y opciones, y concentran el resto del volumen. Su peso se ha mantenido relativamente estable, en torno a dos tercios del volumen, en las dos últimas décadas.

Según su fin: se distinguen cinco categorías, que estructuran el resto de la sección:

- *Obtención o disposición de liquidez*: ligadas a la actividad comercial y de inversión internacional.

- *Cobertura*: cierran posiciones abiertas para eliminar el riesgo cambiario.

- *Arbitraje*: explotan sin riesgo los desajustes transitorios de precios y rentabilidades.

- *Especulativas*: toman posiciones abiertas sobre la base de expectativas, a cambio de asumir riesgo.

- *Oficiales*: mediante ellas, las autoridades monetarias intervienen en el mercado.

2.2.1. Operaciones de obtención o disposición de liquidez

Definición: las *operaciones de obtención o disposición de liquidez* son las más elementales del mercado. Son compraventas de divisa cuyo fin, transaccional y no financiero, es disponer de la moneda extranjera necesaria para atender pagos (o convertir a moneda nacional los cobros recibidos en divisa). Constituyen la función original del mercado de divisas, la que lo conecta con la economía real.

Tipos: según su motivo, se distinguen dos:

- *Actividad comercial*: la necesidad de divisa por parte de una sociedad no financiera en el curso de su actividad comercial internacional.

▪ *Repatriación*: la necesidad de repatriar fondos denominados en divisa, tanto por parte de sociedades no financieras como financieras.

Funcionamiento: estas operaciones persiguen obtener liquidez inmediata en divisa en el segmento *spot* del mercado, donde se realizan las compraventas. La dinámica entre los agentes de este segmento se ilustra con un modelo sencillo de balanza de pagos en el plano (M^*, E), con la cantidad de divisas en el eje horizontal y el tipo de cambio directo en el vertical (figura B13.1):

1. *Demanda de divisas*, $D(M^*)$: la generan las importaciones de bienes y servicios y las inversiones en el exterior (p. ej. inversión directa), que exigen disponer de divisa. Tiene pendiente negativa en el plano (M^*, E) porque una depreciación de la moneda nacional ($\uparrow E$) encarece lo importado, reduce las compras al exterior y, con ellas, las divisas que se piden.

2. *Oferta de divisas*, $S(M^*)$: la generan las exportaciones y las entradas de inversión extranjera, cuyos perceptores quieren convertir en moneda nacional la divisa recibida. Tiene pendiente positiva si la demanda extranjera de exportaciones es elástica, porque entonces una depreciación ($\uparrow E$) eleva el volumen exportado en mayor proporción de lo que baja su precio en divisas, y crecen los ingresos en divisa que se ofrecen.

3. *Equilibrio*: en el punto 0, donde se cortan ambas curvas, la interacción de las dos fuerzas fija el tipo de cambio, E_0, y el volumen de divisas intercambiado, M_0^*. Si el equilibrio es estable[15], a un tipo de cambio superior a E_0 la oferta de divisas excede a la demanda y la moneda nacional se aprecia hasta volver al punto 0, mientras que a un tipo inferior el exceso de demanda la deprecia.

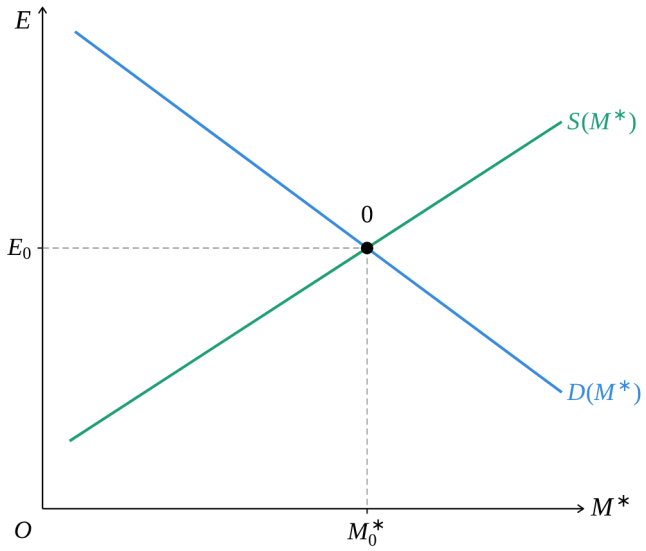

Figura B13.1: Equilibrio en el mercado de divisas. La oferta y la demanda de divisas determinan el tipo de cambio nominal E_0 y la cantidad de divisas intercambiada M_0^*.

Conclusiones:

[15] De acuerdo con la tradición marshalliana, este será estable (cualquier desajuste de la balanza de pagos conducirá de nuevo a él) siempre que se cumpla la condición de Marshall-Lerner.

▪ *Aspectos positivos*: este tipo de operaciones son sencillas y permiten obtener la liquidez en el momento necesario.

▪ *Limitaciones*: muchos flujos comerciales internacionales suponen un intercambio aplazado y se puede incurrir en un riesgo de cambio. Por ejemplo, un exportador europeo que cierra hoy una venta por 1 M de dólares, que cobrará dentro de un año, queda expuesto hasta entonces a las fluctuaciones del tipo de cambio dólar-euro. Ese riesgo puede cubrirse con los instrumentos del segmento *forward*.

2.2.2. Operaciones de cobertura

Definición: las *operaciones de cobertura* buscan mitigar el riesgo cambiario adoptando una posición contraria a la posición abierta. Para ello, el agente acude al segmento *forward* y lleva a cabo una cobertura mediante alguno de los instrumentos anteriores[16].

Tipos: según el instrumento, se distinguen tres:

▪ *Con contratos a plazo*: lo más habitual es utilizar un *forward* de divisas[17]. En el ejemplo del exportador, cabe distinguir dos posiciones:

- *Exportador*: puede vender un *forward* para asegurarse un precio de venta fijo (el tipo *forward*) dentro de un año. En el plano tipo de cambio *spot*-beneficio, la función de beneficios de la posición *forward* es decreciente en el tipo de cambio *spot*, es decir, la posición pierde con la depreciación del euro ($\uparrow E$).

- *Importador*: compra, en cambio, un *forward* para fijar de antemano su precio de compra. La función de beneficios de su posición es creciente en el tipo de cambio *spot*, porque una depreciación del euro encarece la divisa que debe comprar y revaloriza el compromiso de comprarla al tipo pactado.

▪ *Con* swaps *de tipos de cambio*: los FX *swaps* combinan en una sola operación una pata *spot* y una pata *forward* de signo contrario. Se emplean cuando, además de cubrirse, el agente necesita liquidez temporal en la otra divisa o quiere desplazar en el tiempo (*roll over*) una cobertura ya contratada.

▪ *Con opciones*: la estrategia básica consiste en adoptar una posición protectora frente a la posición primaria. Así, el importador compra una *call* para cubrir el riesgo de apreciación de la divisa, es decir, de depreciación de la moneda nacional, y el exportador, una *put* para cubrir el de depreciación de la divisa. Las estrategias más elaboradas, complejas o de diferenciales, se desarrollan en el tema B25, apartado 2.3.

Funcionamiento: las operaciones de cobertura son juegos de suma cero, pues la ganancia del agente que se cubre es la pérdida de su contraparte (véase el tema B25, apartado 1.1.2). La cobertura perfecta exige que la cantidad, la fecha y el par de divisas sean idénticos, para que la posición en el segmento *forward* "case" perfectamente con la posición *spot*.

Conclusiones:

[16] Otras vías de cobertura son abrir posiciones contrarias en pares de divisas correlacionados o mantener cuentas en distintas divisas, en el propio país o en el extranjero.

[17] En la jerga, el exportador, que recibirá divisas, "está largo" de divisa en el mercado *spot* y se cubre abriendo "un corto" en el mercado *forward*.

▪ *Aspectos positivos*: permiten eliminar el riesgo de cambio y reducen la incertidumbre en cuanto a los flujos de caja futuros. Todo ello, a su vez, facilita el acceso al crédito en los mercados financieros y puede ayudar a la planificación empresarial. Las operaciones de cobertura han aumentado desde la crisis financiera.

▪ *Limitaciones*: son operaciones más sofisticadas que pueden requerir más conocimientos técnicos y, en algunos casos, hacer frente a flujos de caja negativos importantes (p. ej. en el caso de los *forwards*).

2.2.3. Operaciones de arbitraje o arbitraje *puro*: arbitraje espacial y paridad cubierta de intereses

Definición: las *operaciones de arbitraje* usan los mismos instrumentos para buscar beneficios. Son estrategias que aprovechan los desajustes transitorios en el valor de las divisas, dado por su tipo de cambio o por su rentabilidad. En sentido estricto, el arbitraje es sin riesgo (y autofinanciado), y se denomina frecuentemente *arbitraje puro*.

Tipos: el arbitraje puro, el arbitraje propiamente dicho, puede ser espacial, si explota diferencias de tipo de cambio entre plazas, o temporal, si explota diferencias de rentabilidad. En cambio, la expresión "arbitraje con riesgo", heredada de la literatura clásica, designa en realidad una operación especulativa, pues quien la realiza toma una posición abierta sobre la base de sus expectativas. Por ello se estudia dentro de las operaciones especulativas, donde el término se conserva por su valor didáctico como puente entre ambas categorías.

Arbitraje espacial sin riesgo o arbitraje puro espacial: consiste en comprar una divisa en la plaza donde cotiza más barata y venderla en el mismo instante en otra donde cotiza más cara. Como las dos operaciones son simultáneas, en ningún momento queda una posición abierta ni se asume riesgo de cambio. La estrategia se limita al segmento *spot* y solo atiende al tipo de cambio nominal[18].

▪ *Supuestos*: el único relevante es la ausencia de costes de transacción. En su presencia, el arbitraje solo se activa cuando el diferencial de cotización supera dichos costes (una banda de no arbitraje en lugar de una igualdad exacta).

▪ *Funcionamiento*: el arbitraje espacial se lleva a cabo de dos formas, con condiciones ligeramente diferentes, y en ambas garantiza el cumplimiento de la ley del precio único en el plano espacial:

- *Bilateral*: la operación implica dos plazas[19], dos monedas y su tipo de cambio. Existe oportunidad de arbitraje si $E^A \neq E^B$, es decir, si el tipo de cambio entre las monedas M^* y M en las plazas A y B no coincide. Su explotación conduce al equilibrio $E^A = E^B$, en el que la misma divisa cotiza igual en todas las plazas.

- *Triangular*: la operación implica tres plazas[20], tres monedas y el tipo de cambio cruzado. Existe oportunidad si $E^A_{M'/M} \neq E^B_{M^*/M} \cdot E^C_{M'/M^*}$, y el equilibrio impone la coherencia de los tipos cruzados, $E^A_{M'/M} = E^B_{M^*/M} \cdot E^C_{M'/M^*}$, con la que desaparecen las oportunidades de arbitraje.

[18] Se omiten otros factores que podrían influir en la decisión (liquidez, riesgo de crédito, inflación o tipo de interés). El tipo de interés se incorpora en el arbitraje temporal.

[19] *Ejemplo*: si el dólar cotiza a 0,93 euros en Fráncfort y a 0,90 en Nueva York (tipo continental), quien compre dólares en Nueva York y los venda a la vez en Fráncfort gana 0,03 euros por dólar sin riesgo. Esas compras encarecen el dólar en Nueva York y las ventas lo abaratan en Fráncfort hasta igualar ambas cotizaciones.

[20] *Ejemplo*: sean $E^{\text{FRA}}_{€/L} = 0{,}95$ (euros por libra en Fráncfort), $E^{\text{TK}}_{€/¥} = 0{,}0077$ (euros por yen en Tokio) y $E^{\text{NY}}_{¥/L} = 121$ (yenes por libra en Nueva York). Si $0{,}95 \neq 0{,}0077 \cdot 121 = 0{,}9317$, existe oportunidad de arbitraje triangular. Solo si $0{,}95 = 0{,}0077 \cdot 123{,}38$ desaparece dicha oportunidad.

▪ *Conclusiones*:

- *Aspectos positivos*: el arbitraje espacial hace que la ley del precio único se cumpla casi al instante. De hecho, con el desarrollo de las operaciones de alta frecuencia[21] y la negociación algorítmica, en el mundo real las posibilidades de llevar a cabo este tipo de arbitraje duran apenas milisegundos[22].

- *Limitaciones*: con costes de transacción, la igualdad de cotizaciones solo se cumple dentro de una banda de no arbitraje, y la brevedad de las oportunidades las reserva a los operadores con sistemas de negociación algorítmica, capaces de explotar ventajas de latencia.

Arbitraje temporal sin riesgo o arbitraje puro temporal: la condición de paridad cubierta de intereses (PCI): el tipo de cambio no agota el valor de una divisa. En sentido amplio, mantener riqueza en una moneda rinde además el tipo de interés de los depósitos denominados en ella. Puede compararse, entonces, la rentabilidad total de invertir en una u otra moneda, formada por el tipo de interés que cada una paga más la variación cambiaria entre el momento de la inversión y el del vencimiento. El arbitraje temporal aprovecha las diferencias entre ambas rentabilidades y elimina el riesgo de cambio mediante la contratación de un *forward*. Las operaciones que las explotan conducen a la paridad cubierta de intereses.

▪ *Supuestos*: agentes racionales (no es preciso suponer neutralidad al riesgo, pues la operación cubierta no entraña riesgo cambiario), que tienen en cuenta el tipo de cambio nominal y el tipo de interés nominal de los depósitos con vencimiento a un año. Se supone también perfecta movilidad de capitales (sin controles ni costes de transacción), cobertura *forward* total[23] y ausencia de endeudamiento o apalancamiento despreciable.

▪ *Funcionamiento*: un residente dispone de 1 u.m. nacional (p. ej. un euro) para colocarla un año en un depósito y puede elegir entre dos alternativas:

- *Activo nacional*: el depósito en moneda nacional le devuelve al cabo del año $r_{€}^{A} = 1 \cdot (1+i)$ u.m. nacionales brutas.

- *Activo extranjero*: la colocación en el exterior exige tres pasos:

  1. *Cambio*: hoy cambia su u.m. nacional al tipo *spot* vigente y recibe $1/E_t^s$ u.m. extranjeras.

  2. *Inversión*: deposita esa cantidad en el extranjero y al cabo del año dispone de $r_{\$}^{B} = (1/E_t^s)(1+i^*)$ u.m. extranjeras.

  3. *Conversión*: pasa ese importe a moneda nacional y obtiene $r_{€}^{B} = (E_{t+1}^s/E_t^s)(1+i^*)$.

▪ *Diferencial de rentabilidades*: *ex post*, vendrá dado por $r_{€} \equiv r_{€}^{B} - r_{€}^{A} \rightarrow r_{€} = (E_{t+1}^s/E_t^s)(1+i^*) - (1+i)$ (1). Si la rentabilidad bruta del depósito extranjero en moneda nacional es superior a la del depósito nacional, el diferencial es positivo y existen ganancias del arbitraje para el residente.

[21] El año 2003 fue decisivo para la popularización de este tipo de operativa, cuando EBS comenzó a ofrecer una interfaz plenamente automatizada a los bancos.

[22] El aprovechamiento del retraso entre el envío de una señal de compraventa por un servidor y su recepción por otro se conoce como *arbitraje por latencia*.

[23] En la práctica, la cobertura nunca es perfecta, aunque en mercados profundos y líquidos se aproxima mucho.

▪ *Aproximación*: la expresión se reduce[24] a $r_{€} \approx (E^s_{t+1} - E^s_t)/E^s_t + (i^* - i) \rightarrow r_{€} \approx g_E + (i^* - i)$ (2), donde g_E es la tasa de depreciación observada de la moneda nacional (tasa de depreciación *ex post*).

▪ *Cobertura forward*: ahora bien, en el momento de invertir (t), el agente no conoce E^s_{t+1}. Si no quiere asumir riesgo cambiario, contratará hoy un *forward*[25] que le permita convertir su retorno futuro a un tipo predeterminado, el tipo de cambio *forward* o a plazo, E^f_t, obteniendo así una ganancia con cobertura total.

▪ *Margen forward*: en estas condiciones, la tasa de depreciación en (2) pasa a ser el margen *forward*, $\mu_f = (E^f_t - E^s_t)/E^s_t$, que indica el porcentaje en que el precio *forward* supera o está por debajo del precio *spot*. Así, si $\mu_f > 0$ la moneda nacional cotiza al descuento en el mercado *forward*[26], si $\mu_f < 0$ cotiza a premio y si $\mu_f = 0$ cotiza a la par. Existen, pues, oportunidades de arbitraje temporal sin riesgo cuando el diferencial de tipos de interés, $(i^* - i)$, difiere del margen *forward* con signo contrario ($r_{€} \neq 0$).

▪ *Ajuste hacia el equilibrio*: la explotación de estas ganancias se ilustra con dos casos, partiendo de una situación en la que ambas monedas cotizan a la par ($E^s_t = E^f_t$, $\mu_f = 0$):

- *Depósito extranjero atractivo* ($i < i^* + \mu_f \rightarrow \Delta\mu_f < 0$): con un interés extranjero mayor que el nacional, a los residentes les conviene comprar hoy divisa y depositarla en el exterior. El rendimiento llegará en divisa, pero los residentes no necesitan esperar al tipo *spot* futuro para convertirlo, porque pueden fijar hoy su conversión con un *forward*. Las compras de divisa al contado deprecian la moneda nacional ($\uparrow E^s_t$) y reducen el margen *forward* ($\downarrow \mu_f$) hasta que desaparece el arbitraje sin riesgo ($r_{€} = 0$).

- *Depósito nacional atractivo* ($i > i^* + \mu_f \rightarrow \Delta\mu_f > 0$): con un interés nacional mayor, son los no residentes quienes compran hoy moneda nacional y aseguran con un *forward* la conversión a divisa de su rendimiento, en un razonamiento simétrico al anterior. La moneda nacional se aprecia al contado ($\downarrow E^s_t$) y el margen *forward* sube ($\uparrow \mu_f$) hasta que el diferencial de rentabilidades se anula ($r_{€} = 0$).

▪ *Equilibrio*: en el límite, en ambos casos el diferencial de rentabilidades es nulo y se cumple la paridad cubierta de intereses, $r_{€} = 0 \leftrightarrow i = i^* + \mu_f$. Con ella, el arbitraje sin riesgo garantiza que se cumple la ley del precio único en el mercado de divisas, ahora en el plano temporal[27].

▪ *Conclusiones*:

- *Aspectos positivos*: la evidencia empírica respalda la paridad cubierta, cuyo cumplimiento avanzó con la liberalización de capitales. Así, el Reino Unido y Japón aún mantenían controles en 1979, y a finales de los años ochenta la liberalización financiera había eliminado casi por completo los diferenciales cubiertos de intereses a corto plazo entre las principales economías industrializadas (Frankel, 1991). Hoy la negociación electrónica permite arbitrar

[24] Escribiendo $E^s_{t+1}/E^s_t = 1 + g_E$ en (1), se tiene $r_{€} = (1 + g_E)(1 + i^*) - (1 + i) = g_E + (i^* - i) + g_E \cdot i^*$, y para variaciones pequeñas el último término, de segundo orden, es despreciable, de donde se obtiene la aproximación $r_{€} \approx g_E + (i^* - i)$.

[25] Nótese que esta operación requiere un contrato *spot* y un contrato *forward*, por lo que ambos podrían combinarse en un FX *swap* o *swap* de tipos de cambio (Mascareñas, 2013, p. 48).

[26] *Ejemplo*: si el par USD/EUR cotiza (tipo continental) a 0,9 en el *spot* y a 0,93 en el *forward*, el margen *forward* es positivo ($\mu_f > 0$). Así, dado que se requieren más euros para adquirir el mismo dólar, el euro cotiza al descuento (con menor poder de compra).

[27] El diferencial de rentabilidades es nulo aunque no lo sean necesariamente el diferencial de intereses ni el margen *forward*. Además, el tipo *forward* se deduce de tres variables observables (Mascareñas, 2013, p. 48), $E^f_t = E^s_t(1 + i)/(1 + i^*)$, lo que ilustra por qué el *forward* es un derivado, pues su precio se deriva del precio del subyacente. La fórmula es el caso cambiario del coste de mantenimiento del tema B25, apartado 1.2.2.

con cobertura *forward* de forma continua y precisa. Por ello, en condiciones normales, una desviación de la paridad desaparece en pocos minutos (Akram, Rime y Sarno, 2008, y tema B14, apartado 1.3.5), salvo las desviaciones persistentes observadas desde 2008.

- *Limitaciones*: antes de 2008, las desviaciones de la PCI se explicaban por los controles de cambio y los costes de transacción. Desde la crisis financiera de 2008, sin embargo, se observan desviaciones persistentes (la llamada base de los *swaps* de divisas o *cross-currency basis*), incluso entre monedas de economías avanzadas sin controles de capital. Se atribuyen a las restricciones de balance de los intermediarios tras la reforma regulatoria. En todo caso, el grado de cumplimiento de la paridad cubierta se sigue usando, hoy en día, como medida del grado de movilidad internacional del capital (más que medidas como la correlación inversión-ahorro o los diferenciales de tipos de interés reales).

2.2.4. Operaciones especulativas: paridad descubierta de intereses, *carry trade*, especulación en el *forward* y *prop trading*

Definición: las *operaciones especulativas* asumen deliberadamente el riesgo de cambio y mantienen posiciones abiertas basadas en las expectativas del agente. En ellas desemboca el arbitraje temporal cuando no hay instrumentos *forward* (como en muchos países en desarrollo) o el agente decide asumir ese riesgo.

Tipos: se distinguen cuatro operaciones, que se presentan por orden de relevancia analítica. La primera es el arbitraje con riesgo, que da lugar a la paridad descubierta de intereses, central en la teoría del tipo de cambio. Siguen el *carry trade*, su explotación práctica más importante, la especulación con derivados y, por último, el *trading* propietario.

Arbitraje con riesgo: condición de paridad descubierta de intereses (PDI): consiste en replicar la lógica del arbitraje temporal sin cobertura *forward*. El agente toma una posición abierta en la divisa que ofrece mayor rentabilidad esperada y asume el riesgo de cambio sobre la base de sus expectativas.

▪ *Supuestos*: los del arbitraje temporal, a los que se añade la neutralidad al riesgo (los agentes no exigen prima por invertir en un sitio o en otro), y las expectativas del agente sobre el tipo de cambio futuro sustituyen a la cobertura *forward*.

▪ *Funcionamiento*: quien no se cubre convierte su rendimiento al tipo *spot* que rija al vencimiento, que desconoce al invertir. Decide, por tanto, con el tipo *spot* esperado para esa fecha, $E_t^e \equiv \mathbb{E}_t[E_{t+1}^s]$, y la tasa de depreciación en (2) pasa a ser la esperada, $g_E^e = (E_t^e - E_t^s)/E_t^s$. Las oportunidades que surgen son de ganancia con riesgo, y su explotación conduce en teoría a la paridad descubierta de intereses por un mecanismo similar al anterior[28].

▪ *Equilibrio*: si se cumple la PDI, $\mathbb{E}_t[r_€] = 0 \leftrightarrow i = i^* + g_E^e$, el diferencial de rentabilidades esperadas es nulo y se han eliminado todas las oportunidades de ganancia con riesgo[29]. Su incumplimiento, en cambio, está estrechamente relacionado con el *puzzle* de Fama, paradoja del margen *forward* o anomalía de la prima *forward* (apartado 3.2). Implica, además, que es posible obtener ganancias asumiendo riesgos, en lo que se basan estrategias como el *carry trade*.

▪ *Conclusiones*:

[28] La diferencia reside en que en el ejemplo anterior ya no subirá o bajará el margen *forward* hasta anular el diferencial de rentabilidades, sino la tasa de depreciación esperada.

[29] Si se cumple la paridad descubierta de intereses, se dispone también de una teoría sobre la determinación del tipo de cambio *spot* (Mascareñas, 2013, p. 52), $E_t^s = E_t^e(1+i^*)/(1+i)$, en la que una de las variables, el tipo de cambio *spot* esperado $E_t^e \equiv \mathbb{E}_t[E_{t+1}^s]$, no es observable.

- *Aspectos positivos*: la paridad descubierta es la condición de equilibrio del mercado de divisas con la que trabajan las teorías de determinación del tipo de cambio (véase el tema B14), lo que la convierte en la más relevante analíticamente de las condiciones de paridad.

- *Limitaciones*: la evidencia empírica muestra que la PDI no se cumple a corto plazo, aunque tiende a cumplirse a horizontes largos (Chinn y Meredith, 2004). Por tanto, existen, por lo general, oportunidades de ganancia con riesgo en los mercados de divisas. El incumplimiento de la PDI puede deberse a tres causas, entre otras:

  - *Dificultad de predicción*: pese a los desarrollos tecnológicos, los agentes anticipan mal la evolución de los tipos de cambio.
  - *Aversión al riesgo*: con agentes aversos, invertir en el exterior solo compensa si la rentabilidad esperada supera la nacional en una prima, $\mathbb{E}_t[r^B_€] = r^A_€ + \rho$.
  - *Expectativas sesgadas*: con datos de encuestas, Froot y Frankel (1989) no hallan prima de riesgo alguna en el sesgo de la prima *forward* y lo atribuyen a errores sistemáticos de expectativas. Esos errores pueden deberse a expectativas sesgadas o, aun siendo estas racionales, a que el suceso que el mercado esperaba no llega a producirse en la muestra, como en el caso de los *peso problems*[30].

Arbitraje con riesgo y apalancamiento: el *carry trade*: consiste en tomar prestado en una divisa de tipo de interés bajo (divisa de financiación, p. ej. el yen) para invertir en activos de otra con tipo alto (divisa de destino, p. ej. el dólar). La estrategia explota de forma apalancada el incumplimiento sistemático de la PDI[31].

▪ *Supuestos*: la estrategia solo es rentable si la divisa de destino no se deprecia (frente a la de financiación) en la cuantía que la PDI predice, lo que exige que esta paridad no se cumpla en el horizonte de la inversión.

▪ *Funcionamiento*: el beneficio de la estrategia es el diferencial de intereses, $i^{\text{destino}} - i^{\text{financiación}}$, menos la depreciación observada de la divisa de destino. Mientras el tipo de cambio permanece estable, la estrategia rinde el *carry*. Sin embargo, su principal riesgo es el desmantelamiento súbito de posiciones (*unwinding*) ante episodios de aversión al riesgo, que provoca movimientos cambiarios bruscos y pérdidas amplificadas por el apalancamiento. En la jerga, el *carry trade* "sube por la escalera y baja en ascensor".

▪ *Conclusiones*:

- *Aspectos positivos*: la estrategia explota de forma sistemática el incumplimiento de la PDI a corto plazo. Por ello se ha aplicado durante décadas con el yen japonés como divisa de financiación, dados los tipos de interés muy reducidos de Japón, para comprar activos en dólares, euros o deuda emergente[32]. Análogamente, es común financiarse en países desarrollados para invertir en deuda de emergentes con rentabilidades más altas.

[30] Este término deriva de la evolución del peso mexicano entre 1955 y 1976. El peso estaba fijado a un tipo de cambio constante con el dólar, pero se vendía al descuento en el mercado a plazo, lo cual sugería que el mercado esperaba una fuerte devaluación, la cual se produjo finalmente en 1976. Esto validó las predicciones del mercado, pero dicha devaluación no se podía predecir utilizando la muestra entre 1955 y 1976 dado que en ese periodo no se había producido ninguna devaluación.

[31] A diferencia del arbitraje cubierto, el *carry trade* no contrata un *forward* (si lo hiciera, la paridad cubierta eliminaría la ganancia).

[32] *Ejemplo*: con un tipo de 108 yenes por dólar y un tipo de interés nulo en Japón, se toman prestados 2.000.000 de yenes, se cambian por 18.519 dólares y se invierten en un bono estadounidense al 2,2 %, que devuelve 18.926 dólares. Si el tipo de cambio no varía, tras devolver el préstamo queda una ganancia de unos 44.000 yenes (2,2 %). Si el yen se aprecia, la ganancia se reduce o se convierte en pérdida.

- *Limitaciones*: el apalancamiento amplifica las pérdidas cuando un episodio de aversión al riesgo obliga a deshacer las posiciones. Ese riesgo se materializó en agosto de 2024. La subida de tipos del Banco de Japón a finales de julio, unida a unos datos de empleo estadounidenses peores de lo esperado, desencadenó un desmantelamiento masivo de posiciones de *carry* en yenes, con una apreciación abrupta de la divisa japonesa y turbulencias bursátiles globales.

Especulación en el segmento *forward*: además de cubrir posiciones, los derivados permiten especular con posiciones abiertas (sin posición contraria en el subyacente asociada a una operación comercial) basadas en las expectativas cambiarias propias.

▪ *Supuestos*: el especulador confía en su expectativa sobre el tipo de cambio futuro y acepta el riesgo de la posición abierta. Su contraparte (p. ej. un agente que se cubre) toma la posición contraria, en un juego de suma cero.

▪ *Funcionamiento*: con expectativas bajistas (*bearish*) sobre la divisa, el especulador vende un *forward* (posición corta), cuyo beneficio es decreciente en el tipo *spot* al vencimiento, porque gana si este queda por debajo del tipo pactado. Con expectativas alcistas (*bullish*), lo compra (posición larga), con un beneficio creciente en el tipo *spot*. Los perfiles de ambas posiciones se representan en el tema B25, apartado 1.1.2.

▪ *Ejemplo*: el caso más conocido de especulación cambiaria en corto es el ataque de George Soros a la libra esterlina en 1992:

- *Contexto*: Soros consideraba que la libra estaba sobrevalorada con respecto al marco alemán, dado que el Gobierno británico, del que dependía entonces el BoE (Bank of England), renunciaba a devaluarla por su pertenencia al Mecanismo de Cambios e Intervención (MCI) del Sistema Monetario Europeo (SME).

- *Primera fase*: Soros tomaba prestado en libras y las vendía para comprar marcos, causando presión depreciatoria sobre la libra y apreciatoria sobre el marco, lo cual intensificaba la divergencia entre ambas monedas.

- *Segunda fase*: Soros anticipaba que el BoE no podría soportar la divergencia, pues defender la paridad exigía mantener tipos de interés muy elevados (en línea con los del Bundesbank tras la reunificación alemana) en plena recesión británica. Con esa expectativa abrió una posición corta sobre la libra por valor de 10.000 M$, a la que siguieron, por comportamiento de rebaño, otras posiciones cortas.

- *Desenlace*: el BoE no pudo aguantar la presión depreciatoria. Así, el 16 de septiembre de 1992, el "miércoles negro", el Reino Unido tuvo que abandonar el MCI del SME.

▪ *Conclusiones*:

- *Aspectos positivos*: el especulador da contrapartida a los agentes que se cubren, pues asume el riesgo cambiario del que estos se desprenden, y con un *forward* puede tomar posición sobre sus propias expectativas sin pagar prima.

- *Limitaciones*: la posición queda abierta a ganancias o pérdidas en un juego de suma cero, y el comportamiento de rebaño puede amplificar la presión sobre una paridad, como en 1992. Por ello, en la práctica, estas operaciones suelen articularse con alguna cobertura parcial mediante otro derivado.

Negociación por cuenta propia (*prop trading*): consiste en la asunción directa de riesgos por parte de las instituciones financieras, que utilizan recursos propios (no de clientes) exclusivamente para operaciones especulativas, incluido el arbitraje con riesgo a gran escala. Esta práctica, también llamada *trading* propietario, abarca los mercados financieros en general, no solo los de divisas.

- *Conclusiones*:

  - *Aspectos positivos*: al incluir arbitraje con riesgo a gran escala, empuja las rentabilidades esperadas hacia la paridad descubierta y aporta liquidez a los mercados.

  - *Limitaciones*: es una práctica de elevado riesgo y, por lo general, menos transparente que otras, pues muchas instituciones no ofrecen información abierta sobre el volumen de su operativa propietaria. En los bancos ha perdido relevancia a raíz de la Gran Recesión (2008-2014), con la aparición de nueva regulación al respecto[33], aunque la han ganado las empresas de negociación algorítmica por cuenta propia (apartado 1.3.3).

2.2.5. Operaciones oficiales: intervención directa e indirecta

> Sobre los regímenes cambiarios y la política de intervención, véase el tema B15, y sobre la determinación del tipo de cambio, el tema B14.

Definición: se habla de intervención cuando las autoridades monetarias actúan para influir en el tipo de cambio, como un operador más en el mercado de divisas o por otras vías. La intervención incluye medidas de distinto tipo (p. ej. anuncios, compraventas…) dirigidas a suavizar los vaivenes del tipo de cambio y a ganar credibilidad. Los controles de cambio también alteran los flujos de divisas, pero no son intervención, sino regulación (apartado 3.3.2).

La intervención también engloba la gestión de reservas, es decir, el proceso de garantizar que los activos en divisas o reservas exteriores de los que dispone el sector público tienen un volumen adecuado y están disponibles en el momento necesario. Las reservas son, en efecto, un seguro frente a *shocks* financieros y paradas súbitas (*sudden stops*) en el acceso a los mercados de capitales, así como un medio para mejorar la credibilidad de la política monetaria.

Tipos: la intervención puede ser directa o indirecta:

- *Intervención directa*: consiste en la compraventa de divisas contra moneda nacional en el mercado, con cargo a las reservas internacionales. Puede ser, a su vez, de dos tipos:

  - *Intervención esterilizada*: el banco central compensa el efecto de la operación sobre la base monetaria mediante operaciones de mercado abierto de signo contrario, de modo que la intervención altera la composición de los activos en manos del público, pero no la oferta monetaria.

  - *Intervención no esterilizada*: la operación se traslada íntegramente a la base monetaria, por lo que equivale, *de facto*, a una operación de política monetaria.

- *Intervención indirecta*: el banco central influye sobre el tipo de cambio sin operar directamente en el mercado de divisas. Puede hacerlo por dos vías:

  - *Política monetaria*: variaciones del tipo de interés que alteran los flujos de capital.

[33] Por ejemplo, la regla Volcker de la Ley Dodd-Frank (2010) en EE. UU.

- *Intervención verbal* (*oral intervention*): los anuncios y declaraciones buscan orientar las expectativas del mercado (efecto señalización).

Funcionamiento: la finalidad de la intervención depende del régimen cambiario:

▪ *Cambio fijo*: la intervención busca mantener la paridad deseada y, con ella, la estabilidad total del tipo de cambio.

▪ *Cambio flexible*: busca contrarrestar los movimientos del tipo de cambio a corto plazo, en lo que se conoce como *lucha contra el viento* (*leaning against the wind*), y garantizar así su estabilidad parcial.

Conclusiones:

▪ *Aspectos positivos*: frente al consenso de los años ochenta, que la tenía por ineficaz, la revisión de Sarno y Taylor (2001) concluye que la intervención puede ser eficaz. Lo es sobre todo como señal de las intenciones futuras de la autoridad, cuando es pública, concertada entre varios bancos centrales (p. ej. los Acuerdos del Plaza de 1985 y del Louvre de 1987) y coherente con la orientación de la política monetaria.

▪ *Limitaciones*: el efecto directo de la intervención esterilizada es limitado, dada la magnitud del mercado en relación con las reservas disponibles.

2.3. Conclusiones

Los instrumentos van del contado a los derivados que trajo la innovación financiera desde los años setenta, y el más negociado es hoy el FX *swap*. Con ellos, las operaciones han pasado de servir al comercio a sostener estrategias complejas. La obtención de liquidez conecta el mercado con la economía real, y la cobertura elimina el riesgo de cambio al trasladarlo a una contraparte. El arbitraje sin riesgo impone la ley del precio único entre plazas y, en el plano temporal, la paridad cubierta de intereses. La especulación, en cambio, toma posiciones abiertas según las expectativas, y el *carry trade* obtiene beneficio porque la paridad descubierta no se cumple a corto plazo. La intervención oficial, por último, es eficaz sobre todo como señal de las intenciones de la autoridad.

La innovación financiera ha producido, así, un cambio estructural en este mercado y un escenario propicio a una mayor volatilidad de los tipos de cambio, con efectos sobre la economía real. La sección 3 examina si un mercado así es eficiente y si debe regularse.

3. Eficiencia y regulación del mercado de divisas

Falta saber si el mercado de divisas es eficiente[34], lo que exige definir antes la eficiencia, contrastarla y examinar el debate sobre su regulación.

3.1. La hipótesis de eficiencia (Bachelier, 1900; Samuelson, 1965; Fama, 1970)

Origen: diversos autores han definido la eficiencia de los mercados financieros, entre ellos Louis Bachelier (1900) y Paul Samuelson (1965), pero la definición de referencia es la de Eugene Fama (1970), Premio Nobel en 2013.

Idea: un mercado es *eficiente* o de juego justo[35] si los precios reflejan plenamente toda la información disponible. En ese caso, la cotización es una estimación fiable del valor intrínseco del activo y ningún agente obtiene de forma sistemática ganancias extraordinarias.

3.1.1. Supuestos

Fama (1970) enuncia tres condiciones suficientes, aunque no necesarias, de eficiencia, que pueden agruparse en dos:

Entorno: no existen costes de transacción ni restricciones a los precios o a la compraventa de activos (p. ej. intervención oficial) y el mercado provee toda la información sin coste.

Agentes: son racionales y tienen expectativas homogéneas. Se supone que los agentes interpretan la información de forma idéntica y actúan en consecuencia.

3.1.2. Razonamiento

Fama distingue tres niveles de eficiencia, según la información que contienen los precios:

- *Eficiencia débil*: los precios recogen toda la información pasada relevante, la contenida en las series históricas de precios. Como corolario, los inversores no pueden predecir los precios actuales o futuros a partir de los precios pasados y, así, obtener un rendimiento superior al rendimiento normal de mercado usando series históricas, pero sí usando información adicional (pública o privada). Dicho de otro modo, si se cumple la eficiencia débil, el mercado "no tiene memoria", en el sentido de que la serie histórica no permite predecir rendimientos extraordinarios (el paseo aleatorio es solo un caso particular).

- *Eficiencia semifuerte*: los precios recogen toda la información pasada y pública. Por ello, con información pública no es posible conseguir rendimientos anormalmente altos, aunque sí cabe alcanzarlos disponiendo de información privada, privilegiada o confidencial (relativa a la empresa o a la coyuntura económica en general).

- *Eficiencia fuerte*: los precios recogen toda la información (pasada, pública y privilegiada). Ningún inversor puede, entonces, obtener rentabilidades superiores a la normal de forma sistemática, ni siquiera utilizando información privilegiada.

[34] La eficiencia se entiende aquí en el sentido financiero y no en el sentido propio de la economía del bienestar.
[35] Un juego (actuarialmente) justo es aquel cuyo valor esperado coincide con su coste.

3.1.3. Conclusiones

Aspectos positivos: la hipótesis ofrece una definición operativa y contrastable de la eficiencia, con tres niveles anidados (la eficiencia fuerte contiene a la semifuerte y esta, a su vez, a la débil). Además, si se verifica, la competencia entre agentes racionales basta para que el precio de cada activo refleje en todo momento su valor intrínseco.

Limitaciones: sus condiciones suficientes no se dan en la práctica, y la hipótesis solo puede contrastarse de forma conjunta con un modelo de equilibrio, que en el mercado de divisas reúne la PDI y la neutralidad al riesgo[36]. El rechazo de esa hipótesis conjunta se estudia a continuación.

3.2. Contraste de la hipótesis de eficiencia: la paradoja del margen *forward* o *puzzle* de Fama (Fama, 1984)

Diversos estudios posteriores rechazan la hipótesis conjunta de eficiencia y neutralidad al riesgo en el mercado de divisas. Se expone primero el contraste de referencia y después las explicaciones de su resultado.

3.2.1. Desarrollo

Entre esos estudios destaca el artículo de Eugene Fama (1984), que contrasta la hipótesis estimando la regresión $g_E = \alpha + \beta\,\mu_f + \varepsilon$, donde la hipótesis nula de eficiencia (junto con la PDI y la neutralidad al riesgo) implica $\alpha = 0$ y $\beta = 1$[37].

Si la hipótesis nula se cumple, el diferencial entre el tipo de cambio *forward* y el tipo de cambio *spot* (el margen *forward*) será un predictor insesgado de la variación (*ex post*) del tipo de cambio *spot* entre t y $t+1$.

Las estimaciones de β, sin embargo, se sitúan de forma reiterada más cerca de -1 que de $+1$, un hecho estilizado de la literatura[38]. De ahí que el tipo *forward* se considere un mal predictor del tipo *spot* futuro[39]. El tipo *forward* falla en la magnitud de la variación (*ex post*) del tipo *spot* e incluso en su signo.

Este resultado se conoce como **paradoja del margen *forward*, *puzzle* de Fama, anomalía de los precios (o prima) *forward***, una de las paradojas clásicas de la literatura sobre el mercado de divisas, junto con la de Meese y Rogoff (1983) del tema B14, apartado 2.4.

Se trata de una paradoja porque, aunque los tipos de interés de un país sean superiores a los de otro y se cumpla la paridad cubierta de intereses, la moneda del país con tipos altos tiende a apreciarse. En cambio, si se cumpliese la hipótesis nula (la PCI junto con la PDI), se depreciaría.

[36] Con neutralidad al riesgo, la eficiencia financiera puede entenderse como una condición de equilibrio en la que se cumplen ambas paridades: $(E_t^f - E_t^s)/E_t^s = (E_t^e - E_t^s)/E_t^s \to E_t^f = E_t^e$, es decir, el tipo *forward* coincide con el tipo *spot* esperado.

[37] Con $\beta = 1$, cada punto de margen se traduce, en promedio, en un punto de depreciación (una simple correlación positiva exigiría solo $\beta > 0$).

[38] Kenneth Froot y Richard Thaler (1990) encuentran que el coeficiente medio de 75 estimaciones publicadas es próximo a $-0{,}88$.

[39] Por mal predictor se entiende que presenta sesgos estadísticos y que no necesariamente es el de menor varianza (no es eficiente u óptimo estadísticamente).

3.2.2. Explicaciones a la paradoja

Algunas de las teorías explicativas de la paradoja son:

Expectativas heterogéneas (Frankel y Froot, 1990; De Grauwe y Grimaldi, 2006): en el mercado conviven *chartistas*, que extrapolan la tendencia reciente, y fundamentalistas, que esperan el regreso del tipo de cambio a su valor de equilibrio. Cuando pesan más los primeros, el tipo de cambio puede alejarse de los fundamentos durante periodos largos. El modelo se desarrolla en el tema B14, apartado 3.2.1.

Agentes aversos al riesgo: el margen *forward* puede recoger, además de la depreciación esperada, una prima de riesgo, la rentabilidad adicional que se exige por mantener la divisa, que crece con la volatilidad de su tipo de cambio. Para explicar $\beta < 0$, esa prima debe variar más que la depreciación esperada de la divisa y estar negativamente correlacionada con ella (Fama, 1984).

Interpretación incorrecta: la paradoja podría desvanecerse si se atiende también al término independiente de la regresión, α. En rigor, un β negativo indica que la depreciación esperada es menor cuanto mayor es el descuento de la moneda nacional (cuanto mayor es μ_f), no que la moneda a descuento vaya a apreciarse.

Con un α positivo y suficientemente grande, la moneda nacional a descuento puede tener todavía una depreciación esperada. Sin embargo, el argumento no resuelve la paradoja con carácter general, porque el signo y la magnitud estimados de α cambian de un par de divisas a otro.

3.3. La situación de la regulación en el mercado de divisas

El rechazo de la hipótesis conjunta no prueba por sí solo que el mercado de divisas sea ineficiente, pues puede reflejar una prima de riesgo variable. Queda abierta, con todo, la cuestión de si debe regularse. El punto de partida es su escasa regulación, frente a la cual se sitúan los controles de cambio y la propuesta de la tasa Tobin.

> Este apartado enlaza el tema con el B15, sobre regímenes cambiarios, y con el B27, sobre regulación financiera.

3.3.1. Contexto: la ausencia de regulación en el mercado de divisas

La mayoría de las operaciones del mercado de divisas están poco reguladas, a diferencia de los mercados de acciones y bonos, y la regulación de cada instrumento corresponde a cada país.

Prácticas que la mayoría de los mercados organizados regulan exhaustivamente, como la inversión ventajista[40] o *front running*, se consideran aquí malas prácticas. Las desalienta el Código Global de Conducta del mercado de divisas (*FX Global Code*, 2017), un conjunto de principios de adhesión voluntaria que mantiene el Comité Global de Divisas (GFXC) y que no constituye regulación. Estas prácticas carecen, por tanto, de una prohibición específica equivalente, aunque pueden perseguirse como fraude o abuso.

En consecuencia, los creadores de mercado tienen una gran facilidad para desplazarse por los distintos mercados que componen el mercado de divisas. Sin embargo, la ausencia de regulación no supone, *a priori*, un problema, dada la naturaleza especial de este mercado. Por ejemplo, las

[40] El creador de mercado compra por cuenta propia justo antes de ejecutar una orden de compra de un cliente, anticipándose a la subida de precio que esta provocará y obteniendo un beneficio a su costa.

restricciones a las ventas en corto que existen en otros mercados no tienen encaje como tal en el mercado de divisas, puesto que la venta en corto de una moneda implica la compra de otra.

No obstante, la escasa regulación del mercado de divisas hace que la información disponible sea insuficiente, pues no se conoce con exactitud el volumen de transacciones de un día concreto. La fuente de información más amplia es el informe trienal del BIS, que complementan las encuestas semestrales de los comités de divisas de las principales plazas (p. ej. Londres o Nueva York) y los datos de liquidación de CLS.

3.3.2. Regulación y controles de cambio

La regulación de los mercados de cambios va de la legislación de base a los límites a los movimientos de capitales, que pueden mover el tipo de cambio porque alteran los flujos de divisas. Se justifica por la teoría del *second best*, según la cual, con fallos de mercado, una restricción adicional a la movilidad del capital puede aumentar el bienestar. La regulación se concreta en los controles de capitales, cuantitativos o basados en precios (como el encaje chileno de 1991, un depósito no remunerado), los tipos de cambio múltiples y los impuestos sobre las transacciones, cuya versión de referencia es la tasa Tobin.

> La justificación, los tipos de regulación con sus ejemplos y su valoración se desarrollan en el tema B15, apartado 3.2.

3.3.3. La tasa Tobin (1972)

Origen: James Tobin la propuso en 1972, en unas conferencias en la Universidad de Princeton, poco después de la suspensión de la convertibilidad del dólar.

Idea: gravar las transacciones cambiarias con una tasa cuya incidencia efectiva sea inversamente proporcional a la duración de la operación, para "echar arena en las ruedas" de la especulación a corto plazo, en expresión del propio Tobin (1978).

Razonamiento: como la tasa se paga en cada operación, apenas encarece las inversiones a largo plazo, mientras que las operaciones de ida y vuelta a muy corto plazo soportan un gravamen anualizado mucho mayor. En consecuencia, desincentiva los flujos especulativos y, con ellos, la volatilidad del tipo de cambio.

Extensiones: acogida al principio con escaso éxito, la propuesta ganó popularidad desde los años noventa en el debate sobre los mecanismos alternativos de financiación de los gobiernos, y se ha planteado también para las transacciones financieras en general.

Conclusiones:

- *Aspectos positivos*: sus partidarios defienden que permitiría reducir la volatilidad de los flujos y obtener importantes ingresos sin alterar notablemente la operativa del mercado. Así, con el volumen actual del mercado de divisas (unos 9,6 billones de dólares diarios, según la encuesta trienal del BIS de 2025), una tasa mundial del 0,05 % recaudaría del orden de 1,2 billones de dólares al año si ese volumen se mantuviera constante. No obstante, la recaudación efectiva sería muy inferior, pues la propia tasa reduciría el volumen negociado.

- *Limitaciones*: sus detractores aducen que es difícil de implementar, porque exige una cooperación internacional intensa, sin la cual las transacciones se desviarían en gran medida hacia las jurisdicciones que no la aplicasen, y que produciría distorsiones en el mecanismo de los precios.

Conclusiones

El mercado de divisas es el mayor mercado financiero del mundo, un mercado OTC global, profundo y concentrado en un núcleo de creadores de mercado, con transacciones sobre todo financieras y con instrumentos que van del contado a los derivados. La cobertura traslada el riesgo de cambio a quien está dispuesto a asumirlo. El arbitraje hace cumplir la ley del precio único entre plazas y, en el plano temporal, da lugar a la paridad cubierta de intereses. La especulación toma posiciones abiertas y obtiene beneficio del incumplimiento de la paridad descubierta, como el *carry trade*, con el riesgo de amplificar la volatilidad cuando se apalanca con derivados OTC, cuya opacidad agravó la crisis financiera global. El contraste de la hipótesis de eficiencia desemboca en el *puzzle* de Fama, porque el margen *forward* predice mal incluso el signo de la depreciación. Queda abierta la regulación de un mercado poco regulado y mal medido, con los controles de cambio y la tasa Tobin como respuestas posibles.

La paridad descubierta es la condición de equilibrio de las teorías de determinación del tipo de cambio, y las expectativas heterogéneas que explican el *puzzle* de Fama reaparecen en ellas (tema B14). Las operaciones oficiales y los controles de cambio son, respectivamente, instrumentos de la intervención y de la regulación con que se gestionan los regímenes cambiarios (tema B15), y la especulación contra una paridad, que puede forzar su abandono aunque los fundamentos sean sanos, es el mecanismo de las crisis de segunda generación (tema B17). Por último, los derivados que se negocian en este mercado tienen su propia teoría de valoración (tema B25), y su opacidad motivó la reforma regulatoria posterior a 2008 (tema B27).

Desde 2008, la paridad cubierta presenta desviaciones persistentes entre las monedas del G10, que Du, Tepper y Verdelhan (2018) documentan y atribuyen a los costes de balance que la regulación poscrisis impone a los bancos que arbitran. Sigue en debate, además, el sentido del *puzzle* de Fama. Brunnermeier, Nagel y Pedersen (2008) interpretan el rendimiento del *carry trade* como compensación por el riesgo de desplomes cambiarios bruscos. Por su parte, Bussière, Chinn, Ferrara y Heipertz (2022) encuentran que el coeficiente de Fama se vuelve positivo y elevado a partir de la crisis financiera, un "nuevo *puzzle*" ligado sobre todo a cambios en la covarianza entre los errores de expectativas y los diferenciales de intereses.

Anexo

A.1. Cómo operan los creadores de mercado

Los creadores de mercado (apartado 1.3.1) cotizan simultáneamente un precio de compra (*bid*) y un precio de venta (*ask*) y operan en ambos lados del mercado con posiciones propias, a diferencia de los *brokers* (apartado 1.3.2). El inversor que quiere vender recibirá el precio *bid*, ligeramente inferior al precio medio de mercado, mientras que el que quiere comprar pagará el precio *ask*, ligeramente superior. El diferencial entre ambos (el *spread*) es un coste para quien opera y una ganancia para el creador de mercado, porque este compra la divisa al *bid* y la vende al *ask*. Esa ganancia le compensa por el servicio de liquidez y de búsqueda que presta.

Ejemplo: el *spread* como coste de la operación

1. *Orden de compra*: un inversor quiere comprar el par GBP/USD cuando la cotización es 1,2756/1,2758. Como compra, el precio de referencia es el *ask*, 1,2758.

2. *Posición inicial*: nada más ejecutarse la operación, su posición está en -2 puntos, porque para deshacerla debe vender, y entonces el precio de referencia es el *bid*, 1,2756.

3. *Coste de la operación*: el inversor ha pagado al creador de mercado 2 puntos, el *spread*.

El *spread* es un coste cada vez más relevante en el mercado de divisas, debido a la proliferación de las operaciones de alta frecuencia y de muy corto plazo gracias al desarrollo de las nuevas tecnologías. Cuando se hacen muchas operaciones en el corto plazo, el coste del *spread* pesa mucho en el beneficio del inversor.

Sin embargo, no debe confundirse el *spread* con la comisión que cobra el *broker*. Cuando se opera en mercados de acciones, la comisión tiene dos tramos, la comisión que cobra el *broker* por el servicio y el canon de bolsa. El canon de bolsa es variable según el importe de la operación y el coste del *broker* puede ser proporcional o fijo. En el caso de las divisas, en cambio, al operarse en un mercado OTC, la comisión solo tiene el tramo del *broker*.

Es común, no obstante, que los *brokers* del mercado de divisas no cobren comisión. En ese caso, suelen cobrar una cantidad fija de puntos que suman al *spread* natural del mercado, y por ello se denominan *brokers* de *spread* fijo. A diferencia de estos, los *brokers* que operan con *spreads* flexibles suelen cobrar una comisión que habrá que sumar al *spread* para calcular el coste real de la operación.

A.2. Obtención de la paridad *cubierta* de intereses (PCI)

Punto de partida: la paridad cubierta de intereses se puede expresar como:

$$0 = \frac{E^f}{E^s}(1+i^*) - (1+i) \rightarrow (1+i) = \frac{E^f}{E^s}(1+i^*) \rightarrow \frac{1+i}{1+i^*} = \frac{E^f}{E^s}$$

Resta de la unidad: sustrayendo 1 de ambos miembros se obtiene:

$$\frac{E^f}{E^s} - 1 = \frac{1+i}{1+i^*} - 1 \rightarrow \frac{E^f - E^s}{E^s} = \frac{i-i^*}{1+i^*}$$

Aproximación: el término $\frac{E^f - E^s}{E^s} \cdot i^*$ toma valores muy pequeños, por lo que se aproxima a cero por simplicidad. Operando se obtiene:

$$\frac{E^f - E^s}{E^s} + \underbrace{\frac{E^f - E^s}{E^s} \cdot i^*}_{\approx 0} = i - i^* \rightarrow \mu_f \approx i - i^*$$

Ejemplo: magnitud del término de segundo orden

Con $E^f = 0{,}9$, $E^s = 0{,}87$ e $i^* = 0{,}03$: $\mu_f = (0{,}9 - 0{,}87)/0{,}87 \approx 0{,}034$, y el término omitido es $\mu_f \cdot i^* \approx 0{,}034 \cdot 0{,}03 \approx 0{,}001$, es decir, un 0,1 % frente a un margen *forward* del 3,4 %.

B14. Determinación del tipo de cambio

Teorías de la determinación del tipo de cambio.

Introducción

Introducción común. Los temas B13 a B15 comparten una introducción, que se desarrolla al comienzo del tema **B13**. Presenta el tipo de cambio, "el precio más importante de una economía" según Volcker, el mercado de divisas en que se determina y el debate recurrente sobre qué determina los tipos de cambio y qué régimen cambiario es preferible.

La **determinación del tipo de cambio** es uno de los problemas más difíciles de la macroeconomía abierta. Las teorías tradicionales la explicaron por los flujos comerciales y por la paridad del poder adquisitivo. Con la libertad de movimientos de capital, los modelos de activos pasaron a tratar el tipo de cambio como el precio de un activo financiero, que depende de las expectativas, y Dornbusch (1976) explicó, con precios rígidos, por qué sobrerreacciona ante una perturbación monetaria (*overshooting*). Sin embargo, Meese y Rogoff (1983) mostraron que, a horizontes cortos, ninguno de esos modelos predice el tipo de cambio mejor que un paseo aleatorio. Esa **desconexión** entre el tipo de cambio y sus fundamentos (Obstfeld y Rogoff, 2000) orienta todavía la investigación, que busca su origen en las expectativas, la microestructura del mercado y las primas de riesgo de los intermediarios financieros.

A partir del artículo de Michael Mussa (1979), pueden sintetizarse cinco regularidades empíricas de los tipos de cambio flexibles que los sucesivos modelos han tratado de explicar:

1. *Imprevisibilidad*: las variaciones mensuales de los tipos de cambio nominales *spot* son grandes y casi enteramente imprevisibles[1].

2. *Tipo de contado y tipo esperado*: las variaciones del tipo de cambio de contado se corresponden bastante de cerca con las del tipo de contado futuro esperado, de modo que los tipos de contado y a plazo se mueven juntos.

3. *Tipo nominal y tipo real*: a corto plazo, las variaciones de los tipos de cambio nominales y reales están muy correlacionadas, lo que indica que los niveles de precios apenas compensan los movimientos del tipo nominal.

4. *Balanza de pagos*: no existe una relación fuerte y sistemática entre los movimientos de los tipos de cambio, nominales o reales, y los saldos por cuenta corriente.

[1] Concretamente, el tipo de cambio (en logaritmos) se aproxima a un paseo aleatorio: $s_t = s_{t-1} + \varepsilon_t$.

5. *Oferta monetaria e inflación*: los movimientos de los tipos de cambio no guardan una relación estrecha con los diferenciales de expansión monetaria, salvo, quizá, en economías de inflación muy alta. A largo plazo, en cambio, las monedas de los países con inflación elevada tienden a depreciarse aproximadamente al ritmo del diferencial de inflación, en línea con la paridad del poder adquisitivo relativa.

El tema estudia los mecanismos de determinación del tipo de cambio en los mercados de divisas en tres etapas:

- *Teorías tradicionales*: el tema parte de las teorías basadas en los modelos de flujos.
- *Desarrollos posteriores*: siguen las teorías basadas en los modelos de activos.
- *Desarrollos recientes*: la última parte se ocupa de los modelos basados en las expectativas, la microestructura del mercado y la intermediación financiera.

1. Teorías tradicionales: teorías basadas en los modelos de flujos de la balanza de pagos

Hasta la década de 1970 se consideraba que las operaciones en divisas se debían a los flujos entre países, principalmente de bienes y servicios y, en mucha menor medida, a los flujos de capital. Esta perspectiva se articula en tres modelos, el de balanza de pagos, el de la paridad del poder adquisitivo (PPA) y el de la paridad de intereses.

1.1. Modelos basados en el ajuste de la balanza de pagos

Origen: el enfoque de flujos carece de un autor único y se construye sobre la llamada *aproximación de las elasticidades*. Su genealogía se remonta a Marshall (1923) y Lerner (1944), de quienes toma su nombre la condición de estabilidad del modelo. La formalización posterior del mercado de divisas se debe a Joan Robinson (1937) y a Machlup (1939), y su integración en el análisis del equilibrio interno y externo, a Meade (1951).

Idea: el tipo de cambio nominal se determina por la interacción de las fuerzas de mercado internacionales que actúan para ajustar las balanzas de pagos de los diferentes países.

El modelo de balanza de pagos permite ilustrar la determinación del tipo de cambio en su forma más simple, es decir, cuando se realizan transacciones *spot* (al contado) en los mercados de divisas para la obtención de moneda.

1.1.1. Supuestos

Entorno: rige un régimen de cambio flexible y no hay barreras al comercio.

Mercados: se cumple la condición de Marshall-Lerner (estabilidad del equilibrio).

1.1.2. Modelo

Planteamiento: el tipo de cambio nominal es el precio que vacía el mercado de divisas (tema B13, apartado 2.2.1), en el que concurren dos fuerzas:

- *Demanda de divisas*: procede de quienes necesitan divisas para pagar importaciones de bienes y servicios y de los no residentes que repatrían fondos y han de convertirlos en divisa. Al depreciarse la moneda nacional ($\uparrow E$) se importa menos y se necesitan menos divisas, por lo que la demanda decrece con el tipo de cambio.

- *Oferta de divisas*: la aportan quienes obtienen divisas con sus exportaciones y desean deshacerse de esa liquidez, y los residentes que repatrían fondos y los convierten a moneda nacional. Crece con el tipo de cambio siempre que la demanda exterior de exportaciones sea elástica. En tal caso, una depreciación de la moneda nacional ($\uparrow E$) eleva las exportaciones en mayor proporción de lo que cae su precio en divisas, y la oferta de divisas aumenta.

Equilibrio: de la interacción de ambas fuerzas se determina el tipo de cambio que equilibra la balanza de pagos. De acuerdo con la tradición marshalliana, este equilibrio es estable siempre que se cumpla la condición de Marshall-Lerner, que garantiza que las cantidades responden a las variaciones del tipo de cambio con la elasticidad suficiente.

Desarrollo gráfico: la figura B14.1 representa el mercado en el plano (M^*, E), con la cantidad de divisas en el eje horizontal y el tipo de cambio directo en el vertical. La curva $D(M^*)$ recoge

la demanda de divisas de los importadores y de los no residentes que repatrían fondos, y es decreciente porque una depreciación ($\uparrow E$) encarece las importaciones. La curva $S(M^*)$ recoge la oferta de divisas de los exportadores y de los residentes que repatrían fondos, y es creciente si la demanda exterior de exportaciones es elástica. El corte de ambas curvas, el punto 0, fija el tipo de cambio E_0 y la cantidad intercambiada M_0^*. A un tipo E' superior, la oferta de divisas excede a la demanda (superávit), y a un tipo E'' inferior la demanda excede a la oferta (déficit). Las flechas grises indican el sentido en que se corrige cada desequilibrio.

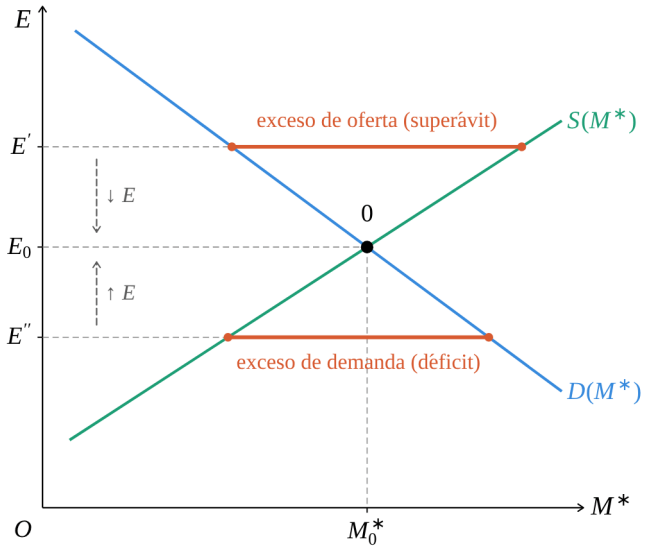

Figura B14.1: Mercado de divisas y tipo de cambio. La demanda, $D(M^*)$, y la oferta de divisas, $S(M^*)$, fijan el tipo de cambio E_0 que equilibra la balanza de pagos (punto 0). A un tipo superior (E') hay exceso de oferta de divisas (superávit) y la moneda nacional se aprecia, y a uno inferior (E''), exceso de demanda (déficit) y se deprecia.

Estática comparativa: todo lo que altera, a un tipo de cambio dado, las importaciones, las exportaciones o los fondos repatriados desplaza una curva del plano (M^*, E) y mueve el equilibrio desde el punto 0:

- *Aumento de la demanda de divisas* (más importaciones o más fondos repatriados por no residentes): $D(M^*)$ se desplaza a la derecha, el tipo de cambio sube ($\uparrow E$, depreciación de la moneda nacional) y la cantidad intercambiada aumenta.

- *Aumento de la oferta de divisas* (más exportaciones o más fondos repatriados por residentes): $S(M^*)$ se desplaza a la derecha, el tipo de cambio baja ($\downarrow E$, apreciación) y la cantidad aumenta.

El modelo predice así que la moneda de un país cuyas importaciones crecen más que sus exportaciones tiende a depreciarse, relación que los datos no confirman de forma sistemática.

Dinámica: fuera del equilibrio, el mecanismo de ajuste opera como sigue:

1. *Superávit comercial*: si BC > 0 (las exportaciones superan a las importaciones), existe un exceso de oferta de divisas, de modo que la divisa se deprecia y, por extensión, la moneda nacional se aprecia ($\downarrow E$) hasta restablecer el equilibrio.

2. *Déficit comercial*: si $\text{BC} < 0$, existe un exceso de demanda de divisas y, a la inversa, la divisa se aprecia y la moneda nacional se deprecia ($\uparrow E$).

1.1.3. Conclusiones

Aspectos positivos: el modelo, muy intuitivo, liga la determinación del tipo de cambio a los flujos comerciales y muestra que la estabilidad del mercado de divisas depende de la condición de Marshall-Lerner.

Limitaciones: omite muchas variables relevantes en la determinación del tipo de cambio (p. ej. la relación de precios). Choca, además, con la cuarta regularidad de Mussa, pues no se observa una relación fuerte y sistemática entre los movimientos del tipo de cambio y los saldos por cuenta corriente. Tampoco explica por qué países desarrollados como Estados Unidos mantienen importantes déficits comerciales sin ver una depreciación sostenida de su moneda.

1.2. Modelos basados en la paridad del poder adquisitivo de la moneda

Origen: el nombre y la formulación moderna se deben a Gustav Cassel (1918a), pero la idea aparece ya en la Escuela de Salamanca[2].

Idea: el tipo de cambio nominal refleja el poder de compra relativo de las monedas.

1.2.1. Paridad del poder adquisitivo absoluta (Cassel, 1918a)

Origen: Cassel (1918a) llama paridad del poder adquisitivo al cociente entre el poder de compra de las monedas de dos países, la formulación original de la que parte la versión absoluta.

Idea: medida en una misma moneda, una cesta de bienes cuesta igual en todos los países, como ocurre si el precio de cada bien, convertido al tipo de cambio, coincide entre países (ley del precio único).

Supuestos:

- *Entorno*: dos países, sin barreras al comercio y con mercados perfectamente competitivos. Los bienes son homogéneos y comerciables, y ambos países los intercambian.

Razonamiento: sin imperfecciones, el arbitraje internacional neutraliza de inmediato cualquier diferencia en el tipo de cambio nominal del par de monedas entre plazas o en el precio del bien entre países, y agota así toda oportunidad de beneficio.

Por ello, el tipo de cambio nominal del par de monedas es el mismo en todas las plazas en que se negocia: $E^A = E^B$, con E el tipo de cambio nominal y $A \neq B$ las plazas.

Dado ese tipo de cambio único, el arbitraje de bienes garantiza que se cumpla la *ley del precio único internacional* o *condición de paridad del poder adquisitivo absoluta*. Según esa condición, el precio del bien, ajustado por el tipo de cambio, es el mismo en todos los países:

$$P_i = E \cdot P_i^*$$

[2] En particular, Martín de Azpilcueta (*Comentario resolutorio de cambios*, 1556), que vinculó la llegada de metales preciosos americanos con la subida de los precios en Castilla y explicó los cambios entre plazas por el distinto valor del dinero según su abundancia o escasez.

donde P_i es el precio nacional del bien i, E es el tipo de cambio nominal idéntico para ambos países y P_i^* es el precio del mismo bien en el extranjero.

Esta ley constituye una condición de equilibrio, según la cual una unidad de moneda nacional tiene el mismo poder de compra en el propio país que en el extranjero, una vez convertida en divisa al tipo de cambio vigente.

De esta condición de equilibrio se deducen dos ecuaciones de determinación del tipo de cambio:

Tipo de cambio nominal PPA: es el tipo de cambio nominal que verifica esta condición, igual a la relación de precios:

$$E = P_i/P_i^*$$

El tipo de cambio nominal acompaña, pues, a la relación de precios. Así, un país con precios más altos que los del extranjero necesita una moneda más depreciada para que esta conserve su poder adquisitivo.

Esta relación también se puede plantear en términos de índices de precios, simplemente aplicando una suma ponderada:

$$\sum_{i=1}^{n} \alpha_i P_i = E \sum_{i=1}^{n} \alpha_i P_i^* \;\rightarrow\; P = E \cdot P^* \;\rightarrow\; E = P/P^*$$

Tipo de cambio real PPA: de todo esto se deduce que es constante e igual a la unidad:

$$e = E \cdot P^*/P = (P/P^*) \cdot (P^*/P) = 1$$

Conclusiones:

▪ *Aspectos positivos*: ofrece un ancla sencilla para el tipo de cambio, basada solo en el arbitraje de bienes, y es la referencia de las comparaciones internacionales de renta y precios en paridad de poder adquisitivo.

▪ *Limitaciones*: sus supuestos rara vez se dan. Hay costes de transporte, bienes heterogéneos y bienes no comerciables, y cada país pondera su cesta de forma distinta (problema de agregación).

1.2.2. Paridad del poder adquisitivo relativa (Cassel, 1918a)

Origen: el propio Cassel (1918a), ante las limitaciones de la versión absoluta, planteó la misma idea en su versión relativa, que exige una condición menos estricta para la equivalencia del poder de compra de las monedas.

Idea: si las imperfecciones que impiden la versión absoluta (costes de transporte, barreras al comercio, problemas de agregación) son estables en el tiempo, no afectan a las variaciones del tipo de cambio, que dependen entonces del diferencial de inflación.

Razonamiento: si el tipo de cambio real permanece fijo en un nivel $\bar{e}$, que no tiene por qué ser la unidad, el *tipo de cambio nominal de PPA relativa* cumple:

$$E = \bar{e} \cdot P/P^* \;(\bar{e} \text{ constante}) \;\rightarrow\; \ln E = \ln \bar{e} + \ln P - \ln P^*$$
$$\rightarrow\; g_E = g_P - g_{P^*} \;\rightarrow\; g_E = \pi - \pi^*$$

La moneda nacional se deprecia, por tanto, al ritmo en que la inflación nacional supera a la extranjera, y las monedas de los países más inflacionistas que el resto del mundo pierden valor. Para el *tipo de cambio real de PPA relativa*, la consecuencia inmediata es que no varía: $g_e = g_E + \pi^* - \pi = 0$. En suma, la PPA hace depender el tipo de cambio nominal del nivel relativo de precios en su versión absoluta y del diferencial de inflación en la relativa.

Conclusiones:

- *Aspectos positivos*: exige supuestos menos estrictos que la versión absoluta y explica la tendencia del tipo de cambio a largo plazo y en las hiperinflaciones.

- *Limitaciones*: el supuesto de que las imperfecciones se mantengan constantes en el tiempo sigue siendo exigente, y la evidencia solo respalda la versión relativa a largo plazo.

1.2.3. Explicaciones al incumplimiento de la PPA (I): hipótesis de Balassa-Samuelson (1964)

Origen: Béla Balassa (1964) y Paul Samuelson (1964) explicaron, de forma independiente, por qué los niveles de precios son más altos en los países ricos. Su *hipótesis de Balassa-Samuelson*, o sesgo de productividad de la PPA, es una de las explicaciones de por qué no se cumple la paridad del poder adquisitivo.

Idea: las diferencias de productividad en la producción de bienes comerciables entre países permiten explicar los diferenciales de inflación y, en último término, las diferencias en los tipos de cambio reales.

Supuestos:

- *Entorno*: dos países abiertos al exterior, con libre movimiento de capital entre países y libre movilidad del trabajo entre sectores, sin fallos de mercado y con dos tipos de bienes, comerciables y no comerciables.

- *Agentes*: la ponderación para la agregación de bienes es idéntica en ambos países (preferencias idénticas).

- *Tecnología*: funciones de producción con rendimientos constantes a escala y la misma participación del trabajo en ambos sectores. En todo momento, la competencia iguala el salario al valor de la productividad marginal del trabajo y la movilidad intersectorial iguala los salarios entre sectores.

- *Sector exterior*: se cumple la PPA para los bienes comerciables: $E = P_C/P_C^* \to s = p_C - p_C^*$ (en logaritmos).

Razonamiento: un aumento de la productividad en el sector comerciable eleva los salarios de dicho sector, puesto que el precio del bien comerciable está fijado internacionalmente (por la ley del precio único) y la competencia iguala el salario al valor del producto marginal. Por la libre movilidad intersectorial del trabajo, la subida salarial se extiende al sector no comerciable y eleva su coste marginal. Sube así el precio de ese sector y, con él, el nivel general de precios, y el tipo de cambio real se aprecia.

Formalmente, el índice de precios (en logaritmos) es una suma ponderada de los índices de precios de cada sector, $p = \alpha p_N + (1-\alpha)p_C$, donde N es el sector del bien no comerciable y C el de los comerciables. Del supuesto tecnológico se deduce que el precio relativo de los bienes no comerciables viene dado por el diferencial de productividades entre sectores:

$$p_N - p_C = a_C - a_N$$

donde a_C y a_N son los logaritmos de la productividad del trabajo en cada sector. Es decir, el precio relativo del bien no comerciable fluctúa en la misma proporción que el diferencial de productividad. Si aumenta la productividad del sector comerciable, sube el precio relativo del no comerciable. Ese encarecimiento relativo tiene efectos sobre la competitividad, que se analizan a través del tipo de cambio real (en logaritmos), $q = s + p^* - p$. Sustituyendo los índices de precios y la PPA de los comerciables:

$$q = \alpha\left[(p_C - p_N) - (p_C^* - p_N^*)\right] = -\alpha\left[(a_C - a_N) - (a_C^* - a_N^*)\right]$$

El tipo de cambio real es, pues, función del diferencial entre países de la productividad relativa entre sectores, y los países con un mayor crecimiento relativo de la productividad en el sector comerciable ven apreciarse su tipo de cambio real ($\partial q/\partial a_C < 0$).

Con ello, el modelo da cuenta de la relación positiva entre la renta per cápita y el nivel de precios. En los países ricos, la mayor productividad relativa del sector comerciable eleva salarios y precios y aprecia el tipo de cambio real, en contra de la PPA. El efecto sobre el nivel de precios es proporcional al peso α del sector no comerciable.

Conclusiones:

- *Aspectos positivos*: ofrece una explicación de las desviaciones persistentes de la PPA basada en factores reales, la productividad relativa entre sectores, y no en perturbaciones monetarias. Permite, además, distinguir una apreciación real de equilibrio, sin pérdida de competitividad, de la que refleja desajustes, una distinción necesaria para interpretar los diferenciales de inflación de una unión monetaria (tema B16, apartado 1.3.4).

- *Limitaciones*: descansa en supuestos fuertes, como la plena movilidad del trabajo entre sectores y el cumplimiento de la PPA en los bienes comerciables, que el *pricing to market* cuestiona. Además, su evidencia es más sólida para las diferencias de nivel de precios entre países de distinto desarrollo que para los movimientos del tipo de cambio real a corto plazo, dominados por los del tipo de cambio nominal.

1.2.4. Explicaciones al incumplimiento de la PPA (II): fijación de precios según mercado o *pricing to market* (Krugman, 1987b; Dornbusch, 1987)

Origen: Krugman (1987b) y Dornbusch (1987) la propusieron ante la escasa traslación a los precios de importación estadounidenses de la fuerte apreciación del dólar en la primera mitad de los años ochenta.

Idea: en competencia oligopolística con mercados segmentados, un mismo bien puede ofrecerse a distintos precios en países diferentes, lo que genera desviaciones persistentes de la ley del precio único y, por extensión, de la PPA.

Razonamiento: cada empresa fija un precio distinto en cada mercado según la elasticidad de su demanda local. Ante una variación del tipo de cambio, absorbe en sus márgenes parte del cambio para no perder cuota, de modo que el precio en moneda local varía menos que el tipo de cambio (traslado o *pass-through* incompleto).

Conclusiones:

- *Aspectos positivos*: explica por qué la ley del precio único falla incluso en bienes comerciables y por qué el traslado del tipo de cambio a los precios de importación es incompleto, como muestra la evidencia.

- *Limitaciones*: exige competencia imperfecta y mercados segmentados, en los que el arbitraje de los consumidores sea limitado.

1.2.5. Conclusiones

Aspectos positivos: Jacob Frenkel (1978) encuentra que la PPA explicaba bien el tipo de cambio durante la hiperinflación alemana, y concluye que se cumple, sobre todo, cuando predominan los *shocks* monetarios frente a los reales. Los contrastes con muestras más largas y técnicas de panel apoyan, además, el cumplimiento de la PPA relativa en el largo plazo. A ese horizonte, la depreciación nominal acompaña al diferencial de inflaciones y el tipo de cambio real revierte hacia su media.

Limitaciones: la volatilidad de los tipos de cambio nominales, comparada con la relativa estabilidad de los índices generales de precios, muestra que el tipo de cambio real no es estable y que la PPA no se cumple de manera continua en el corto plazo. Los primeros contrastes econométricos del periodo posterior a Bretton Woods no lograban rechazar que el tipo de cambio real fuese no estacionario, es decir, que no convergiese hacia ningún nivel de equilibrio, y los posteriores descartan que la PPA, absoluta o relativa, se cumpla a corto plazo. La reversión del tipo real es, además, lenta, con vidas medias estimadas de entre tres y cinco años (Rogoff, 1996).

1.3. Modelos basados en la paridad de intereses

Origen: a diferencia de los modelos anteriores, centrados en los flujos de bienes, la paridad de intereses atiende a los flujos de capitales. Keynes ya había formulado la paridad cubierta en *A Tract on Monetary Reform* (1923), pero estos modelos solo cobraron protagonismo con el aumento de los flujos de capitales en la segunda mitad del s. XX, sobre todo tras el abandono de Bretton Woods y la liberalización de las cuentas de capital.

Idea: el tipo de cambio nominal se mueve hasta que invertir en activos nacionales o extranjeros rinde lo mismo, medido en una misma moneda.

1.3.1. Supuestos

Entorno: no hay fallos de mercado y la movilidad de capitales es perfecta (sin controles ni costes de transacción), con mercados eficientes y completos. Existe sustituibilidad perfecta entre activos nacionales y extranjeros (no hay prima por mantener una divisa u otra)[3].

Agentes: son racionales. La paridad descubierta exige, además, que sean neutrales al riesgo (no exigen prima por invertir en un depósito nacional o extranjero). La cubierta no lo exige, pues la operación cubierta no entraña riesgo cambiario (tema B13, apartado 2.2.3).

[3] Así, los depósitos nacional y extranjero solo difieren en su moneda de denominación, lo que permite un análisis de equilibrio parcial.

1.3.2. Modelo

Un inversor que posee una unidad de moneda nacional (p. ej. euros) elige, con horizonte de un año, la opción de mayor rendimiento entre un activo nacional, que le rinde $1 + i$, y un activo extranjero. Para invertir en el activo extranjero, compra divisa hoy al tipo *spot* E_t^s, la invierte al tipo i^* y convierte el rendimiento a moneda nacional al tipo de cambio del vencimiento. La derivación paso a paso del diferencial de rentabilidades se desarrolla en el tema B13, apartado 2.2.3. En forma aproximada[4], ese diferencial es:

$$r_{€} \approx g_E + (i^* - i)$$

donde $r_{€}$ es el diferencial de rentabilidades *ex post* del activo extranjero frente al nacional, medido en moneda nacional, y $g_E = (E_{t+1}^s - E_t^s)/E_t^s$ es la tasa de depreciación observada de la moneda nacional.

El rendimiento relativo de invertir fuera suma, por tanto, dos términos, el diferencial de intereses, $i^* - i$, y la depreciación de la moneda nacional, g_E. Si los mercados son lo bastante eficientes, los agentes aprovechan las diferencias de rentabilidad hasta anularlas ($r_{€} = 0$), es decir, hasta que se cumple la condición de no arbitraje:

$$E_{t+1}^s = E_t^s \cdot \frac{1+i}{1+i^*} \;\rightarrow\; i \approx g_E + i^*$$

Sin embargo, en el momento de invertir (momento t) el agente no conoce E_{t+1}^s, por lo que existe riesgo de tipo de cambio. La interpretación de esta condición depende, entonces, de si el agente cubre o no ese riesgo, y solo la operación cubierta da lugar a una condición de no arbitraje en sentido estricto.

1.3.3. Condición de paridad cubierta de intereses

Si el agente cubre todo el riesgo de cambio con un contrato *forward*, que le permite convertir su retorno al tipo de cambio *forward* o a plazo, E_t^f, la operación es de arbitraje temporal puro o sin riesgo. La condición de no arbitraje se denomina entonces **paridad cubierta de intereses** (PCI). La tasa de depreciación pasa a ser el margen *forward*, $\mu_f = (E_t^f - E_t^s)/E_t^s$. Si $\mu_f > 0$, la divisa cotiza a premio y la moneda nacional al descuento, y si $\mu_f < 0$ ocurre lo contrario. La PCI exige que el margen *forward* compense el diferencial de tipos de interés, $r_{€} = 0 \leftrightarrow i = i^* + \mu_f$, la condición del tema B13, apartado 2.2.3, donde se explica el ajuste hacia ella.

En su forma exacta, la PCI da una teoría de la determinación del tipo de cambio *forward*, que se calcula a partir de las otras tres variables:

$$E_t^f = E_t^s \cdot \frac{1+i}{1+i^*}$$

Así se fija en la práctica el tipo *forward* en el mercado de divisas, lo que ilustra por qué un *forward* es un derivado, pues su precio se deriva del precio del subyacente.

[4] Desarrollando $(1 + g_E)(1 + i^*) - (1 + i)$ y despreciando el producto $g_E \cdot i^*$, de segundo orden, se obtiene $r_{€} \approx g_E + (i^* - i)$. La condición de no arbitraje $i \approx g_E + i^*$ resulta de igualar la expresión exacta a cero y aproximar $(1 + i)/(1 + i^*) \approx 1 + i - i^*$.

1.3.4. Condición de paridad descubierta de intereses

Si el agente no cubre el riesgo de cambio, estructura la operación solo con contratos *spot* y decide con el tipo de cambio *spot* esperado, $E_t^e \equiv \mathbb{E}_t[E_{t+1}^s]$. La condición resultante es la **paridad descubierta de intereses** (PDI), en la que la tasa de depreciación es una tasa esperada, $g_E^e = (E_t^e - E_t^s)/E_t^s$. Con agentes neutrales al riesgo, se cumple $\mathbb{E}_t[r_\epsilon] = 0 \leftrightarrow i = i^* + g_E^e$. Este tipo de operación la llevan a cabo, generalmente, especuladores que aceptan el riesgo cambiario a cambio de la posibilidad de obtener ganancias (tema B13, apartado 2.2.4).

En su forma exacta, la PDI da también una teoría de la determinación del tipo de cambio *spot*:

$$E_t^s = E_t^e \cdot \frac{1+i^*}{1+i}$$

Sin embargo, la evidencia empírica muestra que la paridad descubierta no se cumple a corto plazo (Fama, 1984). Entre sus explicaciones (tema B13, apartado 2.2.4) destacan tres:

- *Aversión al riesgo*: si los agentes son aversos al riesgo, solo asumen el riesgo cambiario si la rentabilidad esperada de los activos extranjeros supera a la de los nacionales en una prima ρ, es decir, si $i^* + g_E^e = i + \rho$.

- *Expectativas sesgadas*: con datos de encuestas sobre las expectativas, Froot y Frankel (1989) no hallan prima de riesgo alguna en el incumplimiento y lo atribuyen a errores sistemáticos de expectativas.

- *Peso problems*: compatibles con expectativas racionales, son episodios en los que los agentes asignan durante largos periodos una probabilidad positiva a un suceso extremo (p. ej. una devaluación) que tarda en materializarse, lo que sesga en la muestra los errores de predicción.

Ejemplo: PPA relativa y paridades de intereses con cifras

El tipo de cambio al contado es de 0,90 euros por dólar. La inflación esperada a un año es del 2 % en la zona del euro y del 4 % en Estados Unidos, y los tipos de interés a un año son del 3 % y del 5 %.

- *PPA relativa*: el euro, menos inflacionista, debe apreciarse en torno a un 2 %, $g_E = \pi - \pi^* = -2$ %, hasta unos 0,882 euros por dólar.

- *Paridad cubierta*: el tipo *forward* a un año es $E_t^f = 0{,}90 \cdot 1{,}03/1{,}05 \approx 0{,}8829$ euros por dólar, con un margen *forward* de $-1{,}9$ %. El dólar cotiza a plazo con descuento porque rinde más intereses.

- *Paridad descubierta*: con agentes neutrales al riesgo, la apreciación esperada del euro es la misma, en torno al 1,9 %, y coincide casi con la que predice la PPA, porque los tipos de interés reales de las dos áreas son iguales, del 1 %.

1.3.5. Conclusiones

Aspectos positivos: hasta la crisis financiera global, la evidencia empírica mostraba que la paridad cubierta se cumplía de forma continua y con mucha precisión entre las principales monedas. La razón es que el elevado grado de tecnificación de los mercados de divisas hace que los diferenciales de rentabilidad duren, de media, entre medio minuto y cuatro minutos (Akram, Rime y Sarno, 2008). Los incumplimientos de la paridad cubierta se asociaban a los controles de

capital y desaparecieron con la liberalización financiera de los años ochenta (tema B13, apartado 2.2.3). La paridad descubierta, por su parte, tiende a cumplirse a largo plazo (Chinn y Meredith, 2004). A ese horizonte, la tasa de depreciación se aproxima al diferencial de tipos de interés entre las dos economías.

Limitaciones: desde 2008 se observan desviaciones persistentes de la paridad cubierta (la llamada base de los *swaps* de divisas o *cross-currency basis*). Du, Tepper y Verdelhan (2018) las atribuyen a las restricciones de balance de los intermediarios financieros, que encarecen el arbitraje, y a los desequilibrios entre la demanda de inversión y la oferta de financiación en las distintas monedas. La paridad descubierta registra, además, desviaciones importantes a corto plazo, entre ellas el "*puzzle* de la prima *forward*" (Fama, 1984), objeto del apartado 3.4.2.

1.4. Conclusiones

Los modelos de flujos derivan el tipo de cambio de las transacciones que lo originan. En el modelo de balanza de pagos, los flujos comerciales alimentan el mercado de divisas, cuyo precio de vaciado es el tipo de cambio, y el equilibrio es estable si se cumple la condición de Marshall-Lerner. La PPA lo ancla, en cambio, en el poder de compra de las monedas. Su versión absoluta exige supuestos muy restrictivos, y la relativa, que liga la depreciación al diferencial de inflación, solo se sostiene a largo plazo. La hipótesis de Balassa-Samuelson y la fijación de precios según mercado explican sus desviaciones persistentes. Finalmente, la paridad de intereses traslada el análisis a los flujos de capitales. La cubierta, que fija el tipo *forward*, se cumplía con precisión hasta la crisis financiera global, mientras que la descubierta solo tiende a cumplirse a largo plazo.

Los tres modelos atienden a los flujos y no a los *stocks* de dinero y de activos financieros, cuyo peso creció con la liberalización de capitales, y de ese límite parten los modelos de activos de la sección 2.

2. Desarrollos posteriores: teorías basadas en los modelos de activos (*stocks*) de la balanza de pagos

Frente a las teorías tradicionales, los modelos de activos tienen en cuenta el efecto sobre el tipo de cambio de los *stocks* de dinero y de activos financieros, y no solo de sus variaciones, y se agrupan en dos familias. Los modelos monetarios hacen depender el tipo de cambio principalmente del *stock* de dinero, y los modelos de cartera consideran un espectro más amplio de activos financieros.

2.1. Antecedentes de los modelos monetarios: la teoría cuantitativa del dinero y el mecanismo flujo-especie (Hume, 1752)

Origen: los antecedentes de estos modelos se hallan en el mecanismo flujo-especie de David Hume (1752) y en la teoría cuantitativa del dinero.

Idea: ambas teorías ponían ya el énfasis en la oferta de dinero como determinante de los precios y, con ellos, del saldo exterior (con patrón oro, como en Hume) o del tipo de cambio (con tipo flexible).

Supuestos: mundo neoclásico internacional en el que:

- *Mercados*: el mercado de bienes es competitivo, con un bien homogéneo y libre movilidad de bienes, de modo que se cumple la PPA absoluta. El mercado monetario es perfecto, con libre movilidad de capitales, agentes neutrales al riesgo y sustituibilidad perfecta, por lo que se cumple la paridad de intereses.

Razonamiento: en cada país rige la ecuación de cambio, $MV = PY$ y $M^*V^* = P^*Y^*$. Según la teoría cuantitativa, que la lee en sentido causal, la velocidad y la renta no dependen a largo plazo de la cantidad de dinero, y el nivel de precios de cada país queda fijado por su oferta monetaria. A través de la PPA, el tipo de cambio de largo plazo hereda esa dependencia, lo que da la ecuación básica del enfoque monetario del tipo de cambio:

$$E = \frac{P}{P^*} = \frac{M \cdot V \cdot Y^*}{M^* \cdot V^* \cdot Y} \;\rightarrow\; E = f(M^+, V^+, Y^{*+}, M^{*-}, V^{*-}, Y^-)$$

Así, la oferta de dinero y la velocidad de circulación, variables monetarias, influyen en la determinación del tipo de cambio. Un aumento de la oferta monetaria o de la velocidad de circulación nacionales deprecia la moneda, y un aumento de la renta nacional la aprecia.

Conclusiones:

- *Aspectos positivos*: la política monetaria y cambiaria británica hasta la primera mitad del s. XIX estuvo marcada por estas teorías, que se utilizaban para explicar desde los desequilibrios externos y las variaciones de precios hasta la determinación de los tipos de cambio.

- *Limitaciones*: carecen del formalismo matemático de la literatura posterior.

2.2. Modelos monetarios

Origen: los modelos monetarios formalizan, desde mediados de los años setenta, esos antecedentes. Se desarrollan en dos vertientes, la de precios flexibles (Frenkel, 1976; Mussa, 1976) y la de precios rígidos (Dornbusch, 1976), y en una síntesis de ambas (Frankel, 1979).

Idea: en todas ellas, el equilibrio del mercado monetario de cada país determina el tipo de cambio nominal, que depende así de las ofertas y demandas relativas de dinero.

2.2.1. Precios flexibles: modelo de Frenkel (1976) y Mussa (1976)

Origen: Frenkel (1976) y Mussa (1976) formularon la versión de precios flexibles del enfoque monetario.

Idea: con precios plenamente flexibles y PPA continua, el tipo de cambio nominal queda determinado por el equilibrio del mercado monetario de cada país, es decir, por los diferenciales de oferta monetaria, renta y tipo de interés.

Supuestos:

- *Entorno*: dinámico, en el que se cumplen los supuestos anteriores (mundo neoclásico, PPA y paridad de intereses) de manera continua. Se añaden bonos perfectamente sustitutivos entre ambos países[5].

- *Agentes*: la demanda de dinero se modela según una función de elasticidad constante respecto a la renta y semielasticidad constante respecto al tipo de interés, con parámetros idénticos en ambos países.

Planteamiento: se parte de la ley de Walras en los mercados de bonos y de dinero de ambos países, $(M^d - M^s) + (B^d - B^s) = 0$, según la cual, si se vacía el mercado de dinero, también se vacía el de bonos. Basta, por tanto, con analizar el mercado de dinero. La función de demanda de saldos reales se representa por $M^d/P = Y^k \exp(-hi)$ y la condición de equilibrio exige $M^s/P = M^d/P$. Al tomar logaritmos neperianos se llega a la condición de equilibrio monetario de cada economía (en minúscula, el logaritmo de cada variable en mayúscula, salvo el tipo de interés, que va en nivel):

$$m - p = ky - hi \tag{1}$$

$$m^* - p^* = ky^* - hi^* \tag{2}$$

Dado que se cumple la PPA absoluta ($E = P/P^*$), el logaritmo del tipo de cambio nominal es $s \equiv \ln E = p - p^*$ (3). El tipo de cambio queda así ligado a los niveles de precios, que las condiciones (1) y (2) fijan a partir de la oferta y la demanda de dinero de cada país.

Equilibrio: restando (1) y (2) y sustituyendo (3) en la ecuación resultante, se llega a la ecuación básica del modelo monetario:

$$s = (m - m^*) - k(y - y^*) + h(i - i^*)$$

[5] Dado que los bonos de ambos países son sustitutivos perfectos, pueden agregarse en uno solo.

El tipo de cambio nominal depende, por tanto, de los diferenciales de masa monetaria, renta y tipo de interés entre los países. La paridad descubierta ($g_E^e = i - i^*$) y la PPA relativa en términos esperados ($g_E^e = \pi^e - \pi^{*e}$) implican juntas $i - i^* = \pi^e - \pi^{*e}$, con lo que la ecuación básica puede escribirse con la inflación esperada:

$$s = (m - m^*) - k(y - y^*) + h(\pi^e - \pi^{*e})$$

Estática comparativa: de la ecuación básica se derivan las siguientes predicciones:

- *Oferta monetaria*: un aumento relativo de la oferta monetaria nacional debilita la moneda. El aumento eleva en la misma proporción el nivel de precios nacional, de modo que la moneda nacional pierde poder adquisitivo, lo que, según la PPA, se traduce en una depreciación del tipo de cambio nominal.

- *Renta*: un aumento relativo de la renta nacional fortalece la moneda. Al aumentar la renta, sube la demanda de dinero y, con la oferta dada, el equilibrio monetario exige un nivel de precios menor, que, por la PPA, se traduce en una apreciación nominal.

- *Tipo de interés*: un aumento relativo del tipo de interés nominal nacional también debilita la moneda. Intuitivamente, un mayor tipo de interés nominal refleja, vía relación de Fisher, una mayor inflación esperada, que reduce la demanda de dinero, eleva el nivel de precios y, por la PPA, deprecia la moneda hoy.

- *Inflación*: un aumento relativo de la inflación provoca, de acuerdo con la PPA, una depreciación.

Extensiones (formas alternativas de expresar la ecuación básica): definiendo la demanda de dinero (en logaritmos y excluido el nivel de precios) como $l \equiv ky - hi$, la ecuación básica se puede reescribir de dos formas ilustrativas:

- *Ofertas y demandas relativas*: $s = (m - m^*) - (l - l^*)$ indica que el tipo de cambio depende positivamente del diferencial de ofertas monetarias y negativamente del diferencial de demandas de dinero. Cuanto más aumente la oferta monetaria nacional frente a la extranjera, más se depreciará la moneda, al ser menos escasa, y cuanto más aumente la demanda de moneda nacional, más se apreciará.

- *Excesos de oferta*: $s = (m-l)-(m^*-l^*)$ muestra que el tipo de cambio depende positivamente del exceso de oferta monetaria nacional, que deprecia la moneda, y negativamente del exceso de oferta monetaria extranjera, que deprecia la divisa y, por tanto, aprecia la moneda nacional.

Conclusiones:

- *Aspectos positivos*: el modelo explica la parte de largo plazo de la quinta regularidad de Mussa, según la cual las monedas de los países con inflación elevada, fruto de un fuerte crecimiento monetario, tienden a depreciarse aproximadamente al ritmo del diferencial de inflación.

- *Limitaciones*: bajo este enfoque, un diferencial creciente de tipos de interés nominales va asociado a una moneda cada vez más débil, predicción contraria a la evidencia, salvo durante episodios de inflación excepcionalmente alta. Tampoco se observa, fuera de esos episodios, la relación estrecha entre el tipo de cambio y los diferenciales de expansión monetaria que el modelo predice (quinta regularidad de Mussa). Además, el modelo, atractivo por su sencillez, se apoya en supuestos muy exigentes (vaciado continuo de mercados, PPA continua), lo cual fue el punto de partida de las teorías posteriores.

2.2.2. Precios rígidos: modelo de Dornbusch (1976)

Origen: Rudiger Dornbusch (1976), ante las limitaciones del modelo de Frenkel-Mussa, levanta el supuesto de precios flexibles y admite así desviaciones de la PPA a corto plazo, que explican la volatilidad del tipo de cambio nominal.

Idea: mientras los precios de los bienes se ajustan con lentitud, el tipo de cambio nominal, como el resto de precios financieros, responde casi al instante. Por ello, ante un cambio de los fundamentos del tipo de cambio (aquí, la oferta monetaria), la variable rápida compensa la inercia de la lenta, y el tipo de cambio se mueve a corto plazo más de lo que exige su nuevo nivel de largo plazo (sobrerreacción u *overshooting*).

Supuestos: los mismos que en el modelo de precios flexibles, con dos salvedades:

- *Entorno*: los precios son rígidos a corto plazo y flexibles a largo plazo, por lo que la PPA solo se cumple en el largo plazo.

- *Agentes*: los agentes forman sus expectativas racionalmente (hipótesis de expectativas racionales).

Planteamiento: el mercado de dinero se vacía según la condición (1), con la producción fija en su nivel de pleno empleo y y el nivel de precios exterior normalizado a cero ($p^* = 0$). El modelo añade dos relaciones:

$$i = i^* + \eta\,(\bar{s} - s), \qquad \eta > 0 \tag{4}$$

$$\dot{p} = \kappa\,(y^d - y), \qquad y^d = a_0 + \phi\,(s - p) + c\,y - v\,i \tag{5}$$

donde $\bar{s}$ es el tipo de cambio de largo plazo, η la velocidad a la que los agentes esperan que el tipo de cambio converja hacia él, $\kappa > 0$ la velocidad de ajuste de los precios (la pendiente de una curva de Phillips), y^d la demanda de producción nacional, a_0 el logaritmo del gasto autónomo, $s - p = q$ el tipo de cambio real y ϕ, c y v las sensibilidades de la demanda al tipo de cambio real, a la renta y al tipo de interés, como en la IS abierta del tema A32, apartado 1.2.1. La condición (4) es la paridad descubierta con *expectativas regresivas*, pues la depreciación esperada es proporcional a la distancia entre el tipo de cambio de largo plazo y el actual. La condición (5) recoge el ajuste lento de los precios, que suben solo en la medida en que la demanda supera a la producción.

Equilibrio: a largo plazo, $\dot{p} = 0$, $s = \bar{s}$ e $i = i^*$, de modo que la condición (1) fija $\bar{p} = m - ky + hi^*$ y, por la PPA, $\bar{s}$ varía en la misma proporción que $\bar{p}$. A corto plazo, en cambio, con p dado, el mercado de activos se vacía de forma instantánea. Restando a la condición (1), evaluada con (4), su versión de largo plazo, se obtiene

$$s = \bar{s} - \frac{1}{h\eta}\,(p - \bar{p}) \tag{6}$$

donde $p - \bar{p}$ es la distancia entre el nivel de precios y su valor de largo plazo. La ecuación (6) indica que, mientras los precios estén por debajo de ese valor, el tipo de cambio está más depreciado que

a largo plazo. Con expectativas racionales, η coincide con la velocidad de convergencia efectiva de la economía[6].

Desarrollo gráfico: el equilibrio inicial del modelo se representa en el plano saldos reales-tipo de interés (mercado de dinero) y en el plano tipo de cambio-tipo de interés (paridad de intereses), que comparten el eje vertical (figura B14.2):

1. *Oferta de saldos reales* (M_0^s/P_0): refleja la cantidad de dinero en términos reales, es decir, la oferta nominal que fija el banco central dividida por el nivel de precios. Es vertical porque no depende del tipo de interés, y a corto plazo, con los precios rígidos, solo se desplaza si cambia la oferta nominal.

2. *Demanda de saldos reales* (M^d/P): refleja la cantidad de dinero en términos reales que los agentes desean mantener para cada tipo de interés, dada la renta, según la condición (1). Es decreciente porque el tipo de interés es el coste de oportunidad de mantener dinero en lugar de bonos.

3. *Paridad descubierta de intereses* (PDI): recoge las combinaciones de tipo de interés y tipo de cambio actual con las que los bonos nacionales y extranjeros rinden lo mismo. Dado el tipo de cambio esperado, es decreciente, como indica la condición (4), porque un tipo de cambio actual más depreciado reduce la depreciación esperada y, con ella, el tipo de interés que exige la paridad.

4. *Equilibrio*: en el punto 0, la oferta y la demanda de saldos reales fijan el tipo de interés i_0 (panel izquierdo). La PDI asocia a ese tipo de interés el tipo de cambio E_0 (panel derecho), con el que los bonos nacionales y extranjeros rinden lo mismo.

[6] Sustituyendo (4) y (6) en (5) se obtiene $\dot{p} = -\nu\,(p-\bar{p})$, con $\nu = \kappa\,[\phi/(h\eta)+\phi+v/h]$, y, por (6), $\dot{s} = -\nu\,(s-\bar{s})$. Las expectativas regresivas son racionales si $\eta = \nu$, es decir, si $\eta = \frac{\kappa}{2}\left(\frac{v}{h}+\phi\right) + \sqrt{\frac{\kappa^2}{4}\left(\frac{v}{h}+\phi\right)^2 + \frac{\kappa\phi}{h}}$, raíz positiva de la ecuación de segundo grado resultante (Dornbusch, 1976).

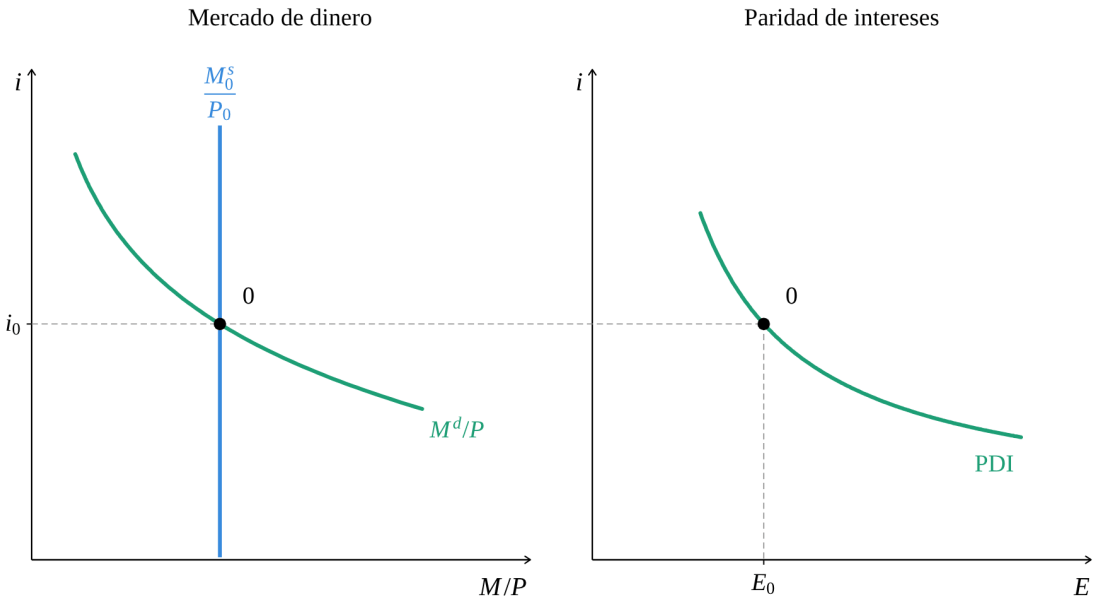

Figura B14.2: Equilibrio inicial del modelo de Dornbusch (1976) en el mercado de dinero y en la paridad descubierta de intereses.

Estática comparativa: se supone un aumento de la oferta monetaria nacional no anticipado y permanente[7], $dm > 0$. A largo plazo, $d\bar{p} = d\bar{s} = dm$ (neutralidad del dinero). A corto plazo, con $dp = 0$, la ecuación (6) da el resultado de sobrerreacción:

$$\frac{ds}{dm} = 1 + \frac{1}{h\eta} > 1 \tag{7}$$

donde el primer sumando es la depreciación de largo plazo y el segundo, el exceso de depreciación en el impacto. Por la condición (1), el tipo de interés cae en $di = -dm/h$. La sobrerreacción (7) es mayor cuanto menor es h, porque el tipo de interés debe caer más para que el mercado de dinero absorba la nueva oferta. También es mayor cuanto menor es η, porque una convergencia esperada más lenta exige una desviación mayor para generar la misma apreciación esperada. En la figura B14.3, el tipo de cambio pasa de E_0 a $E_{0'}$ en el impacto y a E_1 a largo plazo.

Dinámica:

▪ *Largo plazo*: con el tiempo, los precios suben en la misma proporción que la masa monetaria. Los saldos reales vuelven así a su nivel ($M_0^s/P_0 = M_1^s/P_1$), y con ellos el tipo de interés ($i_0 = i_1$). Por la PPA, en cambio, el tipo de cambio de largo plazo queda más depreciado que el inicial. Desde cualquier tipo de cambio actual, la depreciación esperada hacia ese nuevo nivel es entonces mayor ($\Delta E^e > 0$), y la PDI pasa de PDI_0 a PDI_1, desplazada hacia la derecha. El punto 1, con el tipo de cambio E_1, es el nuevo equilibrio de largo plazo.

[7] Si el aumento de la oferta monetaria fuese anticipado y permanente, los resultados cambiarían. Si la expansión se anuncia en t_0, se aplica en t_1 y el estado estacionario se alcanza en t_n, el tipo de cambio se deprecia instantáneamente con el anuncio. Después continúa depreciándose hasta la implementación (momento en el que alcanza su punto de sobrerreacción) y, a partir de ahí, se aprecia gradualmente hasta el nuevo nivel de largo plazo, más depreciado que el original.

▪ *Corto plazo*: como los precios no se mueven, la expansión monetaria eleva los saldos reales (de M_0^s/P_0 a M_1^s/P_0) y, para que el mercado de dinero siga equilibrado, el tipo de interés baja hasta $i_{0'}$. Con un i menor, la PDI solo se cumple si la depreciación esperada se reduce, es decir, si el mercado pasa a anticipar una apreciación futura.

▪ *Sobrerreacción*: esa apreciación esperada solo es posible si la moneda se deprecia hoy ($\uparrow E$) por encima de su nuevo nivel de largo plazo, hasta un punto desde el cual el mercado espera que se aprecie gradualmente. Como ese nivel de largo plazo ya está más depreciado que el original, la depreciación de impacto supera a la de largo plazo, como muestra (7), y la economía salta al punto $0'$ (figura B14.3).

▪ *Medio plazo*: tras el salto, la subida gradual de los precios reduce los saldos reales y devuelve i a su nivel previo, mientras la moneda se aprecia poco a poco. El tipo de cambio converge así a E_1, situado entre el inicial y el de impacto, y la economía pasa del punto $0'$ al 1 (figura B14.4).

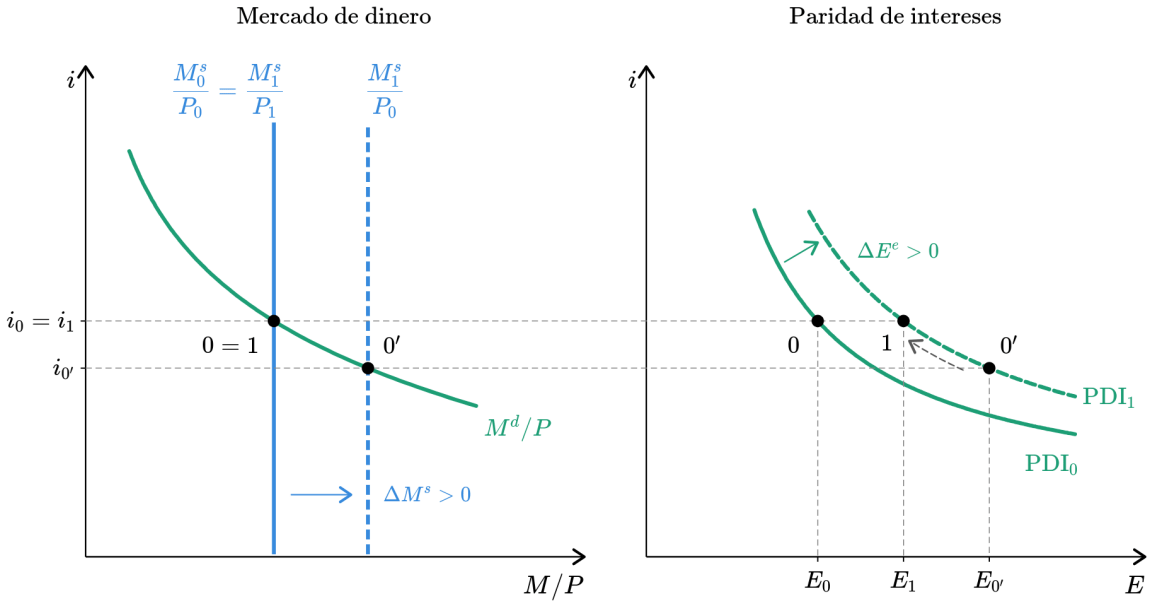

Figura B14.3: *Overshooting* del tipo de cambio en el modelo de Dornbusch ante una expansión monetaria no anticipada.

La figura B14.4 representa las trayectorias temporales de las tres variables, con el tiempo en el eje horizontal y la expansión monetaria en t_0. En el panel superior, el tipo de cambio salta en t_0 de E_0 a $E_{0'}$ y se aprecia después de forma gradual hasta E_1, más depreciado que E_0 y menos que $E_{0'}$. La distancia entre $E_{0'}$ y E_1 mide la sobrerreacción. En el panel central, el nivel de precios no varía en el impacto y sube poco a poco de P_0 a P_1. En el panel inferior, el tipo de interés cae en el impacto de i_0 a $i_{0'}$ y vuelve gradualmente a su nivel inicial ($i_1 = i_0$).

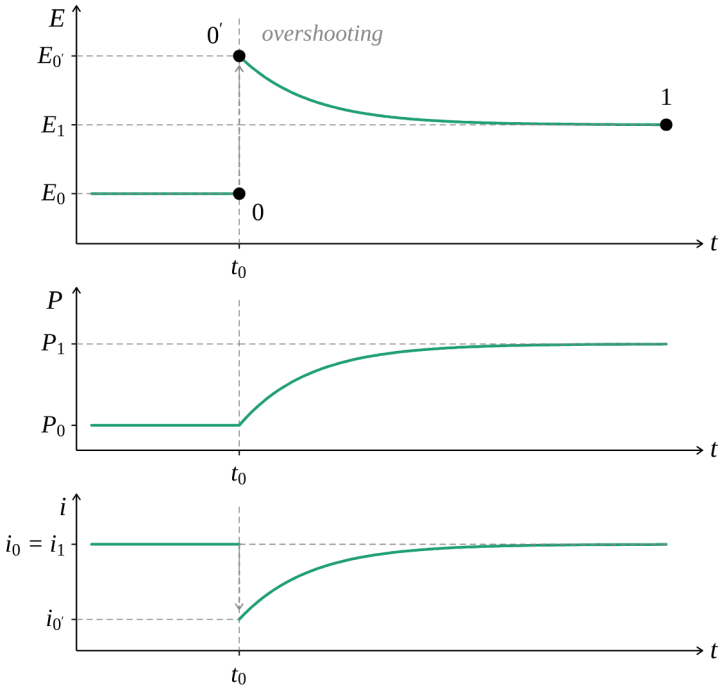

Figura B14.4: Trayectorias temporales del tipo de cambio, los precios y el tipo de interés tras el *shock* monetario en el modelo de Dornbusch.

Ejemplo: cuánto sobrerreacciona el tipo de cambio

El banco central aumenta por sorpresa y de forma permanente la oferta monetaria un 10 % ($dm = 0{,}1$). La semielasticidad de la demanda de dinero al tipo de interés es $h = 5$, es decir, un punto más de tipo de interés reduce los saldos reales demandados un 5 %, y los agentes esperan que cada año se cierre la mitad de la distancia al tipo de cambio de largo plazo ($\eta = 0{,}5$).

- *Largo plazo*: los precios y el tipo de cambio suben un 10 %, $d\bar{p} = d\bar{s} = 0{,}1$.
- *Impacto*: por (7), $ds = [1 + 1/(5 \cdot 0{,}5)] \cdot 0{,}1 = 0{,}14$. La moneda se deprecia un 14 %, cuatro puntos más que a largo plazo, y el tipo de interés cae dos puntos, $di = -0{,}1/5 = -0{,}02$.
- *Coherencia*: con $s - \bar{s} = 0{,}04$, la condición (4) implica una apreciación esperada de $\eta \cdot 0{,}04 = 0{,}02$ al año, justo lo que compensa los dos puntos menos de interés.

Conclusiones:

- *Aspectos positivos*: muestra que un tipo de cambio muy volátil a corto plazo puede ser coherente con sus fundamentos, pues esa volatilidad es la respuesta de equilibrio a los cambios de estos cuando los precios son rígidos. La misma rigidez traslada al tipo real, a corto plazo, las variaciones del nominal, lo que explica la tercera regularidad de Mussa, la estrecha correlación entre ambos. Para la política económica, la consecuencia es que un movimiento brusco del tipo de cambio puede ser racional y no prueba, por sí solo, que haya especulación desestabilizadora.

- *Limitaciones*: son cuatro las principales:

- *Parámetros*: la magnitud de la sobrerreacción depende críticamente de parámetros difíciles de estimar, señaladamente la semielasticidad de la demanda de dinero al tipo de interés y la velocidad de ajuste de los precios.
- *Producción fija*: la sobrerreacción solo está asegurada si la producción no responde a corto plazo. Si responde, el propio Dornbusch (1976) muestra que la sobrerreacción deja de ser necesaria y que el tipo de cambio puede depreciarse menos que la oferta monetaria.
- *Sustituibilidad perfecta*: este supuesto elimina por construcción la prima de riesgo cambiario, que es precisamente lo que el modelo de cartera incorpora.
- *Evidencia empírica*: el contraste empírico del *overshooting* ha resultado esquivo, pues la evidencia sugiere que la máxima depreciación se alcanza con retardo y no de forma instantánea (*delayed overshooting*, según Eichenbaum y Evans, 1995).

2.2.3. Síntesis: diferencial de intereses reales (Frankel, 1979)

Origen: Jeffrey Frankel (1979) parte de que los dos modelos anteriores predicen signos opuestos para el diferencial de intereses nominales. En el de precios flexibles, un diferencial mayor deprecia la moneda, mientras que en el de Dornbusch la aprecia.

Idea: el efecto del diferencial depende de su origen. Si refleja una política monetaria restrictiva, que eleva el tipo de interés real, la moneda se aprecia, y si refleja una mayor inflación esperada, se deprecia.

Planteamiento: Frankel añade a las expectativas regresivas de la condición (4) el diferencial de inflación esperada a largo plazo,

$$g_E^e = \eta\,(\bar{s} - s) + (\pi^e - \pi^{*e}) \tag{8}$$

donde π^e y π^{*e} son las inflaciones esperadas nacional y extranjera. Igualando (8) a la PDI ($g_E^e = i - i^*$) y tomando como $\bar{s}$ la ecuación básica con inflación, resulta

$$s = (m - m^*) - k(y - y^*) - \frac{1}{\eta}\,(i - i^*) + \left(\frac{1}{\eta} + h\right)(\pi^e - \pi^{*e}) \tag{9}$$

donde el diferencial de intereses nominales entra con signo negativo y el de inflación esperada, con signo positivo. Agrupando términos, (9) equivale a $s - \bar{s} = -\frac{1}{\eta}\,[(i - \pi^e) - (i^* - \pi^{*e})]$. Dicho de otro modo, la moneda se aprecia respecto de su nivel de largo plazo en proporción al diferencial de tipos de interés reales, de donde toma su nombre el modelo. Además, (9) anida los dos modelos anteriores, pues con ajuste instantáneo ($\eta \to \infty$) se obtiene el de precios flexibles con inflación y, sin inflación esperada ($\pi^e = \pi^{*e} = 0$), el de Dornbusch.

Conclusiones:

▪ *Aspectos positivos*: reconcilia en un solo marco contrastable los signos opuestos de los modelos monetarios, y su estimación para el marco alemán frente al dólar (1974-1978) obtuvo los signos previstos.

- *Limitaciones*: con el nombre de modelo de Dornbusch-Frankel, figura entre los que Meese y Rogoff (1983) contrastaron sin que esos modelos superaran al paseo aleatorio. Como los demás modelos monetarios, deja además sin explicar la elevada impredecibilidad de corto plazo del tipo de cambio, es decir, la primera regularidad de Mussa, lo que motiva los modelos de expectativas de la sección 3.

2.3. Modelo de cartera (McKinnon y Oates, 1966; Branson, 1968, 1977)

Origen: Ronald McKinnon (con Wallace Oates, 1966) y William Branson (1968) desarrollaron una serie de modelos basados en la aplicación de la teoría de selección de carteras de Harry Markowitz (1952) al estudio de la balanza de pagos. La versión de McKinnon se plantea con tipo de cambio fijo, de modo que los desequilibrios se reflejan en la balanza de pagos y en las reservas. El desarrollo gráfico, con tipo de cambio flexible y las curvas DD, BB y B*B*, corresponde en cambio a la versión de Branson (1977).

Idea: los desequilibrios de la balanza de pagos (con tipo de cambio fijo) o las variaciones del tipo de cambio (con tipo flexible) siguen estando causados por los desequilibrios del mercado monetario. Con todo, los desequilibrios monetarios se ven influidos también por las decisiones de cartera de los agentes económicos.

2.3.1. Supuestos

Entorno: el país es pequeño y abierto, con precios fijos y libre movimiento de capitales. A diferencia de la paridad de intereses (apartado 1.3.1) y de los modelos monetarios, existe sustituibilidad imperfecta entre los activos nacionales y extranjeros[8], debida a factores como la información imperfecta, el riesgo cambiario o el riesgo político. Se definen tres activos financieros:

- *Dinero*: M, que no ofrece rendimiento alguno.

- *Bonos nacionales*: B, que otorgan un rendimiento[9] i y cumplen dos características. No son comerciables internacionalmente, es decir, son irrelevantes para los compradores extranjeros, de modo que los flujos de capital vienen determinados exclusivamente por los flujos de bonos extranjeros. Tampoco son canjeables por bonos extranjeros. Por ello, un aumento de bonos extranjeros se compensa reduciendo la cantidad de dinero y no la de bonos nacionales.

- *Bonos extranjeros*: B^*, que otorgan un rendimiento i^* y sí son comerciables.

2.3.2. Modelo

Planteamiento: la demanda de activos financieros se divide entre el mercado de dinero y los mercados de bonos, y de cada activo se demanda una proporción de la riqueza, que viene dada por la oferta de activos:

- *Demanda de dinero*: $M^d = \omega_M \cdot W$, con $\omega_M = f(i^-, i^{*-})$.

- *Demanda de bonos nacionales*: $B^d = \omega_B \cdot W$, con $\omega_B = f(i^+, i^{*-})$.

[8] Se observa incluso entre países con libre circulación de capitales, como muestra el sesgo doméstico (*home bias*) de los bancos de la zona del euro hacia la deuda soberana de su propio país.

[9] Los bonos ofrecen un tipo de interés nominal, i para los nacionales e i^* para los extranjeros, que, en un contexto de precios fijos, coincide con su rendimiento real.

- *Demanda de bonos extranjeros*: $B^{d*} = \omega_{B^*} \cdot W$, con $\omega_{B^*} = f(i^-, i^{*+})$.

- *Oferta de activos*: la riqueza de la economía viene dada por la oferta de activos financieros, valorada en moneda nacional: $W = M^s + B^s + E \cdot B^{s*}$.

Equilibrio: por la restricción de riqueza se cumple la ley de Walras, $(M^d - M^s) + (B^d - B^s) + (B^{d*} - E \cdot B^{s*}) = 0$. Así, un exceso de demanda en el conjunto de los mercados nacionales se corresponde con un exceso de oferta en el de bonos extranjeros, y viceversa.

Desarrollo gráfico: en el plano tipo de interés nacional-tipo de cambio (figura B14.5), cada curva reúne las combinaciones de tipo de interés y tipo de cambio que vacían uno de los tres mercados:

1. *Curva de dinero* (DD): corresponde al mercado de dinero. Es creciente porque una depreciación ($\uparrow E$) revaloriza en moneda nacional la cartera de bonos extranjeros y, con ella, la riqueza. Con más riqueza se desea más dinero y, con la oferta monetaria fija, solo un tipo de interés más alto, el coste de oportunidad de mantenerlo, restablece el equilibrio.

2. *Curva de bonos nacionales* (BB): corresponde al mercado de bonos nacionales. Es decreciente porque el mismo efecto riqueza de una depreciación aumenta también la demanda de estos bonos. Con su oferta fija, la mayor demanda sube su precio y, por tanto, reduce su rentabilidad, i.

3. *Curva de bonos extranjeros* (B*B*): corresponde al mercado de bonos extranjeros. Es decreciente, porque la depreciación eleva en la misma cuantía el valor en moneda nacional de la oferta de bonos extranjeros y la riqueza. La demanda de estos bonos, en cambio, solo aumenta en la fracción $\omega_{B^*} < 1$ de ese aumento de riqueza[10]. Surge así un exceso de oferta de bonos extranjeros, que solo se absorbe si disminuye el tipo de interés nacional, lo que los hace relativamente más atractivos. Ahora bien, como la demanda de bonos extranjeros es menos sensible al tipo de interés nacional que la de bonos nacionales, la curva B*B* presenta menor pendiente que la BB. Ante una misma variación del tipo de cambio, la rentabilidad debe variar en mayor medida.

4. *Equilibrio*: en el punto 0, la intersección de las tres curvas (DD_0, BB_0 y $B^*B^*_0$) determina el tipo de interés i_0 y el tipo de cambio E_0 a corto plazo. En realidad, si dos de estos mercados están en equilibrio, por la ley de Walras también debe estarlo el tercero[11].

[10] Formalmente, la condición de equilibrio del mercado de bonos extranjeros es $E \cdot B^{s*} = \omega_{B^*}(i^-, i^{*+}) \cdot W$, con $W = M^s + B^s + E \cdot B^{s*}$. Una depreciación $dE > 0$ eleva la oferta, valorada en moneda nacional, en $B^{s*} \cdot dE$, y la demanda solo en $\omega_{B^*} \cdot B^{s*} \cdot dE$. Como $\omega_{B^*} < 1$, surge un exceso de oferta $(1 - \omega_{B^*}) \cdot B^{s*} \cdot dE$ que exige una caída de i para restablecer el equilibrio.

[11] Aquí hay vaciado de mercado porque se supone plena flexibilidad de los precios de los activos y del tipo de interés nacional.

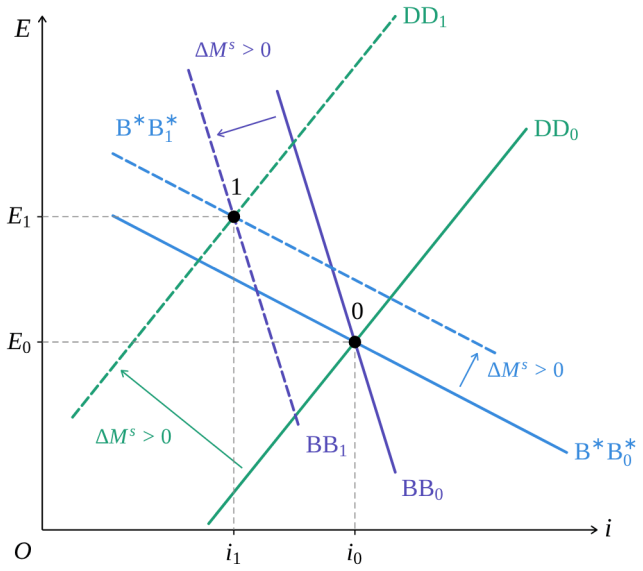

Figura B14.5: Modelo de equilibrio de cartera. El corte de las curvas de dinero (DD), bonos nacionales (BB) y bonos extranjeros (B*B*) determina el tipo de interés y el tipo de cambio a corto plazo (0). Una expansión monetaria reduce el tipo de interés y deprecia la moneda nacional ($0 \rightarrow 1$).

Estática comparativa: las implicaciones de política económica son diversas:

▪ *Política monetaria expansiva*: con más dinero, el mercado monetario solo se vacía, para cada tipo de cambio, con un tipo de interés menor, y DD se desplaza hacia la izquierda. La mayor oferta monetaria es también más riqueza[12], que se dirige en parte a los bonos. La demanda adicional de bonos nacionales sube su precio y baja su rentabilidad, lo que lleva BB hacia la izquierda. La de bonos extranjeros exige vender moneda nacional a cambio de divisa, lo que deprecia la moneda y eleva B*B*. El resultado es un tipo de interés menor y una moneda nacional más depreciada, y la economía pasa del punto 0 al 1 (i_1, E_1).

▪ *Política fiscal expansiva financiada con deuda*: la emisión de nuevos bonos nacionales los abarata y eleva su rentabilidad, y BB se desplaza hacia la derecha. Los nuevos bonos son también riqueza, con la que crece la demanda de dinero y de bonos extranjeros. La demanda de bonos extranjeros presiona hacia la depreciación y eleva B*B*, y la de dinero sube el tipo de interés y lleva DD hacia la derecha. El tipo de interés sube, pero el efecto sobre el tipo de cambio es ambiguo. El efecto riqueza empuja hacia la depreciación, al estimular la demanda de bonos extranjeros, y el mayor tipo de interés nacional, hacia la apreciación, al desviar demanda de los bonos extranjeros a los nacionales.

2.3.3. Conclusiones

Aspectos positivos: las implicaciones son similares a las del modelo monetario, pero, al no ser los bonos nacionales sustitutivos perfectos de los extranjeros, un desequilibrio monetario puede

[12] La riqueza aumenta porque el dinero forma parte de ella ($W = M^s + B^s + E \cdot B^{s*}$). Ello presupone un aumento neto de la oferta monetaria, ya que, si la expansión se instrumentase mediante una operación de mercado abierto (compra de bonos nacionales con dinero), cambiaría la composición de la riqueza, pero no su volumen.

compensarse en parte en el mercado de bonos nacionales. En la versión de tipo fijo de McKinnon, ese desequilibrio no se traslada íntegramente al saldo externo ni se corrige solo vía reservas, sino también mediante el tipo de interés, lo que otorga a la política monetaria un mayor grado de autonomía para corregir los desequilibrios internos.

Limitaciones: el modelo mantiene los precios fijos y las expectativas estáticas, por lo que no explica la dinámica de sobrerreacción ni la impredecibilidad de corto plazo, lo que da paso a los modelos de expectativas.

2.4. El contraste empírico: la paradoja de Meese-Rogoff y el tipo de cambio real de equilibrio

La paradoja de Meese y Rogoff (1983): el contraste sistemático de los modelos estructurales examinados arrojó un resultado muy desfavorable para ellos. Meese y Rogoff compararon su capacidad predictiva fuera de muestra con la de un simple *paseo aleatorio*, es decir, con la regla de predecir que el tipo de cambio de mañana será el de hoy. Concluyeron que ninguno de ellos lo superaba a horizontes de uno a doce meses, ni siquiera alimentándolos con los valores efectivamente observados de los fundamentos, lo que elimina el error de previsión de estos como explicación. El resultado ha resistido cuatro décadas de réplicas y constituye el punto de partida de la literatura posterior.

La paradoja se refiere, en todo caso, al corto plazo. A horizontes largos, algunos estudios (Mark, 1995) encuentran que los modelos basados en fundamentos sí superan al paseo aleatorio, aunque este resultado es menos robusto, de modo que el problema parece ser de horizonte más que de irrelevancia de los fundamentos.

El tipo de cambio real de equilibrio: la dificultad para predecir no impide preguntarse si una moneda está desalineada respecto del valor compatible con los fundamentos, cuestión de primer orden para la vigilancia multilateral que ejercen el FMI y la Comisión Europea. Se emplean con este fin dos enfoques complementarios:

- *FEER* (*fundamental equilibrium exchange rate*, Williamson, 1983): enfoque normativo que define el tipo de cambio de equilibrio como aquel compatible con el equilibrio interno (pleno empleo con inflación estable) y con el equilibrio externo (un saldo por cuenta corriente sostenible). Su dificultad es que exige fijar previamente cuál es ese saldo sostenible, lo que introduce un elemento de juicio.

- *BEER* (*behavioural equilibrium exchange rate*, Clark y MacDonald, 1998): enfoque positivo que estima econométricamente la relación histórica entre el tipo de cambio real y un conjunto de fundamentos (posición de inversión internacional neta, diferencial de productividad, relación real de intercambio, diferencial de tipos de interés). Considera desalineamiento la desviación respecto del valor ajustado.

En el marco del FMI, ambos enfoques se integran en la metodología del *External Balance Assessment*, que combina la estimación de saldos por cuenta corriente y de tipos de cambio reales de equilibrio para evaluar los desequilibrios externos de sus miembros.

2.5. Conclusiones

Los modelos de activos determinan el tipo de cambio por el equilibrio de *stocks*. Sus antecedentes (flujo-especie y teoría cuantitativa) hacían depender los precios, y con ellos el tipo de cambio, de la oferta de dinero. Con precios flexibles, Frenkel y Mussa lo ligan a los diferenciales de masa

monetaria, renta y tipo de interés, lo que explica la depreciación a largo plazo de las monedas con inflación elevada. Con precios rígidos, Dornbusch obtiene la sobrerreacción del tipo de cambio ante un *shock* monetario, y Frankel concilia ambos modelos mediante el diferencial de intereses reales. El modelo de cartera introduce, por último, la sustituibilidad imperfecta entre activos, de modo que un desequilibrio monetario se compensa en parte en el mercado de bonos nacionales.

Ninguno de estos modelos supera, sin embargo, a un paseo aleatorio a corto plazo (Meese y Rogoff, 1983), aunque a horizontes largos algunos estudios dan ventaja a los fundamentos. La dificultad para predecir no impide estimar tipos de cambio reales de equilibrio para medir desalineamientos. De la impredecibilidad a corto plazo parten los modelos de expectativas de la sección 3.

3. Desarrollos recientes: expectativas, microestructura e intermediación financiera

La paradoja de Meese-Rogoff desplaza la pregunta fundamental sobre determinación de tipos de cambio. Si los fundamentos no explican el tipo de cambio a corto plazo, hay que atender a cómo se forman las expectativas y a cómo se negocia efectivamente en el mercado. Los desarrollos que siguen relajan, cada uno, un supuesto distinto de los modelos anteriores.

3.1. Expectativas racionales: noticias y burbujas

El primer bloque conserva la hipótesis de expectativas racionales y muestra que es compatible con un tipo de cambio impredecible a corto plazo, por la llegada de noticias, e incluso con desviaciones respecto de los fundamentos, por la formación de burbujas racionales.

3.1.1. El papel de la nueva información (Frenkel y Mussa, 1980)

Origen: Frenkel y Mussa (1980) parten de las grandes fluctuaciones de los tipos de cambio entre las principales monedas desde la generalización de la flotación en 1973, que interpretan con la teoría de los mercados de activos.

Idea: solo lo que el mercado no esperaba mueve el tipo de cambio, y lo mueve de forma brusca e inmediata. Esa nueva información son las "noticias" sobre sus fundamentos, como una decisión de tipos de interés del banco central que sorprende al mercado.

Razonamiento: en el enfoque de activos (sección 2), el tipo de cambio es un precio de activos y, como tal, descuenta hoy lo que los agentes esperan de sus fundamentos. Lo ya anticipado está, por tanto, en la cotización actual, que solo cambia cuando la información nueva obliga a revisar esas expectativas.

Conclusiones:

- *Aspectos positivos*: explica la impredecibilidad de corto plazo sin abandonar la racionalidad, pues si el tipo de cambio solo se mueve ante lo no anticipado, su variación es por construcción imprevisible.

- *Limitaciones*: es la lectura más benévola de la paradoja de Meese-Rogoff, porque no explica por qué los modelos fallan incluso con los valores observados de los fundamentos (apartado 2.4).

3.1.2. Burbujas especulativas racionales (Blanchard y Watson, 1982)

Origen: Blanchard y Watson (1982) se preguntan si las burbujas de los precios de los activos son compatibles con la hipótesis de expectativas racionales y si pueden detectarse estadísticamente.

Idea: las expectativas de los agentes pueden desviarse de los fundamentos y dar lugar a *burbujas* sin que los agentes dejen de formarlas racionalmente.

Razonamiento: que la burbuja llegue a formarse depende de lo que los agentes sepan sobre su estallido:

- *Fecha de estallido conocida*: no cabe burbuja, pues nadie querría mantener el activo sobrevalorado en el momento del estallido y, por inducción hacia atrás, tampoco en ningún momento anterior, de modo que no llegaría a formarse.

- *Estallido estocástico*: en cambio, sí es posible una burbuja que en cada periodo sobrevive con una probabilidad dada y estalla con la complementaria. Mientras dura, la burbuja crece a un ritmo suficiente para compensar a sus tenedores por el riesgo de estallido, de modo que resulta racional mantener el activo aunque todos los agentes, con la misma información, sepan que está sobrevalorado.

Conclusiones:

- *Aspectos positivos*: demuestra que desviación respecto de los fundamentos y racionalidad no son incompatibles, de modo que la existencia de burbujas no obliga a abandonar el marco de expectativas racionales ni a suponer información imperfecta.

- *Limitaciones*: la idea de que los agentes, sin saber cuándo estallará la burbuja, entran en el mercado confiando en salir antes del estallido no cabe en este marco y exige un desarrollo posterior con información asimétrica.

3.2. Expectativas heterogéneas: chartistas, fundamentalistas y caos

El segundo bloque relaja la homogeneidad de los agentes y la linealidad de la dinámica. La heterogeneidad de expectativas entre chartistas y fundamentalistas explica la persistencia de los desalineamientos, y el caos determinista ofrece otra lectura de la impredecibilidad del tipo de cambio.

3.2.1. Expectativas heterogéneas (Frankel y Froot, 1990)

Origen: Frankel y Froot (1990) parten de la apreciación del dólar entre 1980 y comienzos de 1985, cuya última fase atribuyen a una burbuja especulativa que los fundamentos no explican.

Idea: el mercado reúne dos clases de predictores. Los *chartistas* extrapolan la tendencia reciente de la cotización y los *fundamentalistas* esperan su vuelta al valor que dictan los fundamentos.

Razonamiento: los gestores de cartera, que toman las posiciones y fijan con ellas el tipo de cambio, combinan ambas predicciones y dan más peso a la del grupo que mejor ha acertado recientemente. Por ello, si la apreciación se prolonga, los chartistas ganan peso y alejan aún más el tipo de cambio de su valor fundamental. Los fundamentalistas, que prevén esa conducta de los chartistas, pueden no confiar en una corrección próxima y dejar de apostar por ella.

Conclusiones:

- *Aspectos positivos*: explica la persistencia de los desalineamientos, es decir, por qué el tipo de cambio puede separarse de los fundamentos durante periodos prolongados sin que el arbitraje lo corrija, y anticipa el enfoque de las finanzas conductuales aplicado al mercado de divisas.

- *Limitaciones*: sus resultados dependen de las reglas de predicción, en parte *ad hoc*, que se atribuyen a cada grupo de agentes.

3.2.2. Caos determinista (De Grauwe y Dewachter, 1993)

Origen: De Grauwe y Dewachter (1993) construyen un modelo monetario del tipo de cambio en el que interactúan los chartistas y los fundamentalistas.

Idea: el tipo de cambio puede responder a una dinámica de *caos determinista*, en la que pequeñas variaciones de las condiciones iniciales generan grandes diferencias en el comportamiento futuro, y parecer por ello aleatorio.

Razonamiento: en apariencia, el tipo de cambio nominal es un paseo aleatorio, cuyo mejor predictor es el valor de hoy. Sus perturbaciones no están correlacionadas, pero tampoco son independientes entre sí, y esa dependencia sugiere una dinámica no lineal. En el modelo, la interacción entre ambos tipos de agentes es no lineal y genera movimientos caóticos que reproducen algunos hechos estilizados del tipo de cambio.

Conclusiones:

- *Aspectos positivos*: ofrece una explicación alternativa de la paradoja de Meese-Rogoff, pues una serie generada por reglas deterministas no lineales resulta en la práctica indistinguible de una aleatoria. La impredecibilidad no probaría entonces la ausencia de estructura, sino su complejidad.

- *Limitaciones*: por esa misma razón, la hipótesis es difícil de contrastar con las muestras disponibles.

3.3. Microestructura del mercado de divisas (Evans y Lyons, 2002)

Origen: Evans y Lyons (2002) desplazan el foco desde los fundamentos macroeconómicos hacia la microestructura del mercado. Estiman su modelo con las operaciones entre creadores de mercado del marco alemán y el yen frente al dólar (sistema de Reuters, de mayo a agosto de 1996).

Idea: el *flujo de órdenes* (*order flow*), es decir, la presión compradora o vendedora neta que llega a los creadores de mercado, es el determinante inmediato de las variaciones diarias del tipo de cambio.

Razonamiento: los clientes trasladan a los creadores de mercado cambios de cartera que no son de conocimiento común. Los creadores de mercado observan el flujo de órdenes que intercambian entre sí, infieren esos cambios y ajustan sus cotizaciones. Como los activos en distintas monedas no son sustitutivos perfectos, el tipo de cambio se mueve hasta que el público absorbe la nueva cartera. Con esta variable, el modelo explica el 64 % de las variaciones diarias del marco frente al dólar y el 45 % de las del yen, una fracción muy superior a la que explican las variables macroeconómicas.

Conclusiones:

- *Aspectos positivos*: identifica el mecanismo por el que la información dispersa y heterogénea de los agentes se incorpora al precio, y con ello el eslabón que faltaba entre los fundamentos y la cotización observada.

- *Limitaciones*: el flujo de órdenes es él mismo una variable endógena, de modo que explica el cómo pero no el porqué último de los movimientos.

3.4. Desconexión, primas de riesgo e intermediación financiera

Desde la década de 2000, la literatura explica, sin abandonar los fundamentos, por qué un tipo de cambio que depende de ellos puede comportarse como un paseo aleatorio y apenas relacionarse con la economía real. Este fenómeno se conoce como *desconexión del tipo de cambio* (*exchange rate disconnect*, Obstfeld y Rogoff, 2000).

3.4.1. Fundamentos y paseo aleatorio (Engel y West, 2005)

Origen: Engel y West (2005) reexaminan la paradoja de Meese-Rogoff (apartado 2.4) en los modelos de activos, en los que el tipo de cambio es el valor presente de los fundamentos esperados.

Idea: si los fundamentos siguen un paseo aleatorio y el factor de descuento es próximo a uno, el tipo de cambio se comporta casi como un paseo aleatorio, aunque dependa de ellos.

Razonamiento: en el modelo de precios flexibles, sustituyendo la PDI ($i - i^* = \mathbb{E}_t s_{t+1} - s_t$) en la ecuación básica y resolviendo hacia delante, sin burbujas, se obtiene

$$s_t = (1-\beta)\sum_{j=0}^{\infty} \beta^j \, \mathbb{E}_t f_{t+j}, \qquad \beta = \frac{h}{1+h} \tag{10}$$

donde $f_t = (m_t - m_t^*) - k(y_t - y_t^*)$ recoge los fundamentos y β es el factor de descuento. Por (10), con fundamentos integrados de orden uno y β próximo a uno, la variación del tipo de cambio responde sobre todo a revisiones de expectativas sobre fundamentos lejanos y es casi impredecible.

Conclusiones:

- *Aspectos positivos*: reconcilia la paradoja de Meese-Rogoff con los modelos de fundamentos y explica por qué el tipo de cambio, como precio de un activo, ayuda a predecir los fundamentos, como confirman los datos.

- *Limitaciones*: el resultado exige un factor de descuento muy próximo a uno y deja sin explicar por qué el tipo de cambio es mucho más volátil que los fundamentos observados.

3.4.2. El *puzzle* de la prima *forward* y el *carry trade* (Fama, 1984; Burnside et al., 2011)

Origen: la regresión de Fama (1984), con nueve monedas frente al dólar entre 1973 y 1982, muestra que las monedas con tipos de interés altos tienden a apreciarse, en lugar de depreciarse como exige la PDI, de modo que el *carry trade* obtiene en promedio rendimientos positivos. El *puzzle* y la estrategia se desarrollan en el tema B13, apartados 3.2 y 2.2.4.

Idea: la PDI fallaría porque omite una prima de riesgo cambiario, ligada a un suceso raro, y no por irracionalidad de los agentes.

Razonamiento: Burnside, Eichenbaum, Kleshchelski y Rebelo (2011) muestran que esos rendimientos son elevados y no están correlacionados con los factores de riesgo tradicionales. Más que a pérdidas muy grandes de la estrategia, los atribuyen a un *problema del peso* (*peso problem*, apartado 1.3.4), un suceso raro, apenas presente en las muestras, en el que el factor de descuento estocástico es muy elevado.

Conclusiones:

- *Aspectos positivos*: interpreta el incumplimiento de la PDI como una prima de riesgo, compatible con la racionalidad.

- *Limitaciones*: un suceso raro es por definición difícil de observar, de modo que la explicación es difícil de contrastar.

3.4.3. Intermediación financiera (Gabaix y Maggiori, 2015; Itskhoki y Mukhin, 2021)

Origen: Gabaix y Maggiori (2015) proponen una teoría del tipo de cambio basada en los flujos de capital en mercados financieros imperfectos, cuyas fricciones había puesto de relieve la crisis financiera global. Itskhoki y Mukhin (2021) buscan explicar con un solo modelo los *puzzles* del tipo de cambio.

Idea: el tipo de cambio lo determinan intermediarios financieros con capacidad limitada para asumir riesgo, que absorben los desequilibrios entre la oferta y la demanda de divisas a cambio de una rentabilidad esperada. En consecuencia, la PDI deja de cumplirse y los flujos financieros mueven el tipo de cambio.

Razonamiento:

- *Gabaix y Maggiori (2015)*: los intermediarios exigen una rentabilidad esperada mayor cuanto mayor es la posición en divisas que absorben y cuanto menor es su capacidad de asumir riesgo. Por ello, los flujos comerciales y de cartera afectan al nivel y a la volatilidad del tipo de cambio, el *carry trade* remunera ese riesgo y la intervención en el mercado de divisas es eficaz (tema B15, apartado 3.1). Con capacidad ilimitada, en cambio, se recupera la PDI.

- *Itskhoki y Mukhin (2021)*: integran el mecanismo en un modelo de equilibrio general. Las demandas de divisas de inversores ajenos a los fundamentos (*noise traders*), absorbidas por arbitrajistas aversos al riesgo, generan perturbaciones financieras que dominan la variación del tipo de cambio y que, por el sesgo doméstico del consumo, apenas afectan a la producción y al consumo. El modelo reproduce así la desconexión, la paradoja de Meese-Rogoff y el *puzzle* de la prima *forward*.

Conclusiones:

- *Aspectos positivos*: explican la desconexión y el *puzzle* de la prima *forward* con un único mecanismo y devuelven a los flujos de capital, como el modelo de cartera (apartado 2.3), un papel en la determinación del tipo de cambio.

- *Limitaciones*: las perturbaciones financieras son en buena medida exógenas, de modo que estos modelos, como el flujo de órdenes, desplazan la pregunta hacia el origen de esas demandas.

3.5. Conclusiones

Los desarrollos recientes permiten dar cuenta de la primera regularidad de Mussa, la elevada impredecibilidad de corto plazo, que los modelos de flujos y de activos dejaban sin explicar. Lo hacen relajando supuestos distintos de los modelos anteriores, como la información completa (Frenkel-Mussa), la ausencia de burbujas (Blanchard-Watson), la homogeneidad de los agentes (Frankel-Froot), la linealidad (De Grauwe-Dewachter), la irrelevancia de la estructura del mercado (Evans-Lyons) o la capacidad ilimitada de los intermediarios para asumir riesgo

(Gabaix-Maggiori, Itskhoki-Mukhin). La mayoría conserva la racionalidad de los agentes, que los modelos de chartistas y fundamentalistas sustituyen por reglas de predicción en parte *ad hoc*.

Su carácter fragmentario es, sin embargo, su principal debilidad, pues no constituyen un cuerpo unificado sino explicaciones parciales y en parte alternativas del mismo fenómeno. Ninguno de ellos ofrece, además, una capacidad predictiva sistemáticamente superior al paseo aleatorio, de modo que aciertan al explicar por qué el tipo de cambio es impredecible pero no consiguen predecirlo.

Conclusiones

Los modelos de flujos (balanza de pagos, paridad del poder adquisitivo y paridad de intereses) explican, a través de la PPA relativa, la regularidad de largo plazo identificada por Mussa, la depreciación de las monedas con inflación elevada. Las desviaciones persistentes de la PPA se explican por la hipótesis de Balassa-Samuelson y por la fijación de precios según mercado. Los modelos de activos (monetarios y de cartera) incorporan los *stocks* y, con el *overshooting* de Dornbusch, la sobrerreacción del tipo de cambio. Sin embargo, la paradoja de Meese y Rogoff (1983) muestra que, a horizontes cortos, ninguno de ellos predice el tipo de cambio mejor que un paseo aleatorio, ni siquiera conociendo *ex post* los fundamentos. Los modelos de expectativas y de microestructura abordan esa impredecibilidad. Los enfoques son, así, complementarios, porque las paridades y los modelos monetarios anclan el largo plazo y las expectativas, las noticias y el flujo de órdenes gobiernan el corto.

Las teorías del tipo de cambio parten de las paridades de intereses, que relacionan el mercado de contado con el de plazo y el tipo de cambio con los rendimientos de los activos (tema B13). Comparten, además, la condición de Marshall-Lerner y los enfoques monetario y de cartera con las teorías del ajuste de la balanza de pagos, de las que los modelos de activos son la versión con tipo de cambio flexible (tema B12). Por su parte, la hipótesis de Balassa-Samuelson explica por qué una unión monetaria puede sufrir diferenciales de inflación persistentes entre sus miembros (tema B16). Por último, la dificultad de predecir y de gestionar el tipo de cambio condiciona la elección del régimen cambiario (tema B15).

Itskhoki y Mukhin (2021) atribuyen la desconexión entre el tipo de cambio y sus fundamentos (el *exchange rate disconnect puzzle*) a perturbaciones financieras que separan el tipo de cambio de la paridad descubierta (apartado 3.4.3). Su modelo reproduce, además, la lenta reversión del tipo de cambio real hacia la PPA. Sigue abierto, sin embargo, si la desconexión persiste. Lilley, Maggiori, Neiman y Schreger (2022) encuentran que, desde la crisis financiera global, las compras de activos extranjeros por los inversores estadounidenses explican una parte apreciable de los movimientos del dólar, una "reconexión" con variables financieras que antes no se observaba.

B15. Regímenes cambiarios

Análisis comparado de los distintos regímenes cambiarios. Intervención y regulación de los mercados de cambio.

Introducción

Introducción común. Los temas B13 a B15 comparten una introducción, que se desarrolla al comienzo del tema **B13**. Presenta el tipo de cambio, "el precio más importante de una economía" según Volcker, el mercado de divisas en que se determina y el debate recurrente sobre qué determina los tipos de cambio y qué régimen cambiario es preferible.

El **régimen cambiario** o sistema de tipos de cambio de una economía es el conjunto de reglas con el que un país ordena la relación entre el valor de su moneda y el de las demás divisas.

Desde el punto de vista teórico, existen dos regímenes cambiarios puros o polares:

- *Tipo de cambio fijo*: la autoridad monetaria anuncia una paridad o tipo de cambio nominal[1] frente a una divisa o una cesta de referencia y asume el compromiso de sostenerla.

- *Tipo de cambio flexible*: la autoridad monetaria no se fija ningún objetivo cambiario, y el tipo de cambio nominal lo determinan libremente la oferta y la demanda de divisas.

Los regímenes observados (*de facto*)[2] adoptan, sin embargo, con frecuencia formas intermedias, como el tipo de cambio reptante o las bandas objetivo.

La libertad de elección del régimen cambiario que reconocieron los Acuerdos de Jamaica (1976) ha llevado a algunos autores, como John Williamson (1976), a hablar de un "no sistema". Esa libertad tiene, no obstante, límites, pues cada país debe notificar su régimen al FMI y respetar las obligaciones generales del artículo IV de su Convenio Constitutivo, entre ellas evitar la manipulación del tipo de cambio para obtener ventajas competitivas desleales.

Como consecuencia de esa libertad se observan dos fenómenos:

- *Tendencia hacia la flexibilidad*: respecto de Bretton Woods, han ganado peso los regímenes de tipo de cambio más flexibles. Aunque solo en torno a un tercio de los países tiene un régimen de cambio flexible (apartado 2.3.2), esos países concentran la mayor parte del PIB mundial.

[1] La paridad se entiende aquí como valor comparativo de una moneda con otra (Real Academia Española), no como igualdad, sentido que tiene en la paridad del poder adquisitivo.

[2] La clasificación *de facto* de los regímenes cambiarios del FMI puede consultarse en el *Annual Report on Exchange Arrangements and Exchange Restrictions* (AREAER), que publica anualmente.

- *Crisis cambiarias*: desde la década de los noventa se han producido diversas crisis cambiarias que ponen de manifiesto la relevancia de la política cambiaria en la estabilización de las variables macroeconómicas.

La elección entre unos regímenes y otros se ordena en torno al **trilema de Mundell-Fleming**, según el cual un país no puede tener a la vez un tipo de cambio fijo, libertad de movimientos de capital y una política monetaria autónoma. Tras las crisis de los años noventa, Fischer (2001) sostuvo que solo los extremos, la vinculación fuerte o la flotación, eran sostenibles con libertad de capitales. Rey (2013) ha matizado incluso esa conclusión, al mostrar que el ciclo financiero mundial limita la autonomía monetaria también de los países con tipo de cambio flexible.

Por todos estos motivos, el objetivo de este tema es el estudio comparado de los distintos regímenes cambiarios, en tres partes:

- *Regímenes teóricos*: el tema compara primero los regímenes teóricos o polares.

- *Regímenes observados*: se analizan después los regímenes observados (*de facto*).

- *Intervención y regulación*: la última parte estudia la intervención y la regulación en los mercados de cambio.

1. Análisis comparado de los regímenes cambiarios teóricos

La comparación parte de los dos regímenes polares, el fijo y el flexible, y del problema que plantea elegir entre ellos. Después se contrastan sus efectos por el lado de la oferta, por el de la demanda y en el ajuste externo.

1.1. El problema de la elección del régimen de cambio

Al elegir un régimen cambiario entre los dos polos teóricos, el flexible y el fijo, las autoridades económicas persiguen, generalmente, el doble objetivo de garantizar simultáneamente el equilibrio interno y el equilibrio externo de la economía.

Por **equilibrio interno** se entiende mantener la actividad económica en un nivel próximo a la producción potencial, es decir, con una brecha de producción (*output gap*) reducida, de manera que se garanticen el pleno empleo de los recursos y una inflación estable. Por **equilibrio externo** se entiende mantener el saldo de las transacciones autónomas de la balanza de pagos en una posición equilibrada y sostenible. Para lograr estos objetivos, más allá del recurso a la política fiscal y la política monetaria, las autoridades utilizan la **política cambiaria** (o política de tipos de cambio), que es la que establece, en última instancia, un régimen cambiario u otro.

Estas decisiones remiten al **trilema de Mundell-Fleming** (trinidad imposible), pues en un contexto abierto al exterior existen limitaciones a la aplicación y selección de dichas políticas, especialmente desde el punto de vista monetario. Concretamente, el trilema, formulado a partir de los trabajos de Mundell y Fleming, postula que un país no puede disfrutar simultáneamente de un tipo de cambio fijo, de libre movilidad de capitales y de una política monetaria autónoma. Siempre hay que renunciar a uno de los tres vértices.

El país puede, por tanto, elegir entre tres combinaciones:

- *Tipo fijo y libre movilidad de capitales*: pierde la autonomía monetaria (el caso de los miembros de la zona del euro).

- *Autonomía monetaria y libre movilidad*: debe dejar flotar su moneda (el caso de EE. UU. o el Reino Unido).

- *Tipo fijo y autonomía monetaria*: debe restringir los movimientos de capitales (el caso de China durante décadas).

La elección del régimen cambiario es, pues, inseparable de la elección sobre la apertura financiera y el uso de la política monetaria[3].

De todo lo anterior se desprende que no existe una combinación unívoca que convenga a todos los países, sino que la decisión sobre el régimen cambiario depende de las características de cada economía. Por ello, los apartados siguientes comparan cada régimen y los factores que las autoridades pueden tener en cuenta al elegir, por el lado de la oferta, por el de la demanda y en el ajuste de la balanza de pagos.

[3] Una actualización influyente del trilema es la de Rey (2013), según la cual la existencia de un ciclo financiero global convierte el trilema en un dilema, porque sin gestión de los flujos de capitales la política monetaria pierde autonomía incluso bajo tipos de cambio flexibles.

1.2. Lado de la oferta

Desde el punto de vista de la oferta, los efectos del régimen cambiario se dejan notar, en especial, en el mercado de trabajo, ya que la apertura exterior afecta a la estructura económica, de la que el mercado de trabajo es parte fundamental.

1.2.1. Régimen cambiario flexible

En el corto plazo, los efectos de las variaciones del tipo de cambio sobre el empleo dependen, esencialmente, de tres aspectos:

1. *Existencia de rigideces*: en ausencia de rigideces nominales, una variación del tipo de cambio nominal se traslada íntegramente a precios y salarios, deja inalterado el tipo de cambio real y resulta neutral sobre el empleo. Con precios rígidos, en cambio, las variaciones nominales se trasladan de forma casi unitaria al tipo de cambio real y con notable persistencia, alterando la competitividad-precio y, con ella, la demanda de trabajo de los sectores expuestos a la competencia exterior.

2. *Duración de las perturbaciones*: si las variaciones son transitorias, el efecto recae sobre el componente cíclico del empleo (p. ej. una apreciación puede elevar el desempleo en los sectores exportadores por la caída de las ventas internacionales)[4]. Si, por el contrario, las variaciones son lo suficientemente persistentes como para alterar el tipo de cambio real de equilibrio, pueden afectar al desempleo estructural, e incluso a las decisiones de localización de las empresas.

3. *Volatilidad cambiaria*: cuando es elevada, genera incertidumbre sobre el rendimiento esperado de los proyectos y puede reducir la inversión extranjera y, en último término, el empleo, si bien la evidencia empírica al respecto no es concluyente, en parte por la disponibilidad de instrumentos de cobertura.

En el largo plazo, el nivel de empleo viene determinado por factores reales e institucionales, que fijan la tasa de desempleo estructural o tasa de paro que no acelera la inflación (NAIRU), y no por el tipo de cambio nominal. Este resulta neutral una vez que precios y salarios se han ajustado. El tipo de cambio conserva, no obstante, un papel en la asignación de recursos, ya que su libre determinación por las fuerzas del mercado facilita el ajuste de la economía ante perturbaciones reales y la reasignación de factores hacia sus usos más productivos. La condición es que no concurran distorsiones que impidan ese ajuste.

1.2.2. Régimen cambiario fijo

En el corto plazo, desde el punto de vista de la oferta, un régimen de tipo de cambio fijo ofrece una doble ventaja:

1. *Anclaje de expectativas en la negociación salarial*: las organizaciones empresariales, los sindicatos y demás agentes implicados en la fijación salarial saben que el mantenimiento de la paridad restringe las posibilidades de aplicar una política monetaria expansiva y otras medidas inflacionistas. Por ello, pueden anclar sus expectativas de inflación. La moderación salarial resultante contiene los costes laborales unitarios y, con ellos, la tasa de desempleo estructural.

[4] Con costes de ajuste de plantilla, sin embargo, las empresas no ajustan el empleo ante variaciones transitorias del tipo de cambio, lo que atenúa este efecto.

2. *Disciplina competitiva*: al no disponer de la devaluación como instrumento, las ganancias de competitividad han de proceder de la productividad y de la contención de costes, lo que puede incentivar mejoras en la eficiencia productiva.

Ambas ventajas están, con todo, condicionadas a la credibilidad del compromiso cambiario. Si los agentes no internalizan la restricción y la inflación doméstica supera a la del país ancla, se produce una apreciación real que erosiona la competitividad sin que quepa corregirla por la vía nominal.

En el largo plazo, el principal coste es la renuncia al tipo de cambio como mecanismo de ajuste ante perturbaciones reales asimétricas, de modo que el ajuste recae sobre el empleo y la devaluación interna.

1.3. Lado de la demanda: perturbaciones y eficacia de las políticas

Las ventajas e inconvenientes de cada régimen por el lado de la demanda se resumen en su respuesta ante perturbaciones y en la eficacia de las políticas de demanda en el modelo de Mundell-Fleming con perfecta movilidad del capital. Se compara primero la respuesta de cada régimen ante *shocks* reales y monetarios, y después, la eficacia de las políticas de demanda con tipo flexible y con tipo fijo. La tabla B15.1 recoge sus resultados. Con movilidad imperfecta, en cambio, la eficacia de la política fiscal disminuye con la movilidad en régimen flexible y aumenta con ella en régimen fijo.

> El modelo IS-LM-BP, los cuatro casos de referencia y la tabla de los dieciséis casos según régimen y movilidad se desarrollan en el tema A32, apartado 1.3.

1.3.1. Respuesta ante *shocks* reales y monetarios

En función del régimen cambiario, la economía tiene mayor o menor capacidad para estabilizar la producción ante un *shock* real de demanda. Un ejemplo es el desplome de la inversión por el deterioro de las expectativas, como el asociado a la crisis financiera global de 2008 o a la pandemia de COVID-19.

***Shock* real con régimen cambiario flexible**: gráficamente, en el plano renta-tipo de interés del modelo IS-LM-BP (figura B15.1), la curva IS refleja las combinaciones de renta y tipo de interés que equilibran el mercado de bienes. Tiene pendiente negativa, porque un tipo de interés menor estimula la inversión y la demanda. La LM refleja las que equilibran el mercado de dinero y tiene pendiente positiva, porque una renta mayor eleva la demanda de dinero y, con la oferta dada, el tipo de interés. La BP refleja las que equilibran la balanza de pagos y es horizontal en $i = i^*$ por la perfecta movilidad del capital. La economía parte del punto 0, en el corte de las tres curvas, y el *shock* ($\Delta I_0 < 0$) desplaza la IS a la izquierda, de IS_0 a IS_1:

- *Efecto contractivo interno (desplazamiento de 0 a 0′)*: el *shock* provoca una reducción de la inversión debido a las malas expectativas, lo que supone una caída de la demanda agregada y, con ello, de la renta nacional, hasta $Y_{0'}$. Paralelamente, el tipo de interés se ajusta a la baja en el mercado de dinero, hasta $i_{0'}$, para acomodar el *shock*.

- *Efecto expansivo externo (desplazamiento de 0′ a 1)*: la caída de la renta causa un descenso de las importaciones, que mejora la balanza comercial. Además, la caída del tipo de interés provoca, *ceteris paribus*, una salida de capitales del país, por el cumplimiento de la paridad de intereses,

lo que supone una depreciación de la moneda nacional ($\Delta E > 0$) con la consiguiente mejora de la competitividad-precio de las exportaciones. La IS vuelve a IS_0 y el punto 1 coincide con el 0.

Con un régimen de cambio flexible y perfecta movilidad del capital, la economía es capaz de absorber el *shock*. La razón es que el tipo de cambio flexible lo aísla, al permitir que la mejora de la balanza comercial compense el desequilibrio interno.

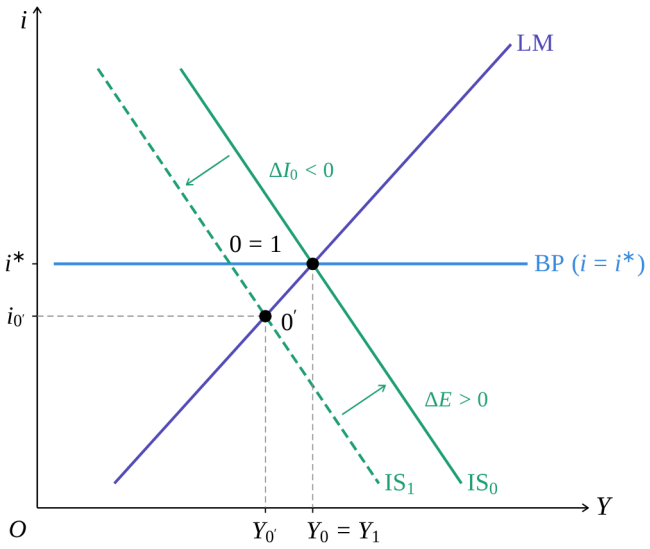

Figura B15.1: *Shock* real negativo con tipo de cambio flexible. La depreciación compensa la caída de la demanda y la producción vuelve a su nivel inicial.

***Shock* real con régimen cambiario fijo**: en el mismo plano (figura B15.2), el *shock* desplaza de nuevo la curva IS de IS_0 a IS_1:

▪ *Efecto contractivo interno (desplazamiento de 0 a 0′)*: como en el caso anterior, la caída de la inversión reduce la demanda agregada y la renta hasta $Y_{0'}$, y el tipo de interés baja hasta $i_{0'}$.

▪ *Efecto contractivo externo (desplazamiento de 0′ a 1)*: la caída del tipo de interés genera una salida de capitales y presiones depreciatorias sobre la moneda. Para defender la paridad, la autoridad monetaria debe vender divisas y comprar moneda nacional ($\Delta R < 0$), lo que contrae la oferta monetaria, desplaza la LM de LM_0 a LM_1 y agrava la caída de la renta, hasta Y_1.

Un régimen de cambio fijo no es, pues, capaz de aislar el *shock*, debido al efecto contractivo de la política monetaria de defensa de la paridad.

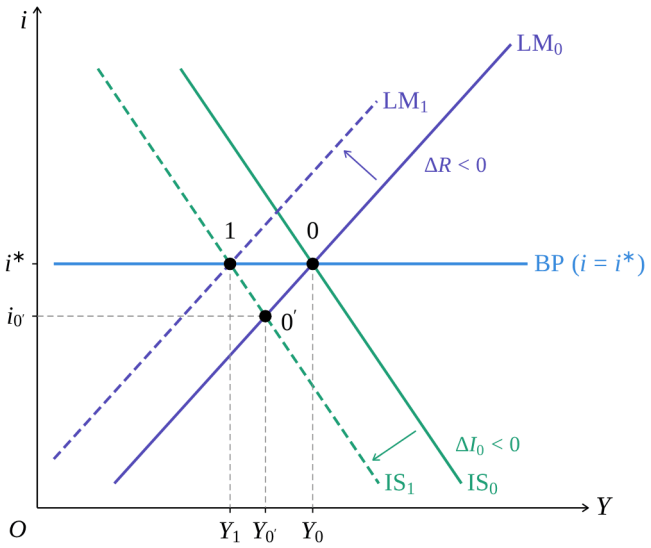

Figura B15.2: *Shock* real negativo con tipo de cambio fijo. El ajuste recae íntegramente sobre producción y empleo.

***Shock* monetario**: ante una perturbación monetaria interna, el resultado se invierte. En el mismo plano (figuras B15.3 y B15.4), un aumento de la demanda de dinero ($\Delta M^d > 0$) desplaza la curva LM a la izquierda, de LM_0 a LM_1. La economía pasa del punto 0 al $0'$, con un tipo de interés $i_{0'}$ superior a i^* y una renta $Y_{0'}$. El ajuste posterior depende del régimen:

- *Tipo de cambio flexible*: la entrada de capitales aprecia la moneda ($\Delta E < 0$), la IS se desplaza de IS_0 a IS_1 y la renta cae hasta Y_1 (punto 1).

- *Tipo de cambio fijo*: el banco central compra divisas para defender la paridad ($\Delta R > 0$), la LM vuelve a LM_0 y la economía regresa al punto de partida (0 = 1).

El tipo de cambio fijo aísla la producción de los *shocks* monetarios internos, porque la variación de reservas acomoda la oferta monetaria a la demanda de dinero. El flexible, en cambio, traslada la perturbación al tipo de cambio y, a través de él, a la producción.

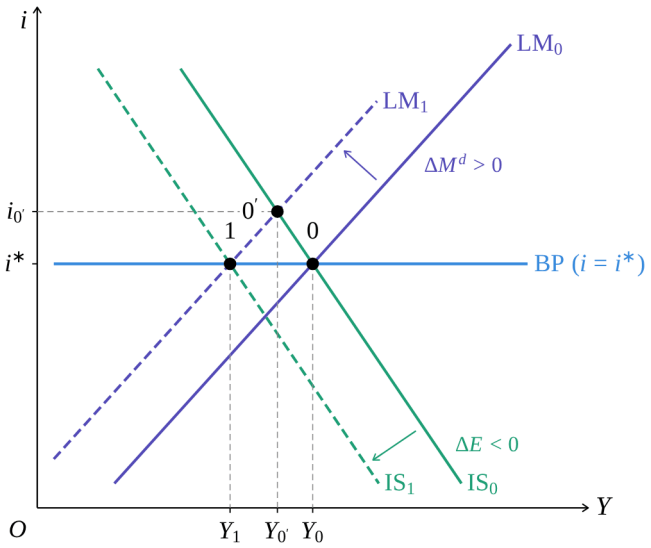

Figura B15.3: *Shock* monetario con tipo de cambio flexible. La perturbación se transmite al tipo de cambio y, a través de él, a la producción.

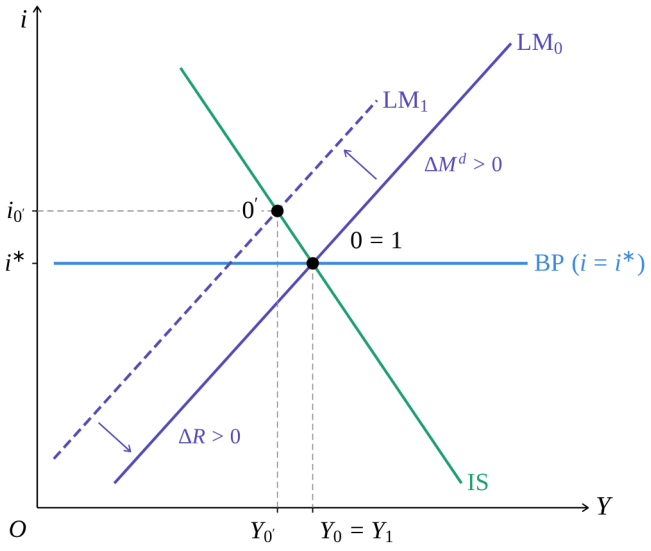

Figura B15.4: *Shock* monetario con tipo de cambio fijo. La variación de reservas aísla a la producción de la perturbación.

1.3.2. Políticas de demanda con régimen cambiario flexible

Política monetaria: gráficamente, en el plano renta-tipo de interés (figura B15.5), con las curvas IS, LM y BP ya descritas, una expansión monetaria ($\Delta M^s > 0$) traslada la curva LM a

la derecha, de LM_0 a LM_1:

- *Efecto expansivo interno (desplazamiento de 0 a 0′)*: el aumento de la masa monetaria desplaza la curva LM. La consiguiente disminución del tipo de interés nominal estimula la inversión, lo que aumenta la demanda agregada y la renta nacional (desplazamiento sobre la curva IS).

- *Efecto expansivo externo (desplazamiento de 0′ a 1)*: la caída del tipo de interés provoca, *ceteris paribus*, una salida de capitales que deprecia la moneda nacional ($\Delta E > 0$), abaratando relativamente las exportaciones y generando una ganancia de competitividad-precio que refuerza la expansión. La IS se desplaza de IS_0 a IS_1 y la renta alcanza Y_1.

En resumen, con un tipo de cambio flexible y libertad de movimiento de capitales, la política monetaria es eficaz, debido al efecto depreciatorio que supone la caída del tipo de interés nacional y al cumplimiento de la condición de paridad de intereses.

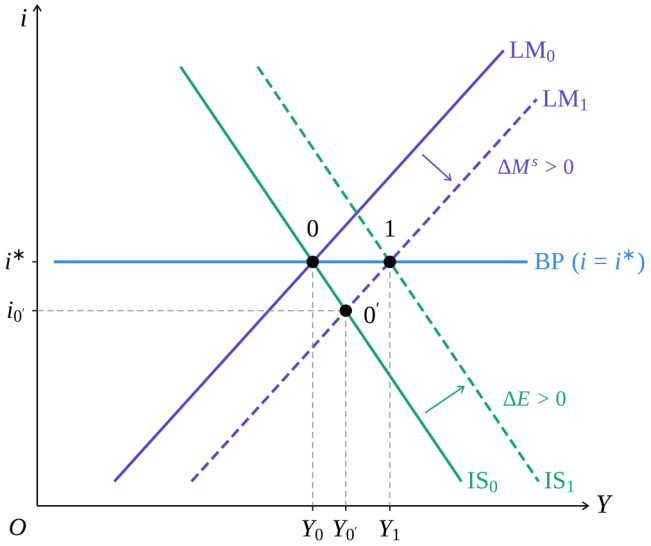

Figura B15.5: Política monetaria con tipo de cambio flexible. La depreciación la hace plenamente eficaz.

Política fiscal: en el mismo plano (figura B15.6), el aumento del gasto público ($\Delta G_0 > 0$) desplaza la curva IS a la derecha, de IS_0 a IS_1:

- *Efecto expansivo interno (desplazamiento de 0 a 0′)*: el aumento de las compras del Gobierno supone un estímulo a la demanda agregada[5], que da lugar a un aumento de la renta nacional, con el consiguiente aumento del tipo de interés para equilibrar el mercado financiero.

- *Efecto contractivo externo (desplazamiento de 0′ a 1)*: el aumento de la renta provoca un aumento de las importaciones, sin que las exportaciones varíen[6], lo que deteriora la balanza comercial. Además, el aumento del tipo de interés provoca, *ceteris paribus*, una entrada de capitales que aprecia la moneda nacional. La apreciación ($\Delta E < 0$) encarece relativamente las

[5] En una economía abierta el estímulo será siempre menor que en una economía cerrada, ya que parte de la demanda interior es de bienes extranjeros.

[6] Las exportaciones dependen de la renta exterior y del tipo de cambio real, no de la renta nacional.

exportaciones, con la consiguiente pérdida de competitividad-precio. La balanza comercial se deteriora aún más y el estímulo inicial queda compensado. La IS vuelve a IS_0 y el punto 1 coincide con el 0.

En resumen, con un tipo de cambio flexible y libertad de movimiento de capitales, la política fiscal es ineficaz. La apreciación que provoca la entrada de capitales reduce la demanda exterior neta en la misma cuantía en que el gasto público eleva la demanda interna.

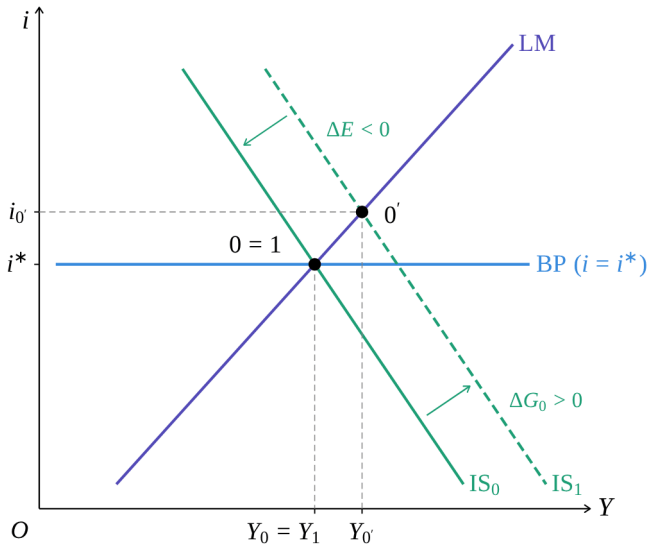

Figura B15.6: Política fiscal con tipo de cambio flexible. La apreciación y el efecto expulsión exterior la hacen ineficaz.

Riesgo de indisciplina y *overshooting*: precisamente porque la política monetaria es eficaz, el Gobierno y el banco central pueden verse tentados a usarla de forma discrecional para reactivar la economía a corto plazo, con un sesgo inflacionista (efecto indisciplina).

Además, la propia dinámica de la política monetaria bajo cambio flexible genera volatilidad cambiaria. Así lo muestra la hipótesis del *overshooting* de Rudiger Dornbusch (1976), según la cual, si los precios de los bienes se ajustan más lentamente que los mercados financieros, el tipo de cambio sobrerreacciona en el corto plazo ante los cambios en la oferta monetaria. Ante una expansión, se deprecia inicialmente por encima de su nuevo nivel de largo plazo, hacia el que converge después apreciándose.

> El modelo de *overshooting* de Dornbusch se desarrolla en el tema B14.

1.3.3. Políticas de demanda con régimen cambiario fijo

Política monetaria: gráficamente, en el plano renta-tipo de interés (figura B15.7), la expansión monetaria ($\Delta M^s > 0$) traslada de nuevo la curva LM a la derecha, de LM_0 a LM_1:

▪ *Efecto expansivo interno (desplazamiento de 0 a 0′)*: el aumento de la masa monetaria y la consiguiente disminución del tipo de interés estimulan la inversión, la demanda agregada y la renta nacional.

▪ *Efecto contractivo externo (desplazamiento de 0′ a 1)*: la caída del tipo de interés provoca una salida de capitales y, con ella, presiones depreciatorias sobre la moneda nacional, reforzadas por el aumento de las importaciones que induce la mayor renta. Para defender la paridad, la autoridad monetaria debe vender divisas y comprar moneda nacional ($\Delta R < 0$), drenando la liquidez que había inyectado, de modo que la oferta monetaria retorna a su nivel inicial y el estímulo se desvanece. La LM vuelve a LM_0 y el punto 1 coincide con el 0.

En resumen, con un tipo de cambio fijo y libertad de movimiento de capitales, la política monetaria es ineficaz, porque el compromiso cambiario obliga a revertir la expansión.

Una política monetaria expansiva solo podría influir en el nivel de renta si la movilidad del capital fuera imperfecta y las autoridades esterilizaran los flujos de reservas exteriores (p. ej. mediante compras adicionales de títulos públicos en operaciones de mercado abierto). El efecto monetario de la pérdida de reservas se compensaría así total o parcialmente y se mantendría el efecto expansivo (es decir, se evitaría que la LM se desplazase hacia la izquierda)[7].

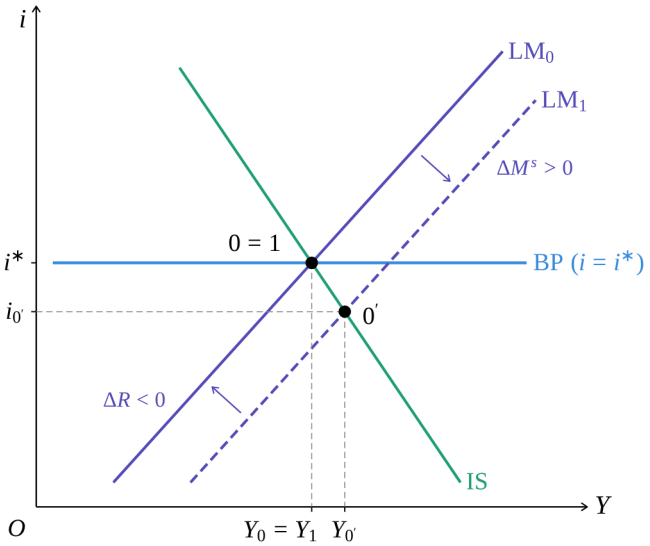

Figura B15.7: Política monetaria con tipo de cambio fijo. La pérdida de reservas la hace ineficaz.

Política fiscal: en el mismo plano (figura B15.8), el aumento del gasto público ($\Delta G_0 > 0$) desplaza la curva IS a la derecha, de IS_0 a IS_1:

▪ *Efecto expansivo interno (desplazamiento de 0 a 0′)*: de nuevo, las compras del Gobierno estimulan la demanda agregada y, con ello, la renta nacional, hasta $Y_{0'}$, y el tipo de interés sube hasta $i_{0'}$.

[7] En todo caso, la esterilización de las pérdidas de reservas no puede mantenerse indefinidamente si no se toman al mismo tiempo medidas de ajuste del déficit exterior. Mientras este persista, continuará la salida de reservas hasta que un ataque especulativo liquide las restantes antes de su agotamiento gradual, lo que dará lugar a una crisis cambiaria de primera generación.

▪ *Efecto expansivo externo (desplazamiento de 0′ a 1)*: la subida del tipo de interés atrae capitales y genera presiones apreciatorias sobre la moneda nacional. Dado que la autoridad monetaria debe mantener la paridad, acomoda estas presiones interviniendo en el mercado de divisas. Compra divisas y vende (es decir, emite) moneda nacional, en una intervención no esterilizada ($\Delta R > 0$) que equivale a una política monetaria expansiva de acompañamiento y desplaza la LM de LM_0 a LM_1, de modo que la renta alcanza Y_1 (punto 1).

En consecuencia, con un tipo de cambio fijo y libertad de movimiento de capitales, la política fiscal es muy eficaz, debido a que la defensa de la paridad obliga a la autoridad monetaria a una intervención que equivale a una expansión monetaria.

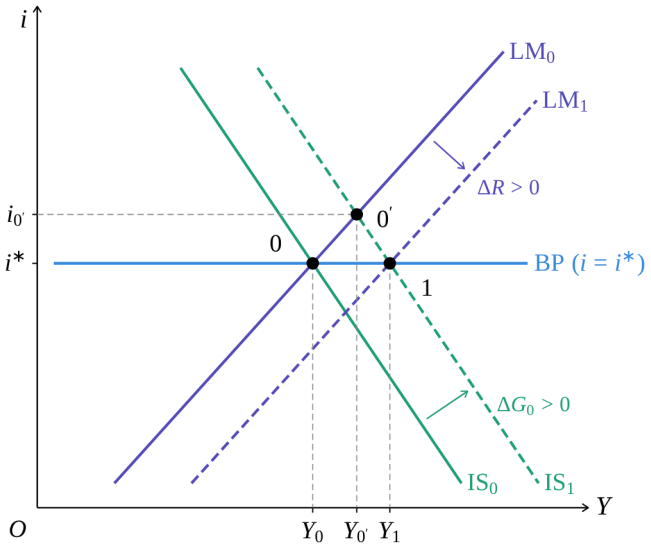

Figura B15.8: Política fiscal con tipo de cambio fijo. La acomodación monetaria la hace plenamente eficaz.

1.4. Diferencias en el ajuste externo

Régimen de cambio flexible:

▪ *Ajuste*: el ajuste de la balanza de pagos es más rápido y automático que en un régimen de cambio fijo, puesto que recae sobre el propio tipo de cambio. Así, un déficit exterior genera presiones depreciatorias que abaratan las exportaciones y encarecen las importaciones, corrigiendo el desequilibrio sin necesidad de intervención. Ahora bien, ese automatismo está sujeto, por un lado, a que se cumpla la condición de Marshall-Lerner y, por otro, a que haya transcurrido el tiempo suficiente para que las cantidades reaccionen. El plazo es necesario porque, en un primer momento, el efecto precio domina sobre el efecto cantidad y la balanza comercial se deteriora (efecto curva J).

▪ *Cuenta corriente*: con un sistema de tipo de cambio flexible, el comercio de bienes y servicios no tiene por qué verse excesivamente afectado por las fluctuaciones del tipo de cambio, siempre y cuando estas sean transitorias y de escasa magnitud. Una volatilidad elevada y persistente, en cambio, opera como un coste añadido para el comercio, si bien atenuado por la disponibilidad de instrumentos de cobertura.

▪ *Cuenta financiera*: bajo este tipo de régimen, el país no incurre en el coste de oportunidad de mantener reservas para defender una paridad, aunque en la práctica muchas economías con flotación acumulan reservas por motivos de precaución. Sin embargo, la flotación no evita otros problemas, como las fugas de capitales, las paradas súbitas (*sudden stops*) o el *pecado original* (Eichengreen y Hausmann, 1999). El término designa la imposibilidad del país de endeudarse en el exterior en su propia moneda, con el consiguiente riesgo de haberse endeudado sin cobertura cambiaria adecuada[8]. En este último caso, por ejemplo, la depreciación podría dejar de ser expansiva, ya que, al elevar el valor en moneda nacional de los pasivos exteriores, deteriora los balances de empresas, bancos y sector público, y puede resultar contractiva.

Régimen de cambio fijo:

▪ *Ajuste*: el ajuste de la balanza de pagos es más lento y costoso, puesto que el tipo de cambio nominal deja de operar como variable de ajuste. La corrección discurre entonces por dos vías. La primera es automática y opera a través de las reservas. Así, un déficit exterior obliga al banco central a vender divisas para sostener la paridad, lo que, si no se esteriliza, contrae la oferta monetaria, reduce la demanda interna y los precios y restablece la competitividad. La segunda vía es la *devaluación interna*, es decir, el ajuste a la baja de precios y salarios nominales, proceso lento por la rigidez a la baja de estos últimos y que recae sobre el empleo. Solo cuando el desequilibrio resulta insostenible se recurre a una devaluación o revaluación de la paridad, que no es propiamente un mecanismo de ajuste, sino una modificación del régimen.

▪ *Asimetría del ajuste*: estos regímenes reparten de forma desigual la carga del ajuste entre países. Los países con déficit exterior se enfrentan a una restricción dura, pues sus reservas son finitas, mientras que los superavitarios pueden acumular reservas indefinidamente. La carga del ajuste recae, por tanto, sobre los primeros.

▪ *Cuenta corriente*: la mayor certidumbre sobre la evolución futura del tipo de cambio puede reducir los costes de información y el riesgo de cambio, lo que puede estimular las operaciones comerciales[9] y la inversión extranjera. El efecto es tanto mayor cuanto más elevado sea el peso del país ancla en el comercio exterior del país que fija, pues el anclaje elimina la volatilidad bilateral, pero no necesariamente la efectiva.

▪ *Cuenta financiera*: el mantenimiento de un volumen de reservas, generalmente invertidas en activos exteriores de bajo rendimiento, supone un coste de oportunidad[10], y pueden producirse crisis cambiarias si el compromiso no resulta creíble, debido a ataques especulativos que fuercen el abandono de la paridad. Ahora bien, según evidencian los modelos de crisis de primera y segunda generación, lo problemático no es el tipo de cambio fijo como tal, sino aquellos factores que restan credibilidad a ese compromiso.

[8] Endeudarse en divisa (en dólares en la mayor parte del mundo, en euros en el este de Europa) suele ser más barato y, con mercados financieros locales poco desarrollados, a veces es la única vía de financiación. A cambio, una devaluación puede desencadenar una crisis de deuda soberana, una crisis bancaria o ambas. Incluso el mero aumento del riesgo percibido de devaluación (o de redenominación, como en la crisis de deuda soberana europea) puede provocar una huida de capitales que deprime el crédito y deteriora las cuentas públicas.

[9] Es más, Rose (2000), a través de un modelo gravitacional aplicado a las uniones monetarias, encuentra que los miembros de una unión monetaria comercian entre sí en torno a tres veces más. Incluso excluyendo el componente de la unión monetaria, concluye que la variabilidad cambiaria influye de manera significativa en el comercio entre los países.

[10] Rodrik (2006) sitúa el coste de oportunidad de las reservas excesivas en torno al 1 % del PIB de los países en desarrollo, y otros cálculos lo elevan hasta cerca del 2 %.

1.5. Conclusiones

La tabla B15.1 resume, por bloques temáticos, las ventajas comparativas de cada régimen polar, a partir de lo examinado por el lado de la oferta, por el de la demanda y en el ajuste externo.

Tabla B15.1: Ventajas comparativas del tipo de cambio flexible y del fijo.

Dimensión	Tipo de cambio flexible	Tipo de cambio fijo
Anclaje nominal e inflación	Sin ancla externa: la credibilidad antiinflacionista ha de proceder de la propia política monetaria (p. ej. objetivos de inflación)	**Ventaja**: la paridad ancla las expectativas y disciplina la negociación salarial y la política monetaria
Autonomía monetaria (trilema)	**Ventaja**: se conserva con libre movilidad de capitales	Se renuncia a ella, salvo que se restrinjan los movimientos de capital
Política monetaria	**Eficaz**: la depreciación inducida refuerza el estímulo	*Ineficaz*: la defensa de la paridad revierte la expansión
Política fiscal	*Ineficaz*: la apreciación inducida compensa el estímulo	**Muy eficaz**: la intervención no esterilizada acompaña al estímulo
***Shocks* reales asimétricos**	**Ventaja**: el tipo de cambio absorbe la perturbación y aísla el producto	El ajuste recae sobre cantidades (empleo) y devaluación interna
***Shocks* monetarios internos**	La perturbación se traslada al tipo de cambio y, a través de él, a la producción (figura B15.3)	**Ventaja**: la intervención automática acomoda la oferta monetaria a la demanda de dinero y aísla la producción (figura B15.4)
Certidumbre, comercio e inversión	Volatilidad cambiaria como coste añadido, atenuable con cobertura	**Ventaja**: menor riesgo de cambio frente a la moneda ancla, lo que favorece comercio e inversión extranjera directa
Ajuste externo	**Ventaja**: rápido y automático, sujeto a Marshall-Lerner y a la curva J	Lento y asimétrico, a través de reservas y devaluación interna
Reservas	**Ventaja**: no se precisan para defender una paridad	Coste de oportunidad de mantener un volumen suficiente
Riesgo de crisis	**Ventaja**: no cabe ataque especulativo contra una paridad, aunque sí depreciaciones bruscas	Riesgo de crisis cambiaria si el compromiso no es creíble
Asignación de recursos a largo plazo	**Ventaja**: los precios relativos reflejan las condiciones del mercado y facilitan la reasignación de factores	Los precios relativos pueden desalinearse si la paridad se aleja del equilibrio

Ningún régimen domina, pues, al otro en todas las dimensiones, y la elección suele atender a cuatro criterios:

- *Naturaleza de los shocks*: el tipo de perturbaciones (reales o monetarias) que predomina en el país.

- *Credibilidad institucional*: su historial de credibilidad e inflación, que determina el valor del anclaje externo.

- *Grado de apertura*: el peso del comercio exterior y la integración comercial con el país ancla, criterio que remite a la teoría de las áreas monetarias óptimas (B16).

- *Descalce*: el desajuste entre la moneda de denominación de los activos y la de los pasivos de sus balances, que condiciona el signo de los efectos de una depreciación.

En la práctica, además, la distinción entre ambos polos es menos nítida de lo que sugiere el análisis teórico, como muestra el miedo a flotar (apartado 2.3.2), y ello da paso a los regímenes intermedios (apartado 2.2).

2. Análisis comparado de los regímenes cambiarios observados

En la realidad no se dan regímenes cambiarios en su forma pura, sino que existen innumerables variantes según el grado de fijación, la frecuencia de las revisiones, la publicidad o el grado de compromiso.

Los regímenes observados se pueden clasificar en tres categorías, de mayor a menor rigidez: **regímenes de vinculación fuerte**, **regímenes intermedios** o de vinculación débil y **regímenes flotantes**. El criterio que separa a los primeros de los segundos no es el grado de fijación de la paridad, que puede ser idéntico, sino la dificultad de salida. En los regímenes de vinculación fuerte, el abandono exige una reforma legal, un tratado internacional o la reintroducción de una moneda propia. En los intermedios, en cambio, basta una decisión administrativa de la autoridad monetaria.

Los regímenes pueden compararse atendiendo a cuatro dimensiones: flexibilidad, independencia de la política monetaria, credibilidad y capacidad de control de la inflación.

2.1. Regímenes cambiarios de vinculación fuerte (*hard pegs*)

Este grupo comprende la unión monetaria, la dolarización o euroización y la junta monetaria. Los tres regímenes comparten el rasgo de hacer costosa o directamente imposible la reversión, que es lo que dota de credibilidad al compromiso, y de renunciar a la política monetaria autónoma. Esa renuncia pesa sobre todo cuando hay asincronías de ciclo con el área o con el país emisor o ancla, porque su política monetaria puede no ser entonces la adecuada para el país.

2.1.1. Unión monetaria

> Sobre la teoría de las áreas monetarias óptimas y la integración monetaria, véase el tema B16.

Definición: varios países comparten una moneda única y una política monetaria común, que dirige un banco central común o un sistema que integra a los bancos centrales nacionales. Los tipos de cambio entre los miembros quedan fijados de forma irrevocable y sus monedas, si las tenían, se sustituyen por la común. Solo queda el tipo de cambio de la moneda común frente a terceras monedas, que los miembros fijan o dejan flotar en común.

Ejemplos: la zona del euro, con veintiún miembros desde la entrada de Bulgaria el 1 de enero de 2026, es la mayor unión monetaria, pero no la única. Funcionan también la Unión Monetaria del Caribe Oriental, cuyo dólar está anclado al dólar estadounidense, y las dos uniones del franco CFA, cuyas monedas están ancladas al euro. Estas dos uniones son la Unión Económica y Monetaria de África Occidental y la Comunidad Económica y Monetaria de África Central (tema B35, apartado 2.4.5). En la clasificación *de facto* del FMI, los miembros de una unión monetaria figuran en el régimen de su moneda común. Así, los de la zona del euro están en libre flotación, los del Caribe Oriental en junta monetaria y los del franco CFA en tipo fijo convencional.

Conclusiones:

- *Aspectos positivos*: su credibilidad supera la de cualquier paridad fija, porque deshacer una unión monetaria cuesta mucho más que abandonar una paridad. Además, a diferencia de la dolarización, los países miembros participan, a través de sus bancos centrales nacionales, en la política monetaria común, que se decide de forma colegiada, y en el reparto del señoraje.

- *Limitaciones*: obliga a los miembros a coordinar sus políticas económicas.

2.1.2. Dolarización o euroización (*exchange arrangements with no separate legal tender*)

Definición: un país utiliza como moneda de curso legal la de otro. La decisión es en principio reversible, aunque revertirla obliga a reconstruir la credibilidad, con un coste elevado. La dolarización es oficial cuando la ley da curso legal a la divisa, por acuerdo con el país emisor o unilateralmente. Es extraoficial (*de facto*) cuando, aunque subsista la moneda nacional, el público hace con la divisa la mayor parte de sus pagos.

Ejemplos: este régimen lo ilustran la dolarización oficial de Ecuador, Panamá y El Salvador; la euroización unilateral de Kosovo y Montenegro; y, ya fuera de la vinculación fuerte, la dolarización extraoficial o *de facto* de economías con historial inflacionista, como Perú o Argentina.

Conclusiones:

- *Aspectos positivos*: es más creíble que una paridad fija y estabiliza mejor los precios. Elimina además por completo el riesgo de cambio frente al país emisor y, con él, la prima de riesgo cambiario incorporada a los tipos de interés.

- *Limitaciones*: supone una renuncia total a la política monetaria propia, dado que la oferta monetaria pasa a depender del saldo exterior (p. ej. del superávit comercial o de la recepción de remesas). A ello se añaden dos costes de primer orden, la pérdida del señoraje y la desaparición del prestamista de última instancia, puesto que la autoridad monetaria no puede crear la moneda en la que están denominados los depósitos del sistema bancario.

2.1.3. Régimen de junta monetaria, panel de convertibilidad o caja de conversión (*currency board arrangements*)

Definición: una junta monetaria asume, por mandato legal explícito, la obligación de cambiar la moneda nacional por una divisa concreta (la moneda ancla) a un tipo fijo. Los países la adoptan, sobre todo, para dar credibilidad a una política antiinflacionista.

Su establecimiento implica la eliminación de las funciones tradicionales del banco central:

- *Respaldo*: la moneda nacional solo puede emitirse contra la moneda ancla, es decir, la base monetaria debe estar respaldada al 100 % por reservas de la moneda ancla. Por ello, aunque se admite la doble circulación monetaria, no puede haber emisión de moneda fiduciaria discrecional. El respaldo íntegro se predica, con todo, de la base monetaria y no de los agregados amplios. El sistema bancario sigue creando depósitos por la vía del multiplicador, de modo que el pasivo bancario excede en todo caso al volumen de reservas.

- *Independencia*: el banco central tiene la prohibición expresa de comprar o mantener títulos del Gobierno, lo que implica que no puede financiar al Gobierno.

- *No prestamista de última instancia*: el banco central no puede conceder liquidez de urgencia a las entidades de crédito[11]. En el modelo ortodoxo de junta monetaria, tampoco regula las

[11] En el balance de la autoridad monetaria, $H = \text{CI} + R$, el respaldo íntegro exige $R \geq H$. Un préstamo de urgencia eleva el crédito interno (CI) y, con él, la base monetaria sin aumentar las reservas, y parte de la moneda quedaría sin respaldo. La junta solo puede prestar con el exceso de reservas sobre la base, si lo tiene.

reservas mínimas que estas deben mantener, aunque algunas juntas, como las de Estonia, Lituania o Bulgaria, sí fijaron coeficientes de reservas obligatorias. Carece, por tanto, de margen de actuación para el control monetario.

Ejemplos: las juntas monetarias tienen su origen en las colonias británicas del s. XIX y principios del XX, con el antecedente doctrinal de la Bank Charter Act británica de 1844. El régimen perdió peso a lo largo del s. XX y volvió a atraer interés en los años noventa. Lo impulsaron los buenos resultados de Hong Kong (desde 1983), de Argentina en sus primeros años (su junta resistió la crisis del Tequila de 1994-1995)[12] y de exrepúblicas soviéticas como Estonia (1992) y Lituania (1994). Ambas repúblicas la mantuvieron hasta su ingreso en la zona del euro (2011 y 2015, respectivamente). Bulgaria, con junta monetaria desde 1997, la mantuvo también hasta su ingreso, el 1 de enero de 2026.

Conclusiones:

- *Aspectos positivos*: el respaldo legal refuerza la credibilidad y, con ella, el control de la inflación. Asimismo, como el país mantiene su moneda y el señoraje, salir del régimen resulta menos costoso que abandonar la dolarización.

- *Limitaciones*: su eficacia depende del origen de la inflación. Funciona cuando la inflación nace de expectativas desancladas, pero puede servir de poco cuando responde a causas institucionales, como sindicatos con gran poder de negociación. Además, la imposibilidad de actuar como prestamista de última instancia hace que las crisis cambiarias puedan derivar en crisis bancarias, precisamente porque el respaldo íntegro de la base monetaria no alcanza a los depósitos.

2.2. Regímenes cambiarios intermedios o de vinculación débil (*soft pegs*)

A diferencia de los anteriores, estos regímenes fijan el tipo de cambio sin comprometerse a la irrevocabilidad de la paridad, que la autoridad monetaria puede modificar. Ganan con ello flexibilidad y una válvula de escape ante perturbaciones adversas, a costa de una menor credibilidad, que es justamente lo que los hace vulnerables a los ataques especulativos.

2.2.1. Tipo de cambio fijado a una divisa o a una cesta de divisas (*fixed peg* o *basket peg*)

Definición: se fija un tipo de cambio contra una divisa o una cesta de divisas sin compromiso de irrevocabilidad de la paridad. En la clasificación *de facto* del FMI, el tipo de cambio puede fluctuar en un margen inferior a ±1 % alrededor de un tipo central, o bien sus valores máximo y mínimo deben mantenerse dentro de un margen del 2 % durante al menos seis meses.

Ejemplos: este régimen lo ilustran la corona danesa, vinculada al euro en el marco del mecanismo de tipos de cambio II (MTC II) con una banda estrecha; las monedas de los países exportadores de petróleo del Golfo (Arabia Saudí, Emiratos Árabes Unidos, Catar), fijadas al dólar; y, entre los anclajes a una cesta, el dólar fiyiano, referenciado desde 1975 a una cesta de monedas ponderada por el comercio.

[12] El caso argentino ilustra, sin embargo, los límites del régimen. La Ley de Convertibilidad (1991) logró abatir una hiperinflación y sostuvo la paridad uno a uno con el dólar durante una década. Con todo, la apreciación real acumulada, la devaluación brasileña de 1999, la rigidez fiscal y la imposibilidad de asistir a un sistema bancario en dificultades desembocaron en la crisis de 2001-2002. La crisis supuso la restricción de la disponibilidad de depósitos (*corralito*), el abandono de la paridad y la suspensión de pagos de la deuda soberana. La enseñanza que suele extraerse es que la credibilidad que aporta el compromiso legal no sustituye a la coherencia de la política fiscal ni resuelve el problema del prestamista de última instancia.

Conclusiones:

- *Aspectos positivos*: proporciona un ancla nominal y reduce el riesgo de cambio frente a la moneda ancla, con los efectos ya examinados sobre el comercio, la inversión y la negociación salarial. Lo consigue, además, sin exigir la renuncia institucional que imponen los regímenes de vinculación fuerte, pues el país conserva su banco central, el señoraje y la función de prestamista de última instancia, y puede realinear la paridad si una perturbación adversa lo hace necesario. El anclaje a una cesta presenta, además, la ventaja de estabilizar el tipo de cambio efectivo, en lugar de un único tipo bilateral, lo que resulta preferible cuando el comercio exterior está diversificado.

- *Limitaciones*: la reversibilidad que le da flexibilidad es también lo que le resta credibilidad. Al ser conocido que la paridad puede modificarse, el régimen queda expuesto a ataques especulativos cuando los fundamentos se deterioran (modelos de primera generación) o incluso por expectativas que se autocumplen (modelos de segunda generación). Exige, además, mantener un volumen de reservas suficiente para defender la paridad, y con libre movilidad de capitales implica la pérdida de la autonomía monetaria por aplicación del trilema. Si el anclaje es a una cesta cuya composición no se hace pública, se gana margen de maniobra a costa de transparencia y, por tanto, de eficacia del anclaje de expectativas.

2.2.2. Tipo de cambio reptante o deslizante (*crawling peg*)

Definición: la paridad se corrige en pequeños pasos, a un ritmo fijado de antemano (p. ej. un porcentaje mensual) o en función de indicadores cuantitativos preestablecidos (p. ej. diferenciales de inflación). Según la referencia del ajuste, hay dos tipos:

- *Pasivo* (*backward looking*): la paridad sigue a la inflación pasada o a la evolución de las reservas.

- *Activo* (*forward looking*): la paridad sigue una senda anunciada con antelación o la inflación prevista, con el fin de guiar las expectativas.

Ejemplos: este régimen lo adoptan países que quieren anclar su moneda a la de otro sin poder igualar la inflación de este, como Nicaragua hasta 2023 o Botsuana, con anclaje a una cesta.

Conclusiones:

- *Aspectos positivos*: es más flexible que una paridad fija. Al corregir el tipo nominal al ritmo del diferencial de inflación, preserva el tipo de cambio real y, con él, la competitividad, y deja a la política monetaria algo más de margen.

- *Limitaciones*: ese margen monetario es, con todo, estrecho. La variante pasiva presenta, además, el inconveniente de indiciar el tipo de cambio a la inflación pasada, con lo que puede realimentar la inercia inflacionista en lugar de corregirla.

2.2.3. Regímenes basados en bandas

Definición: la autoridad monetaria establece un suelo y un techo para el tipo de cambio, que fluctúa libremente dentro de ese intervalo. Las variantes del régimen difieren en cuatro elementos de diseño:

▪ *Frecuencia de las realineaciones*: a mayor frecuencia, menor riesgo de desvío respecto de los fundamentos, pero menor efectividad del régimen para controlar la inflación, promover el comercio y lograr los demás efectos de un tipo fijo[13].

▪ *Amplitud*: al ampliar la banda operan dos fuerzas opuestas. Por un lado, la ampliación tiende a aumentar la volatilidad y, por otro, tiende a aumentar la credibilidad y, con ella, a reducir la volatilidad, como sucedió en el SME en 1993. Esto último puede deberse a que los agentes prevén que, al darle más autonomía al banco central, la política monetaria será más adecuada, o a que la moneda se acerca con menos frecuencia a las bandas, lo que reduce la volatilidad.

▪ *Moneda de referencia*: la fijación puede ser respecto a una moneda o a una cesta. Para cumplir mejor el objetivo de inflación conviene anclarse a una moneda, mientras que para atender mejor a las variables de los principales socios comerciales es preferible una cesta[14].

▪ *Grado de publicidad*: por ejemplo, la composición de la cesta puede ser pública o no.

Ejemplos: estos regímenes los ilustran el patrón oro con los puntos del oro, el sistema de Bretton Woods con bandas de ±1 % frente al dólar o el Sistema Monetario Europeo. En realidad, todos los regímenes cambiarios pueden interpretarse como sistemas más o menos estrictos de bandas.

Conclusiones:

▪ *Aspectos positivos*: los límites fijan las expectativas sobre el tipo de cambio, lo que modera su volatilidad, y la anchura de la banda deja algún margen a la política monetaria. Krugman (1991c) formalizó ese anclaje en su modelo de zonas objetivo, cuyo resultado depende de la credibilidad del compromiso:

- *Compromiso creíble*: la mera expectativa de intervención en los límites estabiliza el tipo de cambio sin necesidad de intervención efectiva (efecto luna de miel o *honeymoon*). Además, el tipo de cambio se vuelve insensible a los fundamentos al aproximarse a los límites (*smooth pasting*), y la banda actúa como un estabilizador automático.

- *Compromiso no creíble*: la aproximación a los límites desestabiliza el tipo de cambio (efecto divorcio, en la expresión de Bertola y Caballero, 1992).

▪ *Limitaciones*: los cambios de la banda o de su centro pueden añadir volatilidad. Los límites de la banda ofrecen, además, un blanco preciso a la especulación, al identificar el nivel exacto en el que la autoridad monetaria habrá de intervenir.

2.3. Regímenes cambiarios flotantes

Dado que el régimen de cambio flexible puro no se da plenamente en la realidad, conviene distinguir dos categorías dentro de la flotación, según la intensidad de la intervención de la autoridad monetaria.

[13] Así, el Sistema Monetario Europeo (SME, 1979) funcionó en sus primeros años más como un *crawling peg* que como un régimen de bandas, debido a la frecuencia de las realineaciones.

[14] Conviene que la cesta incluya al principal socio comercial. Por ejemplo, anclada solo al dólar, Argentina perdió competitividad frente a Brasil tras la depreciación del real a finales de los noventa, una de las causas de la quiebra de su junta monetaria.

2.3.1. Libre flotación (*free floating*)

Definición: el tipo de cambio se determina en los mercados de divisas y la intervención es meramente excepcional, dirigida a corregir situaciones de desorden. El FMI aplica un criterio cuantitativo para reconocer esta categoría, según el cual la intervención no puede producirse en más de tres ocasiones durante seis meses, y ninguna de ellas puede prolongarse más de tres días hábiles.

Ejemplos: el FMI reserva esta categoría, dentro de su clasificación *de facto*, a un número reducido de economías, principalmente avanzadas, entre las que se cuentan Estados Unidos, la zona del euro, Japón o el Reino Unido.

La clasificación descansa, no obstante, en la información que aportan las propias autoridades. Si estas no facilitan datos que confirmen que la intervención se ha mantenido dentro de ese umbral, el régimen se clasifica como flotante (*floating*) y no como de libre flotación. Además, en la práctica algunos de estos países han intervenido de forma no marginal, como Japón, cuyas autoridades han actuado en el mercado de divisas en distintas ocasiones para sostener el yen. La frontera con la categoría inmediatamente inferior (*floating*) es, por tanto, menos nítida de lo que sugiere el criterio cuantitativo.

Conclusiones:

- *Aspectos positivos*: maximiza las dos propiedades examinadas en el apartado 1, la autonomía de la política monetaria y la capacidad del tipo de cambio para absorber las perturbaciones reales asimétricas, pues se ajusta de forma continua y sin coste de reservas.

- *Limitaciones*: el tipo de cambio queda expuesto a una volatilidad elevada y a desalineamientos persistentes respecto de sus fundamentos, dado que la mayor parte de las operaciones del mercado de divisas responde a movimientos de capital y no a transacciones comerciales. Ello impone costes de cobertura y de información a los operadores.

2.3.2. Flotación dirigida o sucia (*managed float*)

Definición: el tipo de cambio flota, pero la autoridad monetaria interviene con regularidad para suavizar sus movimientos, la llamada *lucha contra el viento* (*leaning against the wind*). Aun así, la autoridad no se compromete con ninguna paridad ni senda predeterminada. Lo que separa este régimen de la libre flotación es la frecuencia de esa intervención, superior al umbral señalado.

Aun bajo flotación, el país debe elegir un ancla nominal que oriente la política monetaria y construya la credibilidad de la autoridad. Históricamente lo fueron el precio del oro o los agregados monetarios[15]. En cambio, hoy predominan los objetivos directos de inflación (*inflation targeting*), como en Australia, Canadá, Noruega, Polonia, Suecia o el Reino Unido, o los esquemas que atienden a varias variables al diseñar la política monetaria (como en Estados Unidos o la zona del euro)[16].

Ejemplos: Brasil, Colombia, Corea del Sur, Indonesia o Sudáfrica figuran en la categoría de flotación (*floating*) de la clasificación *de facto* del FMI, la que corresponde a este régimen. Según la clasificación preliminar del FMI a 30 de abril de 2026, esa categoría reúne 34 economías, frente a 31 en libre flotación. Los regímenes intermedios o de vinculación débil siguen siendo, con todo, el grupo más numeroso, con 92 de las 195 economías clasificadas (191 miembros y

[15] Los agregados monetarios, sobre todo en los años setenta y ochenta.

[16] Una propuesta más reciente, que no ha llegado a aplicarse, es la del objetivo de crecimiento de la renta nominal (*nominal GDP targeting*).

cuatro territorios, como Hong Kong y Macao), frente a 27 de vinculación fuerte y 11 en otros regímenes administrados.

Guillermo Calvo y Carmen Reinhart (2002) atribuyen el recurso a estos regímenes al fenómeno del *miedo a flotar* (*fear of floating*). Muchas economías, especialmente las de países en desarrollo que han sufrido crisis financieras recientes, tienen regímenes cambiarios flexibles *de iure*, pero aplican regímenes intervenidos *de facto* por temor a los efectos de las depreciaciones sobre la inflación y sobre los balances endeudados en divisa.

Conclusiones:

- *Aspectos positivos*: la política monetaria conserva un amplio margen propio[17], y el especulador pierde la apuesta en un solo sentido. Al no existir paridad que defender, el tipo de cambio puede moverse en ambos sentidos, de modo que la posición especulativa incorpora un riesgo de pérdida del que carece frente a un compromiso cambiario.

- *Limitaciones*: como la depreciación es siempre posible, las expectativas de inflación pueden quedar mal ancladas, y el riesgo cambiario que subsiste obliga a cubrirse. La intervención discrecional y no anunciada resta, además, transparencia a la política monetaria.

2.4. Conclusiones

Aplicando a los regímenes examinados las cuatro dimensiones anunciadas al comienzo del apartado, se obtiene la comparación de la tabla B15.2.

Tabla B15.2: Comparación de los regímenes cambiarios observados.

Régimen	Flexibilidad	Independencia monetaria	Credibilidad	Control de la inflación
Unión monetaria	Nula	Nula (compartida)	Máxima	Máximo (delegado)
Dolarización	Nula	Nula	Muy alta	Muy alto (importado)
Junta monetaria	Muy baja	Muy baja	Alta	Alto
Fijo convencional	Baja	Baja	Media	Medio-alto
Reptante	Media-baja	Baja	Media-baja	Medio
Bandas	Media	Media	Media (depende de la amplitud)	Medio
Flotación dirigida	Alta	Alta	Depende del ancla nominal	Depende del ancla nominal
Libre flotación	Máxima	Máxima	Depende del ancla nominal	Depende del ancla nominal

Del cuadro se desprende la disyuntiva fundamental que atraviesa todo el apartado. Flexibilidad e independencia monetaria, por un lado, y credibilidad del anclaje, por otro, se mueven en sentidos opuestos, y ningún régimen las maximiza simultáneamente. En los regímenes flotantes, además, la credibilidad no desaparece, sino que se traslada, puesto que deja de descansar en el tipo de

[17] Se requiere, con todo, una elevada credibilidad del banco central para que la financiación exterior sea estable.

cambio y pasa a depender del ancla nominal interna que se adopte, típicamente un objetivo de inflación.

Esta constatación alimentó, tras las crisis cambiarias de los años noventa, la llamada **hipótesis de las soluciones de esquina** o del vaciamiento de los regímenes intermedios (*hollowing out*), formulada por Barry Eichengreen (1994) y popularizada por Stanley Fischer (2001) como "visión bipolar". Según esta hipótesis, con libre movilidad de capitales, los regímenes intermedios serían insostenibles, por ofrecer un blanco identificable a la especulación sin la protección que da un compromiso irrevocable, de modo que los países se verían empujados hacia uno de los dos extremos.

La tesis fue matizada posteriormente, tanto por la evidencia (una proporción muy elevada de países mantiene regímenes intermedios) como por la observación de que la libre flotación tampoco es plena, según muestra el miedo a flotar. La conclusión que mejor resume el apartado es la de Frankel (1999), según la cual ningún régimen cambiario resulta adecuado para todos los países, ni para un mismo país en todo momento.

3. Intervención y regulación en los mercados de cambio

Una vez que las autoridades económicas de un país han elegido un régimen cambiario, establecen relaciones con el mercado de divisas, en cuyo seno aparecen la intervención y la regulación de dichos mercados. Ambos instrumentos son, en el fondo, las dos vías por las que un país gestiona su posición dentro del trilema examinado en el apartado 1.1. La intervención esterilizada es el intento de conservar simultáneamente el objetivo cambiario y la autonomía monetaria sin restringir los movimientos de capital. De ahí que su eficacia sea limitada y condicionada. La regulación, por su parte, actúa directamente sobre el tercer vértice, el de la libre movilidad de capitales.

> Sobre la estructura del mercado de divisas, véase el tema B13. Sobre la determinación del tipo de cambio, véase el tema B14.

3.1. Intervención

Que las autoridades actúen como un agente más en el mercado de divisas obliga a precisar en qué consiste esa actuación, por qué se justifica en cada régimen y con qué instrumentos se lleva a cabo. La experiencia de América Latina, Estados Unidos, Japón y China ilustra después su práctica.

3.1.1. Definición

Hay **intervención** cuando las autoridades monetarias toman parte en el mercado de divisas con acciones y medidas de diverso tipo (anuncios, compraventas, etc.) encaminadas a influir en el nivel del tipo de cambio o a amortiguar sus oscilaciones. La intervención engloba también la gestión de reservas, el proceso de garantizar que las divisas o reservas exteriores de las que dispone la autoridad monetaria tienen un volumen adecuado y están disponibles en el momento necesario.

Para cada país y para cada momento existe un nivel óptimo de reservas. Para concretarlo, la práctica internacional maneja varios criterios de suficiencia de reservas, como son:

- *Cobertura trimestral*: el nivel de reservas debe ser equivalente a tres meses de importaciones, criterio orientado a economías con acceso limitado a los mercados de capitales.

- *Regla de Guidotti-Greenspan*: las reservas deben cubrir la totalidad de la deuda externa a corto plazo por vencimiento residual, es decir, la que habría que refinanciar en el plazo de un año.

- *Métrica compuesta del FMI* (para economías emergentes): pondera exportaciones, agregados monetarios amplios, deuda a corto plazo y otros pasivos, y sitúa el rango de adecuación entre el 100 % y el 150 % de dicho indicador.

Según las medidas o acciones aplicadas, la intervención se clasifica en **intervención directa** e **intervención indirecta**.

3.1.2. Justificación

Régimen cambiario: el objetivo de la intervención depende del régimen:

- *Regímenes de cambio fijo*: el objetivo es sostener la paridad y, con ella, la estabilidad completa del tipo de cambio.

- *Regímenes de cambio flexible*: el objetivo es frenar las oscilaciones de corto plazo sin oponerse a la tendencia, en la *lucha contra el viento*, que da una estabilidad solo parcial.

Acumulación de reservas: entre los motivos que pueden llevar a la intervención y a la acumulación de reservas están:

- *Confianza*: fomentar la confianza de los inversores, dado que el país podrá hacer frente a sus pagos en divisas.

- *Importación*: seguir pagando las importaciones, incluso cuando no sea posible acceder a financiación para ello.

- *Liquidez*: realizar operaciones de gestión de liquidez, dado que los activos en divisas se utilizan con frecuencia como aval o colateral. También podrían realizarse para conceder liquidez al sector bancario en una situación de emergencia.

- *Servicio de la deuda*: asegurar que el Estado puede hacer frente a su deuda pública emitida en moneda extranjera.

- *Rentabilidad*: invertir las reservas en exceso y obtener, así, una rentabilidad. Este caso es menos habitual y, generalmente, va acompañado de la existencia de fondos soberanos (*sovereign wealth funds*).

3.1.3. Tipos de intervención

La intervención puede ser directa (vía compraventas) o indirecta (vía anuncios o variaciones del tipo de interés oficial).

Intervención directa: la autoridad opera en el mercado de divisas con su propio balance. Para frenar una apreciación de la moneda nacional, la vende a cambio de divisas ($\uparrow R$), y para frenar una depreciación, la compra con cargo a sus reservas ($\downarrow R$).

En cualquiera de los casos, las compraventas de divisas suponen una variación de los activos netos mantenidos frente al exterior (ANFE, R) en el balance de la autoridad monetaria y, en consecuencia, pueden suponer una variación de la base monetaria. La autoridad monetaria debe decidir, entonces, si esteriliza o no esa variación, es decir, si permite o no que la base monetaria se mueva. Ambas opciones se pueden ilustrar mediante un doble diagrama.

Intervención (directa) no esterilizada: se permite que la base monetaria varíe junto con las reservas exteriores[18]. Gráficamente, la intervención no esterilizada se representa en el plano cantidad de dinero-tipo de interés y en el plano rendimiento-tipo de cambio (figura B15.9), que recogen el mercado de dinero nacional (izquierda) y el equilibrio en el mercado cambiario (derecha):

[18] En términos del balance del banco central, $H = \text{CI} + R$, la intervención no esterilizada traslada íntegramente la variación de los activos netos frente al exterior a la base monetaria ($\Delta R \to \Delta H$), al mantenerse constante el crédito interno.

1. *Oferta monetaria* (M_0^s): refleja la cantidad de dinero en circulación y es vertical, porque la fija la autoridad monetaria con independencia del tipo de interés.

2. *Demanda de dinero* ($M^d(Y_0)$): refleja la cantidad de dinero que los agentes desean mantener. Es decreciente en el tipo de interés, que es el coste de oportunidad de mantener dinero, y se desplaza a la derecha cuando aumenta la renta, porque crecen las transacciones.

3. *Paridad descubierta de intereses* (PDI): refleja las combinaciones de rendimiento de los activos nacionales y tipo de cambio nominal con las que los activos nacionales y extranjeros ofrecen la misma rentabilidad esperada. Es decreciente, porque un tipo de interés mayor atrae capitales y aprecia la moneda.

4. *Equilibrio*: en el punto 0, el mercado de dinero fija el tipo de interés i_0, que la recta vertical del panel derecho traslada al mercado de divisas, y, a ese rendimiento, la PDI determina el tipo de cambio nominal E_0.

A partir de esa situación, la intervención se desarrolla en dos pasos:

1. *Presión apreciatoria (punto 0′)*: una subida de la renta ($\Delta Y > 0$, p. ej. por una política fiscal expansiva) eleva la demanda de dinero, que pasa de $M^d(Y_0)$ a $M^d(Y_1)$. Como la oferta no cambia, el tipo de interés se eleva hasta $i_{0'}$. Con el tipo extranjero dado y sin trabas a la entrada de capital, el diferencial atrae fondos y la moneda se aprecia hasta $E_{0'}$.

2. *Intervención (punto 1)*: para sostener el tipo de cambio, la autoridad compra divisas con moneda nacional de nueva emisión, que no retira después. La oferta monetaria crece de M_0^s a M_1^s hasta que, en el punto 1, el tipo de interés y el tipo de cambio recuperan i_0 y E_0. La operación equivale a una expansión monetaria, con sus efectos inflacionistas.

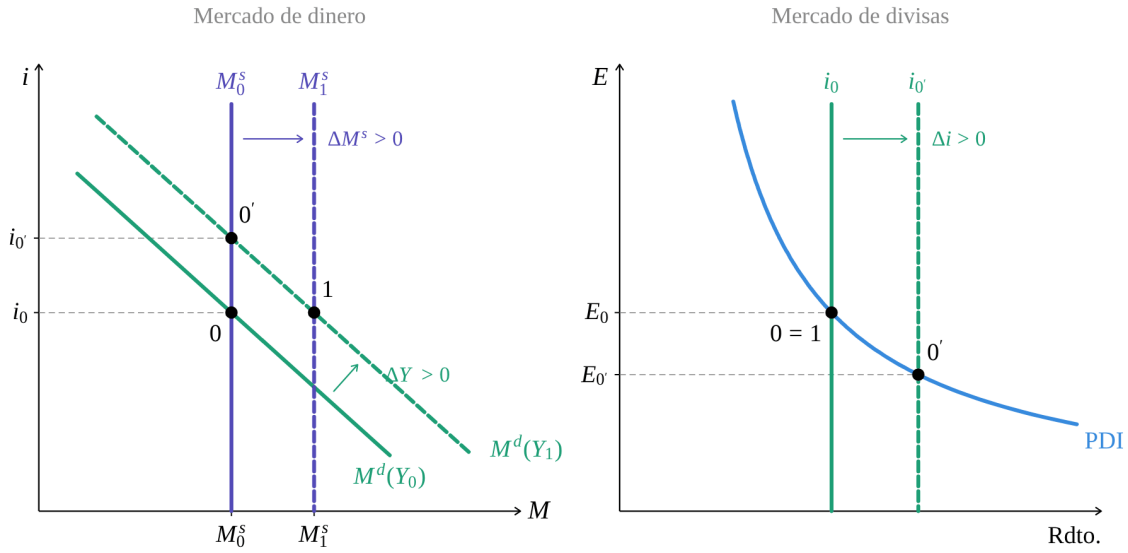

Figura B15.9: Intervención no esterilizada. La compra de divisas amplía la base monetaria y devuelve el tipo de interés y el tipo de cambio a su nivel inicial.

En el caso simétrico de presiones depreciatorias, se venden divisas y se drena moneda nacional, lo que equivale a una política monetaria restrictiva de efectos antiinflacionistas. La intervención

no esterilizada es, en consecuencia, una política monetaria por otro nombre, eficaz sobre el tipo de cambio precisamente porque renuncia a la autonomía monetaria.

Intervención (directa) esterilizada: en este caso no se admiten variaciones de la base monetaria y, por ello, la intervención se acompaña de una esterilización[19]. Por esterilización se entiende la operación de mercado abierto que provoca sobre la oferta monetaria nacional un efecto de signo contrario al que se deriva de la intervención de la autoridad monetaria en el mercado cambiario. La cuestión relevante es entonces si una intervención que deja inalterada la base monetaria puede, aun así, mover el tipo de cambio. La literatura identifica tres canales de transmisión:

- *Canal de cartera* (*portfolio balance*): solo funciona si los activos nacionales y extranjeros son sustitutivos imperfectos. Con sustitución perfecta, a los inversores les es indiferente la composición de su cartera, y una intervención que deja intacta la oferta monetaria no altera el tipo de cambio. Con sustitución imperfecta, en cambio, la venta de bonos en moneda nacional con la que se esteriliza la compra de divisas eleva el peso de esos bonos en las carteras privadas. Los inversores, más expuestos ahora a una depreciación, solo los retienen a cambio de una prima de riesgo, ρ. La prima se suma a la paridad descubierta de intereses, $i = i^* + \frac{E^e - E}{E} + \rho$, y desplaza la curva hacia la derecha, porque para cada tipo de cambio los inversores exigen más rentabilidad.

En las economías avanzadas, donde los activos son muy sustituibles entre sí, la evidencia atribuye la eficacia de la intervención más al canal de señalización que a este (Sarno y Taylor, 2001). En las emergentes, por el contrario, la intervención suele mover el nivel y la volatilidad del tipo de cambio, porque sus bancos centrales disponen de reservas elevadas, no siempre esterilizan por completo y cuentan con controles de capital (Menkhoff, 2013).

- *Canal de señalización*: la intervención transmite información sobre la orientación futura de la política monetaria, alterando las expectativas de los agentes y, con ellas, el tipo de cambio esperado, E^e. La señal resulta creíble porque la autoridad compromete su propio balance. Al tomar una posición en divisas, se expone a pérdidas si su política posterior no resulta coherente con lo señalado, lo que distingue la intervención de una mera declaración.

- *Canal de microestructura*: el mercado de divisas es descentralizado y la información sobre los flujos de órdenes es privada (B13). Por ello, una intervención de volumen apreciable modifica el conjunto de información de los operadores y puede desencadenar ajustes de posiciones superiores a la operación inicial.

Gráficamente, en los mismos planos (figura B15.10), el aumento de la renta del caso anterior ($\Delta Y > 0$) desplaza la demanda de dinero de $M^d(Y_0)$ a $M^d(Y_1)$. Como la oferta monetaria, M^s, no varía, el tipo de interés sube de i_0 a $i_{0'}$ y la moneda se aprecia hasta $E_{0'}$ (punto $0'$). Por el canal de cartera, la prima de riesgo ($\Delta\rho > 0$) desplaza la paridad de PDI_0 a PDI_1 y, con el tipo de interés $i_1 = i_{0'}$, el tipo de cambio vuelve a $E_1 = E_0$ (punto 1).

[19] En términos del balance, la esterilización compensa ΔR con $\Delta\text{CI} = -\Delta R$, de modo que la base monetaria no se mueve (p. ej. compra de divisas, $\uparrow R$, y venta de títulos nacionales de la cartera del banco central, $\downarrow$ CI).

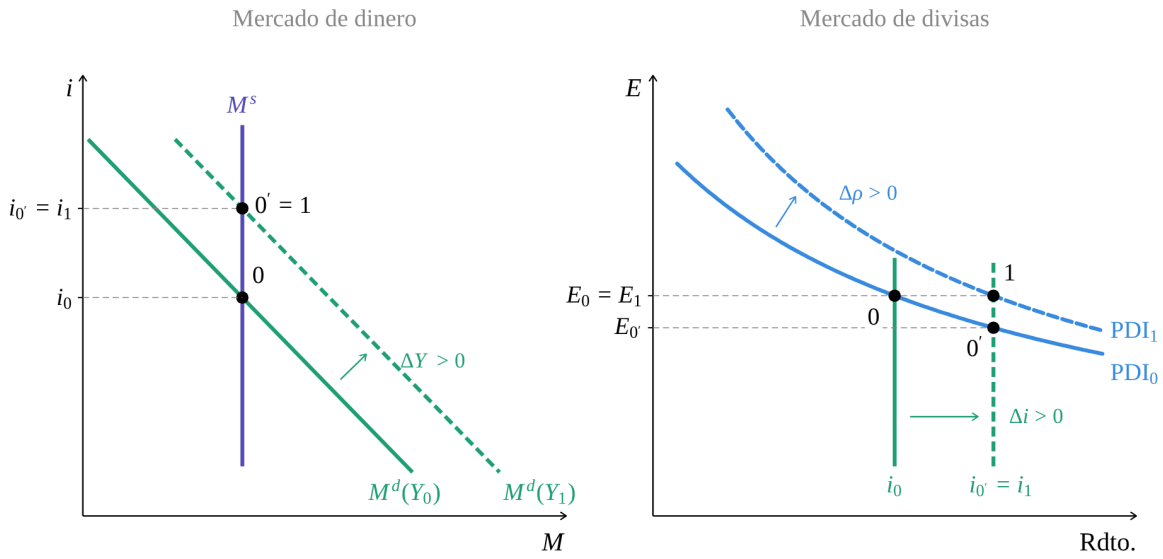

Figura B15.10: Intervención esterilizada. La operación de mercado abierto compensadora mantiene inalterada la oferta monetaria, y la intervención solo actúa vía primas de riesgo o señales.

Intervención indirecta (anuncios, comunicaciones y gestión de expectativas): en un marco de incertidumbre sobre la orientación de la política económica del país, las autoridades pueden recurrir a anuncios públicos o a variaciones del tipo de interés oficial. Una variación del tipo oficial altera el diferencial de intereses y, con él, el tipo de cambio. Ninguna de las dos vías exige operar en el mercado de divisas. Por ello, la intervención indirecta se distingue de la directa, o intervención propiamente dicha, en que no compromete recursos, lo que la abarata pero también la debilita. La intervención indirecta expone, además, a la autoridad monetaria o al sector público a un problema de credibilidad si sus actuaciones no son consecuentes con lo anunciado.

Gráficamente, en los mismos planos (figura B15.11), si se anuncia, por ejemplo, una política fiscal expansiva que pueda generar presiones inflacionistas, el anuncio afecta a las expectativas de los agentes y eleva el tipo de cambio esperado ($\Delta E^e > 0$). La paridad se desplaza de PDI_0 a PDI_1 y, con el tipo de interés i_0 dado, la moneda se deprecia de E_0 a E_1 (del punto 0 al 1), sin intervención cambiaria directa.

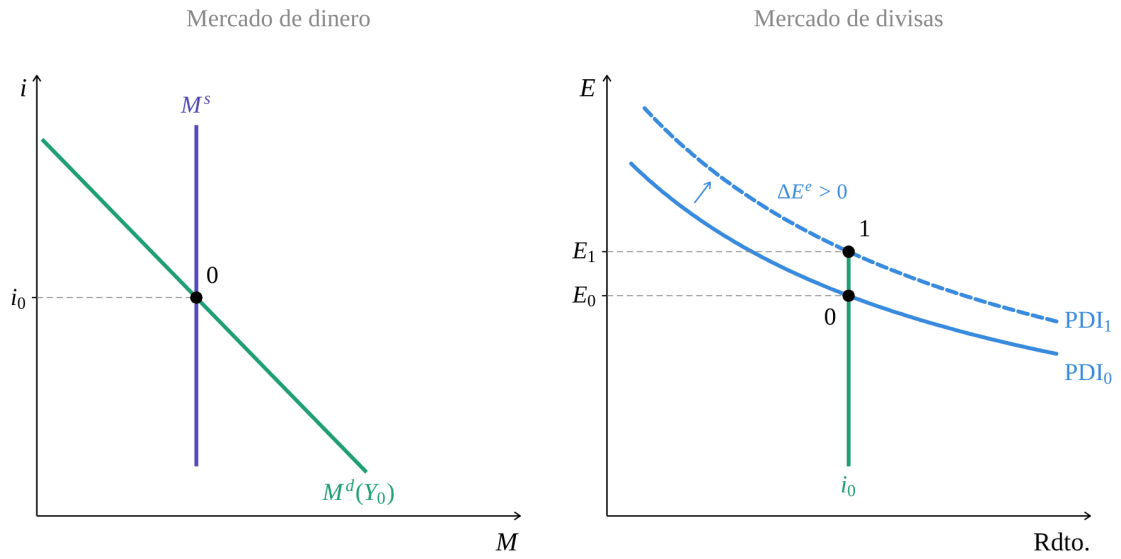

Figura B15.11: Intervención indirecta. Un anuncio que eleva el tipo de cambio esperado (E^e) desplaza la curva de paridad y deprecia la moneda sin intervenir en el mercado de divisas.

3.1.4. Evidencia empírica sobre la intervención: América Latina (1982), EE. UU. (2008-2014), Japón (2010) y China (2015)

> Sobre los modelos de crisis de balanza de pagos, véase el tema B17.

La incompatibilidad de objetivos (crisis de deuda latinoamericana, 1982): la crisis de deuda latinoamericana ilustra la incompatibilidad entre el objetivo cambiario y el resto de objetivos de política económica. Durante los años setenta, numerosos países de la región mantenían tipos de cambio fijos o reptantes frente al dólar mientras financiaban déficits fiscales persistentes con creación de dinero y endeudamiento exterior barato en dólares. El endurecimiento monetario de la Reserva Federal a partir de 1979 (el "*shock* Volcker") elevó los tipos de interés mundiales, encareció el servicio de la deuda y revirtió los flujos de capital. Con ello, la defensa de las paridades exigió cuantiosas pérdidas de reservas, cuyo agotamiento desembocó en devaluaciones forzadas y en la suspensión de pagos de México en agosto de 1982, que se extendió por la región.

El episodio se ajusta a la lógica de los modelos de crisis de primera generación (Krugman, 1979a). Según esa lógica, un tipo de cambio fijo es insostenible cuando la política fiscal y monetaria interna es incompatible con la paridad, y la intervención solo pospone (y encarece) el ajuste.

El caso de Estados Unidos (2008-2014): un debate surgido tras la crisis financiera global es si los países ricos (p. ej. Estados Unidos) intervienen en los mercados de divisas con objetivos cambiarios de manera heterodoxa. Un país grande y de moneda fuerte puede utilizar la expansión cuantitativa (*quantitative easing*, QE) para depreciar su moneda y estabilizar el ciclo. Según sus críticos, ello daría lugar a políticas de empobrecimiento del vecino equivalentes a las devaluaciones competitivas que prohíbe el artículo IV del Convenio del FMI, si bien la política

monetaria orientada a objetivos internos no se considera manipulación cambiaria[20].

El caso de Japón (2010): en septiembre de 2010, el Ministerio de Finanzas japonés ordenó al Banco de Japón una venta de 2,12 billones de yenes (unos 25.000 M$) para debilitar la moneda, que se depreció un 3 % el mismo día. La operación superaba el importe agregado de las intervenciones estadounidenses en el mercado de divisas desde mediados de los años noventa[21]. El contraste ilustra hasta qué punto la práctica de la intervención difiere entre economías avanzadas formalmente clasificadas en el mismo régimen cambiario.

El caso de China (2015): hasta 2014, ante las presiones apreciatorias que generaban sus persistentes superávits por cuenta corriente y las entradas de capital, China intervino de forma masiva y esterilizada para contener la apreciación del yuan frente al dólar. La operación tenía dos fases:

1. *Compra de activos en dólares*: en primer lugar, el banco central compraba activos denominados en dólares entregando yuanes a cambio ($\uparrow R$ con $\uparrow H$), lo que contenía la apreciación del yuan pero expandía la base monetaria. Además, dado que los bonos del Tesoro estadounidense son sustitutivos casi perfectos del dólar por su elevada liquidez, ello suponía retirar liquidez en dólares del mercado internacional.

2. *Esterilización*: en segundo lugar, esterilizaba la operación vendiendo títulos propios denominados en yuanes y elevando el coeficiente de reservas obligatorias de la banca ($\downarrow$ CI neto con $\downarrow H$, y menor multiplicador monetario), lo que devolvía la oferta monetaria a su nivel inicial. El objetivo era neutralizar esa expansión monetaria y sus efectos inflacionistas, y no contener presión depreciatoria alguna.

La cantidad de yuanes en el sistema no variaba, pero la liquidez *de facto* en dólares disminuía, al haberse retirado de la circulación activos en dólares muy líquidos. El "drenaje" de los activos denominados en dólares no es, además, comparable con la "inyección" de activos denominados en yuanes, ya que a estos se les exige una prima por su mayor riesgo.

El propio año 2015 ilustra, sin embargo, la asimetría del ajuste señalada en el apartado 1.4. Tras la devaluación de agosto y el cambio en el mecanismo de fijación de la paridad central, las presiones se invirtieron. Las salidas de capital obligaron al banco central a realizar la operación contraria, es decir, a vender reservas para sostener el yuan. Entre 2014 y 2016 sus reservas exteriores se redujeron en torno a una cuarta parte, y se endurecieron además los controles de capitales.

3.1.5. Conclusiones

Aspectos positivos: la intervención no esterilizada es eficaz sobre el tipo de cambio, aunque a costa de la autonomía monetaria. La esterilizada lo es sobre todo por el canal de señalización en las economías avanzadas, y en las emergentes la intervención suele mover el nivel y la volatilidad del tipo de cambio. Las reservas acumuladas sostienen, además, la confianza de los inversores y los pagos exteriores del país.

Limitaciones: frente a esos beneficios, la acumulación de reservas tiene dos costes, el de oportunidad y, cuando se esteriliza, el cuasifiscal.

[20] En 2010, el ministro de Hacienda de Brasil, Guido Mantega, denunció una "guerra de divisas", pues la expansión monetaria de la Reserva Federal, justificada por la estabilización del ciclo interno, perseguiría en realidad depreciar el dólar.

[21] Desde agosto de 1995, Estados Unidos solo había intervenido en junio de 1998, en apoyo del yen, y en septiembre de 2000, en la actuación concertada en apoyo del euro. Después participó en la venta coordinada de yenes del G7 de marzo de 2011.

El **coste cuasifiscal** surge porque la autoridad monetaria remunera los títulos que emite en moneda nacional para drenar la liquidez mientras percibe la rentabilidad, habitualmente inferior, de los activos exteriores en que materializa las reservas. El diferencial constituye una pérdida recurrente que soporta, en última instancia, el sector público.

En todo caso, la decisión de intervención y gestión de reservas está sujeta a múltiples condicionantes. Entre ellos figuran las restricciones institucionales, el entorno histórico, el posible conflicto con el objetivo último o con los estatutos de la institución monetaria, y la reputación del banco central y su relación con el Gobierno. A ellos se suman las normas internacionales, implícitas y explícitas, sobre el comportamiento de los bancos centrales.

3.2. Regulación

La regulación actúa sobre el tercer vértice del trilema, la libre movilidad de capitales. Su estudio sigue el mismo orden que el de la intervención, desde la definición y la justificación hasta los tipos de regulación.

3.2.1. Definición

La **regulación** reúne las normas que rigen la actividad de los mercados de cambios, desde la legislación general hasta los límites a la entrada y salida de capitales. Estos límites influyen en el tipo de cambio a través de los flujos de divisas que permiten o impiden.

3.2.2. Justificación

El fundamento de la regulación cambiaria procede de la economía del bienestar. Si los mercados fueran perfectamente competitivos y no hubiera fallos de mercado, la libre circulación de capitales sería óptima. En cambio, cuando existen distorsiones que no pueden eliminarse, la teoría del *second best* (Lipsey y Lancaster, 1956) muestra que añadir otra restricción, como un límite a los flujos de capital, puede mejorar el bienestar. Como la competencia perfecta es la excepción, el control de capitales puede estar justificado en casos concretos.

La literatura reciente, con Bianchi (2011) y Korinek (2011), ha precisado el argumento, dentro de la misma economía del bienestar, por la vía de las **externalidades pecuniarias**. Los agentes que se endeudan en divisa no internalizan que su decisión eleva la vulnerabilidad agregada de la economía ante una reversión de los flujos, puesto que la depreciación consiguiente deteriora los balances de todos los demás. Existe, por tanto, un endeudamiento exterior socialmente excesivo, y una restricción que lo corrija mejora el bienestar.

3.2.3. Tipos de regulación

Los movimientos de capitales pueden limitarse de varias maneras, y tres son las más frecuentes[22]: las restricciones cuantitativas (controles de capitales), los impuestos y los tipos de cambio múltiples.

Restricciones cuantitativas o controles de capitales: limitan el volumen de capital que puede entrar en el país o salir de él, para evitar flujos masivos que lo desestabilicen. Recaen sobre todo en los flujos de corto plazo, como la inversión de cartera, y menos en la inversión extranjera directa.

[22] Otras formas incluirían los controles de precios o los requisitos de solicitud de licencia.

- *Controles a la entrada*: suelen consistir en límites cuantitativos o en requisitos que encarecen la entrada. Por ejemplo, Malasia limitó en 1994 los pasivos exteriores de sus bancos y las operaciones de *swap* con no residentes. Chile impuso en 1991 el "encaje", un depósito no remunerado, inicialmente del 20 % de los nuevos créditos externos y del 30 % desde 1992, que operaba como un impuesto implícito más que como un tope.

- *Controles a la salida*: pueden fijar topes anuales o imponer una espera mínima desde que se solicita la repatriación del capital hasta que se autoriza. El caso más citado es el de Malasia en 1998, que en plena crisis asiática fijó el ringgit frente al dólar e impuso un periodo mínimo de permanencia al capital de cartera. La medida, entonces muy criticada, fue después revalorizada por buena parte de la literatura. China, por su parte, limita a 50.000 dólares anuales por persona la compra de divisas, cupo que no puede destinarse a inversión en el exterior. Por encima del cupo, la operación debe justificarse documentalmente ante la autoridad cambiaria.

Impuestos: gravan las transacciones de capital, en la línea de la tasa sobre las operaciones cambiarias que propuso James Tobin en 1972 para penalizar la especulación a muy corto plazo (B13, apartado 3.3.3). Al restar rentabilidad a la especulación, pueden moderar la volatilidad del tipo de cambio, y aportan además recaudación. Sin un acuerdo internacional amplio, sin embargo, las operaciones se trasladan a las plazas que no los aplican.

La experiencia europea ilustra bien ambos extremos. El impuesto sobre transacciones financieras propuesto por la Comisión en 2011 (y reformulado en 2013 como cooperación reforzada entre once Estados) permaneció bloqueado en el Consejo durante más de una década, al exigir unanimidad entre los Estados participantes, sin llegar a aplicarse, y la Comisión retiró formalmente la propuesta en octubre de 2026.

Las figuras nacionales adoptadas entretanto, como las de Francia (2012), Italia (2013) o España (2021), ilustran la otra cara del problema, ya que, aplicadas unilateralmente, gravan sobre todo, con tipos reducidos, la adquisición de acciones de sociedades nacionales cotizadas de gran capitalización. La base resultante es estrecha, y el propio impuesto tiende a erosionarla al reducir el volumen negociado.

Tipos de cambio múltiples: la autoridad aplica tipos de cambio distintos según la clase de operación. Lo habitual es un tipo preferente para las importaciones consideradas esenciales o para el servicio de la deuda y otro depreciado para las demás. Constituyen una práctica de tipos de cambio múltiples sujeta a la aprobación del FMI conforme al artículo VIII de su Convenio Constitutivo. En la práctica, generan oportunidades de arbitraje entre segmentos, actividades de búsqueda de rentas y, con frecuencia, la aparición de un mercado paralelo o informal cuya cotización se aleja progresivamente de la oficial. Han sido empleados, entre otros, por Venezuela, Argentina y Nigeria.

3.2.4. Conclusiones

Aspectos positivos: la regulación encarece la especulación, lo que puede moderar la volatilidad cambiaria, y concede a las autoridades tiempo para corregir sus desequilibrios internos, como en China. En el caso del encaje chileno, el resultado empírico más robusto es que alargó el plazo de las entradas de capital y redujo así la vulnerabilidad del país (De Gregorio, Edwards y Valdés, 2000).

Limitaciones: la inversión se resiente, porque rinde menos y se vuelve más incierta. Además, los controles no distinguen entre la especulación desestabilizadora y el arbitraje, que integra los mercados de divisas y hace cumplir la ley del precio único, y frenan los dos. Distorsionan, por último, las decisiones intertemporales de ahorro e inversión.

En el caso chileno, además, no se aprecian efectos claros y duraderos sobre el volumen total de entradas ni sobre el tipo de cambio real. A ello se añade la constatación de que los controles tienden a erosionarse con el tiempo, a medida que los agentes encuentran vías para eludirlos, salvo que se actualicen al mismo ritmo. Por eso hoy se consideran un instrumento de corto plazo, combinable con otros.

Conclusiones

Elegir un régimen cambiario supone administrar varias disyuntivas, entre credibilidad y flexibilidad, entre estabilidad cambiaria y autonomía monetaria, y entre el aislamiento frente a unos *shocks* y la exposición a otros, bajo la restricción del trilema de Mundell-Fleming. Entre los regímenes polares, el flexible aísla mejor de las perturbaciones reales y refuerza la política monetaria, mientras que el fijo ancla las expectativas y hace muy eficaz la política fiscal. Ningún régimen observado maximiza a la vez la flexibilidad y la credibilidad del anclaje. Los intermedios agrupan a casi la mitad de los países, en contra de la "visión bipolar" (Fischer, 2001), y el *miedo a flotar* (Calvo y Reinhart, 2002) explica que la flotación tampoco sea plena. La intervención y la regulación completan el cuadro, con un nivel óptimo de reservas propio de cada economía y unos controles que alteran sobre todo la composición de los flujos. Como resumió Frankel (1999), ningún régimen es adecuado para todos los países ni para todos los momentos.

La comparación de los regímenes polares aplica el modelo IS-LM-BP y el trilema de Mundell-Fleming (tema A32), y la intervención se apoya en las operaciones oficiales del mercado de divisas (tema B13). En el extremo de la vinculación fuerte, la unión monetaria renuncia por completo al tipo de cambio (tema B16), mientras que las paridades ajustables quedan expuestas a ataques especulativos (tema B17). Por último, la historia del sistema monetario internacional, del patrón oro a Bretton Woods y a la flotación, muestra cómo ha cambiado la forma de resolver esas disyuntivas (temas B20 y B21).

Con el ciclo financiero global se discute el alcance del trilema y, con él, el caso clásico a favor de la flotación. Para Rey (2013), ese ciclo reduce el trilema a un dilema (apartado 1.1). Obstfeld, Ostry y Qureshi (2019) encuentran, en cambio, que en las economías emergentes el régimen sigue importando, porque las de tipo fijo transmiten con más intensidad las condiciones financieras globales. Sigue abierta también la discusión sobre la capacidad aislante del tipo de cambio, ya que, con el comercio facturado en dólares, la depreciación apenas estimula las exportaciones (Gopinath et al., 2020). Por su parte, el marco integrado de políticas del FMI (Basu et al., 2020) concluye que, con mercados cambiarios poco profundos o descalces de moneda, la intervención, los controles de capital y las medidas macroprudenciales pueden complementar a la política monetaria.

B16. Integración monetaria

Teoría de la integración monetaria.

Introducción

> **Introducción común.** Los temas B10, B16 y B35 comparten una introducción, que se desarrolla al comienzo del tema **B10**. Define la integración económica y, ampliando la clasificación de Balassa (1961), los estadios con que se ordena según el grado de compromiso, que van del acuerdo de comercio preferencial y el área de libre comercio a la unión aduanera, el mercado común y el mercado único, la unión económica, la unión monetaria y, por último, la unión fiscal y política.

En la escala de integración descrita, la **integración monetaria** constituye, antes de la unión fiscal y política, el nivel más avanzado y tiene tres características principales:

1. *Política monetaria común*: los países miembros abandonan su autonomía en materia monetaria y cambiaria, cuyo ejercicio suele trasladarse a una autoridad monetaria común encargada de conducir la política monetaria de la unión.

2. *Libre circulación de capitales*: los capitales pueden desplazarse sin restricciones entre la totalidad de los países que integran la unión monetaria.

3. *Tipos de cambio fijos*: el establecimiento de una relación fija entre los tipos de cambio, que puede realizarse mediante:

- *Tipos fijos irrevocables*: la fijación de un sistema de tipos de cambio fijos e irrevocables que garantice la convertibilidad de las divisas, o

- *Moneda única*: la adopción de una moneda única, lo cual suele dotar de mayor credibilidad al proceso de integración.

Su referencia práctica es la zona del euro, que con la entrada de Bulgaria en 2026 reúne a 21 Estados miembros. La teoría que la juzga es la de las **áreas monetarias óptimas**, iniciada por Mundell (1961) y desarrollada por McKinnon (1963) y Kenen (1969), que buscan los criterios, como la movilidad de los factores, la apertura o la diversificación productiva, que permiten a un grupo de países renunciar al tipo de cambio sin grandes costes. La crisis de deuda soberana mostró, además, que una **unión monetaria incompleta**, sin Unión Bancaria ni prestamista de última instancia para los Estados, es vulnerable a crisis que se alimentan a sí mismas (De Grauwe, 2011).

La teoría de la integración monetaria pone en la balanza lo que cuestan y lo que aportan estos procesos para saber cuándo conviene una unión monetaria. El análisis de la optimalidad de un área monetaria ha pasado por tres etapas:

1. *Primera etapa (años sesenta)*: ve en la integración en un área monetaria, sobre todo, un coste de estabilización.

2. *Segunda etapa (años setenta)*: nace con la publicación del *Informe Werner* (1970), que presentaba las líneas maestras para la creación de una unión monetaria en Europa[1], y con la revolución de la hipótesis de expectativas racionales (HER). Pone mayor énfasis en los beneficios.

3. *Tercera etapa (años ochenta en adelante)*: pone énfasis en el estudio de los procesos de creación de uniones monetarias y de la fragilidad de estas mediante el análisis coste-beneficio.

Este tema repasa las teorías de la integración monetaria en tres partes:

- *Costes*: la teoría de las áreas monetarias óptimas trata la integración monetaria como un coste.
- *Beneficios*: los desarrollos posteriores ponen el acento en sus beneficios.
- *Construcción de una unión monetaria*: el tema termina con ese proceso, sus retos y sus soluciones.

[1] Concretamente, se preveía la fijación irrevocable de los tipos de cambio y la plena liberalización de los movimientos de capitales, así como, preferiblemente, la sustitución de las monedas nacionales por una moneda única. Sin embargo, en lugar de un banco central europeo único, contemplaba un sistema comunitario de bancos centrales.

1. Teoría de las áreas monetarias óptimas: la integración monetaria como coste

La sección parte de los antecedentes de la teoría de las áreas monetarias óptimas. En su formulación inicial, la teoría mide sobre todo el coste de renunciar al tipo de cambio, y de ahí deriva los criterios que permiten reducirlo.

1.1. Antecedentes

Debate sobre regímenes cambiarios: la teoría de las áreas monetarias óptimas reformula el debate de los años cincuenta sobre el régimen cambiario. En ese debate, Milton Friedman (1953) y James Meade defendían que los tipos flexibles permiten absorber las perturbaciones exteriores sin deflación ni desempleo. Frente a ellos, los partidarios de los tipos fijos subrayaban la disciplina antiinflacionista y la menor incertidumbre. Sin embargo, ambos bandos compartían el supuesto, nunca examinado, de que el ámbito de circulación de una moneda coincide con la frontera nacional.

Debate sobre uniones monetarias: si bien existían precedentes prácticos de unión monetaria en el siglo XIX, no existía una teoría que explicase qué determina el tamaño óptimo de un espacio monetario. Robert Mundell aborda ese vacío en 1961, invirtiendo el planteamiento del debate. En lugar de preguntarse qué régimen cambiario conviene a un país dado, se pregunta sobre qué territorio conviene que circule una moneda única, con independencia de las fronteras políticas. En su ejemplo de Norteamérica, si las perturbaciones relevantes separan al este del oeste del continente antes que a Canadá de Estados Unidos, la frontera nacional es el lugar equivocado para situar el límite monetario. En ese caso, ni los tipos fijos ni los flexibles entre ambos países resolverían el problema de ajuste.

1.2. El coste de renunciar al tipo de cambio

La postura inicial de Mundell en los años sesenta es que la integración monetaria supone un coste desde el punto de vista de la estabilización macroeconómica[2]. El coste nace de que cada miembro ya no puede recurrir a su tipo de cambio ni a una política monetaria propia para ajustarse. No se trata, sin embargo, de un coste que deba minimizarse sin más, porque el criterio de optimalidad surge del equilibrio entre dos fuerzas opuestas. De un lado, la moneda única priva a cada miembro de esos instrumentos frente a perturbaciones asimétricas. De otro, fragmentar el espacio monetario hasta hacer coincidir cada área con una región elevaría los costes de transacción y erosionaría las funciones del dinero como unidad de cuenta y medio de pago.

Por **área monetaria óptima** se entiende una entidad con varias regiones y moneda única que permite aproximarse simultáneamente al equilibrio interno (pleno empleo y estabilidad de precios) y al equilibrio externo (posición sostenible de la balanza de pagos). Es óptima, en consecuencia, el área que hace mínimo el coste de estabilización, neto de lo que gana en eficiencia con la moneda común. La magnitud de ese coste vendrá determinada por la capacidad de respuesta que tengan el tipo de cambio y la política monetaria propia ante *shocks* reales asimétricos.

[2]Este enfoque nace con la aportación de Robert Mundell en *A Theory of Optimum Currency Areas* (1961).

1.2.1. Escenario base: modelo de oferta y demanda agregada en economía abierta

Los efectos de la integración monetaria pueden ilustrarse con el modelo OA-DA en economía abierta, que determina a la vez la producción y el nivel de precios de cada país.

> El modelo de oferta y demanda agregadas en economía abierta, y la autoría de la curva de oferta agregada, se desarrollan en el tema A32, apartado 2.

Gráficamente, en el plano renta-precios (figura B16.1), el modelo se aplica a dos países, Francia (panel izquierdo) y Alemania (panel derecho), siguiendo el ejemplo clásico de De Grauwe:

1. *Demanda agregada* (DA_0): refleja las combinaciones de renta y precios que equilibran los mercados de bienes y de dinero. Tiene pendiente negativa porque unos precios más altos reducen la oferta real de dinero y elevan el tipo de interés, lo que contrae la inversión, y aprecian el tipo de cambio real, lo que deteriora la balanza comercial.

2. *Oferta agregada a corto plazo* ($\mathrm{OA}_{\mathrm{CP}}$): refleja la producción que ofrecen las empresas a cada nivel de precios con los salarios nominales fijados en los contratos. Tiene pendiente positiva porque, con el salario nominal fijo, una subida de precios abarata el trabajo en términos reales y las empresas producen más.

3. *Oferta agregada a largo plazo* ($\mathrm{OA}_{\mathrm{LP}}$): refleja la producción una vez que las expectativas de precios se cumplen y los salarios se han ajustado a ellas. Tiene mayor pendiente que la $\mathrm{OA}_{\mathrm{CP}}$ porque ese ajuste deja el salario real casi inalterado, y no llega a ser vertical porque, en economía abierta, los salarios se negocian según el IPC, que sube menos que el deflactor.

4. *Equilibrio*: cada país parte de un equilibrio aislado en el punto 0, donde se cortan las tres curvas, con la producción en su nivel de largo plazo, sin presiones sobre los precios y con la balanza comercial equilibrada.

Sobre ese equilibrio incide un *shock* asimétrico de demanda. Un cambio de preferencias de los consumidores conduce a un mayor consumo de productos alemanes en detrimento de los franceses, y cada economía sufre un doble desequilibrio:

- *Economía francesa*:
  - *Desequilibrio interno*: la demanda agregada se desplaza hacia la izquierda, de DA_0 a DA_1 (punto $0'$), con lo que caen la producción y los precios y aumenta el desempleo.
  - *Desequilibrio externo*: la caída de la demanda de sus bienes reduce las exportaciones netas ($\Delta \mathrm{BC}_F < 0$) y genera un déficit comercial.
- *Economía alemana*:
  - *Desequilibrio interno*: la demanda agregada se desplaza hacia la derecha, de DA_0 a DA_1 (punto $0'$), con lo que aumentan la producción y los precios (presiones inflacionistas).
  - *Desequilibrio externo*: el aumento de la demanda de sus bienes eleva las exportaciones netas ($\Delta \mathrm{BC}_A > 0$) y genera un superávit comercial.

El problema lo genera la asimetría del *shock*. En cambio, ante una perturbación simétrica, que afectase a ambas economías en el mismo sentido y con la misma intensidad, una autoridad monetaria común respondería para el conjunto y no existiría coste alguno derivado de compartir la moneda. A partir de aquí, el mecanismo de ajuste depende de la forma de integración entre ambos países.

1.2.2. Mecanismo de ajuste sin unión monetaria

Si ambos países no integran una unión monetaria, conservarán la política monetaria y cambiaria para hacer frente al *shock*:

Tipos de cambio flexibles: el ajuste opera, en primer término, de forma automática. La caída de la demanda de bienes franceses reduce la demanda de su moneda en el mercado de divisas y provoca su depreciación, que abarata las exportaciones francesas y encarece las importaciones francesas procedentes de Alemania, corrigiendo simultáneamente el desequilibrio interno y el externo. La política monetaria refuerza ese mecanismo, no lo sustituye:

- *Francia*: el banco central francés aplicará una política monetaria expansiva para estimular la demanda agregada, cuyo efecto sobre el tipo de cambio se suma a la depreciación ya en curso.

- *Alemania*: el Bundesbank aplicará una política monetaria contractiva para mitigar los efectos inflacionistas, lo que apreciará el marco y reforzará el ajuste en sentido contrario.

El resultado final será un desplazamiento de la demanda agregada francesa hacia la derecha y de la alemana hacia la izquierda, con lo que se restauran ambos equilibrios.

Tipos de cambio fijos: el ajuste ya no es automático y exige una decisión discrecional de las autoridades:

- *Francia*: el banco central francés realizará una devaluación para promover la competitividad de las exportaciones francesas. A diferencia del caso anterior, se trata de un reajuste discreto de la paridad que, en un sistema de tipos fijos negociado, requiere el acuerdo de los demás miembros, lo que limita su disponibilidad efectiva como instrumento.

- *Alemania*: el Bundesbank podrá seguir recurriendo a la política monetaria contractiva para mitigar los efectos de la inflación, si bien lo hará por su condición de país ancla del sistema, y no por conservar autonomía cambiaria. Conforme al trilema de Mundell-Fleming, un régimen de tipos fijos con libertad de movimientos de capital obliga a renunciar a la política monetaria autónoma. En un sistema asimétrico, esa renuncia recae sobre los países periféricos, que importan la política del ancla, y no sobre este.

En ambos regímenes, la no integración permite a los dos países estabilizar el ciclo, de forma automática con tipos flexibles y mediante decisiones discrecionales con tipos fijos (la devaluación francesa y la política monetaria del país ancla).

1.2.3. Mecanismo de ajuste con unión monetaria

Si ambas economías formasen una unión monetaria, ninguno de los dos países podría favorecer el ajuste mediante el uso autónomo de su política cambiaria ni monetaria, al compartir la misma moneda. En su artículo original, Robert Mundell señalaba precisamente este coste como el principal. El ajuste recae entonces sobre precios y salarios, y cada economía pasa del punto $0'$ al 1, sobre la oferta agregada a largo plazo, con precios más bajos en Francia y más altos en Alemania (figura B16.1).

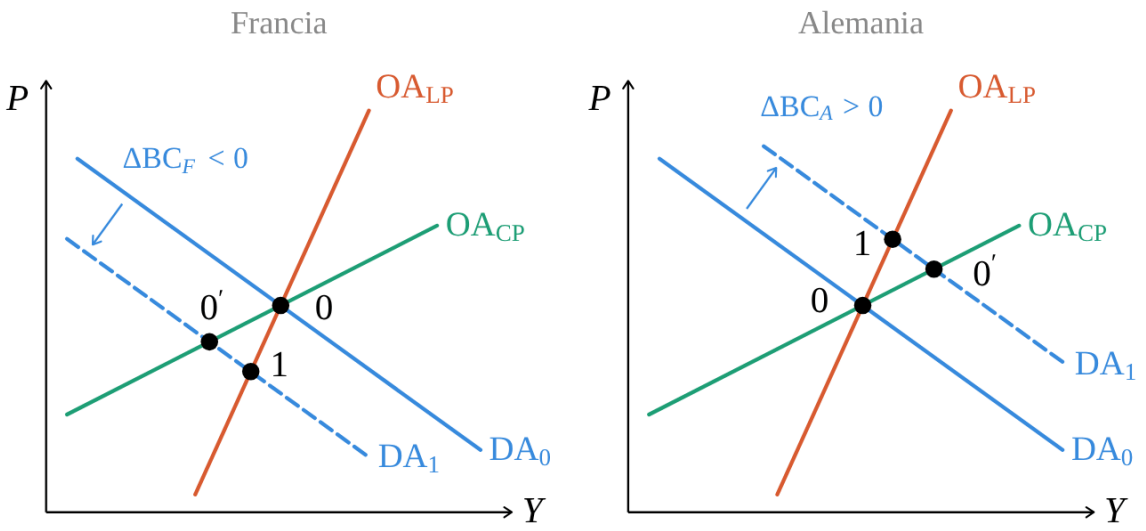

Figura B16.1: Ajuste ante un *shock* asimétrico en una unión monetaria. Sin tipo de cambio propio, el ajuste recae sobre precios y salarios.

El modelo supone que la variación del tipo de cambio nominal se traduce en una variación del tipo de cambio real, lo que exige rigidez nominal de precios y salarios, y que se cumple la condición Marshall-Lerner. Si los salarios se ajustasen con rapidez, la devaluación quedaría erosionada y el instrumento del que se prescinde valdría menos de lo que aquí aparenta. Por tanto, el modelo ofrece una estimación de máximos del coste de la integración monetaria, cuestión que el propio Mundell revisará en 1973 y que se aborda en el apartado 2.1.1.

1.3. Determinantes de la optimalidad

Diversos autores han planteado distintos determinantes que minimizan los costes que supone la integración monetaria. Los siete criterios proceden de los años sesenta y setenta, y cada uno se contrasta con la experiencia de la zona del euro.

1.3.1. Condiciones institucionales del mercado laboral (Mundell, 1961)

Origen: en primer lugar, la propuesta de Mundell en su artículo original es que la flexibilidad de precios y salarios o la libertad de movimiento del factor trabajo permitirán reducir los costes de la integración.

Idea: el autor argumenta que si el mercado de trabajo cumple al menos una de estas características, se puede dar un proceso de ajuste automático que no haga necesaria la intervención. Por lo tanto, el factor determinante en la minimización del coste de la integración son las condiciones institucionales del mercado laboral. Mundell se refiere a la movilidad del factor trabajo y no a la movilidad del capital, que no es un sustituto equivalente, pues permite financiar el desequilibrio pero no corregirlo.

Razonamiento: en el ejemplo anterior, cada condición corrige el doble desequilibrio por una vía distinta:

▪ *Flexibilidad de precios y salarios en ambos países*: el tipo de cambio real puede depreciarse sin que lo haga el nominal. El ajuste pasa así por los precios de los bienes, no por el de las monedas. En el ejemplo, los precios bajan en Francia y suben en Alemania:

- *Economía francesa*:
  - *Ajuste interno*: el desempleo generado por la caída de la demanda, a los salarios iniciales, hace descender el salario, con lo que la oferta agregada y el empleo aumentan y los precios bajan.
  - *Ajuste externo*: la caída de los precios deprecia el tipo de cambio real, y la consiguiente ganancia de competitividad eleva las exportaciones netas y reequilibra el déficit comercial.

- *Economía alemana*: ocurre lo contrario.
  - *Ajuste interno*: al crecer la demanda agregada, las empresas contratan más y pagan salarios más altos, lo que contrae la oferta agregada y eleva los precios.
  - *Ajuste externo*: el tipo de cambio real se aprecia y, con él, caen las exportaciones y se corrige el superávit exterior, aunque al precio de una inflación más alta.

▪ *Libertad de movilidad del factor trabajo entre ambos países*:

- *Economía francesa*:
  - *Ajuste interno*: el paro francés baja porque parte de los desempleados emigra a Alemania, donde el *shock* ha dejado más vacantes que candidatos. Quienes se van dejan de consumir en Francia, con lo que la absorción cae. En cambio, la producción, a la que los emigrantes no contribuían por estar parados, no varía.
  - *Ajuste externo*: con una producción que no cambia, la menor absorción recorta el déficit exterior.

- *Economía alemana*: los trabajadores franceses que llegan cubren las vacantes pendientes y contienen así la subida de salarios y de precios. No obstante, los migrantes pasan a consumir en Alemania, lo que eleva adicionalmente la absorción, con un efecto ambiguo sobre el superávit exterior.

En el país al que el *shock* real de demanda favoreció en un principio, la flexibilidad corrige el superávit a costa de más inflación, y el efecto de la movilidad del trabajo sobre el saldo exterior es ambiguo. En el perjudicado, en cambio, los dos rebajan el paro y el déficit por cuenta corriente.

Extensiones:

▪ *Diferencias institucionales en el mercado laboral* (Calmfors y Driffill, 1988): la relación entre la centralización de la negociación colectiva y los resultados salariales tiene forma de joroba. Como los países difieren en ese grado de centralización, un mismo *shock* real de oferta afecta de forma distinta a la inflación de cada país, lo que eleva el coste de integración (véase el tema A27, apartado 2.1.2).

▪ *Ajuste regional en la práctica* (Blanchard y Katz, 1992; Decressin y Fatás, 1995): ante *shocks* regionales de demanda, en los estados de Estados Unidos la migración interestatal es el mecanismo de ajuste dominante. En las regiones europeas, en cambio, el ajuste discurre sobre todo por la variación de la tasa de actividad (la salida de trabajadores del mercado laboral).

- *Evidencia empírica (zona del euro)*:

  - *Flexibilidad laboral*: en agosto de 2011, en vísperas de empezar a comprar deuda española, el BCE pidió por carta a España que flexibilizara su mercado laboral, una reforma que buena parte de la zona del euro, sobre todo el sur, sigue sin completar.

  - *Movilidad del trabajo*: pese a que la libre circulación de personas permite en teoría una movilidad laboral amplia, las barreras de lengua y cultura, ausentes en Estados Unidos, la restringen notablemente.

Conclusiones:

- *Aspectos positivos*: identifica los mecanismos que corrigen el desequilibrio, y no solo lo financian, como haría la movilidad del capital. El criterio es, además, endógeno a las instituciones, porque la flexibilidad salarial no es un dato sino un objetivo de política económica, como ilustra la condicionalidad impuesta durante la crisis de deuda soberana.

- *Limitaciones*: es el criterio más exigente, pues requiere que el mercado de trabajo sustituya por completo la función estabilizadora del tipo de cambio, y el que la zona del euro cumple con mayor dificultad. Sus efectos sobre el país beneficiado inicialmente por el *shock* son, por otra parte, ambiguos.

1.3.2. Integración fiscal (Kenen, 1969; Informe MacDougall, 1977)

Origen: Peter Kenen (1969) sostiene que la integración fiscal es condición necesaria de una unión monetaria viable.

Idea: renunciar a la política monetaria propia costaría menos si todos los miembros de la unión monetaria compartieran un presupuesto central, porque este redistribuiría recursos de forma automática hacia los afectados por un *shock* asimétrico. El mecanismo opera como un seguro mutuo, dado que la unión fiscal transforma un riesgo idiosincrásico, que cada miembro habría de soportar en solitario, en un riesgo compartido por el conjunto.

Razonamiento: en el ejemplo anterior, el *shock* asimétrico eleva la recaudación fiscal en Alemania (y la reduce en Francia) a través de los estabilizadores automáticos. En el contexto de la integración fiscal, estos recursos podrían transferirse a Francia mediante menores impuestos y mayores transferencias (p. ej. prestaciones por desempleo), con carácter automático y sin necesidad de decisión política *ad hoc*.

Extensiones:

- *Cuantificación del presupuesto necesario*: el Informe MacDougall (1977), elaborado por encargo de la Comisión Europea, constituye el primer intento de cuantificar el criterio. Sus autores estimaron que un presupuesto común del orden del 5 al 7 % del PIB comunitario podría bastar para aportar la nivelación territorial y la amortiguación de perturbaciones necesarias para sostener una unión monetaria, frente al 20-25 % de las federaciones existentes. La conclusión del informe fue que la Comunidad no estaba entonces en condiciones de sostener una moneda única, y su recomendación quedó sin aplicación.

- *Estabilización federal en Estados Unidos*: Sala-i-Martin y Sachs (1992) estiman que el presupuesto federal, mediante impuestos y transferencias, absorbe una fracción significativa de las perturbaciones de renta de los estados. Sin embargo, estudios posteriores, como el de Asdrubali, Sørensen y Yosha (1996), han revisado esa estimación a la baja. Para 1963-1990, estos autores

sitúan en el 13 % la parte de los *shocks* de los estados que absorbe el presupuesto federal (tema B44, apartado 3.3). La comparación sigue ilustrando, con todo, la magnitud de la carencia europea.

▪ *Evidencia empírica (zona del euro)*: la centralización presupuestaria no se cumple. El presupuesto de la Unión, reducido en relación con el PIB comunitario, queda muy lejos del 5-7 % que estimaba necesario el Informe MacDougall, y el refuerzo que ha supuesto NextGenerationEU es temporal y carece de la naturaleza de un presupuesto federal permanente.

> Sobre el presupuesto de la Unión (en torno al 1 % de la renta nacional bruta), el marco financiero plurianual y la financiación de NextGenerationEU, véase el tema B37. La gobernanza fiscal de la Unión Económica y Monetaria (UEM), con el Pacto de Estabilidad y Crecimiento (PEC), el Semestre Europeo y el Mecanismo Europeo de Estabilidad (MEDE), y el debate sobre su capacidad fiscal se desarrollan en el tema B45.

Conclusiones:

▪ *Aspectos positivos*: el presupuesto común actúa como un seguro mutuo automático, a través de los estabilizadores y sin decisión política *ad hoc*, y su dimensión necesaria admite una cuantificación (Informe MacDougall, 1977).

▪ *Limitaciones*: la integración fiscal es el criterio cuya ausencia resulta más visible en el diseño de la zona del euro, creada con moneda y banco central únicos sobre una unión fiscal inexistente, con reglas de disciplina presupuestaria nacional en lugar de seguro mutuo. El ajuste ante perturbaciones asimétricas recae así sobre el presupuesto del país afectado, precisamente cuando se deteriora su capacidad de endeudamiento, lo que introduce un sesgo procíclico.

1.3.3. Integración comercial (McKinnon, 1963)

Origen: Ronald McKinnon (1963) plantea que el criterio relevante es el *grado de apertura* de la economía, que el autor mide por la proporción de bienes comerciables respecto de los no comerciables y que suele aproximarse mediante el índice $(X + M)/Y$.

Idea: una economía muy integrada comercialmente pierde menos al abandonar su política monetaria y cambiaria propia. Hay para ello dos razones principales:

1. *Efectividad de las devaluaciones*: la integración comercial la reduce.

2. *Esfuerzo fiscal*: con más apertura basta un ajuste fiscal menor para corregir la cuenta corriente, con lo que el equilibrio externo depende menos de la política monetaria.

Adviértase que el argumento de McKinnon opera de forma distinta al de Mundell. El coste de la unión monetaria es menor no porque se disponga de un mecanismo sustitutivo, sino porque el instrumento al que se renuncia tenía de partida escasa eficacia.

Razonamiento: la devaluación pierde efectividad por su efecto inflacionario. Cuanto más abierta esté una economía, mayor será la proporción de bienes intermedios importados utilizados en el proceso de producción y también la proporción de bienes comerciables, es decir, bienes potencialmente exportables o importables[3]. En este contexto, una devaluación provoca un doble efecto:

[3] Al estar sometidos a la competencia exterior, los precios interiores de los bienes comerciables no pueden desviarse mucho de los mundiales. Por ello, los diferenciales de inflación entre países proceden sobre todo de los bienes y servicios no comerciables.

- *Costes*: encarece los bienes intermedios importados, elevando los costes de producción y, con ellos, los precios interiores.

- *Precios de consumo*: eleva el precio en moneda nacional de los bienes comerciables, que se trasladará rápidamente al índice de precios de consumo (y, después, a los salarios vía indexación).

El índice de precios combina bienes comerciables y no comerciables, así que el impacto inflacionista de una devaluación crece con la ponderación que tengan los primeros. Cuanto mayor sea esa ponderación, antes se desvanecerá también la ganancia de competitividad real que se buscaba con la devaluación.

El esfuerzo fiscal necesario para equilibrar la cuenta corriente también disminuye con la apertura. Ante un déficit, una política fiscal restrictiva lo recorta al contraer la demanda y, con ella, la renta nacional, que a su vez arrastra a la baja las importaciones. El efecto sobre la renta viene dado por el multiplicador del gasto público, que depende de la propensión marginal a consumir (c), y el efecto sobre la cuenta corriente, por la propensión marginal a importar (m):

$$\Delta Y = \frac{1}{1-c+m}\Delta G$$

$$\text{BC} = X - mY \rightarrow \Delta\text{BC} = -m\Delta Y = \frac{-m}{1-c+m}\Delta G$$

Las economías más abiertas suelen tener una propensión marginal a importar (m) más alta, y cuanto mayor es m, más mejora el saldo exterior con cada unidad monetaria que se recorta del gasto público. Con una integración comercial elevada, la política monetaria resulta, en consecuencia, menos necesaria como instrumento de estabilización.

Gráficamente, en el plano grado de apertura-costes y beneficios en porcentaje del PIB (figura B16.2), el criterio de McKinnon da una curva de costes decreciente con la apertura, $(X+M)/Y$. Los beneficios de la moneda única, como la reducción de los costes de transacción (apartado 2.2.1), crecen en cambio con ella, y el corte de ambas curvas fija un grado de apertura crítico, T^*, a partir del cual la unión monetaria compensa.

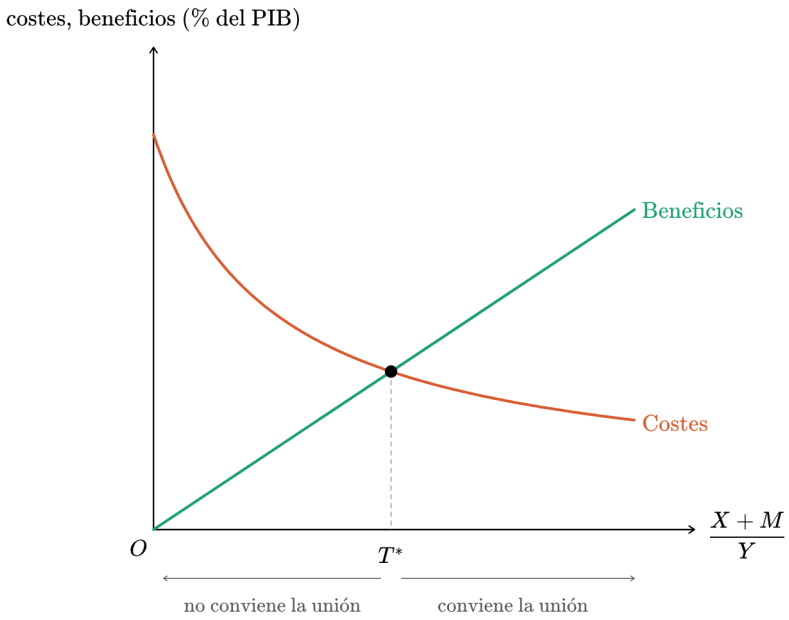

Figura B16.2: Costes y beneficios de la unión monetaria según el grado de apertura. Los costes de renunciar al tipo de cambio decrecen con la apertura, $(X + M)/Y$ (criterio de McKinnon), y los beneficios, como la reducción de los costes de transacción, crecen con ella. A la derecha de su corte, en la apertura crítica T^*, conviene la unión.

Extensiones: en la zona del euro este requisito sí se cumple, y es el criterio que la unión satisface con mayor holgura. Una parte muy sustancial del comercio de los Estados miembros es comercio intracomunitario, lo que eleva el grado de apertura relevante a efectos del criterio de McKinnon.

> Sobre la integración comercial y sus efectos, véase el tema B10.

Conclusiones:

▪ *Aspectos positivos*: el grado de apertura se mide con facilidad, y el criterio muestra que la renuncia al tipo de cambio cuesta menos cuando el instrumento tenía de partida escasa eficacia.

▪ *Limitaciones*: el cumplimiento del criterio no es concluyente, pues Paul Krugman (1993) advierte de que la integración comercial impulsada por economías de escala tiende a concentrar geográficamente las actividades industriales, lo que aumenta la especialización regional y la probabilidad de perturbaciones asimétricas. La integración comercial reduciría así el coste del ajuste al tiempo que incrementa la frecuencia con que resulta necesario, de modo que su efecto neto es ambiguo. El argumento constituye el reverso de la tesis de la endogeneidad (apartado 2.1.3).

1.3.4. Productividad e inflación (Balassa-Samuelson, 1964; Magnifico, 1971)

Origen: en una cuarta línea de aportaciones destacan las de Béla Balassa y Paul Samuelson (1964), Gottfried Haberler (1970), Marcus Fleming (1971) y Giovanni Magnifico (1971), que ponen de manifiesto la importancia de los diferenciales de inflación.

Idea: la estabilidad de los tipos de cambio reales entre los miembros de la unión monetaria aumenta a medida que se estrechan sus diferenciales de inflación, y con ella se contienen los desequilibrios externos.

Razonamiento: según el efecto Balassa-Samuelson, si la productividad crece más deprisa en el sector comerciable, sus salarios suben sin que lo hagan sus precios, fijados en el mercado mundial. La movilidad intersectorial del trabajo traslada esa subida salarial al sector no comerciable, cuyos precios aumentan. Por tanto, el país con mayor crecimiento relativo de la productividad en el sector expuesto registra un diferencial de inflación estructural y una apreciación real sin pérdida de competitividad. Cuanto menores sean las diferencias de crecimiento de la productividad entre los miembros, menores serán los diferenciales de inflación y el coste de renunciar a políticas monetarias autónomas.

> El modelo de Balassa-Samuelson (supuestos, ecuaciones y la expresión del tipo de cambio real en función de los diferenciales sectoriales de productividad) se desarrolla en el tema B14, apartado 1.2.3.

Extensiones:

- *Preferencia nacional entre inflación y desempleo* (Magnifico, 1971): cada país presenta una *propensión nacional a la inflación* diferente, resultado de sus instituciones laborales, de su estructura productiva y de las preferencias sociales sobre el reparto entre inflación y desempleo. No obstante, una unión monetaria impone a todos sus miembros, a medio plazo y salvo diferenciales estructurales, una misma tasa de inflación. Los miembros cuyo óptimo nacional difiera de la media soportarán, por ello, un coste en términos de bienestar aun cuando no se produzca perturbación asimétrica alguna. El coste, en este caso, no procede del ajuste ante *shocks*, sino de la propia uniformidad del resultado.

- *Evidencia empírica (zona del euro)*: el criterio solo se cumple en parte. Antes de la crisis financiera la inflación divergía mucho entre países, y la dispersión se redujo de forma notable después[4]. Volvió a ampliarse con fuerza, sin embargo, en el episodio inflacionista de 2021-2023, pues en 2022 la inflación media osciló entre cerca del 6 % en Francia o Malta y casi el 20 % en Estonia.

Conclusiones:

- *Aspectos positivos*: la convergencia de la inflación reduce el coste de la integración al estabilizar el tipo de cambio real y evitar desequilibrios externos.

- *Limitaciones*: la misma convergencia aumenta el coste en el sentido de Magnifico, pues, mientras exista una disyuntiva entre inflación y desempleo, los miembros no pueden situarse en distintos puntos de la curva de Phillips. Además, por el efecto Balassa-Samuelson, un diferencial de inflación puede ser de equilibrio y no síntoma de pérdida de competitividad, lo que resta precisión al criterio.

[4] Para el criterio de convergencia de precios (art. 140 del Tratado de Funcionamiento de la Unión Europea y Protocolo n.º 13), la referencia son los Estados miembros con mejores resultados en estabilidad de precios, hasta un máximo de tres. La inflación del candidato no puede superar la de ese grupo en más de 1,5 puntos porcentuales. Los informes de convergencia usan en la práctica la media de esos países y descartan los valores atípicos.

1.3.5. Diversificación: sincronización de ciclos (Kenen, 1969)

Origen: Peter Kenen, en el mismo artículo en que formula el criterio de integración fiscal (1969), sostiene que una economía con una producción diversificada depende menos del tipo de cambio como instrumento de ajuste. La razón es que las perturbaciones propias de cada sector tienden a compensarse entre sí y rara vez se traducen en un *shock* agregado asimétrico.

Idea: una misma política monetaria para todos, de talla única (*one size fits all*), resulta menos costosa, en términos de renuncia a la política monetaria propia, cuando los ciclos de los miembros están sincronizados, porque todos atraviesan entonces la misma coyuntura.

Razonamiento: la sincronía de los ciclos exige que se dé al menos una de estas tres condiciones:

1. *Diversificación económica*: que los países presenten tejidos productivos muy diversificados. Cuanto más diverso sea el tejido productivo de cada economía, mayor será la posibilidad de que un *shock* sectorial específico afecte a todos los países a la vez, puesto que en todos ellos ese sector forma parte de su economía.

2. *Concentración sectorial*: a falta de diversificación, que todas las economías se especialicen en un mismo sector, con lo que cualquier *shock* afecta a todas por igual.

3. *Especialización cíclicamente correlacionada*: si no se da ninguna de las anteriores, que los sectores en que se especializa cada país tengan ciclos correlacionados (p. ej. el turismo en un país y la producción de vino en otro).

Las principales fuentes de desincronización del ciclo económico de los países son:

- *Existencia de política fiscal descentralizada*: si la unión monetaria no va acompañada de una coordinación fiscal, las autoridades fiscales de cada país podrían mantener políticas fiscales de distinto tono y ciclicidad y provocar, así, *shocks* asimétricos.

- *Heterogeneidad de las instituciones económicas*: estas diferencias pueden hacer que, ante una misma coyuntura económica o *shock* simétrico, se activen fases expansivas en unos países y fases recesivas en otros. Estas diferencias se suelen reflejar en la divergencia de sus capacidades competitivas.

Extensiones: en la zona del euro el cumplimiento es parcial. Las economías de la zona del euro se parecen bastante entre sí, pero algunas del sur han dependido en exceso de sectores concretos (p. ej. el turismo).

Conclusiones:

- *Aspectos positivos*: el criterio de Kenen señala que la asimetría relevante no es la de los países, sino la de los sectores en que cada uno se especializa.

- *Limitaciones*: es el más ambiguo de los criterios, pues admite tres configuraciones contradictorias entre sí, lo que lo hace difícil de contrastar empíricamente.

1.3.6. Grado de integración financiera (Ingram, 1973)

Origen: James Ingram (1973) estudia la importancia de la integración financiera.

Idea: unos mercados financieros bien integrados entre los miembros abaratan la renuncia a la política monetaria propia, ya que permiten financiar los desequilibrios exteriores sin mover el tipo de cambio.

Razonamiento: las entradas de capital bastan para financiar el desequilibrio por cuenta corriente que provoca un *shock* si la integración financiera es elevada, y el tipo de cambio no necesita ajustarse. En el mismo ejemplo, el ahorro alemán financiaría el déficit exterior francés mientras dura la perturbación.

Extensiones:

- *Diferencias institucionales en el mercado financiero*: la integración financiera exige, además, cierta homogeneidad de las estructuras financieras nacionales. Ante un mismo *shock* (p. ej. un aumento del tipo de interés), los mercados hipotecarios reaccionan de manera distinta según predominen los préstamos a tipo fijo o a tipo variable. El predominio de la banca o de los mercados de capitales determina, a su vez, si el impulso se transmite sobre todo por el canal del crédito, con riesgo de racionamiento (*credit crunch*), o por el efecto riqueza.

- *Evidencia empírica (zona del euro)*: este requisito se cumple con matices. La Unión Bancaria y la Unión de los Mercados de Capitales, relanzada como Unión de Ahorros e Inversiones, han impulsado la integración financiera, con el objetivo de diversificar las fuentes de financiación, reducir la dependencia bancaria y eliminar los obstáculos a la inversión transfronteriza.

> La Unión Bancaria y la Unión de los Mercados de Capitales, hoy Unión de Ahorros e Inversiones, se desarrollan en el tema B44, apartados 2 y 3.

Conclusiones:

- *Aspectos positivos*: la integración financiera permite financiar los desequilibrios por cuenta corriente sin ajustes del tipo de cambio y, cuando se apoya en el capital, reparte además el riesgo entre los miembros, lo que justifica el impulso a la Unión de los Mercados de Capitales.

- *Limitaciones*: la crisis de deuda soberana cambió la lectura del criterio de Ingram. La integración financiera previa a 2008 fue intensa pero de naturaleza bancaria y de corto plazo, de modo que ante la perturbación no absorbió el *shock* sino que lo amplificó. Los flujos se revirtieron con brusquedad (*sudden stop*) y se estableció el círculo vicioso entre bancos y soberanos (*doom loop*, apartado 3.2.1). De ahí que la literatura posterior distinga entre integración por deuda, que financia pero no comparte el riesgo, e integración por capital, la única que cumple la función estabilizadora que Ingram atribuía al criterio.

1.3.7. Diferenciales en las tasas de crecimiento (Magnifico, 1971)

Origen: queda por considerar la relación entre integración monetaria y crecimiento económico, cuestión abordada por Giovanni Magnifico (1971).

Idea: un crecimiento parecido entre los miembros limita los desequilibrios externos dentro de la unión y abarata, por tanto, lo que a cada país le cuesta renunciar a una política monetaria autónoma.

Razonamiento: supóngase una unión monetaria de A y B en la que A crece más deprisa que B. Si la elasticidad-renta de las importaciones es la misma en los dos países, las compras de A al exterior aumentan más que las de B a medida que la renta de A gana peso relativo, y la cuenta corriente de A se deteriora de forma progresiva.

Extensiones:

▪ *Matizaciones al argumento*: el razonamiento anterior descansa en supuestos restrictivos que conviene explicitar:

- *Propensiones marginales a importar*: no tienen por qué coincidir en los dos países.
- *Volúmenes de importación*: lo que cada país importa del otro puede diferir en cuantía.
- *Crecimiento y elasticidad-renta*: el crecimiento rápido suele ir unido a una oferta de productos más diferenciados, cuyas exportaciones tienen mayor elasticidad-renta, lo que protege la cuenta corriente del deterioro (Krugman, 1989). Esta matización enlaza con la ley de Thirlwall (véase el tema B28, apartado 2.1.3). El problema no es el diferencial de crecimiento en sí, sino la composición de la especialización que lo sustenta.

▪ *Evidencia empírica (zona del euro)*: la zona del euro no satisface este requisito, ya que sus diferenciales de crecimiento han sido mayores que los de Estados Unidos, sobre todo durante la crisis de deuda soberana (2010-2014). Después repuntaron con la pandemia y la invasión de Ucrania, antes de volver a niveles relativamente bajos.

Conclusiones:

▪ *Aspectos positivos*: el criterio resulta relevante por su carácter acumulativo. Un diferencial persistente de crecimiento, con elasticidades-renta de las importaciones similares, genera un desequilibrio exterior que se agrava con el tiempo. Sin tipo de cambio, solo puede corregirse mediante devaluación interna o sostenerse mediante financiación exterior continuada, como ilustró la zona del euro entre 1999 y 2010.

▪ *Limitaciones*: es el menos formalizado de los siete criterios y descansa en supuestos restrictivos, como la igualdad de las propensiones a importar y de los volúmenes de importación. Krugman (1989) cuestiona, además, que el crecimiento deteriore la cuenta corriente, pues los países que más crecen elevan la elasticidad-renta de sus exportaciones.

1.4. Conclusiones

Aportación teórica: los costes de la integración monetaria dependen del punto de partida de cada país en los siete criterios anteriores, cuya elevada exigencia dio lugar a una visión pesimista de la integración.

Los siete criterios no son independientes entre sí ni conmensurables. Unos proporcionan mecanismos sustitutivos del tipo de cambio (movilidad del trabajo, integración fiscal y financiera), otros reducen el valor del instrumento perdido (apertura comercial) y otros disminuyen la probabilidad de que la perturbación sea asimétrica (diversificación, convergencia de inflación y de crecimiento). La teoría no ofrece, en consecuencia, un umbral cuantitativo que permita dictaminar si un área es óptima, sino una lista de condiciones cuyo cumplimiento es siempre parcial. A ello se añade su carácter estático, que la literatura posterior corregirá al mostrar que varios de estos criterios son endógenos a la propia unión monetaria.

La zona del euro como área monetaria óptima: difícilmente puede afirmarse que la zona del euro sea un área monetaria óptima, principalmente porque incumple dos de los requisitos de mayor peso, la flexibilidad del mercado laboral y la integración fiscal. Además, el balance por criterios resulta desigual:

- *Criterios que satisface*: la integración comercial, con holgura, y, con matices, la integración financiera y la convergencia de la inflación.

- *Criterios que incumple*: la flexibilidad y la movilidad del trabajo, la integración fiscal y la convergencia de las tasas de crecimiento.

- *Criterio de resultado ambiguo*: la diversificación.

Esa combinación explica por qué la zona del euro ha podido funcionar sin dificultad ante perturbaciones simétricas y ha atravesado, en cambio, una crisis que llegó a poner en duda su continuidad ante la primera perturbación marcadamente asimétrica.

2. Desarrollos posteriores: los beneficios de la integración monetaria

Los costes de la integración se producen, fundamentalmente, en la dimensión macroeconómica (apartado 1). Sin embargo, los beneficios se sitúan tanto en la dimensión macroeconómica como en la microeconómica y en la geopolítica.

2.1. Beneficios macroeconómicos

La revisión de la teoría empezó por su propio fundador. A la visión más optimista del Mundell de los setenta se suman el argumento de la credibilidad, que nace de la inconsistencia dinámica de la política monetaria, y la tesis de que los criterios de optimalidad son endógenos a la unión.

2.1.1. Visión optimista de Mundell (1973)

Origen: las primeras revisiones en la década de los setenta de la teoría de las áreas monetarias óptimas se deben al propio Mundell (1973), que alcanza conclusiones más optimistas acerca de los costes de conformar una unión monetaria.

Idea: el tipo de cambio estabiliza peor de lo que suponía la visión tradicional, por lo que renunciar a él cuesta menos, y la moneda común aporta, además, beneficios propios.

Razonamiento: en su revisión, Mundell añade dos beneficios de las uniones monetarias:

1. *Sistema de aseguramiento privado*: el autor plantea que los flujos de capital, como mecanismo de transmisión de recursos entre los países miembros, permitirían una respuesta más eficiente de la unión ante *shocks* asimétricos[5].

2. *Reducción de la volatilidad cambiaria*: la capacidad de las políticas cambiarias para absorber los *shocks* asimétricos es menor de lo que apuntaba la visión tradicional. De hecho, en un contexto de información imperfecta, la evolución de los tipos de cambio es más bien una fuente de *shocks* asimétricos que un medio para estabilizarlos. Esto se debe a que dicha evolución se guía, en muchas ocasiones, por factores psicológicos que no necesariamente responden a los fundamentos.

Conclusiones:

- *Aspectos positivos*: la revisión desplaza el eje del argumento del propio fundador de la teoría. Si el tipo de cambio no es un instrumento fiable de estabilización, el coste de renunciar a él es menor de lo estimado, y la discusión se traslada de los costes del ajuste a los beneficios de compartir moneda.

- *Limitaciones*: la revisión admite al menos dos críticas.

1. *Sistema de aseguramiento*: el aseguramiento a través del endeudamiento en los mercados de capitales solo es viable ante *shocks* temporales. En cambio, si el *shock* es permanente, el endeudamiento solo pospone (y encarece) un ajuste que resulta inevitable. Incluso en el caso de que el *shock* fuese temporal, la capacidad estabilizadora o aseguradora de los mercados de capitales desaparecería si estos desconfiasen de uno o más países miembros de la unión monetaria, como confirmó la crisis de deuda soberana.

[5] En el ejemplo del tema, los consumidores franceses, ante una caída temporal de su renta, podrían endeudarse con los alemanes, cuya renta ha aumentado, a través de mercados de capitales integrados, y mantener así su consumo.

2. *Volatilidad cambiaria*: si bien esta puede ser una fuente de *shocks* asimétricos, los *shocks* profundos y persistentes pueden ser más fácilmente gestionados fuera de una unión monetaria (p. ej. a través de medidas depreciatorias o devaluaciones) que dentro de ella.

2.1.2. Inconsistencia dinámica de la política monetaria (Kydland y Prescott, 1977; Barro y Gordon, 1983a)

Origen: los problemas de inconsistencia dinámica se han utilizado también para mostrar los beneficios de la unión monetaria. Por *inconsistencia dinámica* se entiende la situación en que una política óptima *ex ante*, cuando se anuncia, deja de serlo *ex post*, una vez que los agentes han formado sus expectativas sobre la base de ese anuncio. Puesto que los agentes anticipan racionalmente el incentivo a desviarse, el compromiso carece de credibilidad desde el primer momento.

Idea: la delegación de la política monetaria en una autoridad común independiente y creíble permite superar los problemas de inconsistencia dinámica, pues los países renuncian a la tentación de políticas monetarias inflacionistas. La ganancia es de credibilidad, no de imposibilidad de expandir, pues la autoridad común conserva esa capacidad. El argumento se ilustra mediante el modelo de Barro y Gordon (1983a), del que aquí interesa su extensión a dos países.

Supuestos:

▪ *Entorno*: juego de una sola vez (modelo estático), autoridad monetaria discrecional, sin mecanismos de compromiso creíble, y economía neoclásica con plena flexibilidad de precios y salarios a largo plazo.

▪ *Agentes*: la sociedad forma sus expectativas de inflación según la HER, $\pi_t^e = \mathbb{E}_{t-1}[\pi_t]$, y sobre ellas fija sus contratos de precios y salarios. El Gobierno (autoridad monetaria), por su parte, minimiza la función de pérdida social $L = a(u - u^T)^2 + b(\pi - \pi^T)^2$, con un objetivo de desempleo inferior a la tasa natural ($u^T < \bar{u}$) por entender que esta resulta ineficientemente elevada. Esa discrepancia origina el sesgo inflacionista. Si u^T coincidiese con $\bar{u}$, el compromiso antiinflacionista sería creíble. Las autoridades son miopes, es decir, actúan racionalmente, pero dan prioridad absoluta al corto plazo debido a los costes políticos.

▪ *Restricciones*: la curva de Phillips ampliada por las expectativas, que implica una disyuntiva entre inflación y desempleo a corto plazo, ante cambios imprevistos, pero no a largo plazo.

> El modelo de Barro y Gordon en economía cerrada (planteamiento, solución con discrecionalidad y con regla, y las soluciones al problema de credibilidad) se desarrolla en el tema A06, apartado 3.1.2. Sobre reglas monetarias, credibilidad y banquero central conservador, véanse los temas A36 y A37; sobre el diseño institucional del Sistema Europeo de Bancos Centrales, el tema B43.

Planteamiento: el Gobierno elige la inflación que minimiza la pérdida social, dadas la curva de Phillips ampliada y las expectativas de la sociedad, en tres escenarios:

▪ *Economía cerrada: el sesgo inflacionista*: la autoridad que anuncia una inflación baja ($\pi = \pi^T$, con $u = \bar{u}$) tiene incentivos a generar una inflación sorpresa que reduzca transitoriamente el desempleo. Sin embargo, bajo la HER los agentes lo anticipan, y el equilibrio discrecional de Nash, en la tangencia entre la curva de indiferencia social (CIS) y la curva de Phillips a corto plazo (CP_{CP}), presenta el mismo desempleo natural con una inflación superior a π^T sin contrapartida real. El sesgo es tanto mayor cuanto más blando sea el Gobierno (*wet*, a elevado, con CIS

alargadas verticalmente, frente a *hard-nosed*, b elevado, con CIS aplanadas). La discrecionalidad no produce menos desempleo, sino más inflación.

▪ *Economía abierta sin integración monetaria*: el marco se replica para dos países con preferencias distintas, Italia (Gobierno blando, con CIS alargadas verticalmente) y Alemania (Gobierno duro, con CIS aplanadas). Se añade, además, un instrumento adicional, el tipo de cambio, cuya fijación, por la vía de la paridad del poder adquisitivo relativa, iguala la tasa de inflación del país que fija a la del país ancla.

▪ *Economía abierta con integración monetaria*: Italia puede, en vez de fijar un tipo de cambio, integrarse en una unión monetaria con Alemania cuya autoridad monetaria común siga los principios del antiguo Bundesbank (en la que prevalezca la política monetaria del país duro). En ese caso, la restricción del modelo se modifica, porque la moneda nacional deja de existir y, con ella, la opción de devaluar por sorpresa.

Desarrollo gráfico: el modelo de Barro y Gordon se representa en el plano desempleo-inflación de cada país (figura B16.3):

1. *Curva de Phillips a largo plazo* (CP_{LP}): refleja las combinaciones de desempleo e inflación con expectativas cumplidas ($\pi = \pi^e$). Es vertical en la tasa natural de desempleo, $\bar{u}$, porque sin sorpresas de inflación el empleo no varía.

2. *Curvas de Phillips a corto plazo* (CP_{CP}): reflejan la disyuntiva entre inflación y desempleo para una inflación esperada dada, π^e, con la que se han fijado los contratos. Tienen pendiente negativa porque una inflación mayor que la esperada reduce transitoriamente el desempleo, y cada una corta la CP_{LP} en $\pi = \pi^e$.

3. *Curvas de indiferencia social* (CIS^A y CIS^I): recogen las combinaciones de desempleo e inflación con igual pérdida social y son elipses centradas en los objetivos (u^T, π^T), con menor pérdida cuanto más cerca de ellos. Las de Italia, Gobierno blando (a elevado), son alargadas verticalmente y las de Alemania, Gobierno duro (b elevado), aplanadas, porque cada Gobierno cede menos en el objetivo que más pesa en su pérdida, L.

4. *Equilibrio*: por sus preferencias, Italia parte de un equilibrio discrecional con inflación elevada, π_0, y desempleo natural (punto 0 de Italia, tangencia de CIS_0^I con $\text{CP}_{\text{CP},0}$). Alemania, con la credibilidad añadida del Bundesbank, parte de otro equilibrio con inflación reducida, π^A, y el mismo desempleo (punto 0 de Alemania).

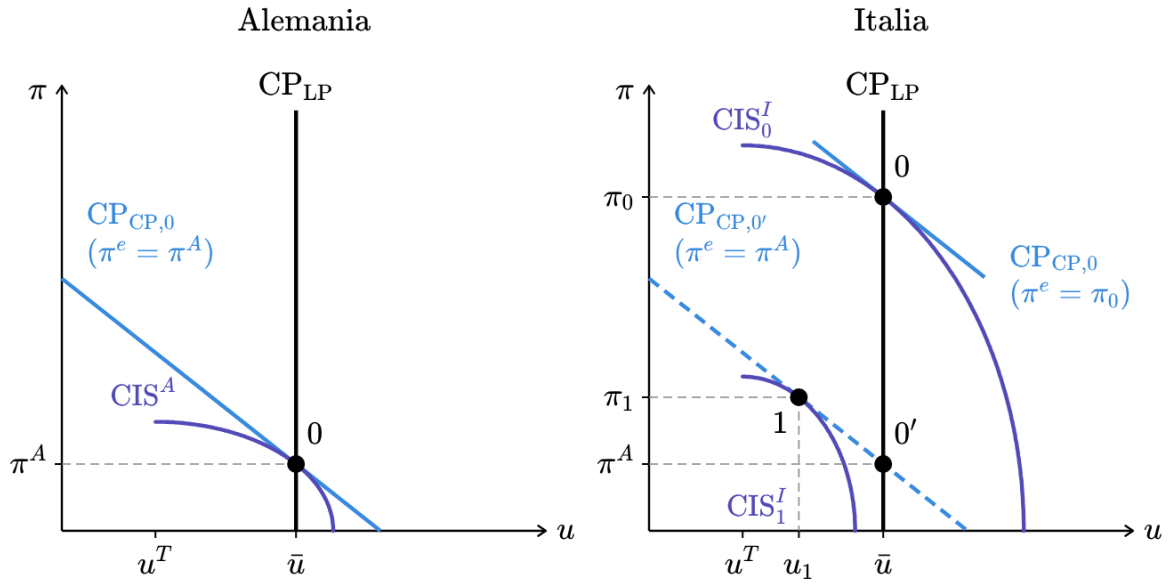

Figura B16.3: Inconsistencia dinámica en dos países (Barro-Gordon). Con un banco central más averso a la inflación, Alemania alcanza un equilibrio con menor sesgo inflacionista que Italia. La unión monetaria permite a Italia importar credibilidad.

Estática comparativa: a partir del equilibrio discrecional 0, Italia puede recurrir al tipo de cambio:

▪ *Fijación del tipo de cambio (0 → 0′)*: Italia vincula su moneda a la alemana con un tipo de cambio fijo para resolver su falta de credibilidad y reducir su inflación. Importa con ello, en la práctica, la tasa de inflación alemana, π^A. La economía italiana se sitúa así en el punto 0′, sobre la curva $\mathrm{CP}_{\mathrm{CP},0'}$.

▪ *Devaluación por sorpresa (0′ → 1)*: sin embargo, esta regla no es creíble porque, una vez situada la economía italiana en 0′, el Gobierno tendría de nuevo incentivos para devaluar por sorpresa su moneda. La devaluación produciría inflación y situaría a la economía en el punto 1, con inflación π_1 y desempleo u_1, en la tangencia con CIS_1^I.

El argumento descansa en que un tipo de cambio fijo es una regla reversible, puesto que el Gobierno conserva la facultad de abandonarla. Como los agentes lo saben, anticipan la devaluación, por lo que la economía italiana no llega a estabilizarse en 0′ y permanece en el equilibrio discrecional 0.

Con integración monetaria, en cambio, el desplazamiento de 0′ a 1 deja de estar disponible y el equilibrio 0′ resulta sostenible. Desaparecería así la inconsistencia dinámica y los objetivos de inflación anunciados ganarían credibilidad.

Bienestar: la pérdida social de Italia disminuirá, puesto que conservará su tasa de desempleo previa con una inflación (y una prima de riesgo) más bajas. El compromiso resulta creíble no porque el Gobierno italiano se haya vuelto más disciplinado, sino porque ha perdido el

instrumento con el que podía incumplirlo, es decir, se ha atado las manos (*tying one's hands*, Giavazzi y Pagano, 1988).

Conclusiones:

▪ *Aspectos positivos*: el modelo ofrece una justificación de la unión monetaria distinta de las examinadas hasta aquí, pues el beneficio procede de la credibilidad y no del ajuste ante perturbaciones. Explica también, mejor que ninguna otra aportación, el diseño institucional de la zona del euro. El mandato de estabilidad de precios y la independencia del BCE responden a la exigencia de que la unión no diluyese la credibilidad de los países que ya la tenían. Con una autoridad común de representación paritaria, en cambio, la inflación resultante sería menor que la italiana inicial pero mayor que la alemana.

▪ *Limitaciones*: la integración monetaria no es imprescindible para resolver el problema, pues bastaría con un banquero central nacional independiente y conservador en la línea de Rogoff (1985), o con la reputación que genera un juego repetido un número indeterminado de veces. Además, su predicción central, un sesgo inflacionista sistemático, encaja bien con los años setenta y mucho peor con las décadas posteriores, en las que numerosos bancos centrales han incumplido su objetivo por defecto y no por exceso.

2.1.3. La endogeneidad de los criterios: integración y simetría de los *shocks* (Comisión Europea, 1990; Frankel y Rose, 1998)

Origen: el estudio *One Market, One Money* (1990) de la Comisión Europea sostiene que los *shocks* asimétricos de demanda son menos frecuentes dentro de una unión monetaria. Jeffrey Frankel y Andrew Rose, en *The Endogeneity of the Optimum Currency Area Criteria* (1998), formalizan esa intuición.

Idea: la aportación cuestiona el propio marco del apartado 1. Los criterios de optimalidad no serían un dato previo que determina si conviene integrarse, sino un resultado que la integración misma contribuye a producir. Un área puede, en consecuencia, no ser óptima *ex ante* y llegar a serlo *ex post*.

Razonamiento: el comercio entre los miembros es predominantemente intraindustrial, es decir, de bienes similares dentro de los mismos sectores. Cuando los países intercambian variedades de un mismo producto en lugar de especializarse en sectores distintos, una perturbación sectorial afecta simultáneamente a todos ellos. Lo que sería un *shock* asimétrico entre economías especializadas se convierte así en uno simétrico, gestionable con una política monetaria común.

En la formulación de Frankel y Rose, al eliminar el riesgo de tipo de cambio y los costes de transacción, la moneda común intensifica el comercio entre sus miembros y, con él, la correlación de sus ciclos económicos. Para un panel de países industrializados, los autores obtienen una relación positiva y significativa entre intensidad comercial y correlación cíclica.

La implicación de política económica es que se invierte la carga de la prueba. El incumplimiento inicial de los criterios deja de ser un argumento decisivo contra la adhesión, pues la propia pertenencia contribuiría a satisfacerlos. Este argumento se esgrimió en el debate previo a la tercera fase de la Unión Económica y Monetaria.

Conclusiones:

▪ *Aspectos positivos*: se trata del único beneficio de carácter dinámico de cuantos recoge el tema, ya que la unión monetaria genera con el tiempo las condiciones de su propia viabilidad. Ello reconcilia la teoría con la práctica de la integración europea, que avanzó sin cumplir los criterios

que la propia teoría exigía, y explica que la correlación de los ciclos entre los miembros de la zona del euro haya aumentado efectivamente desde su creación.

- *Limitaciones*: el argumento descansa por entero en la magnitud del efecto de la moneda común sobre el comercio. Las estimaciones iniciales de Rose (2000), que apuntaban a incrementos muy elevados, fueron revisadas a la baja de forma sustancial en los trabajos posteriores centrados en el euro, donde el efecto resulta modesto. De ahí que, si el canal comercial es débil, también lo sea la endogeneidad que se apoya en él. A ello se añade la objeción de Krugman, examinada en el apartado 1.3.3, según la cual la integración puede favorecer la especialización interindustrial y, con ella, la asimetría, de modo que el signo del efecto es una cuestión empírica antes que teórica.

En esa misma línea escéptica se sitúa la *visión americana* del euro (Friedman, 1997; Feldstein, 1997). Según esa visión, sin unión política, movilidad del trabajo ni integración fiscal, una moneda única entre economías heterogéneas convertiría los *shocks* asimétricos en conflictos políticos y tendría costes superiores a sus beneficios.

2.2. Beneficios microeconómicos

Junto a los beneficios macroeconómicos, la moneda única mejora el funcionamiento de los mercados, porque reduce los costes de transacción, hace más transparentes los precios y refuerza la competencia.

2.2.1. Reducción de los costes de transacción

La moneda única reduce los costes de transacción[6] al eliminar la necesidad de cambiar una moneda por otra. Asimismo, los mercados se fragmentan menos, lo que impulsa su integración y facilita que se cumpla la ley del precio único.

Conviene precisar el alcance de esta ganancia. Lo que el usuario deja de pagar en comisiones de cambio es, en primera instancia, un ingreso que el sistema financiero deja de percibir, es decir, una transferencia antes que una ganancia neta de eficiencia. El beneficio social propiamente dicho se materializa, en cambio, en el medio y largo plazo, conforme los recursos humanos y técnicos empleados en la actividad de cambio se relocalizan hacia actividades de mayor valor añadido.

Se trata, en todo caso, del beneficio más seguro de cuantos se atribuyen a la unión monetaria, por ser permanente e independiente de la coyuntura, si bien también el de menor cuantía. Su relevancia es mayor para las economías pequeñas y abiertas que para las grandes.

2.2.2. Transparencia de precios y atenuación de los problemas de información asimétrica

Mayor transparencia en precios: una misma moneda de referencia reduce los costes de comparar precios en diferentes monedas. Este efecto induce una mayor competencia, con lo que se reducen las ineficiencias que derivan de la discriminación de precios.

Atenuación de los efectos de la información asimétrica: la unión monetaria no mejora aquí el estado de la información. Atenúa, en cambio, las consecuencias de la asimetría preexistente por una vía indirecta, la caída de los tipos de interés nominales. Esta caída sí es atribuible a la moneda única, puesto que la desaparición del riesgo de cambio entre los miembros elimina

[6] Según *One Market, One Money* (Comisión Europea, 1990), entre el 0,25 % y el 0,5 % del PIB comunitario.

tanto la prima por dicho riesgo como la expectativa de devaluación que los tipos nacionales incorporaban. Conforme a la lógica del racionamiento del crédito de Stiglitz y Weiss, unos tipos más bajos mitigan la selección adversa (los prestatarios más seguros dejan de retirarse del mercado) y el riesgo moral (los ya financiados no se ven empujados hacia proyectos más arriesgados).

> El modelo de racionamiento del crédito de Stiglitz y Weiss (1981), con la selección adversa y el riesgo moral que induce una subida de los tipos de interés, se desarrolla en el tema A42, apartado 2.3.1.

Esta convergencia de tipos resultó, no obstante, excesiva en la experiencia europea, ya que la práctica desaparición de las primas de riesgo entre 1999 y 2007 alimentó auges crediticios e inmobiliarios y desequilibrios exteriores en la periferia. La reversión brusca de esos auges y desequilibrios a partir de 2010 convirtió el beneficio en coste.

2.2.3. Efectos procompetitivos

La transparencia de precios y la desaparición del riesgo de cambio facilitan el arbitraje entre mercados nacionales e intensifican la competencia entre empresas de distintos Estados miembros, con efectos precio (menores márgenes), no precio (diferenciación e innovación) y de reducción de la ineficiencia-X.

Estos efectos son, sin embargo, atribuibles en buena medida al mercado único antes que a la moneda única, de la que reciben un refuerzo pero no su causa principal. Su alcance efectivo depende, además, de que se eliminen las barreras no monetarias a la competencia transfronteriza, cuestión especialmente pendiente en el sector servicios.

> Los efectos procompetitivos de la integración económica (efecto precio, efectos no precio y reducción de la ineficiencia-X) se desarrollan en el tema B10, apartado 2.2, y la hipótesis de la ineficiencia-X (Leibenstein, 1966), en el tema A15, apartado 1.1.

2.3. Beneficios geopolíticos

La integración monetaria tiene también una dimensión geopolítica, que opera en tres planos distintos:

1. *Moneda de reserva internacional*: una moneda respaldada por un área económica amplia y por mercados de deuda profundos puede pasar a ser utilizada por terceros países como activo de reserva, unidad de denominación del comercio y moneda de refugio. Ese uso reporta señoraje internacional y la posibilidad de financiar déficits exteriores en moneda propia.

2. *Autonomía estratégica*: una moneda propia con proyección internacional reduce la dependencia de las infraestructuras de pago y de compensación controladas por terceros, cuestión relevante conforme el sistema financiero internacional se ha convertido en instrumento de política exterior mediante las sanciones.

3. *Peso negociador*: la unión monetaria proporciona a sus miembros una representación conjunta y un mayor peso político en las instituciones económicas internacionales, especialmente para las economías de menor tamaño.

Estos beneficios son, sin embargo, en buena medida potenciales antes que efectivos. El euro es la segunda divisa internacional, muy por detrás del dólar, lo que suele atribuirse a la ausencia de un activo seguro común de dimensión comparable a la deuda del Tesoro estadounidense. De ahí que el debate sobre la emisión de deuda común no sea únicamente un problema de estabilización interna, sino también una condición del papel internacional de la moneda.

> Sobre las monedas de reserva, el señoraje internacional y la configuración del sistema monetario internacional, véanse los temas B20 y B21, apartado 2, y, sobre el papel internacional del euro y el proyecto de euro digital, el tema B43, apartado 3.3.1.

2.4. Conclusiones

Frente a los costes de la sección 1, los beneficios de compartir moneda se reparten en tres planos. En el macroeconómico, el propio Mundell (1973) rebaja el coste de renunciar al tipo de cambio, que con información imperfecta es más una fuente de *shocks* asimétricos que un instrumento de estabilización. La delegación de la política monetaria en una autoridad común creíble suprime, además, la opción de devaluar por sorpresa y permite importar credibilidad. La endogeneidad de los criterios añade una dimensión dinámica, pues el comercio que impulsa la moneda común sincroniza los ciclos, si bien el efecto estimado del euro sobre el comercio es modesto.

En el plano microeconómico, la reducción de los costes de transacción es el beneficio más seguro pero también el menor, y los efectos procompetitivos se deben en buena medida al mercado único. Los beneficios geopolíticos son, por su parte, más potenciales que efectivos mientras falte un activo seguro común. La sección 3 estudia los retos de una unión monetaria sin unión fiscal y sus soluciones, entre ellas la deuda común que proporcionaría ese activo.

3. El proceso de construcción de una unión monetaria: retos y soluciones

Dado lo exigentes que son las condiciones de éxito de una unión monetaria, los retos de su construcción pueden derivar, de manera muy simplificada, en uniones monetarias completas o incompletas.

- *Unión monetaria completa*: aquella que va acompañada de una unión fiscal (y, en última instancia, política), con un presupuesto común significativo y capacidad de emisión de deuda común.

- *Unión monetaria incompleta*: aquella en la que no existe unión fiscal. En esta situación, las políticas fiscales de los miembros se llevan a cabo de manera autónoma, si bien estos emiten deuda soberana en una moneda común y tienen un banco central común.

La distinción es relevante porque de ella se deriva la vulnerabilidad característica de la segunda. Un Estado que emite deuda en su propia moneda puede, en último término, monetizarla, de modo que un problema de liquidez nunca se transforma mecánicamente en uno de solvencia. Al integrarse en una unión monetaria incompleta, en cambio, el Estado pasa a emitir deuda en una divisa sobre la que no tiene control, y con ello queda expuesto a que una retirada súbita de los inversores lo aboque al impago aun siendo solvente. Esta asimetría explica por qué economías con niveles de deuda comparables, como el Reino Unido y España en 2011, afrontaron condiciones de financiación radicalmente distintas.

3.1. Unión monetaria incompleta: modelo de equilibrios múltiples (De Grauwe, 2011)

Origen: Paul De Grauwe (2011) aplica a una unión monetaria incompleta la lógica de las crisis autocumplidas y los equilibrios múltiples[7].

Idea: la fragilidad de una unión monetaria incompleta se puede mostrar a través de un modelo sencillo de equilibrios múltiples, que pone de manifiesto la importancia de las expectativas de los agentes económicos.

3.1.1. Supuestos

Entorno: la unión monetaria está abierta al exterior, la componen varias regiones o países y la información es asimétrica. Se considera que los Gobiernos tienen ventajas informacionales respecto al resto de agentes, es decir, conocen sus presupuestos y situación fiscal real en detalle.

Agentes: intervienen una única autoridad monetaria, varios Gobiernos con sus propios presupuestos e inversores privados. Esto supone que la deuda soberana se emite en una moneda común sobre la cual los Gobiernos no tienen control directo, lo que puede dar lugar a una situación de equilibrios múltiples.

[7] La lógica de las crisis autocumplidas y los equilibrios múltiples se remonta a Obstfeld (1986), para las crisis cambiarias, y a Calvo (1988), para la deuda soberana.

3.1.2. Modelo

Planteamiento: uno de los miembros de la unión monetaria sufre un *shock* exógeno asimétrico (p. ej. de solvencia o de pérdida de competitividad), que lleva al Gobierno a plantearse si impagar la deuda o no. El Gobierno no contempla otras alternativas de consolidación fiscal para hacer frente al *shock* debido a sus costes políticos y sociales. Las decisiones se toman en dos etapas:

1. *Inversores*: los agentes privados toman sus decisiones de inversión en deuda soberana antes de que el Gobierno del país tome su decisión de estabilización. Así, forman sus expectativas sobre si el Gobierno impagará la deuda soberana (*default*) o no. Conocen, no obstante, la función de pérdida del Gobierno, es decir, los costes y beneficios que el impago supone para él, y pueden, así, anticipar cómo responderá este a sus expectativas.

2. *Gobierno*: toma su decisión óptima sobre si impagar o no dadas las expectativas de los agentes. Para ello, compara los costes y los beneficios del impago, que dependen de los desequilibrios macroeconómicos que comprometen la estabilidad.

Desarrollo gráfico: el modelo de equilibrios múltiples se representa en el plano intensidad del *shock*-costes y beneficios del *default* (figura B16.4):

1. *Coste del impago* (C_D): dejar de pagar la deuda soberana supone para el Gobierno una pérdida de reputación y de credibilidad que lo excluirá de los mercados de financiación. Esta pérdida se modeliza como un coste fijo, independiente del deterioro de la solvencia del Gobierno, por lo que su curva es horizontal.

2. *Beneficios del impago* (B): se vinculan a la reducción de los gastos asociados al servicio de la deuda o a evitar la aplicación de políticas de austeridad para hacer frente al *shock*. Su curva tiene pendiente positiva, es decir, cuanto mayor sea la intensidad del *shock*, mayores serán los beneficios de impagar, puesto que se evita la aplicación de políticas de austeridad con costes políticos y económicos a corto plazo. Además, los beneficios serán mayores cuanto mayor sea el nivel de deuda inicial, cuanto menor sea la confianza en el Gobierno nacional y cuanto menor sea la eficiencia del sistema tributario.

3. *Beneficios según las expectativas* (B^{NC} y B^{C}): se distinguen, a su vez, dos curvas de beneficios. La curva B^{NC} refleja los beneficios del impago si los inversores anticipan el impago, es decir, no confían, y la curva B^{C}, los beneficios si no lo anticipan, es decir, si confían. Si no confían, los inversores venderán bonos masivamente sin esperar al deterioro efectivo de la solvencia del Gobierno. Esto elevará los costes de financiación y el déficit del Gobierno y, por ello, hará más atractivo el *default* para cualquier intensidad del *shock*. Por este motivo, la curva B^{NC} queda por encima de la curva B^{C}.

4. *Equilibrio*: los equilibrios del modelo dependen de la intensidad relativa[8] del *shock*, S:

- *Shocks de intensidad baja ($S < S_1$)*: los costes del impago son superiores a los beneficios del impago, cualesquiera que sean las expectativas de los inversores, pues incluso la curva B^{NC} se sitúa por debajo de C_D. Por tanto, no se producirá la venta masiva de bonos soberanos, y el equilibrio sin *default*, el "cielo en la tierra", será sostenible en el tiempo. Este podría ser el caso de Alemania o Países Bajos durante la crisis de deuda soberana europea (2010-2014).

- *Shocks de intensidad alta ($S > S_2$)*: los beneficios del impago son superiores a su coste, cualesquiera que sean las expectativas de los inversores, pues incluso la curva B^{C} se sitúa por encima de C_D. Por lo tanto, el impago de deuda resulta inevitable y se produce el "infierno en la

[8] Se entiende que la intensidad es relativa a la fragilidad o robustez de la economía que sufre el *shock*.

tierra". En un marco de HER, los inversores se anticiparán al impago y venderán masivamente sus bonos soberanos, lo que causará una fuerte crisis de liquidez de la economía.

▪ *Shocks de intensidad media ($S_1 < S < S_2$)*: en este tramo la curva B^{C} queda por debajo del coste y la curva B^{NC} por encima, de modo que se llega a un estado de indeterminación (el "purgatorio") en el que existen equilibrios múltiples. Para una intensidad S', el equilibrio que evita el *default* (el "bueno") es el punto N (*no default*), en el cual los inversores confían en el Gobierno, por lo que no anticipan el impago de la deuda y no venden masivamente. El equilibrio que lleva al *default* (el "malo") es el punto D (*default*), en el que los inversores anticipan el impago de la deuda soberana.

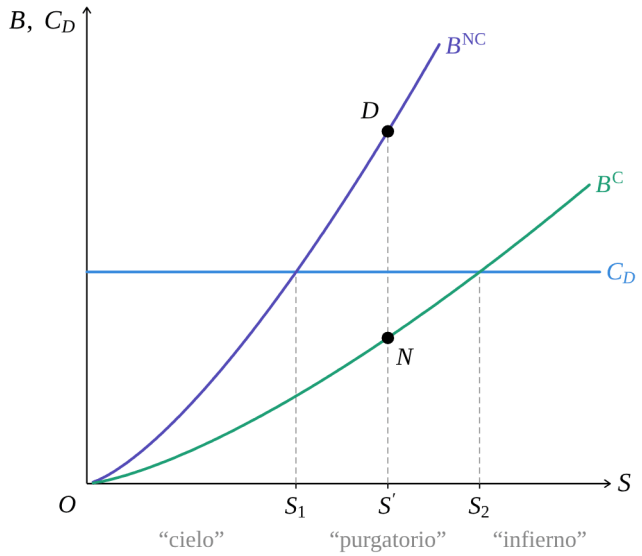

Figura B16.4: Unión monetaria incompleta (De Grauwe, 2011). El mercado de deuda soberana tiene equilibrios múltiples, uno bueno con confianza y otro malo con expectativas de impago.

En el tramo intermedio, el que se alcance uno u otro equilibrio dependerá completamente de las expectativas de los agentes, que actúan como **profecías autocumplidas**. Si esperan un *default*, efectivamente se producirá, y si no lo esperan, no se producirá. Aunque los inversores conocen la función de pérdida del Gobierno, ambos equilibrios son coherentes con ella, y la selección del equilibrio queda, por ello, en manos del sentimiento de mercado. Así, pequeños cambios en este sentimiento pueden llevar a que se pase de una situación a otra. Este podría ser el caso de España e Italia durante la crisis de deuda soberana (2010-2014).

Ello muestra la fragilidad intrínseca de la unión monetaria incompleta, cuyo funcionamiento termina dependiendo de las expectativas de los agentes privados. La unión monetaria puede ser frágil en ausencia de una unión fiscal o de al menos una coordinación fiscal adecuada. La fragilidad no procede del nivel de deuda en sí, sino de la ausencia de un prestamista de última instancia en el mercado de deuda soberana, ausencia que abre el tramo de equilibrios múltiples y convierte un problema de liquidez en uno de solvencia.

Estática comparativa: si la unión monetaria incluye una *cláusula de expulsión* del país miembro que impague la deuda, aumenta el coste reputacional del impago y la curva de costes, C_D,

se desplaza hacia arriba. En consecuencia, la zona de incertidumbre, entre S_1 y S_2, se desplaza hacia la derecha, y los niveles de deterioro de la solvencia necesarios para que se desencadenen profecías autocumplidas son mayores.

3.1.3. Conclusiones

Aspectos positivos: el valor del modelo reside en mostrar que la fragilidad es estructural y no atribuible al comportamiento de un Gobierno concreto.

Limitaciones: el modelo descansa en tres simplificaciones:

- *Ausencia de alternativas*: supone que el Gobierno no dispone de alternativas de consolidación fiscal, cuando en la práctica el ajuste es siempre una opción, si bien costosa.

- *Coste fijo del impago*: lo supone invariable, cuando depende de la exposición del propio sistema bancario nacional a la deuda soberana.

- *Contagio*: no modeliza explícitamente el contagio entre miembros, que fue determinante en la experiencia europea.

> Los modelos de crisis de balanza de pagos y las crisis autocumplidas se desarrollan en el tema B17, apartado 2, y el relato de la crisis de deuda soberana europea, en el tema B43.

3.2. Soluciones a las uniones monetarias incompletas

Para superar los retos que plantea la ausencia de una política fiscal común en una unión monetaria, se han formulado, entre otras, tres líneas de solución.

3.2.1. Unión bancaria

En los casos anteriores de purgatorio "malo" e infierno se produce una venta masiva de bonos soberanos que provoca una caída de su precio. Puesto que los bancos privados nacionales son a la vez los principales tenedores y vendedores de deuda soberana, esa venta masiva deteriora su balance. Existe, así, un vínculo entre riesgo soberano y riesgo bancario.

El vínculo, además, opera en ambos sentidos, y de ahí que se lo conozca como **círculo vicioso entre bancos y soberanos** (*doom loop*), también llamado *bucle diabólico*. De un lado, el deterioro de la solvencia soberana daña el balance bancario por la vía descrita. De otro, el rescate de una banca en dificultades recae sobre el presupuesto nacional y agrava la posición fiscal del Estado, lo que realimenta la desconfianza sobre su deuda. Así sucedió en Irlanda en 2010, donde un rescate bancario transformó una posición fiscal saneada en una crisis soberana.

Junto con las tensiones de liquidez, toda esta situación puede provocar que las crisis de deuda soberana se conviertan en crisis bancarias. Por ello, una posible solución es la creación de una **Unión Bancaria** que traslade la supervisión y la resolución al nivel supranacional, sustituya los esquemas nacionales de garantía de depósitos por uno común y rompa así el círculo vicioso.

Evidencia empírica (zona del euro): la Unión Bancaria, puesta en marcha a partir de 2012, descansa hoy sobre dos pilares operativos, el Mecanismo Único de Supervisión (MUS) y el Mecanismo Único de Resolución (MUR). En cambio, el tercer pilar, el Sistema Europeo de Garantía de Depósitos (SEGD), sigue pendiente por la resistencia a mutualizar el riesgo antes de reducir la exposición bancaria a la deuda soberana nacional, y la Comisión anunció en noviembre

de 2026 la retirada de su propuesta de 2015. El círculo vicioso, por tanto, se ha atenuado pero no se ha roto.

> El código normativo único, el MUS, el MUR y la propuesta de SEGD se desarrollan en el tema B44, apartado 2.

3.2.2. Banco central común como prestamista de última instancia

A priori, parece razonable que la autoridad monetaria común también pueda desempeñar el papel de **prestamista de última instancia**. Podrá entonces dotar de toda la liquidez necesaria a los mercados de deuda si los inversores venden masivamente sus bonos. Así, se podrán prevenir las crisis de liquidez soberana y, con ellas, su contagio a la banca.

La eficacia del instrumento, sin embargo, no depende de su uso efectivo. Si el compromiso resulta creíble, la mera existencia del respaldo elimina el equilibrio "malo" del modelo anterior, pues los inversores dejan de tener motivo para anticipar el impago.

Con todo, hay que tener en cuenta diversas objeciones:

- *Riesgo de inflación*: se ha argumentado que la actuación del banco central como prestamista de última instancia de los soberanos tendría sesgos proinflacionistas, al aumentar la base monetaria. La objeción es discutible, pues la expansión de la base no se traslada mecánicamente a los agregados amplios cuando el multiplicador se contrae, como tras la crisis financiera. El riesgo relevante es, más bien, la deriva hacia la dominancia fiscal.

- *Riesgo moral*: como en todos los mecanismos de aseguramiento, la actuación del banco central como prestamista de última instancia puede incentivar a los Gobiernos a llevar a cabo políticas fiscales poco disciplinadas, con persistentes déficits públicos y abultados niveles de deuda pública[9]. La solución habitual consiste en condicionar el respaldo al cumplimiento de un programa de ajuste, de modo que el seguro no se active de forma automática.

Evidencia empírica (zona del euro): el anuncio de las operaciones monetarias de compraventa (OMT) por el BCE en 2012, tras el "whatever it takes" de Mario Draghi, bastó para reducir sustancialmente las primas de riesgo periféricas sin comprar un solo bono. El resultado es coherente con la lógica de las profecías autocumplidas del modelo de equilibrios múltiples. Las OMT incorporaban, además, la respuesta al riesgo moral, pues su activación quedaba condicionada a un programa de la Facilidad Europea de Estabilización Financiera (FEEF) o del MEDE, y la lógica se reiteró en 2022 con el Instrumento para la Protección de la Transmisión (TPI).

> El papel del BCE como prestamista de última instancia, los programas de compras de activos, las OMT y el TPI se desarrollan en el tema B43, apartado 3.3, y el MEDE, en el tema B45, apartado 2.4.

3.2.3. Reformas fiscales y unión política

Emisión de deuda común: la creación de una autoridad fiscal común que emitiera deuda pública en la divisa controlada por el banco central mutualizaría el riesgo soberano y, sobre todo, eliminaría la raíz de la fragilidad ya descrita. Esa raíz es que los Estados emiten deuda en

[9] Además de reglas fiscales que limiten la emisión de deuda, puede separarse institucionalmente la provisión de liquidez (autoridad monetaria) de la supervisión del cumplimiento de esas reglas (autoridad fiscal independiente).

una moneda sobre la que carecen de control. Un activo seguro común sustraería la deuda de los miembros a los escenarios de restricción de liquidez que abocan a crisis soberanas y bancarias, y proporcionaría además a la unión el activo de reserva del que hoy carece.

Coordinación de políticas presupuestarias y económicas: la contrapartida de la mutualización es el riesgo moral, al incentivarse que los Estados fiscalmente menos disciplinados emitan mayores niveles de deuda soberana. Para evitarlo se requieren marcos presupuestarios nacionales más disciplinados, más homogéneos y vinculados a la puesta en marcha de reformas estructurales que persigan eliminar las diferencias institucionales entre los Estados miembros.

En resumen, completar una unión monetaria exige emitir deuda pública común que internalice los efectos externos nacionales y establecer marcos y reglas fiscales coherentes y creíbles. Estas propuestas exceden, no obstante, el terreno técnico, pues la mutualización de la deuda implica una cesión de soberanía fiscal que exige legitimación democrática, y el obstáculo determinante es, por ello, político.

Evidencia empírica (zona del euro): la zona del euro ha optado hasta ahora por la coordinación mediante reglas (Pacto de Estabilidad y Crecimiento y sus reformas, Semestre Europeo) antes que por la centralización presupuestaria, sustituyendo el seguro mutuo por la disciplina nacional supervisada. La excepción es NextGenerationEU (2020), que, tras el ensayo del instrumento SURE (2020), es la mayor emisión de deuda común de la Unión, si bien temporal y finalista, lo que deja abierta la cuestión de si constituye un precedente o una singularidad.

> La gobernanza fiscal de la UEM (PEC y su reforma de 2024, Semestre Europeo, MEDE) y el debate sobre la capacidad fiscal tras NextGenerationEU se desarrollan en el tema B45; el presupuesto de la Unión y la financiación de NextGenerationEU, en el tema B37.

Conclusiones

La teoría de las áreas monetarias óptimas nace como un análisis de costes. Mundell (1961) muestra que compartir moneda obliga a renunciar al tipo de cambio y a la política monetaria como instrumentos de ajuste, de modo que un área es óptima si dispone de mecanismos sustitutivos ante los *shocks* asimétricos. Los criterios posteriores no ofrecen un umbral cuantitativo, y la zona del euro cumple con holgura el de integración comercial pero no los de movilidad del trabajo e integración fiscal. Los desarrollos posteriores añaden los beneficios, como el aseguramiento a través de los mercados de capitales (Mundell, 1973), la credibilidad antiinflacionista (Barro y Gordon, 1983a) y los menores costes de transacción. La endogeneidad de los criterios (Frankel y Rose, 1998) sugiere que una unión puede llegar a ser óptima *ex post*, aunque la especialización que anticipa Krugman (1993) apunta en sentido contrario. La crisis de deuda soberana mostró que una unión monetaria incompleta es vulnerable a equilibrios múltiples (De Grauwe, 2011), y de ahí sus tres soluciones, la Unión Bancaria, un prestamista de última instancia y la deuda común con reglas fiscales creíbles.

La unión monetaria culmina la escala de integración económica (tema B10) y es el extremo de vinculación fuerte entre los regímenes cambiarios (tema B15). Su beneficio de credibilidad aplica la inconsistencia dinámica, porque delegar la política monetaria en un banco central creíble elimina el sesgo inflacionista (tema A37), y su vulnerabilidad a equilibrios múltiples traslada a la deuda soberana la lógica de las crisis de segunda generación (tema B17). Por último, las soluciones europeas a esa vulnerabilidad se estudian en la política monetaria del BCE (tema B43), la Unión Bancaria (tema B44) y el marco fiscal, con el MEDE y la capacidad fiscal (tema B45).

Completar la unión obliga a combinar el reparto de riesgos con la disciplina de mercado. Bénassy-Quéré et al. (2018) proponen avanzar a la vez en ambos frentes, con un seguro común de depósitos, un fondo de estabilización ante *shocks* grandes, reglas fiscales más sencillas y una vía ordenada de reestructuración de deuda, frente a quienes exigen reducir antes los riesgos nacionales. Sigue abierto, además, el diseño de un activo seguro europeo. Brunnermeier et al. (2016) proponen títulos respaldados por una cartera de deuda soberana de la zona del euro, sin garantía conjunta, cuyo tramo preferente serviría de activo seguro y debilitaría el círculo vicioso entre cada banca y su soberano.

B17. Crisis de balanza de pagos

Teorías explicativas de las crisis de balanza de pagos.

Introducción

> **Introducción común.** Los temas B17 y B26 comparten una introducción, que se desarrolla al comienzo del tema **B26**. Distingue, siguiendo al FMI, las **crisis monetarias o cambiarias**, en las que un ataque especulativo contra el tipo de cambio obliga a renunciar a la paridad o a defenderla a un alto coste, de las **crisis financieras**, un concepto más amplio, y explica por qué su prevención y su gestión están en el centro del debate económico.

Las **crisis cambiarias** son tan antiguas como los tipos de cambio fijos, pero su teoría nació de episodios concretos, y cada generación de modelos respondió a la incapacidad de la anterior para explicar el siguiente. Los colapsos de tipos fijos en varios países en desarrollo durante los años setenta y ochenta dieron lugar a los modelos de primera generación (Krugman, 1979a; Flood y Garber, 1984). Después, la crisis del Sistema Monetario Europeo (1992-1993), que alcanzó a países con fundamentos sólidos como el Reino Unido, llevó a los modelos de crisis autocumplidas (Obstfeld, 1994). Por último, la crisis asiática (1997-1998), en economías con cuentas públicas saneadas, situó el foco en los balances privados y en los descalces de moneda (Krugman, 1999). La lección común es que la prevención depende del origen de la crisis, y por ello sigue abierto el debate sobre las reservas, los controles de capital y la red de seguridad financiera internacional.

El tema estudia las diferentes teorías sobre las crisis monetarias y financieras conforme a la clasificación por generaciones de modelos, cuya terminología se remonta a Eichengreen, Rose y Wyplosz (1995) y que la literatura posterior ha ampliado:

- *Modelos de primera generación* (crisis de los fundamentos): estudian cómo una política económica inconsistente puede ir erosionando los fundamentos de la economía hasta desencadenar una crisis cambiaria.

- *Modelos de segunda generación* (crisis autocumplidas): la crisis cambiaria puede producirse aunque los fundamentos no se hayan deteriorado hasta volverla inevitable. Basta un giro pesimista en las expectativas de los agentes, que termina validándose a sí mismo.

- *Modelos de tercera generación* (crisis gemelas): analizan cómo las crisis cambiarias y las crisis bancarias se retroalimentan.

La crisis financiera global (2008-2014) ha dado lugar, además, a los llamados *modelos de cuarta generación*, cuyo encaje en esta tipología se discute en el apartado 3.4.2.

En todos los casos se exponen la idea principal, el origen, la transmisión y las consecuencias de las crisis, así como las posibles soluciones y los episodios históricos que las ilustran. Además, un rasgo

recurrente de muchas de las crisis monetarias y financieras recientes es su rápida transmisión entre países, lo cual ha suscitado un notable interés en la literatura por el estudio del **contagio**[1]. Por ello, en cada generación se examinan los canales de contagio que el modelo correspondiente permite explicar. En adelante, las crisis monetarias se denominan crisis cambiarias y, entre las financieras, el tema se ocupa sobre todo de las bancarias.

[1] Eichengreen et al. (1996) definen el contagio como una situación en la que la crisis de un país incrementa la probabilidad de ocurrencia de crisis en otros países. Ello significa que no basta con la simultaneidad, sino que es necesario que exista un vínculo de causalidad. Otras definiciones más amplias sí admiten la aparición de crisis simultáneas por una causa común.

1. Modelos de primera generación: modelo de Krugman (1979a) y Flood y Garber (1984). El problema de la incompatibilidad

Origen: los modelos de primera generación tratan de explicar el colapso de los sistemas de tipo de cambio fijo que tuvo lugar durante los años setenta y ochenta en varios países en desarrollo. El modelo canónico se debe a Paul Krugman en *A Model of Balance-of-Payments Crises* (1979a), si bien aquí se sigue la versión linealizada de Flood y Garber (1984) para ilustrar el caso de una monetización del déficit.

Idea: las crisis cambiarias se producen porque la política monetaria y la cambiaria son incompatibles. Un tipo de cambio fijo exige que la política monetaria se dedique ante todo a sostener la paridad. El banco central puede tener, sin embargo, otra prioridad que le lleve a expandir el crédito interno, ya sea cubrir el déficit del Gobierno o apuntalar una banca frágil.

1.1. Supuestos

Entorno:

- *Economía*: pequeña, abierta y sin imperfecciones, que comercia con otro país. Existen dos monedas sustitutivas, la nacional y la extranjera.

- *Sistema de tipo de cambio*: el país mantiene un tipo de cambio fijo y permite la libre circulación de capitales.

- *Cotización directa del tipo de cambio*: un aumento del tipo de cambio equivale a una depreciación de la moneda nacional. La paridad del poder adquisitivo relativa se cumple, y P^*, el nivel de precios exterior, se toma como fijo y normalizado a la unidad. Por tanto, mientras se mantenga el tipo fijo, $g_E = \pi - \pi^* = 0 \rightarrow \pi = \pi^* = 0$.

Agentes:

- *Agentes privados*: tienen expectativas racionales, que les permiten calcular de antemano el tipo de cambio sombra y anticipar la inflación.

- *Gobierno*: su déficit viene dado desde fuera del modelo y lo monetiza el banco central.

- *Banco central*: su balance es $H = \text{CI} + R$, es decir, el pasivo o base monetaria equivale a su activo, dado por el crédito interno y las reservas exteriores que mantiene. El objetivo principal de su política monetaria es la financiación del déficit. En este régimen, solo cabe incidir en el tipo de cambio nominal interviniendo (comprando y vendiendo moneda nacional), y no a través de una regulación basada en controles de capital.

1.2. Modelo

Planteamiento: las autoridades económicas de este país tratarán de alcanzar dos objetivos simultáneamente:

1. *Política cambiaria*: la autoridad monetaria querrá mantener constante el tipo de cambio, $\bar{E}$. Para ello, el banco central deberá poseer un volumen de divisas suficiente.

2. *Política monetaria*: tendrá como objetivo principal financiar el déficit, es decir, atender ↑ AFG (activos del banco central frente al Gobierno, parte de CI), para lo cual dispone de dos opciones:

- *Dejar crecer la base monetaria*: si quiere comprar bonos del Gobierno, deberá permitir que la base monetaria y, por ende, la oferta monetaria se muevan con dichas compras (↑ AFG ↔↑ H).

- *Compensar con venta de divisas*: si quiere comprar bonos del Gobierno sin que se mueva la base monetaria, deberá comprar moneda nacional y vender divisas (↑ AFG, H constante ↔↓ R).

El objetivo es mantener constante el tipo de cambio, y dejar crecer la base monetaria elevaría la inflación y, por el cumplimiento de la paridad del poder adquisitivo relativa, la depreciación. Por ello, el banco central compensará cada compra de bonos con venta de divisas.

Se entienden así los problemas a los que da lugar la búsqueda simultánea de ambos objetivos, pues el proceso provoca una pérdida paulatina de reservas hasta que estas se agotan.

Esta situación supone un deterioro muy notable de los fundamentos y afecta a la propia política de monetización. Como la prioridad sigue siendo financiar el déficit público, una vez agotadas las reservas internacionales solo cabe atenderlo ampliando la oferta monetaria. Sin reservas con las que sostener el tipo fijo, el país ha de abandonarlo y su moneda se deprecia. Esta transmisión íntegra del aumento de la oferta monetaria al tipo de cambio se debe a que se ha supuesto el cumplimiento de la paridad del poder adquisitivo relativa y que $\pi^* = 0$ (nivel de precios exterior constante).

Los agentes privados, capaces de prever esta situación gracias a sus expectativas racionales, saben que las políticas del Gobierno conducirán inexorablemente al agotamiento de las reservas. Por consiguiente, con el fin de evitar pérdidas, actuarán como especuladores. Se desharán de sus tenencias de moneda nacional y comprarán masivamente bonos extranjeros antes de que las reservas se agoten, lo que precipitará el colapso del sistema cambiario.

Equilibrio: el momento concreto del colapso, T, vendrá determinado por la relación entre dos magnitudes. El *tipo de cambio sombra*, $\tilde{E}$, es el que regiría si, agotadas las reservas, el tipo de cambio flotase libremente, es decir, el determinado por los fundamentos. El tipo sombra es creciente porque el crédito interno no deja de expandirse para financiar el déficit[2]. El tipo de cambio oficial, $\bar{E}$, es, en cambio, el tipo intervenido, que se mantiene fijo.

Una vez que el tipo sombra queda más depreciado que el oficial, especular resultaría rentable y la moneda nacional se depreciaría. Al banco central todavía se le pueden comprar divisas al tipo oficial, y esas divisas podrían revenderse luego al tipo sombra[3], ya que, sin el régimen fijo, el mercado las pagaría por encima de lo que costaron.

Como cada especulador quiere hacerse con esa ganancia antes que los demás, con previsión perfecta se cumplen dos condiciones:

- *Beneficio nulo*: la carrera por adelantarse sitúa el ataque, y con él el colapso, justo en T, cuando el tipo sombra alcanza al oficial y la ganancia del especulador es exactamente cero.

- *Agotamiento de reservas*: los especuladores saben que las reservas internacionales se van perdiendo poco a poco y no esperan a que se acaben por esa vía. En T adquieren de una sola vez todas las que aún quedan.

[2] Formalmente, es el tipo que equilibraría el mercado monetario con una base monetaria compuesta solo por crédito interno ($H = \text{CI}$).

[3] *Ejemplo*: antes de T, $\tilde{E} = 1$ €/\$ y $\bar{E} = 2$ €/\$, por lo que el especulador podría comprar 0,5 \$ en el banco central a cambio de 1 €, pero si tratase de venderlos al tipo sombra solo obtendría 0,5 €, con pérdidas. Después de T, en cambio, $\tilde{E} = 4$ €/\$ y $\bar{E} = 2$ €/\$, de modo que compraría 0,5 \$ por 1 € y los vendería por 2 €, obteniendo beneficios.

La versión lineal se desarrolla para que se entienda de dónde salen el tipo sombra, el momento del ataque y el salto de las reservas que muestran las dos figuras que siguen. Para la exposición basta con explicar el modelo gráficamente y enunciar su estática comparativa.

En la *versión lineal* de Flood y Garber (1984), el modelo se resuelve con cinco ecuaciones, el balance del banco central, la demanda de dinero, la PPA, la paridad descubierta de intereses con previsión perfecta y la regla de crecimiento del crédito interno:

$$\begin{aligned} &H_t = \mathrm{CI}_t + R_t, \qquad H_t - P_t = -h\, i_t, \qquad P_t = E_t \\ &i_t = i^* + \dot{E}_t, \quad i^* = 0, \qquad \dot{\mathrm{CI}}_t = \mu \end{aligned}$$

donde $h > 0$ es la semielasticidad de la demanda de dinero al tipo de interés y $\mu > 0$ el crecimiento del crédito interno. Mientras dura el tipo fijo, $\dot{E}_t = 0$ y el tipo de interés es nulo, de modo que la demanda de dinero es constante, $H_t = \bar{H}$, y cada unidad de crédito nuevo se compensa con una pérdida de reservas, $\dot{R}_t = -\mu$. Si las reservas se agotasen y el tipo flotase, la base monetaria sería solo crédito interno y la moneda se depreciaría al ritmo $\dot{E} = \mu$, con lo que el tipo de interés pasaría a ser μ. El tipo sombra es, por tanto,

$$\tilde{E}_t = \mathrm{CI}_t + h\mu$$

y el ataque se produce cuando el tipo sombra alcanza al oficial, $\tilde{E}_T = \bar{E} = \mathrm{CI}_0 + R_0$, es decir, en

$$T = \frac{R_0 - h\mu}{\mu}$$

La fórmula explica el salto de las reservas. En T, la depreciación esperada pasa de cero a μ, el tipo de interés sube en esa cuantía y la demanda de dinero cae en $h\mu$. Esa caída es exactamente el *stock* de reservas que queda, $R_T = R_0 - \mu T = h\mu$, y que los especuladores compran en el ataque. En cambio, sin sensibilidad de la demanda de dinero al tipo de interés ($h = 0$), el colapso llegaría en R_0/μ, cuando las reservas se agotasen solas.

Desarrollo gráfico: la determinación del momento del colapso se representa en el plano tiempo-tipo de cambio (figura B17.1):

1. *Tipo de cambio sombra* ($\tilde{E}$): refleja el tipo de cambio que regiría si, agotadas las reservas, la moneda flotase libremente, el que determinan los fundamentos. Tiene pendiente positiva porque el crédito interno no deja de expandirse para financiar el déficit.

2. *Tipo de cambio oficial* ($\bar{E}$): refleja la paridad que el banco central defiende comprando y vendiendo moneda nacional. Su curva es horizontal porque el tipo intervenido se mantiene fijo mientras queden reservas.

3. *Equilibrio*: ambas curvas se cortan en T, donde el beneficio del ataque es nulo. A la izquierda de T, el tipo sombra queda por debajo del oficial y el ataque daría pérdidas, mientras que a su derecha daría beneficios. Por tanto, hasta T rige el tipo oficial y, desde ese momento, el tipo sombra.

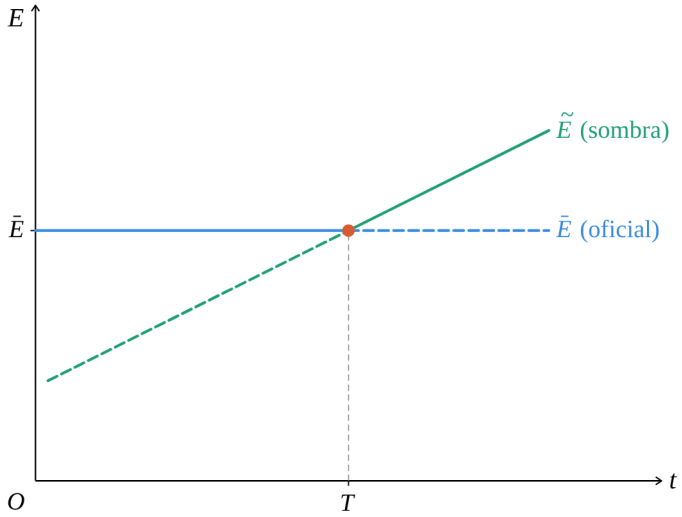

Figura B17.1: Determinación del momento de colapso. Con previsión perfecta, el ataque ocurre exactamente en T, cuando el tipo de cambio sombra iguala al oficial y el beneficio del especulador es nulo.

Gráficamente, la evolución de las reservas se representa en el plano tiempo-reservas (figura B17.2):

1. *Reservas* ($R(t)$): la curva continua refleja el volumen de divisas en poder del banco central en cada momento. Parte de R_0 y cae de forma paulatina por la venta de divisas con que se compensa la financiación del déficit.

2. *Senda sin ataque*: la línea discontinua refleja la evolución que seguirían las reservas si los especuladores esperasen a que se agotaran por esa vía. Prolonga la caída hasta el agotamiento, en un momento posterior a T.

3. *Equilibrio*: en T, el ataque especulativo liquida de golpe las reservas restantes y la curva cae en vertical hasta cero. El colapso se adelanta así respecto a la senda sin ataque, que solo agotaría las reservas más tarde.

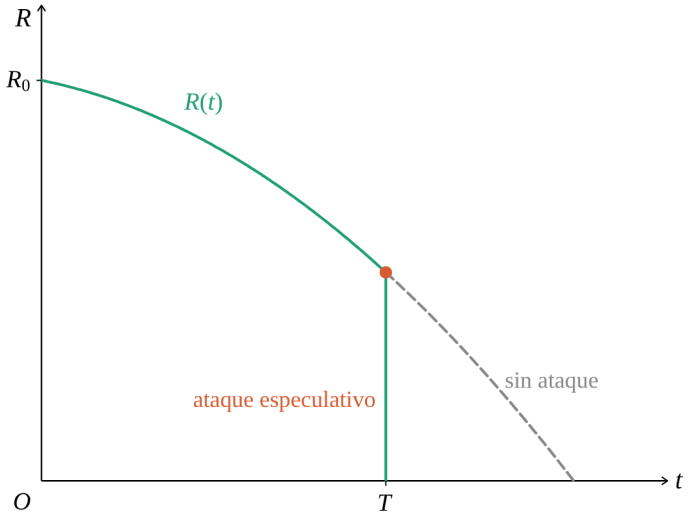

Figura B17.2: Modelo de primera generación (Krugman, 1979a; Flood y Garber, 1984). La financiación del déficit, compensada con venta de divisas, agota paulatinamente las reservas. En T el ataque especulativo compra de golpe las restantes y adelanta el colapso respecto a la senda sin ataque.

Estática comparativa: el ataque se producirá antes (el punto de cruce se situará más a la izquierda) cuanto:

- *Mayor* sea el crecimiento del crédito interno, lo que se traduce en una curva del tipo de cambio sombra con más pendiente.

- *Más apreciado* esté el tipo de cambio oficial, porque estará más cerca del tipo de cambio sombra, que hasta T se sitúa por debajo de él. Gráficamente, la curva del tipo oficial estará más próxima al eje de abscisas.

- *Menor* sea el nivel de reservas internacionales de partida.

- *Más sensible* al tipo de interés sea la demanda de dinero (h), porque mayor es la caída de esa demanda en el momento del colapso y, con ella, el *stock* de reservas que el ataque compra de una vez.

1.3. Contagio: los vínculos comerciales

Los modelos de primera generación permiten explicar el contagio cuando existen **vínculos comerciales**[4]. En estos casos, el mecanismo de transmisión va desde la crisis cambiaria hasta la producción de los socios comerciales. Si los precios internos se ajustan con retraso, la depreciación de la moneda nacional resta competitividad-precio a los productos de los socios frente a los del país en crisis, lo que reduce la producción de los socios y, por tanto, su renta.

[4]Aquí se exponen vínculos comerciales bilaterales. Alternativamente, puede considerarse que los países compiten en un tercer mercado. Este sería el caso de la crisis de deuda latinoamericana de los años ochenta y de la crisis del sudeste asiático de los noventa.

1.4. Evidencia empírica: experiencia histórica

Tres episodios históricos pueden leerse desde la óptica de los modelos de primera generación: la ruptura de Bretton Woods, la crisis de México de los años ochenta y las presiones sobre el yuan entre 2015 y 2017. En los tres, la política monetaria atendía a un objetivo interno incompatible con la paridad.

1.4.1. Ruptura de Bretton Woods (1967-1971)

Causas: la posición exterior estadounidense se deterioró durante los años sesenta a causa del gasto de la guerra de Vietnam y de los programas de la Gran Sociedad, financiados con expansión monetaria mientras el dólar mantenía su convertibilidad fija en oro. El resultado fue una inconsistencia entre el ancla nominal y la política interna, agravada por el dilema de Triffin, según el cual la liquidez internacional exigía déficits estadounidenses que erosionaban la propia confianza en la paridad.

Estallido: la libra esterlina fue atacada primero y hubo de devaluarse en noviembre de 1967. Las presiones se trasladaron después al dólar, cuya cobertura en oro resultaba cada vez más insuficiente, hasta que en agosto de 1971 Estados Unidos suspendió unilateralmente la convertibilidad.

Consecuencias: el Acuerdo del Smithsonian (diciembre de 1971) intentó restablecer paridades ajustadas, pero no resistió, y en 1973 las principales monedas pasaron a flotar.

Soluciones: se abandonó el propio régimen, y los Acuerdos de Jamaica (1976) formalizaron ese abandono al legalizar la flotación y suprimir el papel monetario del oro.

Conclusiones: el episodio ilustra la lógica del modelo, pues la presión sobre la paridad nace de una política monetaria expansiva al servicio de objetivos internos. El episodio muestra, además, que, mientras persiste esa incompatibilidad, el reajuste de las paridades solo aplaza la crisis. Su encaje es, con todo, parcial, porque la moneda sometida a presión era la de reserva del sistema y el dilema de Triffin, un problema estructural del propio régimen, queda fuera del modelo.

> El diseño de Bretton Woods, el dilema de Triffin y el proceso que condujo a su ruptura se desarrollan en el tema B20; el sistema resultante, en el tema B21. Aquí interesa el episodio como ilustración temprana de la lógica de primera generación.

1.4.2. Crisis de México (1982-1989)

Causas: México, donde se inició la crisis de deuda latinoamericana, sostuvo durante la segunda mitad de los años setenta una paridad fija frente al dólar mientras financiaba un déficit público creciente, en parte monetizado. La financiación descansaba en unos ingresos petroleros que se daban por permanentes. La combinación de políticas (*policy mix*) era, por tanto, internamente inconsistente en los términos del modelo, con una expansión del crédito interno incompatible con el mantenimiento de la paridad.

Estallido: el endurecimiento monetario de la Reserva Federal bajo Volcker elevó los tipos de interés internacionales y apreció el dólar, al tiempo que la recesión mundial de 1981 hundía los precios de las materias primas. Con ello, la pérdida de reservas fue continua y obligó al banco central a devaluar en febrero de 1982. En agosto, sin reservas ya, el Gobierno mexicano comunicó que no estaba en condiciones de pagar el servicio de su deuda externa.

Consecuencias: la crisis cambiaria se propagó al conjunto de la región por vínculos comerciales y, sobre todo, por la retirada simultánea del crédito bancario sindicado. Así, lo que había empezado como una crisis cambiaria en un solo país acabó en una crisis de deuda de alcance sistémico. La crisis de deuda dio paso a lo que se llamó "década perdida del desarrollo", unos años ochenta en los que la renta real por habitante retrocedió en muchos países latinoamericanos y africanos.

Soluciones: la respuesta se articuló en tres fases sucesivas. La inmediata consistió en préstamos puente del FMI y del Banco de Pagos Internacionales condicionados a programas de ajuste con enfoque de absorción. Después, el Plan Baker (1985) seguía viendo en la crisis una falta de liquidez y ofreció financiación adicional a cambio de reformas estructurales, sin resultados apreciables. Solo el Plan Brady (1989) reconoció el problema como de solvencia y articuló una quita efectiva mediante el canje de la deuda bancaria por bonos con principal garantizado.

Conclusiones: el episodio se ajusta a las dos predicciones centrales del modelo, que una expansión del crédito interno incompatible con la paridad hace inevitable la crisis y que la paridad cae antes de que se agoten las reservas. Se aparta de él, en cambio, en la forma, porque el desenlace fue una pérdida continua de reservas y no el ataque único que, en la versión con previsión perfecta, liquida de golpe las restantes. Muestra también un límite del modelo, que no recoge cómo la crisis cambiaria se transforma en una crisis de deuda externa con efectos reales duraderos.

> La vertiente de deuda externa de este episodio (sostenibilidad, enfoque de liquidez, Plan Baker, Plan Brady y bonos Brady) se desarrolla en el tema B29, apartado 3.1. El papel del FMI en la gestión del episodio se aborda en el tema B22, apartado 3.3.1.

1.4.3. Caso de China (2015-2017)

Causas: China mantiene un régimen cambiario administrado frente a una cesta de divisas mientras la prioridad de su política monetaria es la estabilidad del PIB y no la del tipo de cambio. Esa jerarquía de objetivos, combinada con una apertura parcial de la cuenta financiera, reproduce la inconsistencia que describe el modelo, pues, si los tipos de interés se fijan atendiendo a la coyuntura interna, no pueden simultáneamente sostener la paridad.

Estallido: entre 2015 y 2017 el yuan sufrió fuertes presiones depreciatorias por tres motivos concurrentes: la necesidad de diversificación de los inversores chinos, la incertidumbre asociada a la desaceleración del crecimiento y la campaña anticorrupción. El banco central respondió comprando yuanes contra reservas en el mercado de divisas. Esas intervenciones sostuvieron la cotización a costa del volumen de reservas.

Consecuencias: entre agosto de 2015 y enero de 2017 el banco central perdió del orden de 600.000 M$ en la defensa de la moneda, si bien sus reservas se han estabilizado desde entonces por encima de 3 billones de dólares.

Soluciones: la autoridad optó por no elevar los tipos de interés, para no comprometer la inversión, y recurrió en su lugar al refuerzo de los controles de capital, la tercera vía del trilema de Mundell-Fleming (trinidad imposible).

Conclusiones: el caso muestra que, con reservas abundantes y controles de capital, la inconsistencia que describe el modelo puede contenerse sin que llegue a producirse el ataque. No obstante, la tensión sigue latente, porque, si persistiera la depreciación del tipo de cambio sombra, el agotamiento previsible de las reservas volvería a hacer racional el ataque. El modelo señala así la fuente de vulnerabilidad, pero no permite anticipar el desenlace, que depende del volumen de reservas y de la vía del trilema que elija la autoridad.

1.5. Conclusiones

El balance de los modelos de primera generación reúne sus resultados, las soluciones que proponen frente a este tipo de crisis y una valoración de sus aciertos y limitaciones.

Resultados:

- *Causas*: las crisis nacen de que las políticas monetaria y cambiaria persiguen objetivos incompatibles, y esa incompatibilidad erosiona los fundamentos.

- *Papel de las expectativas*: resulta esencial, y la depreciación esperada puede hacer que la crisis estalle antes.

- *Predictibilidad*: la crisis puede anticiparse con total exactitud, pero no evitarse, porque las reservas acaban agotándose.

Soluciones: pasan por la aplicación de la *regla de Tinbergen*, según la cual por cada objetivo de política económica debe existir al menos un instrumento disponible. Ello obliga a sacrificar uno de los tres vértices del trilema de Mundell-Fleming:

1. *Régimen cambiario*: dejar flotar la moneda en lugar de sostener el tipo fijo, sin restringir los movimientos de capital.

2. *Política monetaria*: poner la política monetaria al servicio del objetivo cambiario, a costa de su autonomía, sin restringir tampoco los movimientos de capital.

3. *Movilidad de capitales*: limitar la movilidad de capitales, manteniendo el régimen cambiario y la autonomía de la política monetaria, lo que dificulta el ataque, pero solo aplaza la crisis si persiste la monetización del déficit.

Aspectos positivos: el modelo explica con precisión el momento del colapso y demuestra que el ataque especulativo no es un fenómeno irracional ni desestabilizador, sino la consecuencia lógica de una política interna insostenible. La crisis no la causan los especuladores, sino que estos se limitan a anticipar lo que de todos modos habría ocurrido.

Limitaciones:

- *Asimetría en el tratamiento de los agentes*: los agentes privados anticipan con racionalidad plena, mientras que la autoridad sigue una pauta fija y nunca se plantea alternativas a ella. Podría, por ejemplo, endeudarse en los mercados internacionales para cubrir el déficit o subir el tipo de interés para sostener la paridad.

- *Evidencia empírica posterior*: estos modelos no explican los ataques especulativos que sufrió el Reino Unido en la crisis del Sistema Monetario Europeo (1992), puesto que el país no presentaba problemas permanentes en sus fundamentos.

2. Modelos de segunda generación: modelo de equilibrios múltiples (Obstfeld, 1994). Profecías autocumplidas

Origen: los modelos de segunda generación tratan de explicar, fundamentalmente, la crisis del Sistema Monetario Europeo de 1992-1993. Destaca la aportación de Maurice Obstfeld (1994), que formaliza la idea de crisis autocumplidas que él mismo había planteado en 1986.

Idea: abandonar la paridad no es un desenlace forzoso, sino una elección racional del Gobierno[5]. La autoridad dispone aquí de más instrumentos que en la primera generación y podría, por ello, defender el tipo fijo indefinidamente sin agotar sus reservas. Le cabe corregir el desequilibrio exterior con políticas de ajuste (austeridad) y, además, imponer controles de capital o establecer tipos de cambio múltiples.

Cuando el Gobierno renuncia al tipo fijo, lo hace porque mantenerlo le resulta más caro que dejarlo, y, como ese coste varía con lo que esperen los agentes privados, cabe la posibilidad de *profecías autocumplidas*.

2.1. Supuestos

Entorno: se mantiene el del caso anterior, si bien ahora el Gobierno decide después que los agentes privados, de modo que conoce el *shock* y las expectativas de los agentes antes de actuar.

Agentes: el cambio decisivo está en la modelización del comportamiento del Gobierno, que ya no trata de monetizar el déficit y cuyo objetivo es minimizar la pérdida social. Para ello pondera factores como la estabilidad de la brecha de producción (*output gap*), la inflación, los costes de establecer un régimen cambiario u otro, la credibilidad de sus decisiones y los costes políticos y sociales que de ellas se derivan.

2.2. Modelo

Planteamiento: un país abierto al exterior con régimen cambiario fijo sufre un *shock* exógeno asimétrico que deteriora su cuenta corriente, por ejemplo una pérdida de competitividad. Ello lleva al Gobierno a plantearse si abandonar o no el régimen. Se supone que no contempla la contracción de la masa monetaria como alternativa, dados sus costes políticos y sociales.

Los agentes privados toman sus decisiones de compraventa de divisas antes de que el Gobierno adopte las suyas de estabilización, de modo que forman sus expectativas de inflación antes de conocer la inflación efectiva. Conocen, no obstante, la función de pérdida del Gobierno y pueden, así, anticipar cómo responderá este a sus expectativas.

Después, el Gobierno toma su decisión óptima sobre si devaluar o no dadas las expectativas de los agentes, mediante un análisis coste-beneficio basado en los desequilibrios macroeconómicos que comprometen la estabilidad.

Una versión mínima del modelo (Obstfeld, 1996) formaliza ese cálculo. La producción depende de la inflación no anticipada, porque los salarios se fijan con la inflación esperada, y el Gobierno minimiza una pérdida que penaliza la inflación y la desviación de la producción respecto a un objetivo superior a su nivel natural:

$$L = (y - y^*)^2 + \omega\,\pi^2, \qquad y = \bar{y} + \gamma\,(\pi - \pi^e) - S, \qquad y^* > \bar{y},\ \gamma > 0,\ \omega > 0$$

[5] El compromiso cambiario es estado-dependiente y admite una cláusula de escape, que permite devaluar o abandonar la paridad cuando el coste de defenderla resulta excesivo.

donde y es la producción, $\bar{y}$ su nivel natural, y^* su objetivo, π la tasa de devaluación, igual a la inflación por la PPA, π^e la esperada, γ la sensibilidad de la producción a la inflación no anticipada, ω el peso de la inflación en la pérdida y S el *shock*. A la pérdida se suma el coste fijo C_D si el Gobierno devalúa. Con el tipo fijo, $\pi = 0$ y la pérdida es $(y^* - \bar{y} + \gamma\,\pi^e + S)^2$, que crece con el *shock* y con la inflación esperada, porque unos salarios fijados con $\pi^e > 0$ reducen la producción si la paridad se mantiene. Si devalúa, el Gobierno elige la π que minimiza la pérdida, y el beneficio de devaluar, es decir, la pérdida que evita sin contar C_D, es:

$$B = \frac{\gamma^2}{\gamma^2 + \omega}\,(y^* - \bar{y} + \gamma\,\pi^e + S)^2$$

El Gobierno devalúa si $B > C_D$. Como B crece con π^e, la expectativa de devaluación encarece la defensa de la paridad y puede justificar por sí sola su abandono.

Equilibrio: el Gobierno devalúa cuando lo que gana con ello supera lo que le cuesta, y la ganancia depende de lo que esperen los agentes. Por eso el ataque especulativo puede producirse o no, y el modelo presenta equilibrios múltiples y no linealidades. La paridad puede volverse insostenible sin que los fundamentos se hayan deteriorado hasta forzar su abandono. Incluso una economía que podría sostener la paridad puede tener incentivos para dejar el tipo fijo si los agentes esperan lo peor, y ese pesimismo acaba así confirmándose. En el tramo de *shocks* de intensidad media ($S_1 < S < S_2$), la sola expectativa de una devaluación basta para provocarla.

Desarrollo gráfico: la decisión de devaluar se representa en el plano intensidad del *shock*-costes y beneficios de la devaluación (figura B17.3):

1. *Coste de devaluar* (C_D): refleja la pérdida de reputación y credibilidad que sufre el Gobierno si devalúa. Esta pérdida se modeliza como un coste fijo, independiente del deterioro de la cuenta corriente, y por eso su curva es horizontal.

2. *Beneficios de devaluar* (B^{NC} y B^{C}): reflejan el ahorro que supone para el Gobierno evitar las políticas de austeridad con que habría que hacer frente al *shock*. Su pendiente es positiva, porque cuanto más intenso es el *shock*, más costoso es el ajuste que la devaluación evita. Su posición depende de las expectativas de los agentes:

- B^{NC}: recoge los beneficios si los inversores no confían y anticipan la devaluación. En ese caso, los salarios se fijan con una inflación esperada positiva, por lo que mantener la paridad reduce más la producción y devaluar resulta más atractivo para cualquier intensidad del *shock*. Por eso la curva B^{NC} se sitúa por encima de B^{C}.

- B^{C}: recoge los beneficios si los inversores confían y no anticipan la devaluación.

3. *Equilibrio*: según cuáles sean los fundamentos y las expectativas, esta secuencia puede desembocar en más de un equilibrio:

- *Shocks de intensidad baja* ($S < S_1$): con un *shock* así, abandonar el tipo fijo cuesta más que mantenerlo, piensen lo que piensen los agentes. La solidez de los fundamentos basta para disuadir cualquier ataque especulativo. Por tanto, el equilibrio es único y se denomina *cielo en la tierra*.

- *Shocks de intensidad alta* ($S > S_2$): aunque los agentes esperen que se mantenga la paridad ($\pi^e = 0$), el *shock* la hace insostenible. El deterioro de los fundamentos es tan grave que atacar la moneda resulta racional. El equilibrio vuelve a ser único, pero ahora se trata del *infierno en la tierra*.

▪ *Shocks de intensidad media* ($S_1 < S < S_2$): se llega a un estado de indeterminación, el *purgatorio*, en el que existen equilibrios múltiples. En el equilibrio que evita la devaluación (el "bueno"), los agentes esperan una inflación nula ($\pi^e = 0$), confían en el Gobierno, no anticipan la devaluación y no venden masivamente, de modo que lo racional es mantener el régimen cambiario (punto N, para una intensidad S'). En el que lleva a la devaluación (el "malo"), esperan una inflación positiva ($\pi^e > 0$), anticipan la devaluación y venden, y lo racional es entonces abandonar el régimen (punto D, de devaluación).

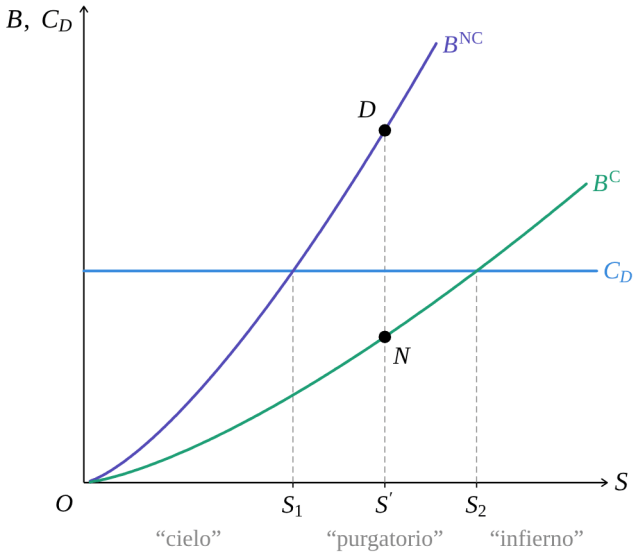

Figura B17.3: Modelo de Obstfeld (1994). El gráfico representa el coste fijo reputacional C_D y los beneficios de devaluar según las expectativas (B^{NC} si los inversores no confían, B^{C} si confían). Para $S_1 < S < S_2$ existen equilibrios múltiples. La profecía autocumplida lleva al punto D.

Estática comparativa: el deterioro de los fundamentos, medido por la ampliación de la brecha de producción negativa, eleva la probabilidad de que el tramo de indeterminación se alcance y de que el ataque llegue a producirse. Gráficamente, equivale a un *shock* más intenso, que sitúa la economía más a la derecha en el eje de abscisas, primero en el tramo de indeterminación ($S_1 < S < S_2$) y, si es grave, en el de devaluación segura ($S > S_2$).

2.3. Contagio: los comportamientos de rebaño

Los modelos de segunda generación, al poner el énfasis en las expectativas, resultan especialmente útiles para explicar los **comportamientos de rebaño**. Como la información es imperfecta y obtenerla cuesta, abundan los inversores poco informados, que suelen imitar las decisiones de un grupo reducido mejor informado.

El mecanismo de transmisión es, en este caso, la retirada de capitales de un país por parte de un grupo de inversores, no por el deterioro de los fundamentos del país, sino por otras razones, como necesidades de liquidez o motivos especulativos. Esa retirada desencadena una reacción en cadena.

2.4. Evidencia empírica: experiencia histórica

Dos episodios históricos pueden leerse desde la óptica de los modelos de segunda generación: la crisis del Sistema Monetario Europeo (1992-1993) y la crisis del Tequila (1994-1995). En ambos, el ataque siguió a sucesos políticos que alteraron las expectativas. En el primero, el rechazo danés al Tratado de la Unión Europea hizo temer el de Francia, y en el segundo, el levantamiento de Chiapas y el asesinato del candidato Colosio precipitaron la salida de capitales.

2.4.1. Crisis del Sistema Monetario Europeo (1992-1993)

> El diseño del Sistema Monetario Europeo, el Mecanismo de Cambios e Intervención y las cuatro etapas de su evolución se desarrollan en el tema B43, apartados 1.2 a 1.4. Aquí interesa el episodio como contraste empírico del modelo de segunda generación.

Causas: el gasto ligado a la reunificación alemana generó tensiones inflacionistas a las que el Bundesbank respondió con subidas de tipos, y los demás países del mecanismo, aunque estaban en recesión, tuvieron que acompañarle. Surgió así el dilema que describe el modelo, pues la defensa de la paridad pedía una política monetaria restrictiva que chocaba con la coyuntura interna. A todo ello se añadió la incertidumbre política, porque en 1992 los daneses habían rechazado en referéndum el Tratado de la Unión Europea y se temía que Francia hiciera lo mismo.

Estallido: los ataques especulativos se desataron en vísperas de la consulta francesa. Francia ratificó finalmente el Tratado, pero con un margen tan estrecho que la presión no cedió. El episodio más conocido es el ataque de George Soros sobre la libra esterlina en septiembre de 1992, articulado mediante una posición corta que anticipaba la depreciación.

Consecuencias: la primera moneda en dejarse flotar fue el marco finlandés, que estaba ligado al ecu por decisión unilateral y no formaba parte del mecanismo. En septiembre de 1992 salieron de este la libra esterlina y la lira italiana. Las presiones siguieron de forma intermitente, con nuevos realineamientos de la peseta y el escudo (noviembre de 1992 y mayo de 1993) y de la libra irlandesa (enero de 1993), hasta recrudecerse en julio de 1993.

Soluciones: la salida inmediata fue el acuerdo de agosto de 1993, que ensanchó al ± 15 % las bandas de fluctuación, y en la práctica el mecanismo dejó de funcionar como se había diseñado. La solución de fondo llegaría por la vía de la integración, con la sustitución del régimen de paridades ajustables por la moneda única.

Conclusiones: el Reino Unido no presentaba una expansión monetaria incompatible con la paridad. Su problema era, en cambio, una recesión que encarecía la defensa de la paridad. Por eso este episodio es el contraste empírico decisivo frente a los modelos de primera generación y el argumento central a favor de la lógica autocumplida.

2.4.2. Crisis del Tequila (1994-1995)

Causas: en 1987 México puso en marcha su propio plan de estabilización, el Pacto de Solidaridad Económica, un programa heterodoxo respaldado desde 1989 por un acuerdo ampliado con el FMI. Bajo el Pacto, las políticas monetaria y fiscal eran restrictivas, y salarios y precios se concertaban con los agentes sociales. El peso servía de ancla nominal, primero con un tipo fijo, desde 1989 con un deslizamiento anunciado de antemano y desde finales de 1991 dentro de una banda cuyo techo se desplazaba, y el sector exterior se abría al comercio y a los capitales.

Al principio el programa funcionó y atrajo capital extranjero. Con todo, al empezar la década de 1990 la economía mexicana tenía dos puntos débiles:

- *Déficit exterior*: un déficit por cuenta corriente próximo al 7 % del PIB, derivado del aumento de la absorción interna.

- *Pecado original*: a lo largo de 1994, buena parte del déficit pasó a financiarse con tesobonos (emitidos desde 1989), deuda a corto plazo indexada al dólar[6]. El país incurría así en lo que Barry Eichengreen y Ricardo Hausmann (1999) denominaron *pecado original*, la incapacidad de usar la moneda propia para endeudarse en el exterior o, incluso dentro del país, a largo plazo. Eichengreen, Hausmann y Panizza (2005) restringieron después el término a la dimensión exterior y lo midieron.

Estallido: la salida de capitales arrancó en 1994 a raíz de dos sucesos. En enero se produjo el levantamiento del Ejército Zapatista de Liberación Nacional en Chiapas, y en marzo fue asesinado el candidato presidencial Luis Donaldo Colosio. Ambos episodios provocaron una fuga de capitales amplificada por los comportamientos de rebaño y por el hecho de que la mayor parte de la inversión extranjera en México era inversión en cartera, con un peso elevado de capital especulativo a corto plazo (*hot money*). Además, el endurecimiento monetario de la Reserva Federal a lo largo de 1994 agravó la reversión de los flujos.

Consecuencias: en diciembre de 1994, con el peso sometido a presiones depreciatorias cada vez más intensas y las reservas muy mermadas, las autoridades tuvieron que renunciar a la banda y dejar flotar la moneda. México sufrió en 1995 una profunda recesión, aunque el producto volvió a crecer con fuerza ya en 1996[7]. La denominación del episodio como *efecto tequila* procede del contagio a Argentina y Brasil, transmitido por comportamientos de rebaño y no por vínculos reales.

Soluciones: México emprendió un programa de reformas estructurales financiado por el FMI y por Estados Unidos, con un paquete de rescate del orden de 50.000 M$, entonces sin precedentes por su magnitud. El episodio motivó además la creación de nuevos instrumentos de financiación de emergencia en el FMI y el debate sobre la arquitectura financiera internacional.

Conclusiones: el episodio responde a la lógica de segunda generación, pues el ataque lo desencadenaron sucesos políticos que alteraron las expectativas y lo amplificaron los comportamientos de rebaño. Sin embargo, el déficit exterior y la deuda a corto plazo indexada al dólar muestran que los fundamentos no eran ajenos a la crisis y anticipan los problemas de hoja de balance que formalizan los modelos de tercera generación.

2.5. Conclusiones

Como en la primera generación, el cierre presenta los resultados del modelo de Obstfeld, las soluciones a las crisis autocumplidas y la valoración del enfoque.

Resultados:

- *Causas*: la crisis cambiaria ya no se explica por una política económica inconsistente ni exige un deterioro previo de los fundamentos que la vuelva inevitable. Surge, por el contrario, de la decisión racional con la que el Gobierno responde al pesimismo de los agentes.

[6] Ello derivaba, en parte, de la decisión gubernamental de 1994 de sustituir deuda denominada en pesos (cetes) por deuda a corto plazo indexada al dólar.

[7] En 1995 el PIB de México cayó en torno a un 6 % y la inflación alcanzó el 52 %.

- *Predictibilidad*: al revés que en la primera generación, la crisis no puede anticiparse pero sí evitarse, porque las reservas internacionales no llegan a agotarse.

Soluciones: dado que el origen de la crisis está en las expectativas y no solo en los fundamentos, la solución pasa por la construcción de credibilidad institucional. Para ello hace falta una política económica ortodoxa que asegure la estabilidad macroeconómica y sirva de ancla a las expectativas de inflación y, por extensión, a las de depreciación. A ello se añaden dos vías complementarias. Una es la acumulación preventiva de reservas, que eleva el coste del ataque y estrecha el tramo de indeterminación, y la otra, el acceso a líneas de liquidez preventivas de carácter multilateral, que operan como respaldo creíble sin necesidad de ser utilizadas.

Aspectos positivos: los modelos de segunda generación explican episodios que la primera generación no podía abordar (señaladamente el ataque sobre la libra esterlina en 1992), y trasladan el análisis desde el automatismo de los fundamentos hacia la interacción estratégica entre autoridad y mercado.

Limitaciones: sin embargo, estos modelos explicaban mal la crisis asiática de 1997, por dos razones. La primera es que el modelo requiere un Gobierno con incentivos para abandonar la paridad, y en aquellas economías, con cuentas públicas equilibradas, inflación baja y sin paro elevado, no los había. En cambio, su vulnerabilidad estaba en el déficit exterior y en el balance de bancos y empresas (apartado 3.4.1). La segunda razón es que estos modelos apenas atienden a la interacción con una crisis financiera que estalle al mismo tiempo. Hay, además, una limitación de método, porque el modelo no determina qué equilibrio se alcanza dentro del tramo de indeterminación, de modo que su capacidad predictiva es, por construcción, limitada.

3. Modelos de tercera generación: modelo de crisis gemelas (Krugman, 1999). Efectos de hoja de balance

Origen: los modelos de tercera generación surgen de las dificultades de los anteriores para explicar la crisis del sudeste asiático (1997-1998). Se apoyan en los trabajos de Douglas Diamond y Philip Dybvig (1983), cuyo modelo de pánicos bancarios se desarrolla en el tema B26, apartado 1.2.4, y de Paul Krugman (1999), cuyo modelo se sigue aquí.

Idea: las crisis financieras y las cambiarias son fenómenos interrelacionados y deben abordarse conjuntamente. La causalidad entre ambas no va en un solo sentido, y de esa interacción surgen las *crisis gemelas*:

1. *Shock financiero*: los problemas del sistema financiero pueden acabar en una crisis cambiaria. Ocurre p. ej. cuando el banco central, como prestamista de última instancia, se ve obligado a crear dinero para salvar a entidades en dificultades.

2. *Shock cambiario*: una crisis cambiaria puede causar una crisis financiera, por ejemplo si la depreciación de la moneda nacional provoca un fuerte aumento del valor en moneda nacional de la deuda denominada en divisas[8].

3. *Shock exógeno*: las dos pueden tener un mismo origen, p. ej. una liberalización desordenada.

La literatura distingue tres variantes de tercera generación, que explican rasgos distintos de la crisis asiática y no se excluyen entre sí:

1. *Riesgo moral y sobreendeudamiento* (McKinnon y Pill, 1996; Corsetti, Pesenti y Roubini, 1999): las garantías públicas implícitas sobre los pasivos de bancos y empresas animan a endeudarse en exceso y a invertir en proyectos arriesgados, lo que infla los precios de los activos y acumula pérdidas latentes. La crisis estalla cuando los acreedores anticipan que el coste fiscal del rescate es incompatible con la paridad. Esta variante explica el auge inmobiliario y bursátil previo y el colapso de las compañías financieras tailandesas.

2. *Iliquidez internacional* (Radelet y Sachs, 1998; Chang y Velasco, 2001): la banca se endeuda a corto plazo en divisas y presta a largo plazo. Si sus pasivos exteriores a corto plazo superan los activos líquidos en divisas del país, un pánico de los acreedores es un equilibrio posible, como en el modelo de Diamond y Dybvig, y con tipo fijo el banco central no puede actuar como prestamista de última instancia en divisas sin agotar sus reservas. Esta variante explica la brusca retirada de los préstamos bancarios en 1997, cuando la deuda externa a corto plazo superaba las reservas en Corea, Indonesia y Tailandia.

3. *Hoja de balance* (Krugman, 1999): la depreciación encarece la deuda en divisas de las empresas, erosiona su riqueza neta y reduce su capacidad de invertir, lo que deprecia aún más la moneda nacional. Esta variante explica la profundidad de la recesión posterior a las devaluaciones, sobre todo en Indonesia, y es la que se formaliza aquí.

En el modelo de Krugman, una reversión de los flujos de capital extranjero puede provocar crisis gemelas por la vía de los *efectos de hoja de balance*. La caída de la inversión deprecia el tipo de cambio real, encarece la deuda en divisas y erosiona la riqueza neta de los empresarios, que pierden capacidad de endeudarse.

[8] La deuda denominada en divisas no debe confundirse con la deuda externa, porque la primera se define por la moneda de denominación y la segunda por la residencia del acreedor, aunque en las economías emergentes ambas suelen solaparse.

3.1. Supuestos

Entorno: la economía es abierta y pequeña, con libertad de circulación de capitales. Los mercados financieros son imperfectos, con información asimétrica, y los prestamistas tienen en cuenta el grado de apalancamiento de los inversores.

El límite máximo a la capacidad de endeudamiento e inversión es la inversión financiable, $I \leq I_f = (1+\theta)W$, donde W es la riqueza neta de los empresarios. El parámetro $\theta > 0$, constante y exógeno, refleja el grado de apalancamiento del sector nacional de empresarios-inversores.

Agentes: los trabajadores destinan toda su renta al consumo y los empresarios, toda la suya al ahorro y la inversión[9].

- *Consumo e inversión*: los bienes nacional y extranjero son sustitutivos imperfectos, con $\sigma_{ij} = 1$. Del gasto en consumo y del gasto en inversión, una parte $m \in (0,1)$ se importa y el resto, $1-m$, se dirige al bien nacional, de modo que m es la propensión marginal a importar de la economía[10].

- *Riqueza neta de los empresarios*: es lo que poseen menos lo que deben, $W = \alpha Y - D - eD^*$, donde $e > 0$ es el tipo de cambio real. Su activo es la parte del producto que les corresponde, αY, y su pasivo es la deuda neta real, con un tramo en moneda nacional, D, y otro en divisa, D^*.

Tecnología: función de producción Cobb-Douglas, $Y = K^\alpha L^{1-\alpha}$, que obtiene un único bien homogéneo a partir de dos factores, el capital y el trabajo. El capital se remunera a αY y tiene una duración de un periodo[11], y el trabajo se remunera a $(1-\alpha)Y$[12].

3.2. Modelo

Planteamiento: el modelo encadena tres efectos de una caída de la inversión (sobre el tipo de cambio real, sobre la riqueza neta de los empresarios y sobre la inversión financiable), cuya conjunción da lugar a las crisis gemelas. El punto de partida es la condición de equilibrio en el mercado de bienes y servicios, que exige que la oferta del bien nacional iguale a su demanda:

$$Y = (1-m)C + (1-m)I + eX^*$$

donde C e I son el consumo y la inversión privados reales y X^* las exportaciones reales, expresadas en unidades del bien extranjero y convertidas al bien nacional mediante el tipo de cambio real directo, e.

El consumo procede solo de los trabajadores, que gastan toda su renta, así que $C = (1-\alpha)Y$ y la condición queda:

$$Y = (1-m)(1-\alpha)Y + (1-m)I + eX^*$$

Equilibrio: despejando e se obtiene la ecuación fundamental del modelo, el tipo de cambio real que equilibra el mercado de bienes:

[9] Se supone implícitamente que los trabajadores carecen de acceso a los mercados de capital, a diferencia de los empresarios.

[10] El problema de transferencia surge porque los residentes dedican al bien nacional una fracción de su gasto, $1-m$, mayor que la que le dedican los no residentes (apartado 3.2, ecuación 1).

[11] Ello implica que $K_t = I_{t-1}$.

[12] Estas remuneraciones resultan de pagar a cada factor su productividad marginal. Con rendimientos constantes a escala, $K \cdot \partial Y/\partial K = \alpha Y$ y $L \cdot \partial Y/\partial L = (1-\alpha)Y$, de modo que el producto se agota (teorema de Euler).

$$e = \frac{Y - (1-m)(1-\alpha)Y - (1-m)I}{X^*} \rightarrow e = f(m^+, \alpha^+, Y^+, I^-, X^{*-}) \tag{1}$$

La ecuación (1) muestra que cuanto mayor es la inversión, menor es el tipo de cambio real. Por ello, una pérdida de confianza repentina que conlleve una caída de la inversión ($\downarrow I$) puede provocar una depreciación real ($\uparrow e$). La intuición es la del problema de transferencia. La caída de la inversión reduce la demanda del bien nacional (en la fracción $1-m$ que recaía sobre él). Con el producto dado, solo una depreciación real, que eleva el valor de las exportaciones en términos del bien nacional, restablece el equilibrio del mercado de bienes.

Estática comparativa: de (1) se desprenden también las relaciones del tipo de cambio real con el resto de variables:

1. *Propensión marginal a importar* (m): cuanto mayor sea, mayor será la demanda de divisas por parte de los residentes y, por extensión, menor la demanda de moneda nacional, lo que debilita la moneda.

2. *Participación relativa de los empresarios* (α): cuanto mayor sea su participación en el producto, menor es la parte de la renta que se destina al consumo, $(1-\alpha)$, y, con ella, la demanda del bien nacional. Para equilibrar el mercado, el tipo de cambio real ha de estar entonces más depreciado.

3. *Renta* (Y): cuanto mayor sea la renta, más se destina al consumo de bienes extranjeros, de modo que la oferta del bien nacional crece más que su demanda interna, lo que exige un tipo de cambio real más depreciado para equilibrar el mercado.

4. *Exportaciones* (X^*): cuanto mayores sean las exportaciones, mayor será la demanda de moneda nacional en los mercados de divisas, lo que fortalece la moneda.

Dinámica: la caída de la inversión y la depreciación real ponen en marcha una secuencia de efectos encadenados. Su primer eslabón es el efecto de hoja de balance, un deterioro de la riqueza neta de los empresarios ($\downarrow W$) que se transmite por dos canales:

1. *Activo*: al caer la inversión cae la renta nacional del periodo siguiente y, con ella, la renta que los empresarios obtienen de la producción, $\alpha \downarrow Y$. Como el producto de cada periodo está predeterminado por el capital instalado, el efecto inmediato sobre la riqueza neta se transmite únicamente a través del pasivo, como recoge la derivada (2).

2. *Pasivo*: al depreciarse la moneda nacional, el valor en moneda nacional de la deuda denominada en divisas aumenta, $\uparrow eD^*$.

Se produce, por todo ello, un deterioro importante de la riqueza neta empresarial. La cadena de acontecimientos va así más allá de la depreciación real y alcanza el balance de las empresas.

La caída de la riqueza neta de los inversores reduce a su vez la inversión financiable ($\downarrow I_f$), debido al empeoramiento de la calidad crediticia de las empresas y a la pérdida de confianza de los bancos en un marco de información imperfecta. La secuencia se realimenta, porque una menor inversión provoca una depreciación real, esta erosiona los balances y unos balances más débiles recortan otra vez la inversión financiable.

La magnitud de la crisis gemela (la retroalimentación entre crisis cambiaria y crisis financiera por conjunción de todos los efectos anteriores) se estudia a partir de la derivada de la inversión financiable respecto de la inversión[13]:

[13] La derivada se obtiene componiendo el efecto de la inversión sobre la riqueza neta y el de esta sobre la inversión financiable: $\partial I_f/\partial I = (\partial I_f/\partial W)(\partial W/\partial I)$, donde $\partial I_f/\partial W = 1 + \theta$.

$$\frac{\partial I_f}{\partial I} = (1+\theta)\frac{(1-m)D^*}{X^*} \rightarrow \frac{\partial I_f}{\partial I} = f(\theta^+, D^{*+}, m^-, X^{*-}) \tag{2}$$

Según (2), su magnitud depende de tres factores:

1. *Grado de apalancamiento* (θ): cuanto mayor sea, mayor será el efecto contractivo sobre la inversión financiable. Así, tanto las economías del sudeste asiático antes de 1997 como el sistema financiero estadounidense antes de la caída de Lehman Brothers en 2008 presentaban grados de apalancamiento privado muy elevados.

2. *Endeudamiento exterior relativo* (D^*/X^*): cuanto más pese la deuda en divisas frente a las exportaciones, mayor será la pérdida de riqueza neta que causa la depreciación y, con ella, la contracción de la inversión financiable. Tailandia, p. ej., mantenía en 1996 una deuda externa a corto plazo superior a sus reservas, lo que la hizo especialmente vulnerable a la reversión de los flujos.

3. *Propensión marginal a importar* (m): cuanto mayor sea, menor será el *shock* sobre la inversión financiable, en la medida en que una mayor parte del ajuste se canaliza vía reducción de la demanda de bienes exteriores.

En consecuencia, pueden darse dos escenarios en función del tamaño del multiplicador:

1. $\partial I_f/\partial I < 1$: la salida de capitales podría generar restricciones crediticias adicionales (racionamiento), pero no una crisis gemela de magnitud significativa, porque el efecto de retroalimentación es decreciente.

2. $\partial I_f/\partial I > 1$: la salida de capitales podría generar una situación de equilibrios múltiples, con una senda explosiva de contracción del crédito, y sí podría dar lugar a una crisis, ya que la retroalimentación crece.

Gráficamente, en el plano inversión-inversión financiable (figura B17.4), la inversión financiable, $I_f(I)$, es nula con una inversión muy baja, porque los empresarios son insolventes, y plana con una muy alta, porque la restricción de financiación deja de ser activa. En el tramo intermedio su pendiente es (2), y si supera la unidad la curva corta tres veces a la recta de 45° ($I_f = I$). Los cortes extremos son estables, el equilibrio alto (punto 0) y el de crisis con inversión nula (punto 1), y el intermedio (A) es inestable, de modo que una pérdida de confianza puede llevar la economía de 0 a 1.

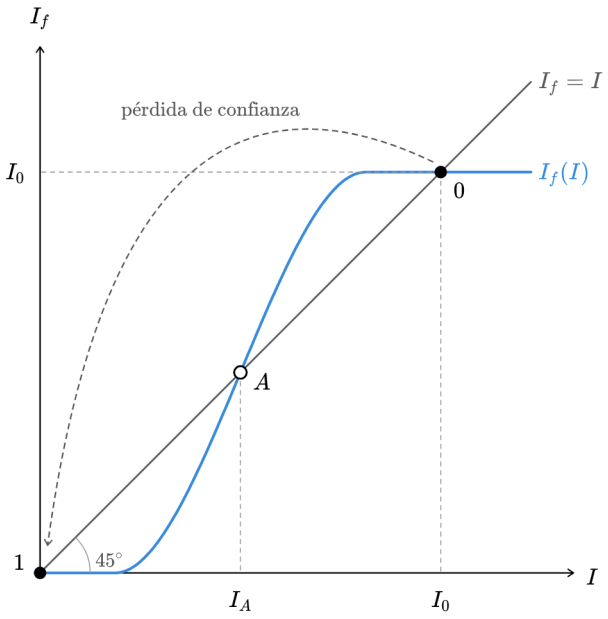

Figura B17.4: Crisis gemelas (Krugman, 1999). La inversión financiable, $I_f(I)$, es nula con inversión muy baja, porque los empresarios son insolventes, y plana con inversión muy alta, porque la restricción de financiación deja de ser activa, y en el tramo intermedio su pendiente supera la unidad. Corta a la recta de 45° en el equilibrio alto (0), en uno intermedio inestable (A) y en el de crisis, con inversión nula (1), al que lleva una pérdida de confianza.

3.3. Contagio: los vínculos financieros

Los modelos de tercera generación explican el contagio que se transmite por **vínculos financieros**. Cuando los inversores[14] tienen posiciones en varios países con fuerte integración financiera, la crisis de uno les lleva a retirar sus fondos tanto de ese país como de los que están ligados a él, porque temen que el problema se propague. Esa misma retirada acaba desencadenando la crisis en los segundos.

Los vínculos financieros dan lugar asimismo al **efecto prestamista común**[15]. Cuando existen bancos internacionales altamente expuestos a países en crisis, esos bancos cortan la financiación a otros países en circunstancias similares para cubrir pérdidas y preservar la calidad de sus carteras, lo que genera una crisis de liquidez. De ahí se deriva la consecuencia más importante, la justificación de los rescates de los organismos multilaterales, que, al frenar la propagación de la crisis, evitan una externalidad negativa.

El canal de los vínculos financieros explica que la crisis de las *subprime* llegara a Europa, cuya banca tenía en cartera deuda hipotecaria estadounidense. Como esos instrumentos eran opacos, no se sabía con certeza cuánto riesgo había asumido cada entidad, y esa incertidumbre paralizó el mercado interbancario y generó problemas de liquidez. Se contrajeron así el crédito y la actividad, a lo que se añadió una demanda agregada más débil por el efecto riqueza y el mayor desempleo, y el resultado fue una recesión en la economía real.

[14] Generalmente, entidades bancarias e inversores institucionales (fondos de inversión, fondos de pensiones y *hedge funds*) con carteras diversificadas internacionalmente.
[15] Kaminsky y Reinhart (2000).

Europa sufrió la crisis con más dureza que Estados Unidos porque su economía dependía más del crédito bancario, mientras que la estadounidense podía recurrir a fuentes más variadas, como la emisión de acciones y bonos o el capital riesgo.

Una modalidad de contagio por vínculos financieros es el **efecto dominó**, es decir, los efectos mecánicos y prácticamente automáticos que se producen en cascada ante la crisis de una región. En este caso, la conexión no reside en la exposición recíproca entre las economías afectadas, que pueden carecer por completo de ella, sino en los balances de los inversores comunes a ambas. Destacan dos canales de transmisión:

- *Canal de liquidez*: el inversor apalancado que sufre pérdidas en una plaza necesita efectivo para reponer garantías o atender reembolsos, y lo obtiene liquidando posiciones en otras, sea cual sea la calidad de esos activos.

- *Canal de balance*: la caída del precio de los activos deteriora el patrimonio de los intermediarios expuestos, lo que reduce su capacidad de asumir riesgo y les lleva a reducir posiciones en el conjunto de sus mercados, propagando la perturbación.

3.4. Evidencia empírica: experiencia histórica

Dos episodios históricos pueden leerse desde la óptica de los modelos de tercera generación: la crisis del sudeste asiático (1997-1998), que motivó su formulación, y la crisis financiera global (2008-2014), cuyo encaje en esta generación se discute. En ambos operó el efecto de hoja de balance.

3.4.1. Crisis del sudeste asiático (1997-1998)

Causas: a mediados de los años noventa, las economías del sudeste asiático presentaban unos fundamentos internos sólidos, con la inflación contenida, un crecimiento rápido, un paro estable, unas cuentas públicas equilibradas y una deuda pública reducida. Esa base, unida a una apertura financiera precipitada que alentó el FMI y al relato del milagro asiático, atrajo mucho capital exterior, sobre todo capital especulativo a corto plazo. Había, no obstante, dos puntos frágiles:

- *Regulación*: el marco de control y supervisión financiera era débil, lo que permitió a los bancos asumir un riesgo cambiario excesivo, tomando prestado en dólares y en yenes y prestando en moneda nacional.

- *Descalce y reservas*: los países mantenían regímenes de tipo de cambio fijo junto a déficits exteriores crecientes y reservas insuficientes frente a su deuda externa a corto plazo.

El resultado fue un elevado endeudamiento a corto plazo en moneda extranjera[16], que, alimentado por el riesgo moral, generaba fragilidad financiera y exposición de hoja de balance. Existía, por tanto, una vinculación muy estrecha entre la situación financiera y la cambiaria, especialmente acusada en Tailandia, cuyo déficit por cuenta corriente rondaba el 8 % del PIB en 1996.

Estallido: la secuencia se inició con el colapso de las compañías financieras tailandesas a comienzos de 1997 (p. ej. Finance One), que evidenció la magnitud de la exposición inmobiliaria del sistema. La pérdida de confianza derivó en ataques especulativos sobre el baht que agotaron las reservas y forzaron al Banco de Tailandia a dejar flotar la moneda en julio de 1997.

[16] Yoshitomi y Ohno (1999) denominan este episodio crisis de la cuenta de capital (la cuenta financiera, en la terminología actual). Las entradas de capital privado, sobre todo préstamos a corto plazo, superan el déficit por cuenta corriente y lo impulsan, y no al revés, hasta su reversión súbita.

Consecuencias: el abandono de las paridades se extendió a la región. La depreciación elevó con fuerza el valor en moneda nacional de la deuda denominada en divisas, deteriorando los balances y provocando pánicos bancarios y racionamiento del crédito en Malasia, Indonesia, Filipinas y Corea del Sur. El contagio alcanzó después a Rusia (1998) y a Brasil (1999) y contribuyó a la caída de los precios del petróleo, desde unos 18,5 $ por barril de Brent a mediados de 1997 hasta un mínimo próximo a 10 $ a finales de 1998, con efectos sobre los países de la Organización de Países Exportadores de Petróleo (OPEP).

Soluciones: la respuesta se canalizó a través del FMI mediante paquetes de rescate de volumen sin precedentes, condicionados a un fuerte ajuste basado en el enfoque de absorción y a la elevación de los tipos de interés para frenar la salida de capitales. La medida buscaba cerrar el canal cambiario, pero agravó el real, al deprimir la inversión y el producto[17]. De hecho, el propio FMI reconoció posteriormente que el ajuste fiscal exigido fue excesivo, al no anticipar la magnitud de la contracción.

Conclusiones: a raíz de este episodio se impuso una gestión más cautelosa de la apertura de la cuenta financiera, una supervisión más estrecha del apalancamiento y la reforma de los instrumentos del FMI. La lección más duradera fue, sin embargo, la acumulación preventiva de reservas por parte de las economías emergentes, que alteraría los flujos financieros internacionales durante la década siguiente.

3.4.2. Crisis financiera global (2008-2014)

> El análisis completo de la crisis financiera global, de los mecanismos de pánico bancario y de la gestión de riesgos se encuentra en el tema B26, apartados 2.2.1, 1.2.4 y 3. La respuesta regulatoria posterior se estudia en el tema B27, apartado 3, y la crisis de deuda soberana europea, en el tema B43, apartado 3.1. Aquí interesa únicamente su encaje en la tipología de modelos de crisis.

Causas: desde el punto de vista teórico, cuatro factores concurrentes generaron la fragilidad previa a la crisis *subprime* (2007-2009), detonante de la crisis financiera global:

- *Riesgo moral*: las garantías implícitas asociadas a las entidades consideradas demasiado grandes para quebrar permitían a estas endeudarse a tipos reducidos y asumir riesgos elevados. En la zona del euro, la falta de credibilidad de la cláusula de no rescate produjo un efecto análogo, con fuertes entradas de capital en la periferia y convergencia de sus condiciones de financiación con las alemanas.

- *Desequilibrios exteriores*: la acumulación de reservas por parte de las economías emergentes tras la crisis asiática se canalizó hacia instrumentos de deuda estadounidense, elevando la liquidez disponible y presionando a la baja los estándares de calidad crediticia para colocar esa mayor oferta de fondos prestables.

- *Laxitud monetaria*: la Reserva Federal mantuvo tipos reducidos desde comienzos de los años dos mil tras el pinchazo de las puntocom y el 11-S. En la zona del euro, la política común resultó además asimétrica respecto de las necesidades de cada miembro.

- *Fragilidad de los mercados financieros*: la desregulación permitió el desarrollo del modelo de originar para distribuir mediante la titulización hipotecaria. Asimismo, la derogación parcial de la Glass-Steagall Act por la Ley Gramm-Leach-Bliley en 1999 permitió la afiliación de la banca comercial con la de inversión y con el negocio asegurador.

[17] En 1998 el PIB de Corea cayó en torno a un 5 %, el de Tailandia un 7,6 % y el de Indonesia un 13 %.

Estallido: la elevación de los tipos de interés por la Reserva Federal a partir de 2004 aumentó la morosidad hipotecaria y precipitó el pinchazo de la burbuja inmobiliaria. El pinchazo se transmitió al sector financiero a través de los productos estructurados y, desde este, a la economía real.

Consecuencias: el efecto de hoja de balance resultó particularmente intenso, dado que prácticamente todos los agentes mantenían exposición al sector inmobiliario, ya fuese mediante la compra directa de activos o mediante hipotecas titulizadas. La contracción del crédito se propagó al conjunto de la economía mundial.

Soluciones: la respuesta combinó rescates bancarios, política monetaria no convencional y estímulo fiscal coordinado en el marco del G20. Le siguió una reforma regulatoria de alcance, con Basilea III, la Ley Dodd-Frank y la Unión Bancaria europea, orientada a elevar los requisitos de capital y a establecer marcos de resolución que impusieran las pérdidas a accionistas y acreedores.

Conclusiones (encaje en la tipología): no existe consenso sobre la clasificación de la crisis financiera global como crisis de tercera generación.

- *A favor*: operan el efecto de hoja de balance y la interacción entre fragilidad financiera y macroeconómica.

- *En contra*: Estados Unidos no sufrió la fuga de capitales ni la depreciación brusca de su moneda que predice el modelo, precisamente por su condición de emisor de la divisa de reserva y destino del vuelo hacia la calidad.

De ahí que se hable de modelos de cuarta generación, anticipados por Krugman (2001) y basados en la pérdida de valor de los activos antes que en la depreciación del tipo de cambio real. Se consideran el marco más adecuado para explicar los problemas de hoja de balance en economías avanzadas.

3.5. Conclusiones

El cierre de la tercera generación sigue el mismo orden que los anteriores, con los resultados del modelo de crisis gemelas, sus soluciones y su valoración.

Resultados:

- *Causas*: las crisis surgen en el sistema financiero-cambiario. Los inversores retiran sus capitales del país en cuanto pierden de repente la confianza, y con esa retirada cae la inversión.

- *Segunda ronda*: la huida de capitales provoca una fuerte depreciación real. Como los agentes deben en moneda extranjera pero poseen activos en moneda nacional, su deuda vale más en moneda nacional y el efecto riqueza negativo deteriora sus balances.

- *Retroalimentación*: con balances más débiles, los agentes pueden endeudarse menos e invierten menos, y el proceso se alimenta a sí mismo. Estos modelos dan cuenta, así, de los graves efectos de la crisis sobre la economía real, que se deben a su impacto en la inversión productiva.

- *Predictibilidad*: las crisis son menos predecibles que las de primera generación, porque parten de unos fundamentos sólidos. El modelo explica cómo se amplifica la crisis, pero no cuándo estalla, porque el detonante es un cambio de expectativas exógeno (véanse las Limitaciones).

Soluciones:

▪ *Solución tradicional*: el FMI recomendaba subir el tipo de interés para retener los capitales. Si estos no salen, la moneda no se desploma y la crisis financiera no llega a convertirse en cambiaria. El remedio tiene, sin embargo, un efecto secundario, porque unos tipos más altos deprimen la inversión productiva, con ella la producción y finalmente la riqueza empresarial, así que el daño se propaga por otra vía con consecuencias similares.

▪ *Soluciones alternativas*: de ahí que se hayan propuesto medidas en dos planos:

- *Medidas preventivas*: actúan sobre θ, m, D^* o X^*, o sobre lo que desencadena la crisis financiera. Entre ellas están una política económica creíble, que disuade la fuga de capitales, una apertura prudente de la cuenta financiera ($\downarrow \theta$, $\downarrow D^*$) y una supervisión financiera que impida el apalancamiento excesivo ($\downarrow \theta$, $\downarrow D^*$). También cuenta la liberalización de la cuenta corriente, que, al elevar las exportaciones, rebaja el peso relativo de la deuda exterior ($\uparrow X^*$). A ellas se suman la vigilancia de los descalces de moneda y de vencimientos y la acumulación preventiva de reservas.

- *Medidas paliativas*: ayuda financiera multilateral que evite al país un ajuste interno desmedido.

Aspectos positivos: el modelo de hoja de balance explica por qué una depreciación, en vez de estimular la actividad, puede deprimirla (véanse los Resultados), algo que las dos generaciones anteriores no podían explicar. De ahí se siguen los equilibrios múltiples y la simultaneidad de crisis cambiarias y bancarias.

Limitaciones: el modelo describe el mecanismo de amplificación pero no determina el momento del colapso, que sigue dependiendo de un cambio de expectativas exógeno al modelo. Tampoco especifica por qué las economías se endeudan en moneda extranjera, cuestión que la literatura posterior aborda bajo la noción de pecado original (apartado 2.4.2).

Conclusiones

Cada generación de modelos de crisis surge de la insuficiencia de la anterior para explicar un episodio concreto. Así, en la primera generación (Krugman, 1979a; Flood y Garber, 1984), la crisis es el desenlace predecible e inevitable de una política fiscal y monetaria incompatible con la paridad. La segunda (Obstfeld, 1994), en cambio, introduce una autoridad que compara los costes y los beneficios de defender la paridad, de modo que unos mismos fundamentos admiten equilibrios múltiples y la crisis puede ser una profecía autocumplida. Por su parte, la tercera (Krugman, 1999) se fija en el balance privado y explica por qué, con descalces de moneda y de vencimientos, la depreciación es contractiva y las crisis cambiaria y bancaria llegan juntas. Por ello, la prevención depende del origen de la crisis y pasa por la disciplina macroeconómica, por la credibilidad y las reservas o por la regulación prudencial y una apertura ordenada de la cuenta financiera. No obstante, en la práctica, los tres mecanismos rara vez actúan por separado.

Los modelos de crisis descansan en las operaciones del mercado de divisas (tema B13) y en la intervención y la gestión de reservas, de cuyo nivel depende el éxito del ataque (tema B15). Además, la lógica de los equilibrios múltiples reaparece en una unión monetaria incompleta, donde la deuda soberana ocupa el lugar de la paridad (tema B16). Asimismo, los descalces de balance de la tercera generación anticipan la crisis financiera global (tema B26) y la reforma regulatoria posterior (tema B27). Por último, la prevención y la resolución de las crisis son la razón de ser de los programas del FMI y de la red de seguridad financiera global (tema B22).

La gestión de los flujos de capital ha ganado respaldo en el FMI. En su visión institucional de 2012, el Fondo aceptó los controles como instrumento temporal ante oleadas de entradas, y su revisión de 2022 permite aplicarlos de forma preventiva a las entradas de deuda cuando hay descalces de moneda elevados. En cambio, el reparto entre el autoseguro y el seguro colectivo sigue en discusión. Obstfeld, Shambaugh y Taylor (2009) muestran que en el pánico de 2008 las economías con reservas escasas en relación con su sistema financiero sufrieron depreciaciones mayores. Estos autores destacan, además, el papel de las líneas *swap* de la Reserva Federal. Queda abierto, así, si la red de seguridad debe descansar en reservas propias, que son costosas, o en las líneas entre bancos centrales, los acuerdos regionales y el FMI.

Anexo. Contagio puro (Calvo y Reinhart, 1996)

Cada una de las tres generaciones de modelos explica un canal de contagio distinto: los vínculos comerciales en los de primera generación (§1.3), los comportamientos de rebaño en los de segunda (§2.3) y los vínculos financieros en los de tercera (§3.3). Existe, sin embargo, una modalidad que no se deja adscribir a ninguna de ellas y que la literatura ha tratado de forma autónoma, razón por la cual se recoge aquí.

Siguiendo la clasificación de Calvo y Reinhart (1996), se denomina **contagio puro** al que se transmite de una economía a otra por meros cambios en las expectativas, sin variación alguna en los fundamentos del país receptor ni vínculo real o financiero que lo justifique. Su relevancia analítica es considerable, porque si la crisis puede propagarse en ausencia de deterioro objetivo y de conexión económica, ni la disciplina macroeconómica ni la diversificación de los vínculos exteriores bastan para prevenirla. Ello refuerza el argumento a favor de las redes de seguridad financiera de carácter multilateral.

Los motivos que conducen al contagio puro son de carácter informativo. Las imperfecciones informativas dificultan la valoración de los fundamentos económicos y abarcan los fallos de extracción de señales, el efecto *wake-up call*[18] y las manchas solares. El efecto dominó, que a veces se incluye en esta categoría, opera en realidad a través de los balances de los inversores comunes y se ha tratado, por ello, entre los vínculos financieros (§3.3). La crisis de deuda soberana europea (2010-2014) ilustra, al menos en parte, el contagio puro.

[18] Hace referencia a cómo la crisis de un país lleva a una revisión de los fundamentos de otros.

B18. Globalización

La nueva globalización económica y financiera. Determinantes y efectos de los movimientos internacionales de factores productivos: movilidad de trabajadores, inversión de cartera e inversión directa, con especial referencia a las cadenas globales de valor.

Introducción

El **sistema económico internacional** (SEI) puede definirse como el orden[1] que alberga las relaciones económicas entre países y que integra tres dimensiones interrelacionadas: el sistema comercial, el sistema monetario y el sistema financiero. Como conjunto de "reglas del juego" de la economía mundial, todo SEI tiene encomendadas tres funciones básicas: repartir de forma simétrica entre países el **ajuste** de los desequilibrios, ofrecer **liquidez**, es decir, activos de reserva en cantidad bastante para facilitar las transacciones internacionales, e inspirar **confianza** en su propia estabilidad, de modo que no haya trasvases bruscos entre esos activos que lo desestabilicen. El grado en que las cumple determina, en buena medida, la estabilidad y la prosperidad de la economía mundial.

La posguerra, a diferencia del periodo de entreguerras, se abrió con una clara voluntad de cooperar, pese a que la Segunda Guerra Mundial había dejado en ruinas la economía internacional. En julio de 1944, cuarenta y cuatro países firmaron en Bretton Woods el Acta Final de la conferencia[2]. De sus acuerdos salieron dos de los tres pilares institucionales del orden económico de la posguerra, y el tercero se negoció aparte:

1. *Monetario*: el **Fondo Monetario Internacional** (FMI), encargado de velar por la estabilidad cambiaria y de financiar a los países con dificultades de balanza de pagos.

2. *Financiero*: el **Banco Internacional de Reconstrucción y Fomento** (BIRF, hoy integrado en el Grupo Banco Mundial), encargado de canalizar capital a largo plazo hacia la inversión productiva, primero para la reconstrucción y después para el desarrollo.

3. *Comercial*: debía recaer en la Organización Internacional del Comercio, negociada en La Habana (1947-1948), que no llegó a crearse, y quedó en manos del **GATT** (1947), con carácter provisional, hasta que nació la **Organización Mundial del Comercio** (OMC) en 1995.

[1] Ese orden se articula en instituciones formales, como el FMI y el Grupo Banco Mundial (GBM), en tratados de integración económica, desde los acuerdos de libre comercio (ALC) hasta la unión económica, y en acuerdos con contenido económico (sobre emisiones, datos, etc.).

[2] España no participó en la conferencia. Ingresó en el FMI y en el BIRF el 15 de septiembre de 1958, en vísperas del Plan de Estabilización de 1959.

Muy transformado, ese entramado sigue siendo la columna vertebral del sistema.

Desde el fin del sistema de Bretton Woods, el SEI ha vivido un proceso de **globalización** comercial y financiera sin precedentes, pero también crisis recurrentes, como la de la deuda latinoamericana, la asiática, la financiera global de 2008-2014 y la de deuda soberana europea. Más recientemente, la pandemia de 2020, la guerra de Ucrania, la rivalidad entre grandes potencias y el cambio climático han recrudecido la **fragmentación geoeconómica** abierta tras la crisis financiera global. Esa fragmentación tensiona la relación entre apertura y autonomía estratégica, entre multilateralismo y bilateralismo, y entre crecimiento y sostenibilidad.

El proceso de globalización y fragmentación se manifiesta en los movimientos internacionales de bienes y de factores (tema B18) y hace necesaria la coordinación internacional de las políticas económicas (tema B19). Culmina una evolución histórica que va del patrón oro al "no sistema" surgido en 1976 de los Acuerdos de Jamaica (temas B20 y B21), y hoy se refleja en los flujos comerciales y financieros, en las nuevas áreas emergentes y en las perspectivas de la economía mundial (tema B32). A ese proceso han tenido que adaptarse los tres pilares del orden de Bretton Woods, el FMI (tema B22), el Grupo Banco Mundial (tema B30) y el GATT y la OMC (temas B33 y B34).

El estudio de los **movimientos internacionales de factores** comprende el análisis de los flujos entre países de trabajo, capital y otros factores productivos (p. ej. recursos naturales[3] o incluso la iniciativa empresarial y el conocimiento).

Generalmente, estos flujos internacionales ocurren de dos formas principales:

1. *Migraciones*: en el caso del factor trabajo, los flujos son desplazamientos geográficos de los trabajadores de un país a otro.

2. *Transferencias*: en el caso del factor capital, los flujos tienen lugar mediante transacciones con instrumentos financieros y contratos que, generalmente, implican la transmisión de derechos de propiedad.

Las fases de mayor apertura comercial han tendido a coincidir con las de mayor movilidad de los factores, con la salvedad de los controles de capital de Bretton Woods, y las de repliegue, con restricciones a ambas (tema B21, apartado 1.2, y tema B32, apartado 1.1).

Richard Baldwin (2016) lee la globalización en términos de dos "desagregaciones" (*unbundlings*), analizadas en el apartado 1.3. La de producción y consumo, en la primera globalización, movilizó capital y trabajo hacia las economías de frontera. La segunda, la de las etapas de un mismo proceso productivo, en la tercera globalización, traslada tareas y capital productivo hacia los eslabones de menor coste y da lugar a las cadenas globales de valor.

Por todo ello, a menudo es difícil distinguir el comercio internacional de los movimientos internacionales de factores, ya que ambos están ligados y la teoría neoclásica del comercio los ha considerado sustitutivos (teorema de igualación del precio de los factores, tema B05, apartado 2.4.2). No obstante, los movimientos de factores plantean cuestiones de índole social que no están presentes en los de bienes y servicios.

Hoy, los flujos internacionales de factores están de máxima actualidad, porque la pandemia de COVID-19 desorganizó las cadenas globales de valor (CGV), y la guerra de Ucrania y la de Siria han generado desplazamientos forzosos masivos hacia Europa. Según el Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR), el número de desplazados forzosos en el mundo,

[3] Aunque los recursos naturales son inmóviles, los derechos de propiedad y las licencias sobre ellos (p. ej. sobre zonas de extracción de crudo) sí se negocian internacionalmente.

incluidos los internos, alcanzó un máximo histórico de 123 millones a finales de 2024 y bajó a 117,8 millones a finales de 2025, primer descenso en una década.

El tema estudia las causas y efectos de los movimientos internacionales de factores productivos según el siguiente esquema:

- *Caracterización*: rasgos de la nueva globalización económica y financiera, situación de los flujos de factores y cadenas globales de valor.

- *Factor trabajo*: determinantes y efectos de sus movimientos internacionales.

- *Factor capital*: determinantes y efectos de sus movimientos internacionales, con especial atención a la inversión extranjera directa.

1. La nueva globalización económica y financiera y las cadenas globales de valor

La sección caracteriza la nueva globalización por sus rasgos, describe la situación actual de los flujos de trabajo, capital y energía y estudia las cadenas globales de valor.

1.1. Características de la nueva globalización

Cinco rasgos, que explican las características del sistema comercial y la paradoja de la integración financiera (mercados integrados, carteras marcadamente nacionales), distinguen la globalización nueva de las oleadas anteriores y condicionan el comportamiento de los factores productivos:

1. *Ventaja comparativa en eslabón*: el objeto del intercambio es a menudo la tarea y no el producto terminado, lo que da lugar a las cadenas globales de valor.

2. *Comercio electrónico*: una parte creciente de los intercambios se entrega por vía electrónica, de modo que la competencia internacional en el mercado de trabajo ya no exige el desplazamiento físico del trabajador cuando la tarea puede prestarse a distancia.

3. *Flujos de capital y organización*: el capital se desplaza también para organizar cadenas de suministro, y no solo para explotar diferenciales de rentabilidad, lo que altera la naturaleza de la inversión directa.

4. *Seguridad económica*: desde la segunda mitad de la década de 2010, la organización de la producción responde, además de a la eficiencia de costes, a criterios de seguridad y de afinidad geopolítica. La fragmentación por bloques se ha convertido en un determinante autónomo de los flujos.

5. *Globalización financiera*: las posiciones exteriores brutas, activos y pasivos, han crecido mucho más que el PIB y que los saldos netos (Lane y Milesi-Ferretti, 2018). Tras la crisis financiera global, la banca transfronteriza se retiró y ganaron peso la inversión en cartera y la intermediación no bancaria. Los flujos brutos se mueven, además, a la vez en todo el mundo, con la aversión al riesgo y la política monetaria de Estados Unidos (*ciclo financiero global*, Rey, 2013), lo que limita la autonomía monetaria aun con tipos de cambio flexibles (tema B21, apartado 2.2.4).

> Las cuatro características del sistema comercial contemporáneo (desaceleración de la globalización, comercio intraindustrial e intraempresa, cadenas globales de valor y digitalización) y los cuatro rasgos del sistema financiero (digitalización de activos, paradoja de la integración, innovación financiera y endeudamiento) se desarrollan en el tema B21 §1.1 y §3.1, y las disrupciones comerciales recientes, en su §1.2.3.

1.2. Situación actual de los flujos

Con ese marco, los flujos actuales se describen por factores, primero el trabajo, después el capital y, por último, los recursos naturales, ilustrados con el caso de la energía.

1.2.1. Flujos internacionales de factor trabajo

Concepto: los movimientos internacionales de factor trabajo comprenden los desplazamientos geográficos de personas, y se miden por el número de migrantes que entran y salen de un país durante un periodo determinado, generalmente un año.

Tipos: según la clasificación de las Naciones Unidas, la delimitación estadística distingue entre *movimientos a corto plazo*, con estancias de entre tres meses y un año, y *movimientos de largo plazo*, con estancias superiores al año.

Desde el punto de vista teórico cabe distinguir además la *migración de retorno*, cuando el trabajador regresa a su país tras el desplazamiento, y la *migración repetida*, cuando reitera la salida. Una tercera clasificación, relevante para el análisis de los efectos, separa la migración cualificada de la no cualificada.

Estructura geográfica: conviene distinguir entre el *stock de migrantes*, es decir, las personas que residen fuera de su país de nacimiento en un momento dado, y los *flujos anuales de entrada y salida*, de magnitud muy inferior. Según el Departamento de Asuntos Económicos y Sociales de las Naciones Unidas (DAES), el *stock* de migrantes internacionales ascendía en 2024 a unos 304 millones de personas[4], en torno al 3,7 % de la población mundial, frente al 2,8 % del año 2000.

El principal corredor migratorio, medido igualmente en términos de *stock*, es México-Estados Unidos, con unos 11 millones de personas, seguido de los corredores Afganistán-Irán, Siria-Turquía[5], Rusia-Ucrania e India-Emiratos Árabes Unidos. El patrón dominante combina la migración económica hacia las economías avanzadas con desplazamientos forzosos de carácter regional y con corredores heredados de la historia política.

Estructura sectorial: la inserción laboral de los migrantes presenta una marcada polarización. Los menos cualificados se concentran en la agricultura, la construcción, la hostelería y los servicios de cuidados, actividades intensivas en trabajo y difícilmente automatizables. Los cualificados, por su parte, se dirigen a la sanidad, la investigación y los servicios tecnológicos y son objeto de políticas de captación selectiva por parte de los países receptores, lo que plantea a los de origen el problema de la fuga de cerebros. El segmento intermedio, por el contrario, está infrarrepresentado.

Conclusiones: la evolución reciente muestra un cambio cualitativo respecto de las migraciones masivas del s. XIX, dirigidas entonces a economías de frontera con abundancia de tierra. Hoy, en cambio, el desplazamiento responde a diferenciales de productividad y de renta antes que a la disponibilidad de recursos, y se orienta hacia trabajadores cualificados en una proporción creciente.

1.2.2. Flujos internacionales de capital

El capital puede hacer referencia al capital físico productivo, al financiero o al humano. Desde el punto de vista empírico, los flujos de capital físico son de escasa magnitud, y el capital humano se desplaza incorporado al factor trabajo, de modo que los modelos se centran en el capital financiero. Este capital se canaliza a través de la cuenta financiera de la balanza de pagos en dos modalidades principales, la inversión extranjera en cartera y la inversión extranjera directa.

[4] DAES, *International Migrant Stock 2024*.

[5] Turquía llegó a acoger a más de 3,5 millones de sirios, cifra que se redujo a unos 2,3 millones a finales de 2025 tras los retornos posteriores a la caída del régimen de Al Asad (diciembre de 2024).

> Las magnitudes actuales de la inversión extranjera directa se recogen en el tema B32 §1.3. Los determinantes de los flujos de capital y sus efectos se desarrollan en la sección 3 de este mismo tema.

Inversión extranjera en cartera (*foreign portfolio investment*): comprende la inversión transfronteriza en títulos de deuda o de participación en el capital distinta de la directa y de los activos de reserva, caracterizada por ser fácilmente negociable. Persigue optimizar el binomio rentabilidad-riesgo sin buscar un interés duradero y, por ello, tiene vocación de corto plazo.

▪ *Tipos*: según el instrumento, cabe distinguir entre acciones, participaciones en fondos de inversión y títulos de deuda, tanto privados como públicos.

▪ *Estructura geográfica*: la inversión en cartera se concentra en las economías avanzadas con mercados de capitales profundos y en los centros financieros que domicilian fondos de inversión. Según el FMI, la posición mundial de activos de cartera alcanzaba a mediados de 2024 unos 71 billones de dólares. Estados Unidos es el primer destino, pues los no residentes poseían a mediados de 2025 valores estadounidenses por unos 35,3 billones[6]. Los cinco primeros tenedores de sus valores a largo plazo son el Reino Unido, las islas Caimán, Canadá, Japón y Luxemburgo, lo que refleja el papel de los centros financieros como intermediarios.

▪ *Estructura sectorial*: por instrumentos, las participaciones en el capital y en fondos de inversión sumaban unos 40 billones de dólares, algo más de la mitad de la posición mundial, y el resto correspondía a títulos de deuda. Entre estos, la deuda pública de las economías avanzadas actúa como activo seguro, y la del Tesoro estadounidense a largo plazo en manos de no residentes rondaba los 7,7 billones de dólares a mediados de 2025. Por tenedores, las instituciones oficiales extranjeras poseían en torno a una quinta parte de los valores estadounidenses en manos de no residentes, y los inversores privados, el resto.

▪ *Conclusiones*: la inversión en cartera es mucho más volátil y procíclica que la inversión directa y constituye, junto con los flujos bancarios (otra inversión), la principal vía de las interrupciones súbitas de capital (apartado 3.2.2).

Inversión extranjera directa (IED, *foreign direct investment*): comprende la inversión transfronteriza realizada por un residente de una economía con el objetivo de establecer un interés duradero (el control o un grado significativo de influencia en la gestión de una empresa no residente). Convencionalmente se emplea para ello el umbral del 10 % de los derechos de voto, conforme a la sexta edición del Manual de Balanza de Pagos del FMI (MBP6) y a las directrices de la OCDE. Sus principales actores son las grandes empresas multinacionales, las instituciones financieras y los fondos de capital privado.

▪ *Tipos*: cabe distinguir dos criterios de clasificación:

- *Destino de los fondos*:
  - *Nueva planta* (*greenfield*)[7]: crea capacidad productiva adicional en el país receptor.
  - *Fusiones y adquisiciones* (*M&A*): la adquisición de empresas existentes transfiere la propiedad sin ampliar el capital instalado.
- *Lógica de la operación*:

[6] FMI, *Portfolio Investment Positions by Counterpart Economy* (antigua Encuesta Coordinada de Inversión de Cartera, CPIS), datos de junio de 2024 ajustados por cobertura (marzo de 2025), y Departamento del Tesoro de EE. UU., *Foreign Portfolio Holdings of U.S. Securities as of June 30, 2025* (abril de 2026).

[7] La matriz crea la filial desde cero, lo que supone mayor riesgo y compromiso que cualquier otra forma de IED.

- *Horizontal*: replica la producción en el mercado de destino para servirlo localmente.
- *Vertical*: fragmenta el proceso productivo entre países según sus ventajas de coste, dando lugar a las cadenas globales de valor que se examinan a continuación.

▪ *Estructura geográfica*: según la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD, *World Investment Report* 2026), las economías en desarrollo recibieron en 2025 algo más de 900.000 M$, más de la mitad del total mundial, y las desarrolladas, 723.000 M$. Estados Unidos fue a la vez el primer receptor (277.000 M$), por delante de Singapur, Hong Kong (China) y China, y el primer emisor (263.000 M$), por delante de Japón y China. No obstante, una parte de los flujos transita por centros financieros sin contrapartida productiva (apartado 3.1.3), lo que obliga a leer con cautela los datos por país.

▪ *Estructura sectorial*: el valor de los proyectos de nueva planta anunciados en 2025 creció sobre todo por los centros de datos, seguidos del petróleo y el gas y de los semiconductores, mientras que retrocedieron las energías renovables, las demás infraestructuras y las manufacturas. Las fusiones y adquisiciones transfronterizas cayeron, en cambio, un 7 %. Esa composición refleja el peso creciente de la economía digital y de la seguridad de suministro en las decisiones de localización.

▪ *Conclusiones*: la IED es la forma más estable de financiación exterior y la que mayor transferencia de tecnología y de conocimiento organizativo genera, si bien su recuperación de 2025 ha sido desigual y muy concentrada, y la IED se reorienta hacia países políticamente afines[8].

1.2.3. Flujos internacionales de recursos naturales: el caso de la energía

Concepto: junto al trabajo y al capital, los recursos naturales constituyen un tercer factor cuyos flujos internacionales condicionan de forma notable el comportamiento de las economías. A diferencia de aquellos, no se desplaza el factor sino su producto, de modo que el comercio de recursos opera como sustituto de una movilidad factorial imposible, porque el yacimiento no puede trasladarse.

Tipos: comprenden recursos energéticos, minerales industriales y metales, y minerales críticos y tierras raras, de relevancia estratégica creciente.

Estructura geográfica: es el rasgo definitorio de estos flujos y la fuente de su relevancia geoeconómica. La oferta presenta una concentración muy superior a la de cualquier otro flujo internacional, tanto en los hidrocarburos, con la Organización de Países Exportadores de Petróleo (OPEP) y Rusia, como en los minerales críticos. En estos, China controla una proporción dominante de la capacidad de procesamiento con independencia de dónde se extraiga el mineral. La demanda, en cambio, está mucho más distribuida, asimetría que confiere a los productores una capacidad de influencia desproporcionada respecto de su peso económico.

Estructura sectorial: las elasticidades de sustitución de la demanda energética son muy reducidas a corto plazo, lo que explica la volatilidad de precios característica de este mercado.

Conclusiones: episodios como la guerra de Ucrania, que llevó a la UE a sustituir el gas ruso por gas natural licuado y renovables, han consolidado la seguridad de suministro, junto al coste, como criterio de política económica (véase B21 §1.2.3).

[8] *World Investment Report* 2026, de la Conferencia de las Naciones Unidas sobre Comercio y Desarrollo (UNCTAD). La IED mundial creció un 6 % en 2025, hasta 1,6 billones de dólares, y puso fin a dos años de descenso. Los flujos hacia las economías en desarrollo, sin embargo, solo aumentaron un 2 %, y más del 80 % del total se dirigió a las veinte principales economías receptoras, en un contexto de fragmentación geoeconómica y de política industrial.

1.3. Cadenas globales de valor

Concepto: una CGV es una cadena formada por etapas que van desde la producción de un bien o servicio hasta la venta al cliente final, de tal manera que dicha cadena tiene al menos dos eslabones en países distintos. Se dice que una empresa participa en una cadena global de valor si realiza al menos una etapa de dicha cadena. El rasgo distintivo de una CGV es que el objeto del comercio pasa a ser la tarea, y no el producto terminado.

Desde 1990 hasta 2008 se produce un avance muy significativo de la globalización, acompañado de la fragmentación de los procesos productivos a raíz de fenómenos como la subcontratación internacional (*outsourcing*) y la expansión de las franquicias. Como resultado, se han ido configurando cadenas cada vez más largas y complejas (p. ej. la fabricación de aviones o de teléfonos inteligentes).

Causas: la fragmentación internacional de la producción responde a la caída de los costes de transporte y, sobre todo, al abaratamiento de la coordinación a distancia que explica Baldwin (2016). A ello se sumaron la liberalización comercial y de la inversión, la incorporación al comercio mundial de grandes economías con abundante mano de obra y la posibilidad de aprovechar la ventaja comparativa en cada tarea, y no ya en el producto completo.

▪ *Desagregaciones* (Baldwin, 2016): las dos de la Introducción responden a restricciones distintas. La caída de los costes de transporte en el s. XIX separó la producción del consumo, pero no las etapas de la producción, porque coordinar procesos complejos exigía proximidad. Solo cuando las tecnologías de la información y las comunicaciones abarataron la coordinación a distancia, a partir de 1990, pudieron las empresas del G7 trasladar etapas a países de salarios bajos, llevando consigo su conocimiento técnico y organizativo. Esa deslocalización del conocimiento explica la "gran convergencia". La participación de los actuales países ricos en la renta mundial, que había pasado de en torno al 20 % en 1820 a casi el 70 % en 1990, ha retrocedido desde entonces, según Baldwin (2016), a su nivel de 1900.

▪ *Comercio de tareas* (Grossman y Rossi-Hansberg, 2008): cada factor realiza un continuo de tareas más o menos fáciles de trasladar. La empresa deslocaliza la tarea i si $w^* c_i < w$, donde w y w^* son el salario nacional y el extranjero del factor que la realiza y $c_i \geq 1$ el coste de deslocalizarla, mayor cuanto más exige proximidad o trato personal. Al caer esos costes, el ahorro que permite la deslocalización actúa como una mejora de la productividad del factor cuyas tareas se trasladan con más facilidad, y ese *efecto productividad* puede elevar su remuneración. Abaratar el comercio de tareas puede beneficiar, así, a todos los factores nacionales, a diferencia del conflicto distributivo que acompaña al abaratamiento del comercio de bienes.

▪ *Fabricar o comprar* (Antràs, 2003; Antràs y Helpman, 2004): decidida la localización, la empresa elige entre producir el insumo en una filial propia (integración, que en el exterior adopta la forma de IED vertical y genera comercio intraempresa) o comprarlo a un proveedor independiente (externalización). Con contratos incompletos, sede y proveedor realizan inversiones específicas de la relación, y la propiedad conviene a quien más aporta. Así, la integración predomina en los sectores intensivos en capital, en los que pesa más la inversión de la sede.

Antràs y Helpman (2004) añaden la heterogeneidad de las empresas y cruzan propiedad y localización en cuatro formas organizativas (integración o externalización, en el país o en el exterior). Como los costes fijos son mayores en el exterior y con integración, en los sectores intensivos en servicios centrales y con una ventaja salarial del exterior suficiente, solo las empresas más productivas integran en el exterior mediante IED vertical. Esta ordenación por productividad traslada a la IED vertical la jerarquía de Helpman, Melitz y Yeaple (apartado 3.1.3) y la ventaja de internalización de Dunning.

- *Otros determinantes*: la localización de cada eslabón depende, además, de cuatro factores:

  - *Costes laborales*: atrae a los eslabones intensivos en trabajo un coste laboral unitario bajo (el salario corregido por la productividad), no el salario bajo sin más.

  - *Instituciones*: los países que hacen cumplir los contratos se especializan en los bienes cuyos insumos exigen inversiones específicas.

  - *Acuerdos comerciales*: como los aranceles se acumulan cada vez que un insumo cruza una frontera, su eliminación y los acuerdos profundos (inversión, servicios, reglas de origen) favorecen la fragmentación.

  - *Logística*: la producción "justo a tiempo" exige entregas rápidas y fiables, de modo que los retrasos en frontera y en puerto penalizan la participación en las cadenas.

- *Riesgo geopolítico*: el riesgo de interrupción del suministro por aranceles, controles a la exportación, sanciones o conflictos entra en el coste esperado de cada localización. Por ello, la mejor opción deja de ser simplemente la más barata y pasa a ser la que mejor combina coste y seguridad del suministro, lo que favorece a los países próximos (*nearshoring*) o afines (*friendshoring*). Las disrupciones recientes han desplazado el paradigma del "justo a tiempo" al "por si acaso", con diversificación de proveedores (p. ej. la estrategia *China+1*) más que relocalización (*reshoring*), lo que alarga en apariencia las cadenas sin desglobalización sustantiva. Los efectos de esa reconfiguración sobre los flujos se examinan en los temas B21, apartado 1.1.3, y B32, apartados 3.1 y 1.3.1.

> El modelo de Antràs y Helpman (2004) se desarrolla en el tema B06, apartado 3.2.2. Aquí se recoge su papel en la organización de las cadenas globales de valor.

Tipos: las CGV pueden configurarse de maneras muy distintas. Las más destacables son:

1. *Araña* (*spider*): los componentes fabricados en varios países se envían a una única planta de ensamblaje (p. ej. la automoción).

2. *Serpiente* (*snake*): el valor añadido se genera de forma secuencial, a medida que la fabricación se encadena en eslabones situados en diferentes países.

Magnitud: según el Grupo Banco Mundial (GBM, 2020), el comercio vinculado a las CGV creció con rapidez entre 1990 y 2008, año en que alcanzó su máximo, en torno al 52 % del comercio mundial. Desde entonces se ha estancado en torno a la mitad del total[9].

- *Medición de la participación*: existen varios indicadores:

  - *Participación hacia atrás* (*backward*): mide el valor añadido extranjero incorporado en las exportaciones del país (cuánto importa un país para poder exportar). Esta participación es característica de las economías situadas "aguas abajo" (ensamblaje).

  - *Participación hacia delante* (*forward*): mide el valor añadido nacional incorporado en las exportaciones de terceros países (cuánto exporta un país en forma de insumos que otros incorporan a sus propias exportaciones). Esta participación es característica de las economías situadas "aguas arriba" (proveedores de materias primas o de bienes intermedios).

[9] Las bases de datos de CGV, como la TiVA de la OCDE y la OMC, construidas sobre tablas *input-output* globales, permiten analizar el comercio en términos de valor añadido. Las estadísticas oficiales en términos brutos sobreestiman el volumen real al contabilizar los bienes intermedios tantas veces como cruzan una frontera.

- *Medición agregada*: la suma de la participación hacia atrás y hacia delante ofrece un índice global de participación en las CGV, mientras que la relación entre ambas (índice de posición) permite situar a cada país en la cadena.

▪ *Participantes*: según el GBM (2020), forman cuatro grupos principales:

- *Exportadores de materias primas* (p. ej. Rusia): participan sobre todo "hacia delante", suministrando recursos naturales y bienes primarios que otros incorporan a sus exportaciones.
- *Manufactura e industria limitada* (p. ej. Argentina): integrados en tareas de ensamblaje y manufactura básica, con participación "hacia atrás" y valor añadido nacional reducido.
- *Manufacturas avanzadas y servicios* (p. ej. China): producen bienes complejos y servicios especializados y captan tramos de mayor valor añadido.
- *Actividades innovadoras* (p. ej. EE. UU.): concentran el mayor valor añadido en los extremos de la "curva de la sonrisa" y ejercen de empresas líderes de las cadenas.

Efectos: la fragmentación internacional de la producción afecta al reparto del valor añadido entre las etapas y opera en dos planos, el microeconómico y el macroeconómico:

▪ *Curva de la sonrisa y gobernanza*: el valor añadido no se reparte de forma homogénea a lo largo de la cadena. La "curva de la sonrisa" (atribuida a Stan Shih, de Acer) describe, en el plano etapas de la cadena-valor añadido (figura B18.1), cómo las etapas situadas en los extremos, la preproducción (I+D, diseño, propiedad intelectual) y la posproducción (marca, mercadotecnia, servicios posventa), concentran la mayor parte del valor añadido. La fabricación y el ensamblaje intermedios, más fácilmente deslocalizables, generan, en cambio, el menor. Gereffi, Humphrey y Sturgeon (2005) complementan este análisis con una tipología de la gobernanza de las cadenas (de mercado, modular, relacional, cautiva y jerárquica), según el grado de coordinación y de asimetría de poder entre empresas líderes y proveedores.

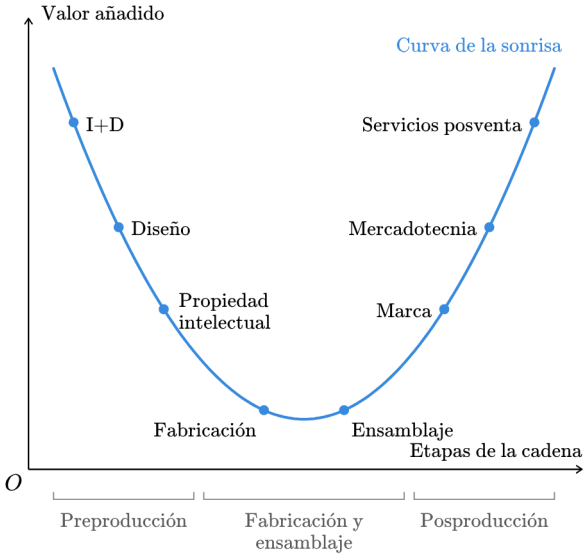

Figura B18.1: Curva de la sonrisa. El valor añadido de cada etapa de una cadena global de valor es máximo en los extremos, la preproducción (I+D, diseño, propiedad intelectual) y la posproducción (marca, mercadotecnia, servicios posventa), y mínimo en la fabricación y el ensamblaje, las etapas más fáciles de deslocalizar.

▪ *Efectos microeconómicos* (sectoriales): especialización más fina y por tareas, apertura del comercio internacional a las pymes al reducir la escala mínima necesaria para participar en él, establecimiento de relaciones duraderas cliente-proveedor, difusión tecnológica y efectos sobre la desigualdad salarial.

▪ *Efectos macroeconómicos*: se manifiestan en varios frentes:

- *Interdependencia*: generación de interdependencias entre países, con transmisión prácticamente inmediata de las perturbaciones, pues la interrupción de un eslabón detiene el conjunto de la cadena.
- *Inflación*: intensificación de la sincronización internacional de la inflación (p. ej. guerra de Ucrania).
- *Comercio bilateral*: intensificación de los flujos comerciales bilaterales (p. ej. China-EE. UU.).
- *Medio ambiente*: efectos sobre la sostenibilidad medioambiental (p. ej. los *pollution havens* o refugios de contaminación).
- *Fiscalidad*: retos para la recaudación tributaria (el problema de los precios de transferencia).

Este tema es la fuente del análisis de las cadenas globales de valor. Su recapitulación como característica del sistema comercial contemporáneo figura en el tema B21 §1.1.3.

1.4. Conclusiones

La globalización actual se distingue de las anteriores en que el objeto del intercambio es a menudo la tarea y no el producto terminado, y en que el capital se desplaza también para organizar cadenas de suministro. La seguridad económica compite, además, con la eficiencia de costes en la localización. Los flujos de cada factor reflejan ese cambio. Así, las migraciones se dirigen menos que en el s. XIX a economías con tierra abundante y cuentan con una proporción creciente de trabajadores cualificados. La volatilidad de la inversión en cartera contrasta con la estabilidad de la directa, y los recursos naturales, inmóviles, circulan a través del comercio de su producto, con una oferta muy concentrada.

Las cadenas globales de valor condensan esa transformación. El abaratamiento de la coordinación a distancia permitió trasladar etapas y conocimiento a países de salarios bajos, y el riesgo geopolítico pesa, junto al coste laboral unitario, en la localización de cada eslabón. Detrás de estos flujos operan determinantes económicos, geográficos e institucionales, que la sección 2 analiza para el trabajo y la sección 3, para el capital.

2. Determinantes y efectos de los movimientos internacionales de factor trabajo

El estudio del factor trabajo contrasta primero la predicción neoclásica con los hechos y examina después qué determina los flujos de trabajadores y qué efectos producen.

2.1. El punto de partida teórico: la sustituibilidad neoclásica (Mundell, 1957)

Origen: Robert Mundell (1957) estudia en el modelo de Heckscher-Ohlin la relación entre el comercio de bienes y la movilidad internacional de los factores.

Idea: el punto de partida es la *sustituibilidad* entre comercio de bienes y movilidad de factores que se deriva del marco neoclásico (Mundell, 1957).

> El teorema de igualación del precio de los factores, del que la sustituibilidad de Mundell constituye el reverso, se desarrolla en el tema B05 §2.4.2. Aquí interesa su lectura en clave de movilidad de factores.

Razonamiento: en el marco neoclásico, bajo sus supuestos[10] y en ausencia de especialización completa, el libre comercio de bienes iguala los precios de los factores entre países aun cuando estos no se muevan. Mundell formalizó la relación inversa, según la cual, si un obstáculo (p. ej. un arancel) impide el comercio de bienes, la movilidad de factores tiende a suplirlo. Comercio de bienes y movilidad de factores son así sustitutivos perfectos, y el estudio separado de los movimientos de factores resultaba redundante.

Conclusiones:

- *Aspectos positivos*: el resultado de Mundell sirve de referencia para contrastar los desarrollos posteriores y anticipa que los obstáculos al comercio estimulan la movilidad de factores, la lógica que recogen las teorías comerciales y de localización de la IED (apartado 3.1.3).

- *Limitaciones*: la realidad contradice la predicción de sustituibilidad, ya que los flujos comerciales coexisten con las migraciones y con los flujos de capital, en lugar de sustituirlos. En el plano teórico, Markusen (1983) mostró que la relación se invierte cuando el comercio no se debe a diferencias de dotaciones, sino de tecnología, de rendimientos a escala o de impuestos, porque entonces la movilidad de factores amplía el comercio (tema B05, apartado 3.3.3). Esa discrepancia motiva los desarrollos que siguen, basados en el levantamiento de los supuestos neoclásicos (información imperfecta, costes de transporte, heterogeneidad y dinámica intertemporal).

2.2. Determinantes de los flujos internacionales de trabajadores

Los flujos de trabajadores responden a diferencias salariales, que formaliza el modelo de Roy y Borjas, y a factores socioeconómicos y geográficos.

[10] En especial, tecnologías idénticas, ausencia de reversiones en la intensidad factorial, ausencia de costes de transporte, rendimientos constantes a escala y competencia perfecta.

2.2.1. Determinantes económicos: diferencias salariales (Hicks, 1932; Roy, 1951; Borjas, 1987)

Origen: *Self-Selection and the Earnings of Immigrants* (1987), de George Borjas, analiza las causas y la composición de los movimientos migratorios internacionales a partir de la heterogeneidad de las habilidades individuales.

Idea: los migrantes realizan un análisis coste-beneficio, comparando el salario que obtendrían en cada país dadas sus habilidades, para tomar su decisión.

Supuestos:

▪ *Entorno*: dos países o regiones, el de origen de la migración, A (p. ej. Ucrania), y el de destino, B (p. ej. Polonia).

▪ *Agentes*: se supone que la decisión de migración es irreversible y, dada la heterogeneidad de habilidades en la población, los salarios se comportan como una variable aleatoria. El modelo considera además que existen unos costes de migración entendidos en sentido amplio, $c > 0$. Estos incluyen desde los *costes directos* de transporte de pertenencias y personas hasta los *costes de oportunidad* de renuncia a ingresos durante la migración o los *costes psicológicos* derivados de la desutilidad de dejar atrás a la familia y a los amigos.

Planteamiento: el modelo adopta la estructura del modelo de Roy (1951), en la que el salario que un individuo obtendría en el país j, $w^{e,j}$, se descompone en dos componentes. El primero, observable y común a todos los residentes, es el salario medio del país, $\bar{w}^j$. El segundo, individual y no observable, ε^j, recoge la remuneración que ese país concede a las habilidades del individuo:

$$w^{e,j} = \bar{w}^j + \varepsilon^j, \qquad \varepsilon^j \sim N(0, (\sigma^j)^2), \qquad j = A, B$$

donde $\bar{w}^A \neq \bar{w}^B$ son los salarios medios de cada país y ε^j es una perturbación aleatoria de media nula. Su varianza, $(\sigma^j)^2$, es específica de cada país, pues mide cuánto premia cada uno la habilidad (el rendimiento de la cualificación). Precisamente esa asimetría entre σ^A y σ^B gobierna los resultados de selección que se obtienen en el apartado 2.3.1.

Se sigue de ello que a cada individuo le importa el salario que prevé obtener dadas sus propias habilidades, y no el salario medio *per se*.

La decisión de migrar sigue la lógica de la obra de John Hicks (1932), según la cual los trabajadores deciden si migrar o no sobre la base de los diferenciales salariales. En este modelo, el trabajador decide emigrar desde A hacia B si:

$$(\bar{w}^B - \bar{w}^A) + (\varepsilon^B - \varepsilon^A) > c$$

Equilibrio: de esa condición se obtiene la probabilidad de migrar, que mide la proporción de trabajadores de A que emigran:

$$p_m = P(\bar{w}^B - \bar{w}^A + \varepsilon^B - \varepsilon^A - c > 0)$$

Se tipifica la expresión anterior definiendo la variable $z \equiv \dfrac{\bar{w}^A - \bar{w}^B + c}{\sigma}$, con lo que se obtiene:

$$p_m = 1 - \Phi(z) = 1 - \Phi\left(\frac{\bar{w}^A - \bar{w}^B + c}{\sigma}\right)$$

donde Φ denota la función de distribución de una normal tipificada y σ mide la dispersión del diferencial de habilidades, es decir, la desviación típica de $(\varepsilon^B - \varepsilon^A)$. La tipificación no es un mero recurso de cálculo, ya que, al reducir los cuatro parámetros del modelo a un único índice z, muestra que solo importan las magnitudes relativas. Así, un aumento proporcional del diferencial salarial, del coste de migrar y de la dispersión deja inalterada la probabilidad de migrar.

Estática comparativa: de todo lo anterior se desprende que la probabilidad de migrar se puede resumir en la función:

$$p_m = p_m(\bar{w}^B - \bar{w}^A,\ \sigma,\ c)$$

Según el modelo, los determinantes de los flujos internacionales de trabajo son:

1. *Brecha salarial*: diferencia de salario medio entre destino y origen, $\bar{w}^B - \bar{w}^A$ (signo positivo). Cuanto mayor es, mayor es la probabilidad de migrar.

2. *Dispersión del diferencial de habilidades*: σ (signo ambiguo). Al situarse en el denominador de z, un aumento de la dispersión acerca z a cero y, por tanto, p_m a 0,5, de modo que eleva la probabilidad de migrar cuando el diferencial salarial no la hacía atractiva y la reduce en caso contrario. Su efecto principal recae, por ello, sobre la composición de los migrantes antes que sobre su volumen, cuestión que se desarrolla en el apartado 2.3.1.

3. *Costes de migración*: c (signo negativo). En el país de origen pueden existir restricciones de crédito o trampas de pobreza que impidan afrontar el coste migratorio pese a que migrar sea deseable. En sentido contrario, el tamaño de la colonia de migrantes en el país de destino puede contribuir a reducir los costes, gracias a la información, el alojamiento y la intermediación laboral que esta proporciona (efecto red o efecto diáspora, coloquialmente "efecto llamada").

Conclusiones:

- *Aspectos positivos*: el marco de Hicks-Borjas ordena en una sola expresión los determinantes de la decisión migratoria y deriva de forma explícita la probabilidad de migrar. Su anclaje en el modelo de Roy permite además determinar quiénes migran.

- *Limitaciones*: el modelo trata la decisión como irreversible e individual, dejando fuera la migración de retorno, la repetida y la decisión familiar. Tampoco incorpora los determinantes no económicos ni las restricciones legales de entrada, que en muchos corredores son vinculantes antes que los diferenciales salariales.

2.2.2. Otros determinantes: factores socioeconómicos y geográficos

Influyen también factores no estrictamente económicos, como las razones políticas, la proximidad cultural, la búsqueda de seguridad jurídica o la reunificación familiar[11].

Por último, intervienen los factores geográficos. Según los modelos gravitatorios (§3.1.1), que también se aplican a las migraciones, los flujos bilaterales de personas crecen con el tamaño de las poblaciones de origen y destino y decrecen con la distancia, entendida en sentido amplio (física, cultural, lingüística e institucional). A ello se añade el efecto red, que perpetúa los corredores históricos (dependencia de la trayectoria), muchos de ellos nacidos de vínculos coloniales, lingüísticos o fronterizos.

[11] Entre los países de la OCDE, la migración familiar fue en 2024 el primer motivo de la migración permanente (OCDE, *International Migration Outlook 2025*).

2.3. Efectos de las migraciones internacionales

Explicada la decisión de migrar, sus consecuencias se ordenan desde la composición de los flujos (selección) hasta los efectos redistributivos y los específicos sobre el país de destino y sobre el de origen.

2.3.1. Efectos de selección: autoselección positiva y negativa (Roy, 1951; Borjas, 1987)

Origen: el modelo de Andrew Roy (1951), concebido para explicar la elección entre ocupaciones, que Borjas traslada a la elección entre países. En él importa no solo la media del salario en cada país, sino la distribución completa de las remuneraciones a la habilidad y la correlación entre las capacidades del individuo en el origen y en el destino.

Idea: el apartado anterior determina cuántos trabajadores migran, y este, quiénes lo hacen y de qué tramo de la distribución de habilidades del país de origen proceden. De ese tramo dependen tanto el efecto sobre el origen como la inserción laboral en el destino.

Supuestos:

- *Entorno*: se mantienen los supuestos del modelo de diferencias salariales, con las desviaciones típicas de cada país, y se añade un parámetro.

- *Rendimiento de la habilidad*: por un lado, σ^A y σ^B, las desviaciones típicas de los componentes de habilidad en cada país, que miden el rendimiento de la cualificación.

- *Correlación*: por otro, ρ, el coeficiente de correlación entre ambos componentes, que recoge en qué medida las capacidades valoradas en el origen lo son también en el destino.

Planteamiento: de los parámetros anteriores se obtiene la varianza del diferencial, cuya desviación típica aparece en el denominador de la variable tipificada:

$$\sigma^2 = \mathrm{Var}(\varepsilon^B - \varepsilon^A) = (\sigma^A)^2 + (\sigma^B)^2 - 2\rho\,\sigma^A\sigma^B$$

Equilibrio: de esta expresión se derivan los tres casos de selección que contempla el modelo:

1. *Autoselección positiva*: se produce cuando la remuneración de la habilidad es mayor en el destino ($\sigma^B > \sigma^A$) y la correlación de capacidades es suficientemente alta ($\rho > \sigma^A/\sigma^B$). Quienes más ganan con la migración son entonces los trabajadores situados en la cola superior de la distribución, los más cualificados. Este caso corresponde a las migraciones desde países igualitarios hacia países desiguales.

2. *Autoselección negativa*: se produce cuando la remuneración de la habilidad es mayor en el origen ($\sigma^A > \sigma^B$), manteniéndose una correlación suficientemente alta ($\rho > \sigma^B/\sigma^A$). Los que más ganan migrando son los de la cola inferior, que en su país están "penalizados" por la elevada dispersión salarial. Un país de origen muy desigual "retiene" así a sus trabajadores más capaces, que allí obtienen una prima elevada, y "expulsa" a los de menor cualificación, que en un destino más igualitario mejoran su posición relativa.

3. *Selección de refugiados* (*refugee sorting*): se produce cuando la correlación es baja (inferior al menor de los cocientes σ^A/σ^B y σ^B/σ^A), es decir, cuando las habilidades que se premian en el origen y en el destino guardan poca relación entre sí. La selección es entonces negativa respecto del origen, sea cual sea el país de mayor dispersión. Emigran quienes ocupan posiciones bajas en la distribución del país de origen, que sin embargo pueden situarse en tramos altos de la del destino, precisamente porque uno y otro país valoran atributos distintos. Este caso describe los flujos originados por rupturas políticas o institucionales, en los que las capacidades valoradas en el origen pierden valor al cruzar la frontera.

Desarrollo gráfico: la autoselección se representa en el plano habilidad-salario (figuras B18.2 y B18.3, panel izquierdo), con los niveles (s) de una habilidad laboral, que se supone perfectamente transferible entre ambos países, en el eje de abscisas:

1. *Rectas de remuneración* ($w^A(s)$ en el origen y $w^B(s)$ en el destino): para cada país, una recta que asocia a cada nivel de habilidad el salario que percibe el trabajador con ese nivel. Su pendiente mide la remuneración por unidad de habilidad, es decir, el rendimiento de la cualificación en ese país, y se corresponde con σ^A y σ^B respectivamente.

2. *Umbral*: el corte de ambas rectas define un umbral de habilidad, s_P o s_N, en el que, sin contar el coste de migrar, el trabajador percibe el mismo salario en los dos países. Migran quienes quedan del lado del umbral en que la recta del destino va por encima de la del origen, y permanecen los demás. El área sombreada entre ambas rectas mide la ganancia salarial de los que migran.

3. *Distribución de la habilidad* (panel derecho): la densidad de la habilidad en el país de origen, $f(s)$, sobre el mismo eje y con el mismo umbral, muestra de qué tramo proceden los que migran (área sombreada), la cola superior con autoselección positiva y la inferior con autoselección negativa.

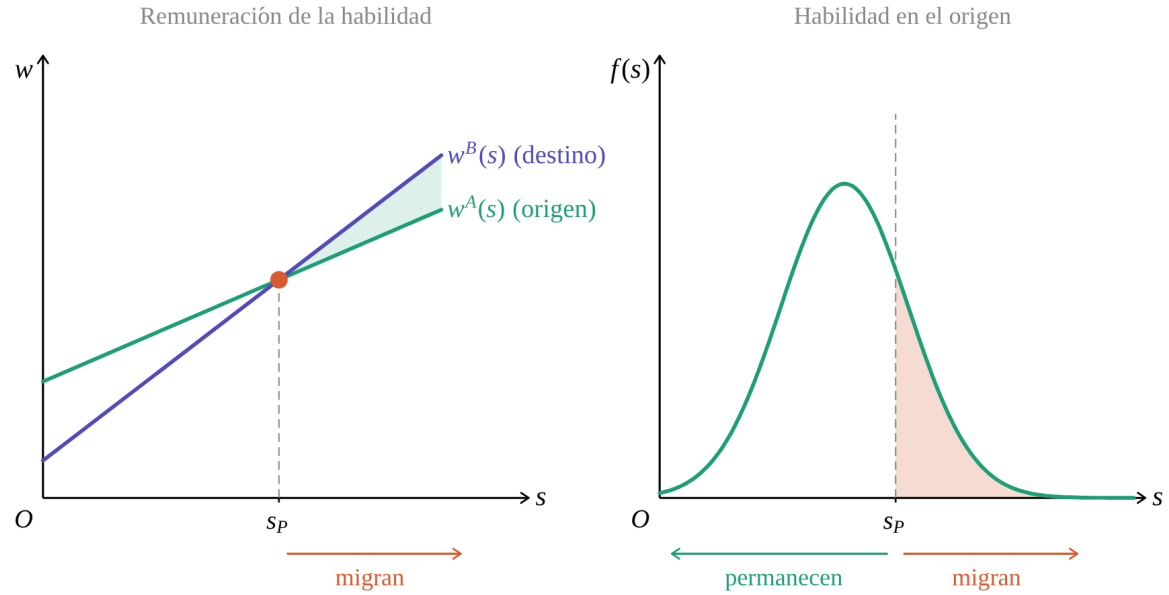

Figura B18.2: Autoselección positiva (Roy-Borjas). El destino remunera más cada unidad de habilidad y migran los trabajadores con cualificación superior al umbral s_P (panel izquierdo), es decir, la cola superior de la distribución de habilidades del país de origen (panel derecho).

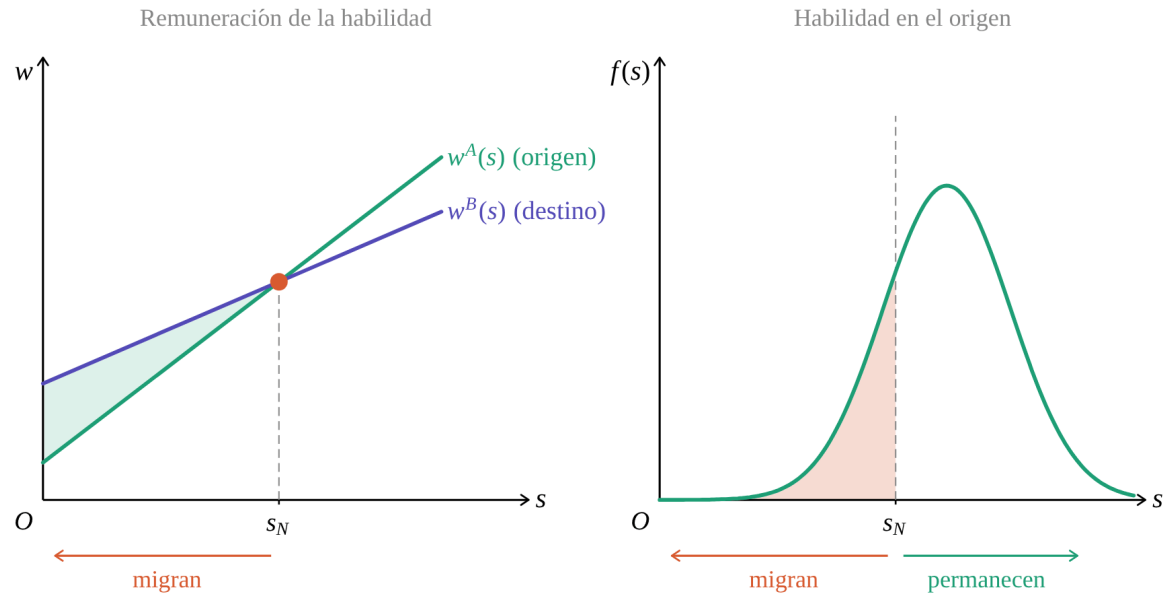

Figura B18.3: Autoselección negativa. El origen remunera más la habilidad y migran los trabajadores con cualificación inferior al umbral s_N (panel izquierdo), es decir, la cola inferior de la distribución de habilidades del país de origen (panel derecho).

Estática comparativa: el signo de la selección depende de la inclinación relativa de las rectas:

- *Recta del destino más inclinada* (figura B18.2): migran quienes superan el umbral s_P (autoselección positiva, p. ej. desde Suecia).

- *Recta del destino menos inclinada* (figura B18.3): migran quienes quedan por debajo de s_N (autoselección negativa, la que el modelo predice desde países en desarrollo más desiguales que el destino).

Extensiones: además de la selección positiva y negativa, hay otros dos efectos:

1. *Ciclos*: los ciclos económicos pueden influir sobre los efectos de autoselección, ya que los auges y las recesiones provocan desplazamientos hacia arriba y hacia abajo de las rectas de remuneración de la habilidad.

2. *Selección de retorno*: una vez en el destino, no todos los migrantes permanecen indefinidamente, sino que muchos regresan a su país de origen. La extensión de Borjas y Bratsberg (1996) predice que los afectados por la selección de retorno serán aquellos que se encuentran en el margen (los que dudaban entre migrar y no migrar). Por ello, la migración de retorno intensifica los patrones iniciales de selección, al hacer más positiva la selección positiva y más negativa la negativa.

Conclusiones:

- *Aspectos positivos*: el modelo ofrece el marco de referencia para estudiar la selección de migrantes en el corredor bilateral México-Estados Unidos. La evidencia es, no obstante, discutida, pues unos estudios hallan selección negativa y otros, una selección intermedia o positiva en términos de cualificación observable. La mayor virtud analítica del modelo es haber demostrado que el signo de la selección no depende del nivel de desarrollo de los países implicados, sino

de la desigualdad relativa entre ellos y de la correlación ρ. El resultado es contraintuitivo y de considerable alcance para el diseño de las políticas migratorias.

- *Limitaciones*:

  - *Otros determinantes*: el modelo no contempla determinantes distintos de la educación (p. ej. motivos políticos, guerras o la edad). Empíricamente, los migrantes son en su gran mayoría jóvenes de entre 16 y 49 años.

  - *Selección en países en desarrollo*: se le ha criticado que, en muchos países en desarrollo en los que el rendimiento de la educación es mayor en el origen que en el destino, la selección no es negativa, sino positiva. Por ello, Chiquiar y Hanson (2005) sostienen que los costes migratorios (p. ej. las restricciones de crédito) influyen mucho más de lo que supone el modelo, pues, al pesar más sobre los menos cualificados, favorecen una selección intermedia o positiva.

  - *Transferibilidad de la habilidad*: supone que la habilidad es perfectamente transferible, cuando la evidencia muestra una depreciación sustancial del capital humano al cruzar la frontera (Friedberg, 2000), por el idioma, el reconocimiento de titulaciones o la especificidad de la experiencia. Esa depreciación puede invertir el signo de la selección observada respecto de la prevista.

2.3.2. Efectos redistributivos: eficiencia y equidad (MacDougall, 1960)

Origen: Donald MacDougall (1960) analiza los efectos de la entrada de capital extranjero sobre el país receptor. Su modelo se aplica aquí, en versión simplificada, a los movimientos del factor trabajo.

Idea: además de la selección, interesa el impacto sobre el bienestar de los movimientos migratorios, que el modelo permite analizar en un solo gráfico.

Supuestos:

- *Entorno*: dos países (origen A y destino B), inicialmente no integrados comercialmente, crean un mercado común y liberalizan los flujos internacionales de factor trabajo entre ellos (podría ser el caso de EE. UU. y México). Se considera que no hay fallos de mercado (competencia perfecta en todos los mercados). Las dotaciones factoriales no son necesariamente iguales, de modo que las productividades marginales (y, por ello, la posición de las curvas) pueden ser diferentes en cada país. La dotación de trabajo (la oferta de trabajo) se considera constante y exógena.

- *Tecnología*: cada país produce un bien (X, Y) con una tecnología no lineal neoclásica de buen comportamiento y dos factores de producción homogéneos, idénticos y perfectamente divisibles. Se supone un entorno de corto plazo en el que el factor capital permanece fijo en ambos países.

De todo ello se deriva que los factores productivos son remunerados de acuerdo con el valor de su producto marginal y que los resultados del modelo dependerán exclusivamente de las diferencias en dichos rendimientos entre los países.

Planteamiento: el modelo compara el equilibrio de autarquía con el del mercado común, y sus principales intuiciones se ilustran gráficamente.

Desarrollo gráfico: los efectos de la movilidad del trabajo se representan en el plano cantidad de empleo[12]-salario nominal (figura B18.4), con el empleo de cada país medido desde su origen (O_A u O_B):

1. *Curva de oferta de trabajo*: al ser la oferta de trabajo exógena, es una recta vertical (no dibujada para facilitar la ilustración gráfica).

2. *Curva de demanda de trabajo* ($\text{VPmg}_L^A(L^A)$ y $\text{VPmg}_L^B(L^B)$): refleja, para cada nivel de empleo, el salario que las empresas de cada país están dispuestas a pagar, que en un entorno competitivo coincide con el valor de la productividad marginal del trabajo. Tiene pendiente negativa respecto de su propio origen, dado que la productividad marginal laboral es decreciente, y, con las tecnologías neoclásicas habituales (p. ej. Cobb-Douglas), es convexa. La curva de B, medida desde O_B, asciende por ello de izquierda a derecha en la figura.

3. *Equilibrio*: en autarquía, antes de la integración y apertura de los flujos de trabajadores (puntos 0), el salario mexicano es w_0^A, mientras que el salario en EE. UU. es w_0^B. El nivel de empleo en México es (O_A, L_0^*) y en EE. UU., (L_0^*, O_B). Los trabajadores mexicanos tienen un incentivo a desplazarse hacia EE. UU., pero todavía no lo hacen. El área sombreada mide la pérdida de eficiencia derivada de esa inmovilidad.

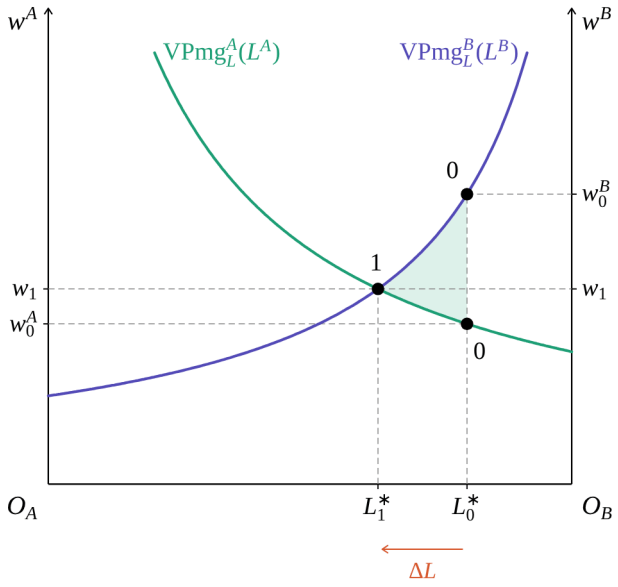

Figura B18.4: Modelo de MacDougall (1960). La libre movilidad del trabajo produce convergencia salarial y recupera el triángulo de eficiencia.

Estática comparativa: si se permite la movilidad internacional de trabajadores (integración en un mercado común), los mexicanos se desplazarán hacia EE. UU., dado que allí su productividad se remunera a un salario mayor. Así, $\Delta L = L_0^* - L_1^*$ es el flujo de trabajadores desde México hacia EE. UU., y los salarios convergen en w_1 (punto 1).

Bienestar: esos movimientos producen efectos de eficiencia y de distribución:

[12] La longitud del eje de abscisas ($O_A O_B$) es la población activa conjunta de ambos países. Con pleno empleo, coincide con el empleo total.

▪ *Efectos sobre la eficiencia*: en el agregado, y al margen del reparto de las remesas, se produce un aumento de la producción mundial (por el aprovechamiento de las ganancias potenciales de eficiencia) y convergencia salarial entre ambos países.

▪ *Efectos sobre la distribución*: esa ganancia agregada oculta, sin embargo, una redistribución de renta entre factores y entre países:

- *México (país de origen)*: los trabajadores que se quedan ganan, los migrantes obtienen la masa salarial que perciben en el destino, los propietarios del capital pierden y la producción interna cae. Intuitivamente, los trabajadores ganan porque suben sus salarios, mientras que los propietarios del capital, que son demandantes de empleo, pierden por esa misma subida. El efecto neto sobre el bienestar de México depende de lo que ocurra con la masa salarial de los migrantes. Si vuelve al país en forma de remesas[13], se produce una ganancia neta de bienestar, y si no vuelve, una pérdida neta.

- *EE. UU. (país de destino)*: los trabajadores estadounidenses pierden, los propietarios del capital ganan y la producción total sube.

En consecuencia, la libre movilidad del trabajo es eficiente, pues aumenta el bienestar global, pero da lugar a ganadores y perdedores, lo que solo desde el punto de vista de la equidad podría justificar las políticas de restricción de flujos.

Extensiones: como la ganancia de eficiencia es, en principio, suficiente para compensar a los perdedores (criterio de Kaldor-Hicks), el debate se traslada al terreno de la economía política. Si las transferencias compensatorias no se articulan, la resistencia de los perdedores puede bloquear una apertura globalmente eficiente.

Conclusiones:

▪ *Aspectos positivos*: el modelo muestra en un único gráfico los efectos de eficiencia y redistributivos de los flujos migratorios. Además, el mismo esquema se puede aplicar a los movimientos de capital, sin más que reinterpretar los ejes en términos de productividad marginal y remuneración del capital.

▪ *Limitaciones*: parte de supuestos muy restrictivos, por lo que no recoge algunos efectos observados:

- *Homogeneidad*: al considerar trabajadores homogéneos, no tiene en cuenta las diferencias en la cualificación, siendo así que la composición del flujo determina buena parte de sus efectos, como muestran los efectos de selección.

- *Tratamiento de las remesas*: el modelo trata el destino de la masa salarial de los emigrantes como un supuesto binario (vuelve o no vuelve) y no analiza los efectos de las transferencias.

- *Predicción salarial*: la caída del salario nativo que el modelo predice para el país de destino ha resultado, en el contraste empírico, menor de lo esperado en la mayoría de los estudios, cuestión que se examina a continuación.

[13] En el caso de los movimientos del factor capital, las remesas serían repatriaciones de rendimientos del capital en el extranjero.

2.3.3. Efectos sobre el país de destino: salarios nativos e impacto fiscal

La predicción de MacDougall de que la inmigración deprime el salario de los nativos ha sido objeto de uno de los contrastes empíricos más intensos de la economía de las migraciones, con resultados discutidos.

Efectos sobre los salarios: el estudio de referencia, de David Card (1990), aprovecha como experimento natural el *éxodo del Mariel*, la llegada repentina de unos 125.000 cubanos a EE. UU. en 1980, cerca de la mitad de los cuales se estableció en Miami (con un aumento de la población activa local de en torno al 7 %). Sin embargo, Card no encuentra efecto apreciable sobre los salarios ni el empleo de los nativos.

La controversia se abre con Borjas (2003), quien, al sustituir el enfoque de mercados locales por el de celdas nacionales de cualificación y experiencia, sí halla efectos negativos significativos.

La conciliación más aceptada de ese debate procede de Ottaviano y Peri (2012), según la cual nativos e inmigrantes no son sustitutivos perfectos aun dentro de un mismo nivel formativo. Por ello, la inmigración deprime sobre todo el salario de los inmigrantes previos y resulta complementaria de buena parte del trabajo nativo.

Con todo, el propio Borjas (2017) reabrió la cuestión al hallar caídas salariales acusadas en el Mariel para los varones sin estudios secundarios, resultado que trabajos posteriores atribuyen al reducido tamaño y a la composición de esa submuestra, de modo que el debate permanece abierto.

Impacto fiscal: se mide por el saldo entre lo que los inmigrantes aportan en impuestos y cotizaciones y lo que reciben en prestaciones y servicios. Ese saldo depende sobre todo de la estructura por edades, de la tasa de empleo de la población inmigrante y de la generosidad del Estado del bienestar.

Las estimaciones comparadas de la OCDE (25 países, 2006-2018) sitúan ese saldo entre el -1 % y el $+1$ % del PIB en la mayoría de ellos, con una aportación media positiva, y muestran que la variable decisiva es la inserción laboral. El efecto es, por tanto, reducido y depende más de las políticas de integración que del volumen del flujo.

2.3.4. Efectos sobre el país de origen: remesas y capital humano

Dos fenómenos determinan el efecto neto sobre el país emisor que el modelo de MacDougall deja abierto. Las remesas deciden si la masa salarial de los migrantes revierte en el origen, y la emigración cualificada altera, en un sentido u otro, su dotación de capital humano.

Remesas: tienen un impacto macroeconómico de primer orden sobre los países de origen. Según el *World Migration Report* 2024 de la Organización Internacional para las Migraciones (OIM), las remesas internacionales pasaron de 128.000 M$ en 2000 a 831.000 M$ en 2022, y superan ya a la inversión extranjera directa como fuente de financiación exterior para los países de renta baja y media.

Las remesas sostienen el consumo de los hogares, relajan las restricciones de crédito y financian educación y pequeños negocios, y presentan además una notable estabilidad anticíclica, al aumentar precisamente cuando la economía de origen se deteriora. Sus contrapartidas son la posibilidad de generar dependencia, la reducción de la oferta laboral de los hogares receptores y las presiones de apreciación real que pueden erosionar la competitividad de los sectores comerciables (síntomas de *enfermedad holandesa*).

Fuga de cerebros y ganancia de cerebros: cuando la selección es positiva, el país de origen pierde precisamente a los trabajadores en cuya formación ha invertido, con el consiguiente perjuicio en sectores críticos como la sanidad o la investigación (*brain drain*). Ante ello, Jagdish Bhagwati (1972, 1976, 1989) propuso la fijación de un *impuesto a la fuga de cerebros* (*brain-drain tax*), con el objetivo de repatriar institucionalmente la renta obtenida por los trabajadores cualificados que han debido emigrar.

Frente a esta visión tradicional, parte de la literatura subraya la posibilidad de una ganancia de cerebros (*brain gain*). Según esta tesis, la perspectiva de emigrar eleva el rendimiento esperado de la educación e incentiva la formación de capital humano en el origen, parte del cual no llega a emigrar. A ello se suman los retornos de migrantes cualificados, que regresan con capital y experiencia acumulados, y la transferencia de conocimiento y de contactos a través de las diásporas.

2.4. Conclusiones

Efectos: la movilidad eleva el producto mundial pero genera ganadores y perdedores (MacDougall). La evidencia apunta a efectos sobre los salarios nativos menores que los del modelo, con el debate abierto, y a un impacto fiscal reducido que depende sobre todo del empleo de los inmigrantes.

Magnitud potencial: las estimaciones de Michael Clemens (2011) sitúan la ganancia de eficiencia derivada de una liberalización plena de los movimientos internacionales de trabajo en un orden de magnitud muy superior al de la liberalización comercial pendiente. La razón es que la brecha de productividad entre un mismo trabajador en origen y en destino (*place premium*[14]) supera con mucho a cualquier distorsión arancelaria subsistente.

[14] Clemens, Montenegro y Pritchett (2019, *Review of Economics and Statistics*) hallan que un mismo trabajador puede multiplicar varias veces su salario por el mero hecho de cruzar una frontera hacia una economía de mayor renta.

3. Determinantes y efectos de los movimientos internacionales de capital

Al igual que las migraciones, los flujos de capital, en sus distintas modalidades (apartado 1.2.2), coexisten con el comercio en contra de la predicción neoclásica.

3.1. Determinantes

Algunos determinantes, como la rentabilidad y la distancia, actúan sobre cualquier flujo de capital. Otros son propios de la inversión en cartera o de la inversión extranjera directa.

3.1.1. Determinantes comunes: rentabilidad y distancia

> El modelo intertemporal de la balanza de pagos (Obstfeld y Rogoff, 1995), la condición de sostenibilidad del déficit exterior y la paradoja de Lucas se desarrollan en el tema B12 §3, y los modelos gravitatorios y su formulación estructural, en el tema B06 §2.3.2.

Diferencial de rentabilidades: el determinante económico fundamental de los flujos internacionales de capital es la diferencia de rendimiento esperado entre economías. Según el modelo intertemporal de la balanza de pagos (tema B12, apartado 3.1.2, y su figura), una economía con escasa dotación de capital presenta una productividad marginal, y por tanto una remuneración, superior a la internacional ($r > r^*$), lo que atrae inversión exterior. La apertura financiera permite entonces separar las decisiones de consumo e inversión, que en autarquía estaban ligadas por la restricción del ahorro nacional. Así, el país puede invertir por encima de su ahorro, financiar la diferencia con un déficit por cuenta corriente y devolver el préstamo con la mayor producción futura.

De ese planteamiento se derivan dos corolarios relevantes para este tema:

- *Dirección de los flujos*: el capital debería fluir de las economías abundantes en capital a las escasas, predicción que la evidencia contradice de forma sistemática (paradoja de Lucas). La discrepancia se atribuye a la calidad institucional, al riesgo de expropiación y a las diferencias de capital humano.

- *Sostenibilidad exterior*: depende de la relación entre el tipo de interés y la tasa de crecimiento, de modo que el endeudamiento que financia inversión productiva, y no consumo, puede atenderse con el crecimiento posterior.

Fuerzas gravitatorias y distancia: junto a la rentabilidad, los flujos bilaterales de capital responden a la misma lógica que los comerciales. El modelo gravitatorio predice que aumentan con el tamaño económico de origen y destino y disminuyen con la distancia, entendida no solo en sentido físico sino también informativo, cultural e institucional. Su aplicación a los flujos financieros tiene una implicación notable. Si el capital fuese perfectamente móvil, la distancia sería irrelevante, de modo que su significación estadística constituye una medida indirecta del grado efectivo de integración financiera y conecta con la paradoja de Feldstein-Horioka[15].

[15] La paradoja de Feldstein y Horioka (1980) constata que, entre los países de la OCDE, el ahorro y la inversión nacionales están fuertemente correlacionados, lo que sugiere una movilidad internacional del capital inferior a la que cabría esperar bajo integración financiera plena.

3.1.2. Determinantes específicos de la inversión en cartera

> La teoría de selección de carteras (Markowitz, 1952; Tobin, 1958; Sharpe, 1963), el modelo de valoración de los activos de capital (CAPM) y la teoría de valoración de activos por arbitraje (APT) se desarrollan en el tema B23 §2 y §3.

La inversión en cartera no persigue el control de la empresa participada, sino la optimización del binomio rentabilidad-riesgo, de modo que sus determinantes son los de cualquier decisión de cartera aplicados en un contexto internacional.

El riesgo de una cartera depende de la correlación entre sus componentes, y la diversificación lo reduce tanto más cuanto menor sea esta. De ahí el argumento clásico a favor de la **diversificación internacional**, según el cual los ciclos económicos nacionales no están perfectamente correlacionados y los activos extranjeros permiten desplazar la frontera eficiente.

Dos hechos matizan esa conclusión y explican la magnitud observada de los flujos:

- *Sesgo doméstico* (*home bias*): los inversores mantienen una proporción de activos nacionales muy superior a la que recomendaría la teoría, lo que se atribuye a los costes de información, al riesgo de tipo de cambio y a las barreras regulatorias.

- *Correlación intermercados*: aumenta precisamente en los episodios de tensión, de modo que el beneficio de la diversificación se desvanece cuando más se necesita, lo que constituye una de las manifestaciones del contagio financiero.

Factores de empuje y de atracción: la literatura empírica y los análisis del FMI distinguen dos tipos de determinantes de los flujos de cartera hacia las economías emergentes (Calvo, Leiderman y Reinhart, 1993; Fernández-Arias, 1996). Los de *empuje* (*push*) son externos al país receptor, como los tipos de interés de las economías avanzadas, la aversión global al riesgo (medida, p. ej., por el índice VIX) y la liquidez mundial. Los de *atracción* (*pull*) son internos, como el crecimiento, la solidez macroeconómica y la calidad institucional. Forbes y Warnock (2012) muestran que los episodios extremos (oleadas, paradas súbitas, fugas y repliegues de capital) dependen sobre todo de factores globales, en especial del riesgo, lo que enlaza con el ciclo financiero global de Rey (2013; tema B21, apartado 2.2.4).

3.1.3. Determinantes específicos de la inversión extranjera directa

A diferencia de la inversión en cartera, la inversión directa busca el control y una relación estable con la empresa participada. La teoría debe responder, por ello, a dos preguntas. La primera es por qué una empresa produce en el exterior en lugar de exportar, y la segunda, por qué lo hace mediante filial propia en lugar de licenciar su tecnología a un productor local.

> Los desfases tecnológicos y el ciclo del producto, junto con su papel en la evolución dinámica de las ventajas comparativas, se desarrollan en el tema B06 §1.2.1.

Teorías tecnológicas (Posner, 1961; Vernon, 1966): el primer grupo de explicaciones vincula la inversión directa a la ventaja tecnológica de la empresa que la realiza. Según Posner (1961), la innovación otorga al país innovador un monopolio temporal que se erosiona con la imitación (*imitation lag*). Vernon (1966) traslada esa intuición al ciclo de vida del producto (nuevo, maduro y estandarizado). En ese ciclo, la innovación, y con ella el origen de la IED, se concentra en los países desarrollados, y la producción se desplaza hacia economías de menor coste a medida que el producto se estandariza. Queda sin explicar, sin embargo, por qué hay IED bidireccional entre

economías desarrolladas y por qué las empresas eligen invertir directamente antes que conceder licencias.

Teorías de organización industrial (Hymer, 1960; Caves, 1971; Buckley y Casson, 1976): el segundo grupo desplaza el foco desde la tecnología hacia las imperfecciones de mercado y las ventajas específicas de la empresa:

- *Poder de mercado*: según Hymer (1960), los flujos de IED obedecen al poder de mercado, ya que las empresas acceden a mercados extranjeros para proteger su cuota. Para ello se sirven de determinadas ventajas con las que contrarrestan los costes de entrada y que, por lo general, guardan relación con imperfecciones de mercado, como las economías de escala. Charles Kindleberger subrayó que esas ventajas específicas deben bastar para superar la "desventaja del extranjero" (lo que después se denominó *liability of foreignness*), es decir, el desconocimiento del marco institucional, cultural y comercial del país de destino frente a los competidores locales.

- *Información imperfecta*: Caves (1971) vincula la IED a la diferenciación de producto, pues la empresa dispone de un conocimiento sobre cómo servir el mercado que puede trasladar a otros países. Buckley y Casson (1976) explican, a su vez, por qué la empresa explota ese conocimiento mediante filiales propias. La fabricación exige combinar muchos activos intangibles, como las patentes, la marca, la reputación, el *know-how* o el capital humano, y con información imperfecta las partes no logran pactar el reparto de las rentas que generan. La empresa prefiere entonces invertir directamente a ceder licencias de fabricación (teoría de la internalización). Esta línea responde a la pregunta que Vernon dejaba abierta, la de por qué la empresa internaliza la operación en lugar de licenciarla.

Teorías financieras y fiscales (Aliber, 1970): el tercer grupo atiende no a dónde se produce, sino a dónde se financia la inversión y dónde se declaran sus beneficios:

- *Moneda fuerte*: según Robert Aliber (1970), las empresas domiciliadas en áreas de moneda fuerte se financian con una menor prima de riesgo cambiario y capitalizan los flujos de renta de sus filiales a un tipo de descuento inferior. De ahí que valoren una misma corriente de beneficios en el exterior más que las empresas de áreas de moneda débil. Ello explicaría oleadas históricas de IED asociadas a las monedas de referencia de cada época.

- *Disparidad fiscal*: cuando los sistemas tributarios difieren, las empresas invierten en el exterior para obtener el mayor beneficio neto de impuestos posible y sitúan sedes y filiales allí donde el impuesto sobre sociedades es más bajo. El mecanismo consiste no solo en trasladar actividad real, sino sobre todo beneficios, mediante los precios de transferencia y la localización de los activos intangibles. Por ello, una parte no menor de la IED registrada corresponde a flujos financieros de tránsito sin contrapartida productiva y el dato agregado debe interpretarse con cautela (apartado 1.2.2).

Teorías comerciales y de localización (Markusen, 2002): el cuarto grupo explica la IED como respuesta a los obstáculos al comercio, pues cuando exportar resulta caro por los aranceles o los costes de transporte, producir en destino se convierte en la alternativa. Estas barreras tienden, por ello, a estimular la IED horizontal, que replica plantas para atender cada mercado desde dentro, y a desincentivar la vertical, que divide el proceso productivo entre países según los precios de los factores y multiplica los cruces de frontera.

James Markusen sintetizó ambas formas en el modelo de "capital-conocimiento" (*knowledge-capital*), que hace depender la elección entre una y otra de tres características:

- *Tamaño y dotación*: su similitud entre países favorece la IED horizontal.

- *Precios de los factores*: sus diferencias favorecen la vertical.

- *Activos del conocimiento*: la intensidad de la empresa en activos de conocimiento fácilmente trasladables entre plantas.

La predicción principal del modelo es que la IED horizontal predomina entre países desarrollados de renta y tamaño similares, mientras que la vertical vincula economías con dotaciones dispares. La primera predicción concuerda con la evidencia de que el grueso de la IED mundial es Norte-Norte.

Síntesis ecléctica de Dunning o teoría OLI (1981): John Dunning reúne las explicaciones anteriores en un marco común. Una multinacional elige la IED cuando dispone a la vez de tres ventajas, que se designan con las iniciales de sus nombres en inglés (OLI):

1. *Propiedad* (*ownership*): la empresa posee activos intangibles de los que carecen las empresas del país de destino. Además, llevarlos a cualquier otro lugar le resulta fácil y barato, y como rinden lo mismo aunque se empleen una y otra vez en emplazamientos distintos, expandirse horizontalmente le cuesta muy poco.

2. *Localización* (*location*): producir en destino es preferible a exportar cuando la exportación resulta cara, p. ej. por los aranceles o el transporte.

3. *Internalización* (*internalization*): las dos ventajas anteriores podrían, en principio, aprovecharse mediante licencias. La empresa prefiere, sin embargo, producir ella misma, porque con información asimétrica es difícil vigilar al licenciatario y redactar un contrato completo que intercambie tecnología por rentas. Por la misma razón, cuesta mucho evitar que este actúe de forma oportunista (p. ej. copiando la tecnología).

En el marco de Dunning, las tres ventajas, combinadas, determinan cuál de las tres formas de servir un mercado exterior elige la empresa:

- *Solo propiedad*: cederá licencias.

- *Propiedad e internalización, sin localización*: exportará, produciendo en casa e internalizando.

- *Propiedad y localización, sin internalización*: cederá licencias a un productor local.

- *Triple ventaja*: solo cuando concurren las tres la IED domina a la exportación y a la licencia como modo de entrada.

Así, el paradigma ecléctico integra en un único esquema los grupos anteriores. Las teorías tecnológicas y el poder de mercado de Hymer explican la ventaja de propiedad. Las comerciales y de localización, junto con la disparidad fiscal, explican la de localización, y la información imperfecta (Buckley y Casson), la de internalización.

- *Conclusiones*:

  - *Aspectos positivos*: el paradigma responde a la vez a las dos preguntas del apartado, por qué producir en el exterior y por qué hacerlo con filial propia, y deduce de ellas el modo de entrada. Esa capacidad de síntesis lo ha convertido en el marco de referencia de la literatura sobre la empresa multinacional.

  - *Limitaciones*: las tres ventajas suelen identificarse después de observar la inversión, por lo que el paradigma funciona más como un esquema de clasificación que como un modelo con predicciones contrastables. Además, es estático y no explica por qué, dentro de un mismo sector, unas empresas invierten en el exterior y otras solo exportan, laguna que cubre la jerarquía de Helpman-Melitz-Yeaple.

El modelo de competencia monopolística con heterogeneidad de empresas (Melitz, 2003) y su extensión a la elección entre exportar e invertir (Helpman, Melitz y Yeaple, 2004) se desarrollan en el tema B06 §3.1 y §3.2.1.

Heterogeneidad de las empresas y elección entre exportar e invertir (Helpman, Melitz y Yeaple, 2004): la aportación más reciente desplaza el foco hacia la empresa individual. Como invertir exige costes fijos superiores a los de exportar pero suprime los costes variables de transporte y aranceles, resulta una *ordenación por niveles de productividad* (jerarquía de Helpman-Melitz-Yeaple). Según esa ordenación, las empresas menos productivas atienden solo el mercado interior, las intermedias exportan y solo las más productivas invierten en el exterior. Su implicación para este tema es que los flujos de inversión directa no responden únicamente a diferenciales de rentabilidad entre países, sino a la composición del tejido empresarial de origen. Una economía con pocas empresas de productividad elevada generará, por ello, escasa inversión exterior por muy atractivo que resulte el destino.

3.2. Efectos

Los efectos siguen la misma división, con unos comunes a todos los flujos y otros específicos de la inversión en cartera y de la inversión directa.

3.2.1. Efectos comunes: balanza de pagos, localización y distribución

Efectos sobre la balanza de pagos:

- *Efecto positivo*: tanto el modelo de media-varianza como el intertemporal de la balanza de pagos indican que la movilidad del capital puede elevar el bienestar. Este último lo atribuye a que amplía las opciones para distribuir en el tiempo consumo e inversión, y el primero, a que facilita diversificar la cartera y rebajar su riesgo.

- *Efecto negativo*: el signo del efecto puede invertirse cuando la cuenta financiera se abre mientras la cuenta corriente sigue sin estar liberalizada de forma plena, porque, con aranceles vigentes, los precios internos dan una señal falsa sobre la ventaja comparativa del país. Si los capitales se mueven guiados por esas señales, pueden ahondar la mala asignación de los recursos y alejar todavía más la especialización del país de la que le sería óptima (argumento de *second best*).

Efectos sobre la decisión de localización: los flujos internacionales de capital inciden sobre la decisión de dónde localizar la producción. En presencia de economías de aglomeración y rendimientos crecientes, la nueva economía geográfica (Krugman, 1991a, desarrollada en el tema B06 §2.3.3) muestra que, por debajo de un umbral de costes de comercio, la actividad tiende a concentrarse en los núcleos ya industrializados (causación circular acumulativa). Ello explica por qué el capital no siempre fluye hacia la periferia pese a su menor dotación.

Efectos redistributivos: basta con poner el capital en lugar del trabajo en el esquema de MacDougall (figura B18.4, con el capital y su remuneración en los ejes) para obtener una conclusión paralela. La movilidad del capital aumenta el bienestar mundial, pero no todos salen beneficiados, pues en cada país ganan quienes ofrecen el factor abundante y pierden quienes ofrecen el escaso.

3.2.2. Efectos de los flujos internacionales de inversión en cartera

Beneficios de la liberalización: a las ganancias de la diversificación se suma una asignación más eficiente del ahorro a escala mundial, puesto que la inversión se dirige hacia las combinaciones rentabilidad-riesgo óptimas.

Profundidad de mercado como factor de atracción: un mercado profundo, líquido y con un amplio abanico de activos poco correlacionados atrae capital por sí mismo, y este, a su vez, lo profundiza (círculo virtuoso). En cambio, los mercados poco desarrollados, con escasas oportunidades de diversificación, atraen menos. Ello explica la concentración de los flujos de cartera en un número reducido de plazas financieras, así como la dificultad de los mercados emergentes para retener capital estable.

Volatilidad y comportamientos de rebaño: la contrapartida es la inestabilidad. Al carecer de vocación de permanencia, la inversión en cartera es especialmente susceptible de comportamientos de rebaño, en los que los inversores imitan las decisiones ajenas en lugar de basarse en información propia. De ahí resultan fases de entrada masiva de capital (*bonanzas*) seguidas de interrupciones abruptas de esas entradas (*sudden stops*), así como la propagación del episodio entre países (contagio), a menudo con independencia de los fundamentos de la economía receptora.

> Los comportamientos de rebaño como mecanismo de contagio, y los vínculos comerciales y financieros a través de los que este opera, se desarrollan en el tema B17 §1.3, §2.3 y §3.3. Los episodios históricos correspondientes se analizan en ese mismo tema.

Extensiones: la evidencia de las crisis financieras ha matizado la visión benévola de la apertura. Las ganancias de la apertura dependen de forma crítica de la secuenciación y de la existencia de precondiciones (un sistema financiero interno sólido y bien supervisado, un marco macroeconómico estable y una regulación prudencial adecuada). Una apertura prematura puede amplificar la volatilidad, alimentar auges de crédito y exponer a la economía a reversiones súbitas. De ahí que los beneficios de liberalizar la IED y la inversión de cartera de largo plazo se consideren más claros que los de abrirse a los flujos de corto plazo. Por la misma razón, el FMI admite, en determinadas circunstancias, medidas de gestión de los flujos de capital (véase B22).

3.2.3. Efectos de la inversión extranjera directa

> Los efectos de selección, de productividad agregada y de aprendizaje (*learning by exporting*) que la apertura genera sobre el tejido empresarial se desarrollan en el tema B06 §3.1 y §3.4. Explican qué empresas llegan a invertir en el exterior, no los efectos de esa inversión, que son los que siguen.

Los efectos de la inversión directa sobre las economías de origen y destino operan en cuatro planos:

- *Trabajo*: el impacto de la IED sobre el empleo del país receptor ha sido siempre objeto de debate. Para sus críticos, las multinacionales se instalan fuera para aprovecharse de la mano de obra local. Sus defensores sostienen, por el contrario, que la IED crea empleo y eleva los salarios, porque con más capital sube la productividad marginal del trabajo. En el país de origen, la IED es polémica cuando hay paro.

- *Balanza de pagos*: para el país de destino, la IED supone nuevos pasivos financieros y, con ellos, un menor saldo de la cuenta financiera. Baja también el saldo de la cuenta corriente, salvo que lo compensen otras partidas de la propia cuenta financiera, como las reservas. En el país de origen (emisor) el efecto es el inverso.

- *Estructura industrial*: la IED impulsa la industrialización del país de destino, aunque también puede crear multinacionales con mucho poder de mercado, que repiten sus plantas en numerosos países mediante inversión horizontal. Además, cuando la inversión es vertical (fragmenta la cadena de producción entre países), las actividades de mayor valor añadido tienden a permanecer en el país de origen.

- *Crecimiento económico*: la IED puede acelerar el crecimiento del país de destino gracias a sus externalidades positivas, como la difusión de tecnología, la acumulación de capital humano o la cesión de activos intangibles, sobre todo cuando ese país tiene suficiente capacidad de absorción. Para que esas externalidades se materialicen, muchos receptores han exigido que la inversión extranjera adopte la forma de empresas conjuntas (*joint ventures*) con socios locales (p. ej. China, que con la Ley de Inversión Extranjera de 2019, en vigor desde 2020, limitó ese requisito a los sectores de su lista negativa).

Todo ello sitúa a los Estados ante una tensión característica. El país receptor tiene incentivos a competir por atraer inversión mediante ventajas fiscales, lo que, llevado al extremo, degenera en una carrera a la baja de la imposición societaria. El emisor y la comunidad internacional, por su parte, afrontan el traslado artificial de beneficios vía precios de transferencia[16]. La respuesta multilateral ha sido el proyecto de la OCDE y el G20 contra la erosión de bases imponibles y el traslado de beneficios (BEPS), con el acuerdo de 2021 sobre un impuesto mínimo global (véase B19, apartado 3.2.2).

3.3. Conclusiones

Efectos: la liberalización mejora la asignación internacional del ahorro y permite suavizar el consumo ante perturbaciones transitorias, pero introduce volatilidad y riesgo de reversión súbita, especialmente en la inversión de cartera. La inversión directa presenta un perfil distinto, con mayor estabilidad y efectos de arrastre tecnológico y organizativo sobre la economía receptora.

Comparados ambos factores, el capital es hoy el factor más móvil internacionalmente y el trabajo el menos móvil, pese a que liberalizar la movilidad laboral reportaría las mayores ganancias potenciales (apartado 2.4), discrepancia que constituye la principal anomalía de la globalización contemporánea de los factores.

[16] Precios de las transacciones entre entidades vinculadas de un mismo grupo (p. ej. matriz y filial), que determinan el reparto del beneficio entre países y pueden utilizarse para trasladarlo a jurisdicciones de menor tributación o para eludir aranceles. Su regulación internacional se coordina en el seno de la OCDE.

Conclusiones

Los movimientos internacionales de factores coexisten con el comercio de bienes, en contra de la sustituibilidad neoclásica (Mundell, 1957). La decisión de migrar responde a la brecha salarial, a la remuneración relativa de la habilidad, que fija el signo de la autoselección (Roy y Borjas), y a la distancia y las redes. La movilidad eleva el producto mundial pero crea ganadores y perdedores (MacDougall), y su efecto sobre el país de origen depende de las remesas y de la fuga de cerebros. Los flujos de capital, por su parte, siguen al diferencial de rentabilidad corregido por la distancia, la cartera busca además diversificar el riesgo y la inversión directa combina ventajas de propiedad, localización e internalización. La liberalización mejora la asignación del ahorro a costa de volatilidad, mientras que la inversión directa aporta estabilidad y tecnología. De ahí una anomalía, pues el capital es hoy el factor más móvil y el trabajo, cuya liberalización daría las mayores ganancias, el menos móvil.

La coexistencia de los movimientos de factores con el comercio de bienes contradice la sustituibilidad que se deriva del teorema de igualación del precio de los factores, y la complementariedad que le opone Markusen (1983) amplía el modelo de dotaciones (tema B05). La elección entre exportar e invertir en el exterior procede de la heterogeneidad de las empresas, porque solo las más productivas cubren el coste de establecer filiales (tema B06). En los flujos de capital, el enfoque intertemporal explica el saldo exterior como reparto del ahorro entre países (tema B12), y la diversificación de carteras, la inversión de cartera (tema B23). Por último, las cadenas globales de valor y la integración financiera caracterizan el sistema posterior a Bretton Woods (tema B21), y la competencia fiscal por la inversión exige coordinación multilateral (tema B19).

El efecto de la inmigración sobre los salarios nativos es objeto de controversia. Card (1990) no halla efectos apreciables y Borjas (2003, 2017) sí, con estimaciones que dependen del método y de la muestra (apartado 2.3.3). Tampoco está resuelto el peso de la geopolítica en la inversión directa. El FMI (2023) documenta que la parte de la IED entre economías geopolíticamente afines sigue creciendo, más que entre economías próximas, y que las emergentes y en desarrollo serían las más perjudicadas por esa fragmentación. La seguridad económica compite ya con la eficiencia como criterio de localización.

B19. Coordinación de políticas económicas internacionales

La coordinación internacional de políticas económicas. Aspectos teóricos y prácticos. El G20, la OCDE y otros foros internacionales.

Introducción

Introducción común. Los temas B18, B19, B20, B21 y B32 comparten una introducción, que se desarrolla al comienzo del tema **B18**. Define el sistema económico internacional como el orden de las relaciones económicas entre países, con sus dimensiones comercial, monetaria y financiera y sus funciones de ajuste, liquidez y confianza, recuerda los tres pilares del orden de Bretton Woods (FMI, Banco Mundial y GATT-OMC) y sitúa el sistema en la etapa actual de globalización y fragmentación geoeconómica.

En su mensaje a la sesión inaugural de la Conferencia de Bretton Woods (1 de julio de 1944), Franklin Roosevelt señaló: "las enfermedades económicas son muy contagiosas y, por ello, la salud económica de cada país es asunto que concierne legítimamente a todos sus vecinos, próximos o lejanos"[1].

La **coordinación de políticas económicas** comprende modificaciones mutuamente acordadas de las políticas interiores de diferentes naciones debido a la interdependencia que existe entre ellas. En el campo macroeconómico involucra, generalmente, acuerdos explícitos sobre la conducta monetaria y fiscal de los países.

Es más rigurosa que la mera "cooperación", pues la coordinación exige el ajuste mutuo de las políticas en un momento concreto y conforme a unos compromisos establecidos. La coordinación internacional puede dividirse en diferentes niveles:

1. *Intercambio de información*: los países se informan mutuamente de su situación coyuntural y de las políticas que aplican (p. ej. en los foros de la OCDE, el FMI o el G7).

2. *Coordinación temporal* o *ad hoc*: suele surgir como reacción ante una crisis (p. ej. los Acuerdos del Plaza de 1985, los Acuerdos del Louvre de 1987 o la cumbre de Londres de 2009[2]).

3. *Coordinación permanente*: es el caso de la Unión Europea.

Si se opta por la coordinación, existen dos enfoques para llevarla a cabo:

[1] Mensaje de Franklin D. Roosevelt a la Conferencia de Bretton Woods (julio de 1944), recogido por el Centre virtuel de la connaissance sur l'Europe (CVCE).
[2] Sobre la cumbre del G20 de Londres (abril de 2009) y sus acuerdos, véase el apartado 3.2.1.

1. *Enfoque hegemónico*: se reconoce la existencia de un líder o de una autoridad capaz de imponer la coordinación y a la que los demás países subordinan parte de la autonomía de su política económica.

2. *Enfoque simétrico*: admite dos formas:

- *Enfoque discrecional*: consiste en la actuación sin mecanismos de coacción[3]. Corresponde a los dos primeros niveles y permite una mayor flexibilidad de acción.

- *Enfoque normativo*: consiste en establecer reglamentaciones que imponen una cierta disciplina a los países involucrados. Supone, en cambio, una reducción notable de la flexibilidad de actuación.

En los años setenta, con el desmoronamiento del sistema de Bretton Woods y una interdependencia mundial en aumento, la coordinación de las políticas económicas entre países suscitó un interés cada vez mayor. Aunque los resultados de los modelos son muy sensibles a sus supuestos, cada vez son más los sectores que la asocian con ganancias potenciales de bienestar para todos los implicados[4].

El tema responde al enunciado en tres secciones, dos dedicadas a los aspectos teóricos y una a los prácticos:

- *Fundamentos teóricos*: la interdependencia y los canales de transmisión de los efectos desbordamiento, de los que dependen las ganancias potenciales de coordinar.

- *Justificación teórica*: la teoría de juegos y la de los bienes públicos globales explican por qué esas ganancias exigen un diseño institucional, y la literatura señala los límites de la coordinación.

- *Aspectos prácticos*: las características de los acuerdos, el G20, la OCDE y otros foros internacionales (apartado 3.2), la experiencia y las dificultades del proceso.

[3] Los juegos repetidos muestran que la cooperación puede sostenerse sin coacción (apartado 2.1.3).

[4] Aun así, la comunidad internacional llegó a la crisis financiera global (2008-2014) con escasa coordinación previa, pues los desequilibrios globales apenas ocuparon al G8, pese a las advertencias del FMI y del Banco de Pagos Internacionales (BIS). Solo la gravedad de la crisis impulsó la concertación en el G20.

1. Fundamentos teóricos: interdependencia y transmisión internacional

El análisis parte de la interdependencia y de los efectos desbordamiento que genera, y sigue con los canales por los que estos se transmiten, el de la renta y el monetario, el de la inflación y el de las expectativas.

1.1. Interdependencia, efecto desbordamiento y canales de transmisión

Interdependencia y efecto desbordamiento: al coordinar sus políticas, los países reconocen que dependen unos de otros. Las decisiones de cada uno generan externalidades sobre los demás, los llamados efectos desbordamiento (*spillover effects*), con tres consecuencias:

- *Eficacia de la política económica*: queda alterada en cada país.

- *Autonomía de las autoridades*: se reduce el margen de decisión de cada país.

- *Pérdida de eficiencia*: la actuación por separado conduce a resultados subóptimos, porque cada país decide atendiendo únicamente a los efectos que su política produce sobre sí mismo y desatendiendo los que produce sobre los demás. El problema se agrava en presencia de bienes públicos globales, como el libre comercio, la estabilidad cambiaria o un sistema financiero regulado, cuya naturaleza no excluyente y no rival genera incentivos al comportamiento de *free rider*. Ante esta situación, la coordinación no elimina las externalidades, que son consustanciales a la interdependencia, sino que permite internalizarlas en la decisión de cada país.

Canales de transmisión: los efectos desbordamiento se propagan internacionalmente por cuatro canales:

- *Canal de la renta*: permite que las economías con mayor peso relativo induzcan estímulos en la renta internacional a través del aumento de su propia renta nacional. La transmisión exterior opera, principalmente, a través de dos cuentas de la balanza de pagos:

  - *Cuenta corriente*: el aumento de la renta nacional supone un aumento de las importaciones de bienes y servicios, lo cual induce un aumento de las exportaciones de otros países y, por tanto, de la renta internacional.

  - *Cuenta financiera*: el deterioro de la balanza comercial tiene como contrapartida una entrada neta de capitales que lo financia, de modo que el impulso de renta se refleja simultáneamente en ambas cuentas.

- *Canal monetario*: los impulsos monetarios se transmiten internacionalmente a través de dos subcanales vinculados al mercado monetario y de divisas:

  - *Tipo de interés*: sus variaciones mueven los flujos de capitales, y estos repercuten en la inversión, nacional y extranjera, y en el tipo de cambio. Este subcanal actúa sobre todo por la cuenta financiera.

  - *Tipo de cambio nominal*: su influencia se transmite, a su vez, por dos vías, la cuenta corriente y la cuenta financiera:

    - *Cuenta corriente*: cuando se verifica la condición de Marshall-Lerner, la depreciación del tipo de cambio nominal, a precios dados, hace crecer las exportaciones netas, es decir, mejora la balanza comercial.

- *Cuenta financiera*: quien cambia moneda nacional por divisa para comprar un activo extranjero obtiene, por una parte, el rendimiento de ese activo y, por otra, lo que varíe el tipo de cambio hasta el vencimiento. Bajo la paridad descubierta de intereses, la suma de ambos componentes iguala, en valor esperado, el tipo de interés nacional.

▪ *Canal de la inflación*: las variaciones de los precios relativos modifican el tipo de cambio real, que condiciona el comercio y los movimientos de capital. Su efecto se deja sentir, por ello, tanto en la cuenta corriente como en la financiera.

▪ *Canal de las expectativas*: opera a una escala más abstracta y amplia, en la medida en que los cambios en las expectativas de los agentes pueden desencadenar por sí solos crisis de deuda y cambiarias.

1.2. Canal de la renta y canal monetario: modelo IS-LM-BP

Origen: el modelo IS-LM-BP de Fleming (1962) y Mundell (1963), cuyo caso estándar de economía pequeña se desarrolla en el tema A32, apartado 1, se aplica aquí a dos economías grandes.

Idea: el modelo ilustra de forma intuitiva los efectos que se transmiten por el canal de la renta y el canal monetario.

1.2.1. Supuestos

Entorno: dos economías abiertas y grandes, A y B, cuyas decisiones alteran las variables exteriores de la otra y, en particular, el tipo de interés mundial, i^*. El horizonte es el corto plazo, con precios fijos ($\bar{P}$), lo que tiene dos implicaciones:

▪ *Oferta horizontal*: la oferta de bienes es perfectamente elástica al nivel de precios dado y, por tanto, la renta queda determinada por la demanda agregada.

▪ *Inflación nula*: $\pi = 0$ y, por tanto, $i = r$. Además, el tipo de cambio nominal, E (unidades de moneda nacional por unidad de moneda extranjera, de modo que un aumento de E es una depreciación), y el real, $e = EP^*/P$, varían en la misma proporción.

Agentes: son neutrales al riesgo, de modo que no exigen prima de riesgo por mantener activos en distintas monedas, y mantienen expectativas estáticas sobre el tipo de cambio.

Mercados: el modelo incluye los de bienes (IS), dinero (LM) y sector exterior (BP). En el mercado financiero se comercian dinero y bonos, nacionales y extranjeros, que son sustitutivos perfectos. Con perfecta movilidad del capital se cumple la paridad descubierta de intereses, que, con expectativas estáticas, iguala en equilibrio el tipo de interés de cada país al mundial ($i = i^*$).

Sector exterior: la cuenta corriente se identifica con la balanza comercial ($\text{SB}_{\text{CC}} = \text{BC}$), sin rentas primarias ni secundarias ni cuenta de capital, y se cumple la condición Marshall-Lerner ($\varepsilon_X + |\varepsilon_M| > 1$)[5], de modo que una depreciación ($\uparrow E$) eleva las exportaciones netas.

[5]Se supone, además, que el país parte de una situación de equilibrio externo y las elasticidades de oferta de importaciones y exportaciones son infinitas.

1.2.2. Modelo

Planteamiento: las curvas IS, LM y BP de cada país son las del tema A32, apartado 1.2, con su misma notación. En el plano (Y, i), la IS refleja las combinaciones de renta y tipo de interés que equilibran el mercado de bienes y tiene pendiente negativa, porque un menor tipo de interés estimula la inversión. La LM refleja las que equilibran el mercado de dinero y tiene pendiente positiva, porque una renta mayor eleva la demanda de dinero y, con una oferta dada, el tipo de interés. La BP refleja las que equilibran la balanza de pagos y es horizontal al nivel de i^*, por la perfecta movilidad del capital.

La diferencia con el caso estándar es que, al ser ambos países grandes, el tipo de interés mundial, i^*, responde a sus políticas, y con él se desplaza la BP de cada país.

Estática comparativa (política monetaria expansiva): el BCE aumenta la oferta monetaria ($\Delta M^s > 0$). Gráficamente, en el plano (Y, i) de cada país (figura B19.1), ambos parten del punto 0, en el corte de las curvas IS, LM y BP.

▪ *Canal de la renta*:

- *Zona del euro (país A)*: el aumento de la masa monetaria produce un desplazamiento hacia la derecha de la curva LM, de LM_0 a LM_1, y una caída del tipo de interés nominal. De esta forma aumenta la renta nacional y se pasa al punto $0'$.

- *EE. UU. (país B)*: el aumento de la renta nacional en A eleva las exportaciones de B a A ($\Delta X > 0$, de IS_0 a $\text{IS}_{0'}$), lo que aumenta la renta nacional de B y, con ella, el tipo de interés (punto $0'$).

▪ *Canal monetario*: el desequilibrio entre el tipo de interés en la zona del euro y en EE. UU. ($i^A_{0'} < i^B_{0'}$) produce movimientos de capitales:

- *Zona del euro (país A)*: la libre movilidad de capitales y el menor tipo de interés en la zona del euro que en EE. UU. dan lugar a una salida de capitales con dos efectos:
  - *Tipo de interés*: la salida de capitales se traduce en un aumento de la demanda de bonos estadounidenses, lo cual hace subir su precio y bajar el tipo de interés extranjero (dado que el país es grande). El descenso del tipo de interés extranjero desplaza hacia abajo la BP de ambos países, de BP_0 a BP_1 ($\Delta i^* < 0$)[6].
  - *Tipo de cambio*: la salida de capitales produce una depreciación de la moneda que permite obtener ganancias en la competitividad-precio de las exportaciones. Así, se produce un desplazamiento hacia la derecha de la curva IS ($\Delta E > 0$), de IS_0 a IS_1, y la economía pasa al punto 1, con una renta Y_1.

- *EE. UU. (país B)*: la entrada de capitales provoca una apreciación de la moneda ($\Delta E < 0$), lo cual hace perder competitividad-precio a las exportaciones del país. En consecuencia, la IS se desplaza de $\text{IS}_{0'}$ a IS_1 y la renta de la economía cae hasta Y_1, inferior a la inicial (punto 1).

En resumen, una política monetaria expansiva con tipos de cambio flexibles, libre movilidad del capital e interdependencia produce un efecto empobrecimiento del vecino (*beggar-thy-neighbour*).

[6] En la ecuación de la BP del tema A32, apartado 1.2.3, $i = \frac{1}{\gamma}\left[m\,Y - x\,Y^* - \phi\,e\right] + i^*$, el tipo de interés mundial, que el caso estándar trata como exógeno, desplaza la curva. Aquí m es la propensión marginal a importar, x la propensión marginal a exportar, ϕ el efecto neto del tipo de cambio real e sobre la balanza comercial (positivo por la condición Marshall-Lerner) y γ el grado de movilidad del capital.

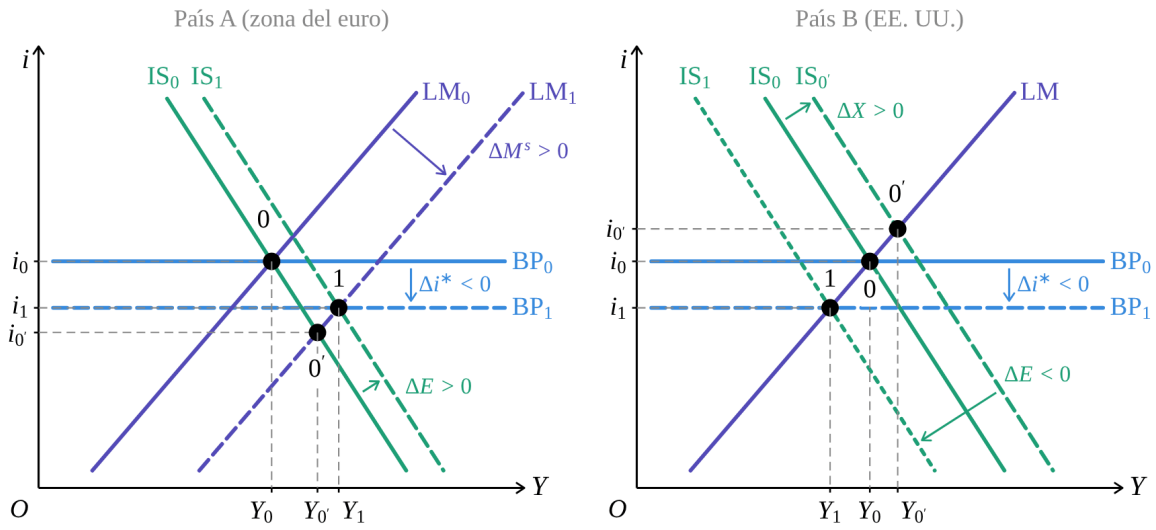

Figura B19.1: Efecto empobrecimiento del vecino de una política monetaria expansiva con dos países grandes.

- *Otros casos*:

- *A pequeño y B grande*: corresponde al caso estándar del modelo IS-LM-BP (tema A32, apartado 1.3.1). En este caso, el punto de cruce entre la IS y la BP queda más hacia la derecha, porque la curva BP no se desplaza, dado que el país no tiene capacidad para alterar el tipo de interés del resto del mundo. En consecuencia, el hecho de que el país A sea más grande hace que la política monetaria expansiva pierda efectividad.

- *A grande y B pequeño*: siguiendo la misma lógica, cuanto mayor sea el país A respecto de B, mayor será su capacidad para alterar el tipo de interés internacional y, con ella, mayor el efecto desbordamiento que genera. En el polo opuesto, si el país A es "enorme" con respecto al país B, determinará por sí solo el tipo de interés mundial y, por tanto, la efectividad de su política monetaria será equivalente a la de una economía cerrada.

Estática comparativa (política fiscal expansiva): si un país con margen presupuestario, como Alemania, aplica un estímulo fiscal, se da la situación opuesta. Gráficamente, en el plano (Y, i) de cada país (figura B19.2), con las curvas IS, LM y BP descritas en el Planteamiento, ambos parten de nuevo del punto 0.

- *Canal de la renta*:

- *Alemania (país A)*: el estímulo fiscal ($\Delta G_0 > 0$) supone un aumento de la demanda agregada y de la renta nacional (de IS_0 a $\text{IS}_{0'}$), lo cual hace subir el tipo de interés (punto $0'$).

- *Polonia (país B)*: se produce un aumento de la renta nacional debido al aumento de las exportaciones netas hacia Alemania, en la medida en que el aumento de la renta en dicho país estimula la compra de productos polacos ($\Delta X > 0$, de IS_0 a $\text{IS}_{0'}$, punto $0'$).

- *Canal monetario*:

- *Alemania*: con libre movilidad de capitales, el diferencial de tipos de interés provoca una entrada masiva de capitales en el país. La consiguiente apreciación del euro da lugar a una pérdida de competitividad-precio de las exportaciones, que desplaza hacia la izquierda la

curva IS ($\Delta E < 0$), de $\text{IS}_{0'}$ a IS_1, hasta el punto 1. Además, dado que el país es grande, la entrada masiva de capitales eleva el tipo de interés internacional ($\Delta i^* > 0$), lo que desplaza hacia arriba la BP, de BP_0 a BP_1.

- *Polonia*: se da el efecto inverso, es decir, una salida de capitales que deprecia la moneda. Así, mejora la competitividad-precio de las exportaciones ($\Delta E > 0$, de $\text{IS}_{0'}$ a IS_1), lo que acaba expandiendo la renta nacional hasta Y_1 (punto 1).

En resumen, con tipos de cambio flexibles, la política fiscal expansiva de un país grande expande la renta de ambos países, lo que se conoce como efecto locomotora (*fiscal spillover* o *locomotive policy*).

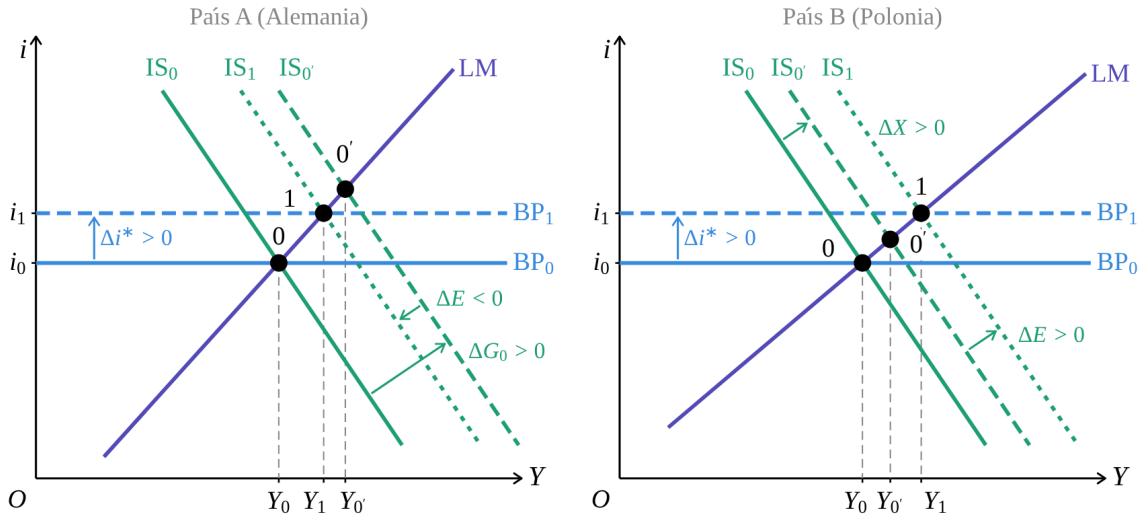

Figura B19.2: Efecto locomotora de una política fiscal expansiva con dos países grandes.

Si, en lugar de aplicar estímulos de manera autónoma, los países coordinan sus políticas y es uno con suficiente peso relativo el que lleva a cabo el estímulo fiscal, pueden existir ganancias mutuas de bienestar. Esta lógica tuvo su principal aplicación en la cumbre de Bonn de 1978, cuyos límites se examinan en el apartado 3.3.1.

▪ *Otros casos*:

- *A pequeño y B grande*: corresponde al caso estándar del modelo IS-LM-BP (tema A32, apartado 1.3.1), en el que la política fiscal con tipos de cambio flexibles y perfecta movilidad del capital no es efectiva. En consecuencia, el hecho de que el país A sea más grande hace que la política fiscal expansiva gane efectividad.

- *A grande y B pequeño*: cuanto mayor sea el tamaño de A en relación con B, mayor resultará la efectividad de la política para ambos países.

Extensiones: el signo y la intensidad del efecto desbordamiento dependen del régimen cambiario y del tamaño relativo de las economías. Por ejemplo, con tipos fijos, a diferencia de los casos anteriores, la expansión monetaria se transmite positivamente a través del canal de la renta.

1.2.3. Conclusiones

Aspectos positivos: el modelo muestra que, cuanto mayor es la interdependencia, mayor es la divergencia entre el óptimo nacional y el óptimo conjunto y, con ella, la ganancia potencial de la coordinación. Muestra además que la asimetría de tamaños importa, porque un país pequeño apenas genera desbordamientos, mientras que las decisiones de una economía grande condicionan al resto sin que esta internalice el coste. Este resultado es el argumento central a favor de mecanismos formales de coordinación.

Limitaciones: hereda las del modelo IS-LM-BP (tema A32, apartado 1.4). En particular, al suponer precios fijos y prescindir de las expectativas, ignora los canales de la inflación y de las expectativas, que se examinan a continuación.

1.3. Canal de la inflación: modelo de oferta y demanda agregadas

Origen: la evidencia empírica muestra que en muchos casos el efecto desbordamiento de la política monetaria expansiva de los países grandes sobre el resto del mundo es positivo (Dedola, Rivolta y Stracca, 2017, para la de EE. UU.). El resultado contradice el empobrecimiento del vecino que predice el modelo IS-LM-BP con tipos de cambio flexibles.

Idea: la explicación está, en parte, en el canal de la inflación, que el modelo anterior ignora. Al atenuar el desvío del gasto, este canal deja prevalecer el canal de la renta y el financiero (menores tipos de interés globales).

> Para ilustrarlo se recurre al modelo de oferta y demanda agregadas en economía abierta, desarrollado en el tema A32 §2, en su representación de dos países, empleada en el tema B16 §1.2.1 para un *shock* asimétrico de demanda en una unión monetaria. Aquí se aplica a una expansión monetaria con tipos de cambio flexibles.

Planteamiento: se consideran dos economías abiertas de tamaño comparable, con tipos de cambio flexibles. El país A adopta la política monetaria expansiva y el país B sufre el efecto desbordamiento.

Desarrollo gráfico: en el plano (Y, P) de cada país (figura B19.3), el panel izquierdo representa el país A y el derecho, el país B. En cada panel se distinguen:

1. *Demanda agregada* (DA_0): refleja las combinaciones de renta y nivel de precios que equilibran a la vez los mercados de bienes y de dinero. Tiene pendiente negativa porque un mayor nivel de precios reduce la oferta monetaria real, lo que eleva el tipo de interés y contrae la inversión, y porque el mayor nivel de precios aprecia el tipo de cambio real, lo que deteriora la balanza comercial.

2. *Oferta agregada a corto plazo* ($\mathrm{OA}_{\mathrm{CP}}$): refleja la producción que ofrecen las empresas a cada nivel de precios con los salarios nominales dados. Tiene pendiente positiva por la rigidez de esos salarios, pues un mayor nivel de precios reduce el salario real y estimula la contratación.

3. *Oferta agregada a largo plazo* ($\mathrm{OA}_{\mathrm{LP}}$): refleja la producción una vez que los salarios nominales se han ajustado a los precios. Tiene pendiente positiva y mayor que la de corto plazo, pero no es vertical, porque en economía abierta el deflactor y el IPC difieren (tema A32, apartado 2.2.4).

4. *Equilibrio*: ambas economías parten del punto 0, donde se cortan las tres curvas, con un tipo de cambio real dado que determina su competitividad relativa.

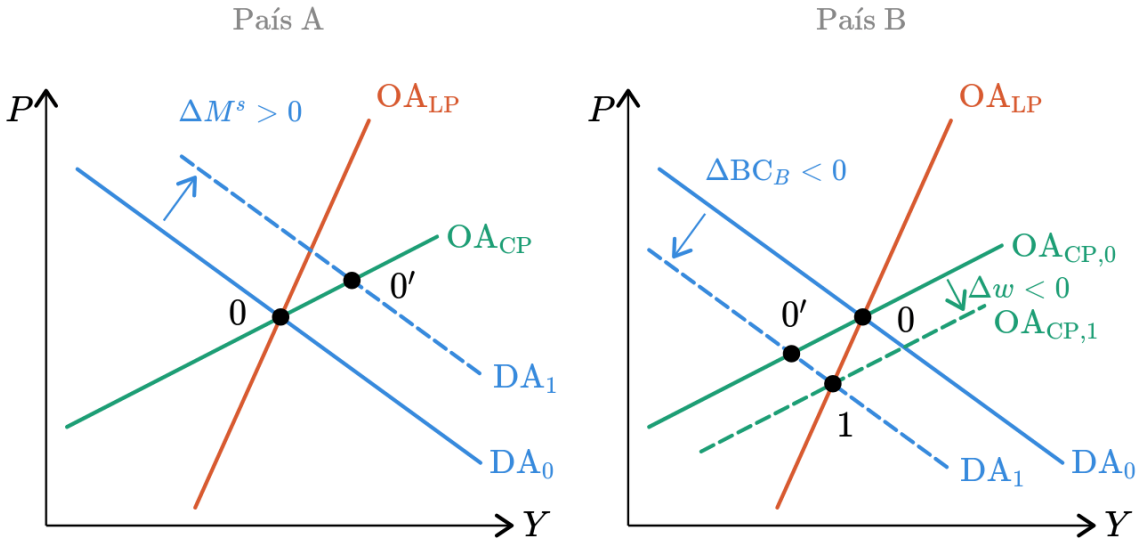

Figura B19.3: Canal de la inflación con dos países. La expansión monetaria de A contrae la demanda de B (punto $0'$), y la caída de precios y salarios en B desplaza su oferta a corto plazo hacia la derecha (punto 1).

Estática comparativa: la perturbación se propaga en cuatro momentos sucesivos:

1. *Impulso en A*: la expansión monetaria ($\Delta M^s > 0$) desplaza su demanda agregada de DA_0 a DA_1, con lo que aumentan su producción y sus precios (punto $0'$), y deprecia su moneda.

2. *Efecto desbordamiento sobre B*: la apreciación de su moneda le hace perder competitividad-precio, por lo que sus exportaciones netas caen ($\Delta \mathrm{BC}_B < 0$) y su demanda agregada se desplaza de DA_0 a DA_1, con menor producción y menores precios (punto $0'$). Este es el efecto de empobrecimiento del vecino que predice el modelo IS-LM-BP, y en él se detiene aquel modelo.

3. *Ajuste vía precios en B*: la caída del nivel de precios en B supone, para un tipo de cambio nominal dado, una depreciación de su tipo de cambio real, es decir, una recuperación parcial de la competitividad que acababa de perder. La pérdida inicial queda así amortiguada, y tanto más cuanto más flexibles sean los precios.

4. *Traslación a salarios*: si la caída de los precios se traslada además a los salarios nominales ($\Delta w < 0$), se reducen los costes laborales unitarios y la oferta agregada a corto plazo de B se desplaza de $\mathrm{OA}_{\mathrm{CP},0}$ a $\mathrm{OA}_{\mathrm{CP},1}$ (punto 1). La apreciación nominal refuerza ese desplazamiento, porque reduce el IPC esperado (tema A32, apartado 2.3.1), y la recuperación de la competitividad puede llegar así a compensar íntegramente el impulso negativo inicial.

Conclusiones:

▪ *Aspectos positivos*: el modelo muestra que el canal de la inflación opera en sentido contrario al canal monetario y atenúa el empobrecimiento del vecino, que puede llegar a neutralizar. Para que el desbordamiento sea positivo, como a menudo se observa, han de dominar además el canal de la renta y el financiero, que el modelo no recoge. Que domine el empobrecimiento del vecino o el canal de la inflación depende del grado de rigidez nominal de la economía receptora. Así, con

precios y salarios muy rígidos prevalece el empobrecimiento del vecino, mientras que con precios y salarios flexibles el ajuste por la vía de los precios relativos neutraliza buena parte del impulso adverso. De ahí que la conveniencia de coordinar la política monetaria no pueda establecerse en abstracto, sino que dependa de las características estructurales de los países implicados.

- *Limitaciones*: hereda las del modelo de oferta y demanda agregadas (tema A32, apartado 2.5), en particular que determina el nivel de precios y no la inflación y que trata la política monetaria como un control de la cantidad de dinero.

1.4. Canal de las expectativas: la fragilidad de una unión monetaria incompleta (De Grauwe, 2011)

Origen: el último canal opera a través de las expectativas de los agentes, que pueden por sí solas desencadenar una crisis sin que cambien los fundamentos, siempre que estos se sitúen en un tramo intermedio. Su ilustración más clara es el modelo de equilibrios múltiples de Paul De Grauwe (2011). La lógica de De Grauwe se remonta a los modelos de crisis autocumplidas de Obstfeld (1986) para las crisis cambiarias y de Calvo (1988) para la deuda soberana.

> El modelo se desarrolla íntegramente, con sus supuestos, su representación gráfica, sus limitaciones y las soluciones a las uniones monetarias incompletas, en el tema B16 §3.1.

Idea: en una unión monetaria sin unión fiscal, los Estados emiten deuda en una moneda sobre la que no tienen control, de modo que no pueden monetizarla y un problema de liquidez puede transformarse en uno de solvencia.

Planteamiento: el Gobierno de un país que sufre un *shock* decide si impaga comparando el coste reputacional del impago con sus beneficios, que dependen de que los inversores anticipen o no ese impago. De esa comparación se obtienen tres tramos de intensidad del *shock*.

Desarrollo gráfico: en el plano intensidad del *shock*-costes y beneficios del impago (figura B19.4), la curva C_D refleja ese coste, la exclusión de los mercados de financiación, y es horizontal porque se supone fijo. Las curvas de beneficios reflejan el ahorro en el servicio de la deuda y en el ajuste fiscal que permite el impago, y crecen con la intensidad del *shock*. Son mayores si los inversores anticipan el impago (B^{NC}), porque su desconfianza encarece la financiación, que si confían (B^{C}). Los cortes de ambas curvas de beneficios con la de costes, en S_1 y S_2, delimitan los tres tramos:

- *Cielo* (*heaven on Earth*, $S < S_1$): ante perturbaciones leves, el equilibrio sin impago es sostenible con independencia de lo que esperen los inversores (p. ej. Alemania y los Países Bajos durante la crisis de deuda soberana).

- *Infierno* (*hell on Earth*, $S > S_2$): ante perturbaciones graves, el impago resulta inevitable, también con independencia de las expectativas.

- *Purgatorio* (*purgatory*, $S_1 < S < S_2$): en un tramo intermedio existen **equilibrios múltiples**, de manera que el resultado depende enteramente de las expectativas de los inversores y se producen profecías autocumplidas (p. ej. España e Italia durante la misma crisis). Para una intensidad S', el punto N es el equilibrio con confianza y el punto D, el de impago.

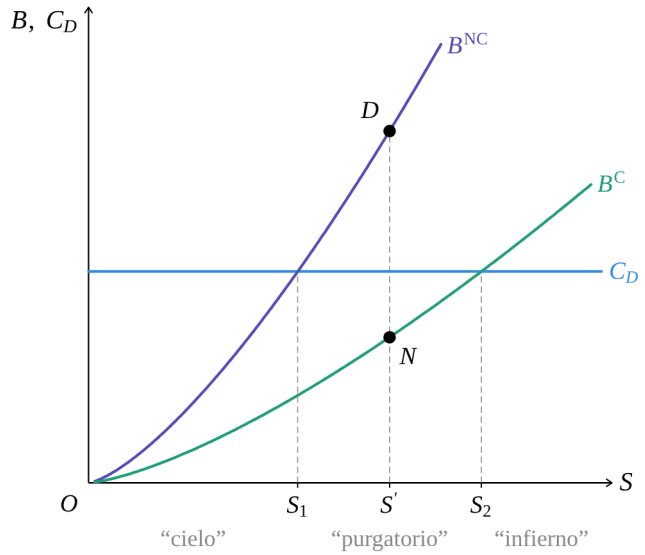

Figura B19.4: Canal de las expectativas (De Grauwe, 2011, a partir de Obstfeld, 1986). Ante un *shock* de solvencia hay equilibrios múltiples, N (confianza, sin impago) y D (profecía autocumplida, *default*).

Conclusiones:

▪ *Aspectos positivos*: el modelo muestra que el canal de las expectativas convierte la falta de coordinación fiscal en una fuente autónoma de inestabilidad, pues la fragilidad no procede del nivel de deuda de un país concreto, sino del diseño institucional del conjunto. Por tanto, su corrección no puede ser nacional, sino que exige mecanismos comunes, es decir, un prestamista de última instancia creíble en el mercado de deuda soberana y reglas fiscales que contengan el riesgo moral que aquel introduce.

▪ *Limitaciones*: supone un coste del impago fijo y ninguna alternativa de consolidación fiscal, y no modeliza el contagio entre los miembros de la unión (tema B16, apartado 3.1.3).

1.5. Conclusiones

La interdependencia genera efectos desbordamiento que se propagan por cuatro canales de signo distinto. El de la renta transmite el impulso de forma positiva, mientras que el monetario puede invertirlo (empobrecimiento del vecino), según el régimen cambiario y el tamaño relativo de las economías. El de la inflación opera en sentido contrario al monetario, y el de las expectativas introduce una inestabilidad que depende del diseño institucional del conjunto y no solo de los fundamentos de cada país.

El impacto efectivo de estos efectos depende de la naturaleza de la perturbación y del grado de apertura de la economía, junto con las variables asociadas a ella (elasticidad de las importaciones y exportaciones, competitividad-precio, estructura exportadora…). Depende asimismo del grado

de desarrollo y estabilidad de los mercados financieros, del grado de rigidez nominal y de la credibilidad del marco institucional[7].

De todo ello se desprende que la existencia de efectos desbordamiento es condición necesaria para que la coordinación resulte ventajosa, pero el análisis realizado hasta aquí ha supuesto implícitamente que los países no reaccionan ante las decisiones ajenas.

[7] Influyen también la posición cíclica de cada país y su margen de maniobra. Un Gobierno muy endeudado no puede recurrir a la política fiscal, y un banco central con objetivo cambiario ve limitada su política monetaria.

2. Justificación teórica de la coordinación

Queda por ver si esa condición necesaria es también suficiente. Para ello se considera de manera explícita la reacción de los países desde dos perspectivas complementarias, la teoría de juegos y la de los **bienes públicos globales**, y se examinan después los límites de la propia coordinación.

2.1. Perspectiva de la teoría de juegos

Se comienza por los juegos estáticos, en los que los jugadores deciden una sola vez y de forma simultánea, primero con estrategias discretas (dilema del prisionero) y después con estrategias continuas (modelo de Corden), y les siguen los juegos repetidos y los cooperativos.

> Los conceptos de solución que se aplican (equilibrio en estrategias dominantes, equilibrio de Nash, equilibrio de Nash perfecto en subjuegos, núcleo y valor de Shapley) se desarrollan en el tema A14, apartado 1.4, y el dilema del prisionero, en su apartado 2.1.1.

2.1.1. Juegos estáticos con estrategias discretas: el dilema del prisionero (Tucker, 1950)

Planteamiento: el dilema del prisionero, formulado por Albert Tucker (1950), es el marco de referencia para analizar por qué la coordinación fracasa. En esta aplicación, dos países sufren un *shock* y pueden escoger entre política fiscal restrictiva y política fiscal expansiva (laxa). Los pagos de la tabla B19.1 representan ganancias de bienestar, que pueden medirse como reducciones del índice de miseria de Okun[8], es decir, cuanto mayor es el pago, menor es el índice.

Tabla B19.1: Matriz de pagos del dilema del prisionero en la coordinación de la política fiscal.

	PAÍS 2	
PAÍS 1	Política fiscal laxa	Política fiscal restrictiva
Política fiscal laxa	A: (2,2)	B: (0,3)
Política fiscal restrictiva	C: (3,0)	D: (1,1)

Equilibrio en estrategias dominantes (EED): el país 1 elegirá siempre una política fiscal restrictiva, su *estrategia dominante*, porque le reporta una ganancia mayor sea cual sea la política del país 2, y, por simetría, el país 2 hará lo propio:

- *País 2 laxo*: el país 1 obtiene una mayor ganancia con una política fiscal restrictiva (3 frente a 2).

- *País 2 restrictivo*: el país 1 obtiene igualmente una mayor ganancia con una política fiscal restrictiva (1 frente a 0).

Equilibrio de Nash (EN): todo EED es un EN, y aquí es el único. La celda D es, por tanto, el único perfil en el que la estrategia de cada país es *mejor respuesta* a la del otro, de modo que ninguno tiene incentivos a modificar unilateralmente su decisión.

[8] Suma de la tasa de desempleo y la tasa de inflación (IPC). Así, un paro del 5 % y una inflación del 1 % dan un índice del 6 %.

Bienestar: existe, sin embargo, una situación Pareto-superior al equilibrio de Nash. Si ambos países eligen una política fiscal laxa (es decir, celda A), ambos obtienen pagos superiores a los del equilibrio de Nash. Por ello, cuando los países actúan de acuerdo con su interés individual, se produce una pérdida de eficiencia.

Conclusiones: conviene precisar que no toda situación de coordinación fallida responde a esta estructura. En el dilema del prisionero el obstáculo es el incentivo a desviarse de un acuerdo mutuamente ventajoso, de modo que la solución pasa por hacer creíble el castigo. En otros juegos, en cambio, como la batalla de los sexos (tema A14, apartado 2.1.5), existen varios equilibrios y el problema no es disuadir la deserción, sino seleccionar entre ellos. Ello exige un mecanismo de comunicación o un liderazgo que actúe como punto focal antes que un régimen sancionador.

2.1.2. Juegos estáticos con estrategias continuas: el modelo de Corden (1994)

Origen: el dilema anterior puede formalizarse con estrategias continuas, en lugar de una elección binaria, mediante el modelo de Corden (1994), que desarrolla el enfoque de Oudiz y Sachs (1984) y cuya representación gráfica se conoce como diagrama de Hamada (1976).

Idea: el modelo permite representar gráficamente tanto el coste de la no coordinación como el reparto de las ganancias de coordinar.

Supuestos:

- *Entorno*: dos países, A y B, en un juego estático y no repetido.

- *Agentes*: cada país se enfrenta a una disyuntiva entre inflación y desempleo y minimiza una función de pérdidas definida sobre ambas variables. Se supone que B presenta una mayor aversión a la inflación que A, que prioriza el empleo. Ambos países comparten un mismo tipo de cambio compatible con sus respectivos objetivos internos, $E_{A/B,0}$.

- *Restricciones*: en cada país, tanto la depreciación de su moneda como el aumento de su oferta monetaria elevan la inflación y reducen el desempleo.

- *Sector exterior*: rigen la paridad descubierta de intereses y el enfoque monetario, según el cual el tipo de cambio $E_{A/B}$ (unidades de la moneda de A por unidad de la de B) es función del cociente entre ambas ofertas monetarias, $E_{A/B} = f(M^{s,A}/M^{s,B})$. La moneda de A se deprecia cuando ese cociente aumenta.

Planteamiento: cada país elige su oferta monetaria, $M^{s,A}$ o $M^{s,B}$, para minimizar su función de pérdidas, dada la del otro. Como el tipo de cambio depende de ambas ofertas monetarias, la elección de cada país altera la inflación y el desempleo del otro.

Desarrollo gráfico: la interacción entre los dos países se representa en el plano de ofertas monetarias $M^{s,A}$-$M^{s,B}$ (figura B19.5):

1. *Rayos vectores*: cada recta que parte del origen representa un tipo de cambio, determinado por el cociente entre ambas ofertas monetarias. Cuanto mayor es su pendiente, es decir, cuanto mayor es $M^{s,B}$ respecto de $M^{s,A}$, más apreciada está la moneda de A.

2. *Puntos de saturación*: el punto a es la combinación que maximiza la utilidad de A, con ofertas monetarias holgadas en los dos países, que reducen el paro al precio de una inflación mayor. El punto b es el óptimo de B, con una política más restrictiva. Los dos quedan sobre un mismo rayo vector, el de $E_{A/B,0}$, porque ambos países comparten el objetivo cambiario.

3. *Curvas de indiferencia* (CI^A y CI^B): cada una reúne los pares de ofertas monetarias que dan a un país la misma utilidad. Son circunferencias concéntricas en torno a su punto de saturación y representan menor utilidad cuanto más se alejan de él.

4. *Curvas de reacción* (R^A y R^B): dado que las curvas de indiferencia son circulares, a A le conviene la misma oferta monetaria $M^{s',A}$ con independencia de lo que haga B. Por ello, la *función de reacción* de A (o *función de mejor respuesta*, tema A14, apartado 1.3.3) es una recta vertical en el plano $(M^{s,A}, M^{s,B})$ y, por simetría, la de B es una recta horizontal en $M^{s',B}$. Cada país tiene, por tanto, una estrategia dominante, exactamente como en el dilema del prisionero.

5. *Equilibrio*: de lo anterior se derivan dos soluciones posibles.

- *Equilibrio no cooperativo*: corresponde al cruce de las dos funciones de reacción, el punto 0 (N en la figura). Como cada país responde de la mejor forma posible a lo que hace el otro, se trata de un equilibrio de Nash, y es estable. Deja, sin embargo, margen para que los dos mejoren a la vez, porque el tipo de cambio que resulta, $E_{A/B,N}$, se aparta del objetivo común $E_{A/B,0}$. No es, por tanto, un óptimo de Pareto.

- *Solución cooperativa*: las curvas de indiferencia trazadas por el punto 0 encierran una lente, y cualquier combinación situada dentro de ella acerca a los dos países a sus puntos de saturación, con lo que ambos ganan. El óptimo está en el *núcleo*, la parte de la *curva de contratos* que cae dentro de esa lente. La curva de contratos, formada por los pares de ofertas monetarias Pareto-eficientes, es el segmento recto que une a y b, pues las curvas de indiferencia son concéntricas. Esa misma propiedad, junto con el objetivo cambiario común, hace que el bienestar conjunto alcance su máximo precisamente en $E_{A/B,0}$. Qué punto del núcleo se alcance, p. ej. el punto 1, con ofertas monetarias $M^{s*,A}$ y $M^{s*,B}$, depende del poder negociador de cada parte.

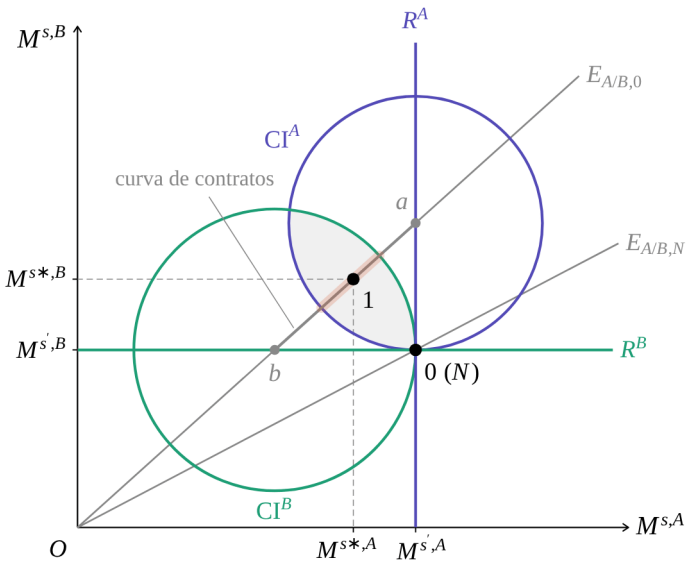

Figura B19.5: Modelo de Corden (diagrama de Hamada). Las curvas de reacción se cruzan en el equilibrio de Nash (punto 0, N). La solución cooperativa (punto 1) se sitúa en el núcleo (tramo sombreado de la curva de contratos ab) y mejora el bienestar de ambos.

Extensiones: en una lectura fiscal, si se sustituyen las ofertas monetarias por el tono de la política fiscal, el equilibrio no cooperativo resulta excesivamente restrictivo, pues cada país prefiere recibir el impulso de la demanda de exportaciones del socio sin soportar el déficit propio. Esta lectura formaliza el efecto locomotora (apartado 1.2.2) y la lógica de la cumbre de Londres de 2009, cuando EE. UU. impulsó un estímulo fiscal concertado para acercarse a la solución de coordinación[9].

Conclusiones:

▪ *Aspectos positivos*: el modelo muestra que la pérdida de bienestar no procede de un error de política, sino de la propia interacción estratégica, pues cada país elige racionalmente y, precisamente por ello, ambos acaban en una posición peor de la alcanzable. Muestra además que la ganancia de la coordinación no es un punto sino un rango de reparto, lo que explica que las negociaciones internacionales versen tanto sobre cuánto se gana como sobre cómo se distribuye lo ganado.

▪ *Limitaciones*: la solución cooperativa no es un equilibrio de Nash y, por tanto, no es estable. En efecto, si A coopera situándose en $M^{s*,A}$ (su oferta monetaria en la solución cooperativa, punto 1), B obtiene mayor utilidad desviándose hacia $M^{s',B}$. Así, el modelo identifica como problema decisivo que no basta con demostrar que coordinar es ventajoso, sino que es preciso un mecanismo que haga cumplir el acuerdo. A ello se añaden el carácter estático del análisis y el supuesto de un objetivo cambiario común, que elimina por construcción el conflicto de intereses sobre el nivel deseado del tipo de cambio.

2.1.3. Juegos dinámicos

La inestabilidad de la solución cooperativa puede resolverse si se abandona el supuesto de juego único y la relación se repite, de modo que la desviación de hoy pueda castigarse mañana. Ello hace posible la **estrategia del gatillo** (*trigger strategy*), en la que cada país coopera mientras el otro coopere y, ante una sola desviación, revierte para siempre al equilibrio no cooperativo.

> Los juegos repetidos, tanto en horizonte finito como infinito, y la estrategia del gatillo se desarrollan en el tema A14 §2.3.2 y §2.3.3. Aquí se recogen sus dos resultados, con la misma notación, y su traducción a la coordinación internacional.

Equilibrio de Nash perfecto en subjuegos (horizonte finito): cuando el número de repeticiones es conocido y el juego de etapa tiene un único equilibrio de Nash, como el dilema del prisionero, la cooperación se desmorona por inducción hacia atrás y los acuerdos no se sostienen. El razonamiento parte del último periodo, en el que no existe futuro ni, por tanto, castigo creíble, y se extiende a todos los anteriores. Por ello, el único equilibrio de Nash perfecto en subjuegos consiste en T repeticiones del equilibrio de Nash del juego de etapa.

La implicación directa para la coordinación internacional es que los acuerdos concebidos como respuesta puntual a una crisis, con una fecha de terminación conocida (p. ej. la propia cumbre de Londres de 2009), son estructuralmente frágiles. La razón es que cada participante anticipa el final del compromiso y, con él, la desaparición del castigo. De ahí un argumento de peso a favor de las instituciones permanentes frente a las cumbres *ad hoc*.

[9] Una visión alternativa asocia la cooperación a la disciplina fiscal y no al estímulo. En los juegos de disciplina, los déficits de cada país compiten por el ahorro internacional, elevan los tipos de interés y perjudican a todos, lógica próxima a la posición alemana en la zona del euro.

Equilibrio de Nash perfecto en subjuegos (horizonte infinito): si el juego se repite un número indeterminado de veces[10], la cooperación puede sostenerse (Friedman, 1971), porque los agentes pueden preferir la corriente de ganancias futuras de cumplir frente a la ganancia inmediata de incumplir. La estrategia del gatillo sostiene la cooperación como equilibrio de Nash perfecto en subjuegos si el factor de descuento, β, supera un umbral:

$$\beta \geq \frac{\pi_{nc,c} - \pi_{c,c}}{\pi_{nc,c} - \pi_{nc,nc}}$$

donde $\pi_{c,c}$ es el pago de cooperación mutua, $\pi_{nc,c}$ el de traicionar cuando el otro coopera y $\pi_{nc,nc}$ el del equilibrio no cooperativo. El numerador recoge la tentación de desviarse y el denominador suma a esa tentación la severidad del castigo, $\pi_{c,c} - \pi_{nc,nc}$. La cooperación es tanto más sostenible cuanto menor sea la ganancia de la traición, mayor la pérdida asociada a la represalia y más valoren los agentes el futuro.

Ejemplo: en el contexto de la coordinación internacional, la ruptura del acuerdo tácito puede dar lugar a guerras de divisas (temas B15 y B20, apartado 1.2.2). Tras la crisis financiera global, la expansión cuantitativa (*quantitative easing*, QE) de la Reserva Federal (2008-2014) y las políticas excepcionalmente expansivas del Banco de Japón (*Abenomics*[11]) y del BCE (ampliación del programa de compras de activos, 2015) depreciaron sus monedas. La depreciación provocó la disconformidad de los socios comerciales. Brasil, por ejemplo, calificó esas políticas de empobrecimiento del vecino y Dilma Rousseff se refirió en 2012 a la expansión monetaria de los países desarrollados como un "tsunami monetario".

Conclusiones: de los dos resultados se derivan los cuatro requisitos que debe satisfacer todo mecanismo de coordinación:

1. *Horizonte indefinido*: los participantes no deben conocer cuándo termina la relación.

2. *Supervisión*: debe poder detectarse la desviación, sin lo cual el castigo no puede activarse.

3. *Represalia creíble*: se excluyen las amenazas vacías y, en la práctica, también el castigo eterno, que ambas partes tendrían interés en renegociar, de modo que se recurre a una fase de castigo de duración limitada (el *castigo óptimo* del tema A14, apartado 2.3.3).

4. *Paciencia*: los participantes no deben descontar en exceso el futuro, condición que comprometen los ciclos políticos cortos.

2.1.4. Juegos cooperativos: el reparto de las ganancias

Los apartados anteriores responden a si los países cooperan. Queda por responder cómo se reparten las ganancias, cuestión que el modelo de Corden deja abierta al remitirla al poder negociador de las partes y de la que se ocupa la teoría de juegos cooperativos, que presupone la existencia de acuerdos vinculantes.

[10] Admite dos lecturas. En la primera, el juego se repite infinitos periodos y los pagos se descuentan con $\beta = 1/(1+\rho)$, $0 < \beta < 1$, siendo ρ la tasa de preferencia temporal. En la segunda, su duración es desconocida y en cada etapa continúa con probabilidad $0 < p < 1$, que actúa como un factor de descuento adicional.

[11] El término *Abenomics* se refiere a las políticas económicas propulsadas por el primer ministro de Japón, Shinzo Abe, desde las elecciones generales de 2012. La *Abenomics* se basa en "tres flechas" de expansión monetaria, estímulo fiscal y reformas estructurales.

El núcleo y el valor de Shapley se desarrollan en el tema A14, apartados 1.4.2 y 2.4. Aquí se recoge su aplicación a la coordinación.

Núcleo (Edgeworth, 1881; Gillies, 1959): conjunto de repartos eficientes que ninguna coalición de participantes puede bloquear. El modelo de Corden recurre a él al situar la solución cooperativa en el tramo de la curva de contratos interior a la lente. La limitación del núcleo es que suele ser amplio y no señala un punto concreto.

Valor de Shapley (Shapley, 1953): frente a esa indeterminación, ofrece un reparto único, que asigna a cada país una parte de la ganancia igual a su contribución marginal media al resultado conjunto sobre todos los órdenes posibles de incorporación a la coalición.

Conclusiones:

- *Aspectos positivos*: el enfoque ofrece un criterio de reparto equitativo al que apelar en la negociación y, con el valor de Shapley, concreta el reparto que el modelo de Corden deja al poder negociador de las partes.

- *Limitaciones*: su interés es más normativo que descriptivo, porque su supuesto central, el carácter vinculante de los acuerdos, no se satisface en un sistema de Estados soberanos, en el que no existe autoridad capaz de imponer su cumplimiento. Esa inaplicabilidad directa explica por qué la coordinación internacional descansa en el diseño institucional del apartado 3.2.

2.2. Perspectiva de los bienes públicos globales

Frente a la interacción estratégica entre dos autoridades, esta perspectiva atiende a la provisión de un recurso compartido por muchos países, los bienes públicos globales mencionados en el apartado 1.1. Como en la interacción estratégica, la actuación racional individual impide alcanzar el óptimo colectivo, pero el remedio que sugiere esta perspectiva es distinto, pues no descansa en el castigo sino en el diseño de las reglas de uso.

2.2.1. La tragedia de los comunes y los límites de las soluciones tradicionales (Hardin, 1968)

Origen: la formulación clásica del problema es la *tragedia de los comunes*, propuesta por el ecólogo Garrett Hardin en *The Tragedy of the Commons* (*Science*, 1968). Se refiere a los problemas medioambientales que surgen del uso excesivo de recursos escasos de libre acceso (p. ej. caladeros de pesca o pastos para la ganadería).

Idea: ante un bien no excluible, rival y con una capacidad de regeneración limitada, un agente racional que persigue su propio interés puede terminar haciendo un uso por encima del socialmente óptimo, generando así una externalidad negativa.

Razonamiento: en el ejemplo de Hardin, un pasto comunal, la falta de coordinación agota inexorablemente el recurso, es decir, se alcanza un equilibrio de Nash que no es socialmente óptimo.

Extensiones:

▪ *Aplicación a la coordinación macroeconómica*: la analogía es estrecha. La estabilidad cambiaria, la disciplina fiscal en una unión monetaria o la integridad del sistema comercial multilateral son bienes públicos. Cada país tiene, sin embargo, incentivos a hacer de ellos un uso excesivo, como si fueran un recurso común (devaluar competitivamente, endeudarse en exceso, elevar aranceles), y el resultado es la degradación del bien del que todos dependen.

▪ *Límites de las soluciones tradicionales*: las propuestas convencionales se han agrupado en dos campos, ambos con graves dificultades en el plano internacional y ninguno basado en el acuerdo voluntario entre los propios afectados, el vacío que aborda Ostrom:

- *Asignación de derechos de propiedad*: la principal dificultad reside en que el bien en cuestión suele ser demasiado extenso y, por tanto, resulta complicado definir el reparto de los derechos. En el ámbito internacional se añade que no existe una autoridad con competencia para asignarlos.

- *Intervención pública*: supondría la regulación del acceso y el establecimiento de un sistema de supervisión. La objeción es aquí más radical, pues la intervención presupone la existencia de un poder coercitivo supranacional, que no existe en un sistema de Estados soberanos.

Conclusiones:

▪ *Aspectos positivos*: condensa en un ejemplo sencillo la divergencia entre la racionalidad individual y el resultado colectivo en el uso de los recursos comunes, y se ha convertido en una referencia habitual para analizar los problemas de acción colectiva a escala global, como el clima.

▪ *Limitaciones*: supone usuarios aislados, sin comunicación ni reglas de uso, y equipara el libre acceso con la propiedad comunal. De hecho, la evidencia reunida por Ostrom sobre comunidades que gestionan sus recursos comunes durante largos periodos muestra que la tragedia no es inexorable.

2.2.2. La solución de Ostrom (1990)

Origen: Elinor Ostrom (*Governing the Commons*, 1990), premio Nobel en 2009, responde a Hardin con la evidencia de comunidades que gestionan sus recursos comunes durante largos periodos sin privatizarlos ni someterlos a una autoridad coercitiva.

Idea: el problema puede resolverse mediante acuerdos libres alcanzados de forma voluntaria entre los usuarios directos de los recursos comunes, siempre que dichos acuerdos cumplan ocho principios de diseño.

Razonamiento: los ocho principios son los siguientes:

1. *Delimitación*: definición de fronteras claras para la comunidad.

2. *Proporcionalidad*: debe existir una equivalencia proporcional entre beneficios y costes.

3. *Elección colectiva*: quienes fijan y modifican las reglas son los propios afectados.

4. *Supervisión*: debe existir una vigilancia continua del cumplimiento de las reglas.

5. *Sanciones graduales*: su aplicación depende de la gravedad del incumplimiento.

6. *Celeridad*: mecanismos rápidos y baratos de resolución de conflictos.

7. *Autonomía*: reconocimiento, por parte de otras autoridades, del derecho de la comunidad a organizarse.

8. *Gobernanza policéntrica* (*polycentric governance*): relaciones adecuadas con otros niveles de los organismos de decisión.

Estos principios son la traducción institucional de los requisitos del modelo de Friedman, de modo que ambas perspectivas convergen. Así, la supervisión continua cumple el requisito de detectar la desviación, las sanciones graduales sustituyen el castigo eterno por una respuesta proporcionada, y la elección colectiva y la autonomía dotan al acuerdo de la legitimidad que sostiene su cumplimiento voluntario.

Conclusiones:

- *Aspectos positivos*: describe con precisión el modo en que operan los foros examinados en el apartado 3, que combinan la deliberación entre iguales con mecanismos de evaluación mutua (*peer review*) antes que con sanciones formales.

- *Limitaciones*: sus principios proceden de comunidades pequeñas, cuyos usuarios se conocen e interactúan de forma repetida, de modo que su traslado a recursos comunes globales, compartidos por muchos Estados heterogéneos, no es inmediato.

2.3. Los límites de la coordinación

El análisis anterior podría sugerir que la coordinación es siempre preferible a la actuación autónoma y que el único obstáculo es hacerla sostenible. La literatura ha identificado, sin embargo, costes de la coordinación y circunstancias en las que resulta contraproducente.

Coordinación contraproducente (Rogoff, 1985): en presencia de inconsistencia dinámica, el comportamiento no cooperativo puede resultar preferible. La razón es que, actuando por separado, cada banco central se ve disciplinado por el temor a que una expansión monetaria deprecie su moneda y eleve los precios de las importaciones. En cambio, si todos se coordinan para expandir simultáneamente, esa restricción cambiaria desaparece, con lo que se refuerza el sesgo inflacionista analizado en el modelo de Barro-Gordon (véase B16 §2.1.2). La ausencia de coordinación operaba, en este caso, como un mecanismo de compromiso, y su eliminación empeora el resultado.

Ganancias reducidas (Oudiz y Sachs, 1984): las estimaciones cuantitativas de los beneficios de la coordinación entre EE. UU., Japón y Alemania arrojan magnitudes inferiores al medio punto porcentual del PIB en EE. UU. y Alemania y en torno a un punto en Japón. Las cifras quedan muy por debajo de la importancia que el debate político atribuye a estos acuerdos y contrastan con el coste político y administrativo de negociarlos y sostenerlos.

Incertidumbre sobre el modelo (Frankel y Rockett, 1988): la coordinación exige que los Gobiernos compartan un diagnóstico común sobre cómo funciona la economía y sobre la magnitud de los efectos desbordamiento. Cuando cada uno opera con un modelo distinto, y ninguno es necesariamente el correcto, un acuerdo negociado sobre esa base puede dejar a los participantes en peor situación que la actuación autónoma, y el problema no es entonces de voluntad política, sino de conocimiento.

Otros costes: a lo anterior se añade la pérdida de autonomía en la fijación de los objetivos internos, que puede resultar especialmente costosa cuando las perturbaciones son asimétricas. Existe también el riesgo de que el compromiso internacional se emplee para eludir la responsabilidad ante el electorado nacional, atribuyendo a la coordinación decisiones impopulares.

2.4. Conclusiones

Ambas perspectivas responden a la pregunta que dejaba abierta el apartado 1. La existencia de efectos desbordamiento es condición necesaria pero no suficiente para que se produzca la coordinación. De ahí que la coordinación no surja espontáneamente del reconocimiento del beneficio mutuo, sino que requiera un diseño institucional que la sostenga. Los requisitos que identifican los modelos anteriores explican la forma que adoptan las instituciones examinadas en el apartado 3.2.

Conviene evitar, en todo caso, la conclusión de que más coordinación es siempre mejor, porque su valor depende de la naturaleza del problema. Así, el valor de la coordinación resulta indiscutible ante bienes públicos globales con fuertes externalidades (el sistema comercial multilateral, la estabilidad financiera, el clima) y mucho más discutible en la gestión coyuntural de la demanda, donde las ganancias medidas son reducidas y la incertidumbre sobre el modelo es elevada.

3. Aspectos prácticos de la coordinación internacional

Mostradas las ganancias de bienestar que puede reportar la coordinación, quedan sus aspectos prácticos. La sección examina las características de los acuerdos, el G20, la OCDE y otros foros, la experiencia acumulada y los obstáculos del proceso.

3.1. Características

Los acuerdos de coordinación presentan las siguientes características:

Amplitud: cantidad de materias que abarca la negociación. Una agenda extensa facilita el acuerdo, ya que, al negociar muchas cuestiones a la vez, cada parte puede compensar lo que cede en unas con lo que obtiene en otras, y aumentan así las opciones de beneficio mutuo[12]. El inconveniente es que una agenda muy amplia encarece la negociación.

Profundidad: nivel de detalle al que desciende la discusión de cada materia. Un acuerdo más detallado abarata su vigilancia y seguimiento posterior y hace más creíbles los compromisos asumidos, pero cuesta más negociarlo y deja menos flexibilidad.

Grado de coordinación: abarca desde el simple intercambio de información hasta la armonización plena de las políticas, con cesiones de soberanía y costes de negociación crecientes a medida que se profundiza.

Frecuencia: las reuniones frecuentes ayudan a aproximar posiciones y hacen menos probable el incumplimiento. Conviene, con todo, reservar algunas para las ocasiones en que el acuerdo sea muy probable (p. ej. las cumbres del G20).

Tamaño: cuantos más países participan, más externalidades quedan internalizadas, pero también resulta más costoso coordinarlos, como ilustra el contraste entre el G7, homogéneo, y el G20, más representativo.

Enfoque: el acuerdo puede descansar en la hegemonía o en la simetría (véase la Introducción). Con un hegemón el acuerdo es más fácil, ya que su poder le permite inclinar el resultado hacia sus preferencias. Un sistema simétrico, en cambio, conjura el peligro de que ese mismo poder lleve al líder a imponer una solución socialmente subóptima, como en la ruptura del sistema de Bretton Woods[13] o del Sistema Monetario Europeo[14].

3.2. El G20, la OCDE y otros foros internacionales

La coordinación se instrumenta a través de foros e instituciones de naturaleza muy distinta. Las organizaciones con capacidad de condicionar o sancionar y otros foros de cooperación se estudian en otros temas:

[12] Las rondas comerciales de la OMC son un ejemplo, porque en ellas se negocian a la vez los obstáculos al comercio de casi todos los productos. Los países desarrollados tienen entonces interés en rebajar su protección agrícola si, en contrapartida, los países en desarrollo recortan sus aranceles industriales.

[13] Sus promotores confiaban en que EE. UU., por ser el miembro más poderoso, atendería al bienestar del conjunto. Sin embargo, la política expansiva estadounidense de finales de los años sesenta y comienzos de los setenta obligó a sus socios a elegir entre importar inflación y abandonar la paridad fija, y muchos optaron por lo segundo.

[14] Tras la reunificación, Alemania aplicó políticas fiscales expansivas para atender las necesidades de Alemania Oriental. Sin embargo, a partir de 1991 adoptó una política monetaria contractiva para contener la inflación, lo que obligó a los demás países europeos a elevar sus tipos de interés en plena recesión.

- *Instituciones de Bretton Woods y OMC*: el FMI, el Grupo Banco Mundial y la OMC, en los temas B22, B30 y B33.

- *Otros foros*: los BRICS+, en el tema B32, apartado 2.3, y el Foro Económico Mundial y la Iniciativa de la Franja y la Ruta, en el tema B35 §3.

- *Marco general*: la evolución del sistema económico internacional que los enmarca, en los temas B20 y B21.

A efectos de este tema interesa su tipología: foros de *coordinación negociada* con reglas y sanciones (FMI, OMC), foros de *concertación* sin capacidad coercitiva (G7/G20, OCDE) y foros *alternativos* surgidos de la insatisfacción con la gobernanza existente (BRICS+, nuevos bancos de desarrollo). Los de concertación, basados en la deliberación entre iguales y la evaluación mutua, son los que mejor responden a la lógica del apartado 2, y a ellos se dedican los tres subapartados siguientes. Como muestra el caso del G20 (apartado 3.2.1), su eficacia depende menos de su forma que de una emergencia que alinee los intereses de los participantes.

3.2.1. El G20

Definición: es el foro principal de la cooperación económica internacional desde la cumbre de Pittsburgh (2009). Propuesto por la cumbre del G7 de Colonia (junio de 1999) tras las crisis asiática y rusa, nació en septiembre de 1999 como reunión de ministros de Finanzas y gobernadores de bancos centrales de los países de importancia sistémica. La crisis financiera global lo elevó al nivel de jefes de Estado y de Gobierno (cumbre de Washington, noviembre de 2008).

Objetivos: persigue la estabilidad financiera internacional y un crecimiento mundial "fuerte, sostenible y equilibrado", con una agenda ampliada después al desarrollo, la fiscalidad, la salud y el clima.

Miembros: lo forman diecinueve países (Alemania, Arabia Saudí, Argentina, Australia, Brasil, Canadá, China, Corea del Sur, EE. UU., Francia, India, Indonesia, Italia, Japón, México, Reino Unido, Rusia, Sudáfrica y Turquía), la Unión Europea y, desde la cumbre de Nueva Delhi (2023), la Unión Africana. En conjunto, representan en torno al 85 % del PIB mundial y el 75 % del comercio internacional. España asiste como invitado permanente.

Funcionamiento: sin tratado constitutivo, personalidad jurídica ni secretaría permanente, decide por consenso mediante declaraciones no vinculantes. Una presidencia rotatoria anual fija la agenda y aporta una secretaría temporal, y la continuidad se confía a la *troika* (presidencias saliente, actual y entrante). El G20 trabaja en dos vías. La de sherpas prepara la cumbre y coordina los grupos sectoriales, y la financiera, de ministros y gobernadores, constituye su núcleo histórico.

A falta de burocracia propia, se apoya en el FMI, la OCDE y el Consejo de Estabilidad Financiera (FSB, por su sigla inglesa) para el análisis y el seguimiento. Así, en el *proceso de evaluación mutua* (Mutual Assessment Process, MAP), lanzado en Pittsburgh con el apoyo técnico del FMI, cada miembro somete sus políticas al examen de los demás.

Evolución: su etapa más eficaz fue la respuesta a la crisis financiera global. En Londres (abril de 2009) acordó, junto al estímulo fiscal concertado (apartado 2.1.2), triplicar los recursos del FMI hasta 750.000 M$ y una asignación de derechos especiales de giro (DEG) de 250.000 M$. Creó además el FSB, cuarto pilar de la arquitectura junto al FMI, el Banco Mundial y la OMC. En Seúl (2010) respaldó la reforma de cuotas del FMI, que trasladó más del 6 % de las cuotas a

las economías emergentes y en desarrollo. Después, la concertación perdió intensidad, salvo en la fiscalidad y en la deuda de los países pobres durante la pandemia.

Su dependencia de la presidencia de turno se ha hecho patente desde 2025. EE. UU. boicoteó la cumbre de Johannesburgo (noviembre de 2025), la primera celebrada en África. La presidencia estadounidense de 2026 ha excluido a Sudáfrica de sus reuniones y de la cumbre de Miami (14 y 15 de diciembre de 2026), con lo que la presidencia saliente ha quedado fuera de la troika. Tampoco contó con España, invitado permanente, en el arranque de sus reuniones preparatorias.

Conclusiones: el G20 gana en representatividad frente al G7, pero, sin reglas ni sanciones, su eficacia depende de una emergencia compartida y de la presidencia de turno, como anticipaba el apartado 2.1.3. Su aportación más duradera es el mandato a los organismos técnicos para convertir compromisos políticos en estándares sometidos a evaluación mutua.

3.2.2. La OCDE

> La financiación de la OCDE y los acuerdos que se negocian en su seno, de los Códigos de liberalización al Consenso sobre crédito oficial a la exportación, se recogen en el tema B35, apartado 3.1. Aquí se desarrolla la organización como foro de coordinación de políticas.

Definición: la Organización para la Cooperación y el Desarrollo Económicos es un organismo intergubernamental creado por el Convenio de París (1960), en vigor desde 1961, como heredero de la Organización Europea de Cooperación Económica (OECE, 1948).

Objetivos: además de promover el crecimiento, el empleo y el comercio, su Convenio obliga a los miembros a mantenerse informados, consultarse de forma continua y cooperar estrechamente, con una acción coordinada "cuando proceda" (art. 3). Esas obligaciones corresponden a los dos primeros niveles de coordinación de la Introducción.

Miembros: la forman 38 países, el último Costa Rica (2021). La OCDE tiene, además, ocho candidatos a la adhesión: Argentina, Brasil, Bulgaria, Croacia, Perú y Rumanía, desde 2022, e Indonesia y Tailandia, desde 2024. Como deja fuera a China y a la India, colabora estrechamente con el G20 y abre algunos de sus marcos a países no miembros.

Funcionamiento: el Consejo decide por acuerdo mutuo, el trabajo se realiza en comités y el método es el *examen entre pares* (*peer review*). El Comité de Examen Económico y de Desarrollo (EDRC, por su sigla inglesa) discute aproximadamente cada dos años el estudio económico de cada miembro. El Grupo de Trabajo n.º 3 del Comité de Política Económica analiza desde los años sesenta las políticas macroeconómicas y los desequilibrios exteriores de las principales economías.

Evolución: donde más ha avanzado desde el intercambio de información hacia los compromisos es la fiscalidad. A petición del G20, lanzó en 2013 el proyecto contra la erosión de bases imponibles y el traslado de beneficios (Base Erosion and Profit Shifting, BEPS), aplicado desde 2016 a través del Marco Inclusivo, abierto a países no miembros (148 jurisdicciones en 2026). Su acuerdo de octubre de 2021 tiene dos pilares:

- *Pilar 1*: reasigna parte de los beneficios de las mayores multinacionales a los países de mercado. Su Importe A sigue bloqueado, ya que su convenio multilateral no se ha abierto a la firma y EE. UU. se desvinculó del acuerdo en enero de 2025. En cambio, el Importe B, un enfoque simplificado de precios de transferencia, es aplicable de forma opcional desde 2025.

- *Pilar 2*: impone una tributación efectiva mínima del 15 % a los grupos con ingresos de al menos 750 M€, aplicada desde 2024 en la mayor parte de la UE y en más de cincuenta jurisdicciones. Sin embargo, en enero de 2026, un régimen "lado a lado" (*side-by-side*) eximió de parte de sus reglas a los grupos con matriz en países con un impuesto mínimo equivalente, como EE. UU.

Conclusiones: la OCDE es el ejemplo más acabado de coordinación por presión entre pares, la solución de Ostrom (apartado 2.2.2) aplicada a las políticas públicas. El Pilar 2 muestra también su límite, ya que un acuerdo sobre un recurso común ha tenido que acomodarse a la retirada de su principal miembro.

3.2.3. Otros foros: G7, Consejo de Estabilidad Financiera y BIS

Completan la arquitectura de concertación un núcleo político de economías avanzadas, el G7, y dos redes técnicas de autoridades financieras, el FSB y el BIS. Ninguno adopta decisiones vinculantes, y su influencia descansa en el consenso y en la presión entre pares.

G7 (1975): nacido en la cumbre de Rambouillet tras el fin de Bretton Woods y el primer *shock* del petróleo, reúne a Alemania, Canadá (desde 1976), EE. UU., Francia, Italia, Japón y el Reino Unido, con participación de la UE. Rusia lo integró entre 1997 y 2014 (G8). La homogeneidad del G7 facilita el acuerdo, como mostraron los acuerdos del Plaza, en el G5, y del Louvre, en el G6, pero le resta representatividad, por lo que el G7 cedió la primacía al G20 en 2009. Conserva, aun así, un papel de impulso, como muestra su acuerdo de junio de 2025, que abrió el camino al régimen "lado a lado" del Pilar 2.

> El FSB y el BIS se desarrollan, como instituciones de la regulación financiera internacional, en el tema B27, apartado 2.1.1. Aquí se recoge su papel en la coordinación.

Consejo de Estabilidad Financiera (FSB, 2009): se creó en la cumbre de Londres a partir del Foro de Estabilidad Financiera del G7 (1999). Reúne a ministerios de Finanzas, bancos centrales y supervisores de 24 jurisdicciones, entre ellas España, y de la UE, junto con las instituciones financieras internacionales y los organismos que fijan estándares. Decide por consenso y rinde cuentas al G20. Sus miembros se comprometen a aplicar los estándares internacionales y a someterse a exámenes entre pares y, cada cinco años, al Programa de Evaluación del Sector Financiero (FSAP) del FMI y el Banco Mundial.

Banco de Pagos Internacionales (BIS, por su sigla inglesa, 1930): reúne a 63 bancos centrales. Sus reuniones bimestrales de gobernadores son el principal foro de intercambio de información entre bancos centrales. No fija estándares por sí mismo, pero alberga los comités que los elaboran, como el de Supervisión Bancaria de Basilea, y la secretaría del FSB.

Conclusiones: estos foros completan una gobernanza en red que separa el impulso político (G7, G20), la elaboración de estándares (los comités con sede en el BIS, como el de Basilea, y la OCDE, coordinados por el FSB) y la vigilancia de su aplicación (FMI, FSB). Apoyada en la presión entre pares, esa gobernanza ha logrado una notable armonización regulatoria sin tratados vinculantes.

3.3. Experiencia y perspectivas de la coordinación

La experiencia de la coordinación se recorre en el tiempo, desde los acuerdos históricos y la evidencia empírica sobre ellos hasta el consenso y el marco actuales y las tensiones que condicionarán su futuro.

3.3.1. El pasado: experiencia histórica y evidencia empírica sobre la coordinación

Regulación cambiaria: ningún vínculo entre países es tan visible como el tipo de cambio, y por eso la mayoría de los acuerdos de coordinación han girado en torno a él y a la balanza de pagos. La necesidad de cooperar en este terreno quedó clara tras la Gran Depresión (1929-1933), cuando se multiplicaron las políticas de empobrecimiento del vecino.

> El sistema económico internacional anterior a 1944 y el propio sistema de Bretton Woods se desarrollan en el tema B20; los acuerdos posteriores a su colapso, en el tema B21; y la integración monetaria europea, en el tema B43. Aquí se recapitulan como experiencias de coordinación.

- *El patrón oro* (1870-1914): constituye una coordinación implícita más que negociada, pues el contenido en oro de cada moneda determinaba las paridades bilaterales sin acuerdo formal ni institución supervisora. La disciplina la imponía el mecanismo de flujo-especie, a costa de subordinar la política monetaria interna al equilibrio exterior.

- *Los Acuerdos de Bretton Woods* (1944-1971): primera coordinación cambiaria explícita y multilateral, con paridades fijas pero ajustables frente al dólar, convertibilidad de este en oro y una institución encargada de supervisar el cumplimiento y de proveer liquidez, el FMI. Su colapso ilustra que la coordinación no es sostenible cuando el país emisor de la moneda de reserva afronta un conflicto entre sus objetivos internos y su obligación externa (el dilema de Triffin).

- *La Serpiente Europea* (1972-1979): tras la ruptura del sistema, los países de la Comunidad Económica Europea intentaron preservar la estabilidad cambiaria intraeuropea limitando al 2,25 % el margen máximo de fluctuación bilateral entre sus monedas. La experiencia fracasó por la ausencia de una coordinación efectiva de las políticas monetarias nacionales, de modo que las monedas más débiles abandonaron el mecanismo sucesivamente.

- *Sistema Monetario Europeo* (1979-1998): coordinación cambiaria regional articulada sobre una moneda cesta, la unidad de cuenta europea (ECU), y un mecanismo de bandas con intervención obligatoria. Funcionó como importación de credibilidad antiinflacionista desde el Bundesbank y se quebró en 1992-1993 cuando la asimetría del ajuste, agravada por la reunificación alemana, hizo insostenible la defensa simultánea de varias paridades.

- *Los Acuerdos del Plaza* (1985): acuerdo del G5 para depreciar de forma concertada un dólar sobrevalorado mediante intervención coordinada en los mercados de divisas. Constituye el ejemplo canónico de coordinación eficaz en materia cambiaria, ya que el dólar se depreció con intensidad frente al yen y al marco en los dos años siguientes.

- *Los Acuerdos del Louvre* (1987): acuerdo del G6 (el G7 sin Italia, que no lo suscribió) para detener la depreciación desencadenada en el Plaza y estabilizar los tipos de cambio en torno a bandas de referencia no publicadas. Ilustra la dificultad simétrica de la coordinación, que consiste en frenar un movimiento previamente inducido, y suele señalarse su contribución a la política monetaria expansiva japonesa que alimentó la burbuja de finales de los ochenta.

- *Zona del euro* (1999-): la forma más intensa de coordinación posible, en la que la política monetaria deja de coordinarse para unificarse. Elimina el problema de coordinación en el ámbito monetario, pero lo traslada al fiscal, donde persisten veintiuna autoridades nacionales (tras la incorporación de Bulgaria en 2026) sujetas a reglas comunes.

Regulación bancaria y financiera: el aumento del riesgo sistémico derivado de la globalización financiera ha favorecido acuerdos como los de Basilea. En ellos, el Comité de Supervisión Bancaria de Basilea, con sede en el BIS, establece estándares de capital mínimo de los bancos en proporción a sus activos ponderados por riesgo. Sin embargo, el Comité carece de capacidad normativa, de modo que sus acuerdos son recomendaciones que cada jurisdicción incorpora a su ordenamiento, arquitectura de derecho blando que explica tanto la flexibilidad de la coordinación financiera internacional como su desigual grado de aplicación efectiva.

> El contenido de los acuerdos de Basilea y el conjunto de la arquitectura regulatoria internacional se desarrollan en el tema B27.

Cuestiones de deuda: la coordinación en materia de deuda soberana carece de un marco jurídico equivalente al concursal interno. Descansa, por ello, en foros de acreedores, como el Club de París (1956), de acreedores oficiales bilaterales, y el Club de Londres (1976), de bancos comerciales. Recurre también a planes negociados caso por caso: el Plan Baker (1985), el Plan Brady (1989), la Iniciativa para los Países Pobres Muy Endeudados (HIPC, 1996), la Iniciativa para el Alivio de la Deuda Multilateral (MDRI, 2005) y, más recientemente, el Marco Común del G20. La dificultad recurrente es la coordinación entre acreedores heterogéneos, agravada desde que China y los tenedores de bonos han desplazado a la banca sindicada como principales financiadores.

> El análisis detallado de la deuda externa, del Club de París y de las iniciativas de alivio se encuentra en el tema B29; los bancos multilaterales de desarrollo, en el tema B30.

Cuestiones fiscales: la coordinación fiscal es la más difícil de todas, porque la política presupuestaria constituye el núcleo de la soberanía parlamentaria y porque sus beneficios se reparten de forma asimétrica entre quien expande y quien se beneficia del arrastre.

Lo ilustra la cumbre de Bonn (1978), principal aplicación del efecto locomotora (apartado 1.2.2). En ella, en el marco del G7, Alemania y Japón se comprometieron a aplicar una política fiscal expansiva, a cambio de que EE. UU. liberalizase los precios internos del petróleo para reducir sus importaciones de crudo. Sin embargo, el estímulo se adoptó poco antes del segundo *shock* del petróleo, y las autoridades alemanas concluyeron después que las había llevado a una expansión que no convenía a su situación interna, episodio que alimentó su reticencia posterior a este tipo de acuerdos.

De ahí que los avances se hayan concentrado en dos frentes acotados, la fiscalidad internacional (convenios de doble imposición y tipo mínimo global del 15 % del Marco Inclusivo de la OCDE) y las reglas fiscales vinculantes de la zona del euro.

> La gobernanza fiscal de la Unión Económica y Monetaria, el Pacto de Estabilidad y Crecimiento y el Semestre Europeo se desarrollan en el tema B45.

Medioambiente: el clima constituye el caso más puro de bien público global, y la atmósfera, como sumidero de emisiones, el de tragedia de los comunes a escala planetaria, ya que los beneficios de la mitigación son difusos y globales y sus costes, concentrados y nacionales. Tras la Convención Marco de las Naciones Unidas sobre el Cambio Climático (1992), el Protocolo de Kioto (1997) intentó resolverlo mediante compromisos vinculantes de reducción, y fracasó por la ausencia de los mayores emisores. El Acuerdo de París (2015) invirtió la lógica, con contribuciones determinadas a nivel nacional sujetas a revisión periódica y a presión reputacional, en la línea de la gobernanza policéntrica de Ostrom.

El análisis de los bienes públicos globales y de la cooperación medioambiental se desarrolla en el tema B31.

Evidencia sobre el bienestar: las ganancias de la coordinación son positivas pero moderadas (apartado 2.3). Además, la evidencia empírica apenas logra probar que coordinar internacionalmente las políticas aporte algo más que compartir información entre Gobiernos, e indica que las ganancias crecen durante los periodos de inestabilidad, precisamente cuando los efectos desbordamiento se intensifican.

3.3.2. El presente: consenso y marco de la coordinación

Consenso: hoy existe cierto acuerdo sobre la tarea que corresponde a cada política económica:

- *Política monetaria*: ha de orientarse a la estabilidad de precios.
- *Política fiscal*: debe dejar que los estabilizadores automáticos amortigüen las fluctuaciones cíclicas.
- *Reformas estructurales*: a ellas corresponde elevar el crecimiento económico[15].

Marco internacional: descansa en los foros e instituciones ya examinados. En la fijación de normas intervienen, junto a las autoridades públicas que las redactan y supervisan su cumplimiento, organismos privados como el Consejo de Normas Internacionales de Contabilidad (IASB) en materia contable. El contenido de la reforma regulatoria posterior a la crisis financiera global se recoge en el tema B27.

3.3.3. El futuro

El futuro de la coordinación internacional depende de la resolución de tres tensiones:

1. *Erosión del consenso técnico*: mientras existió acuerdo sobre los objetivos de la política económica, la discusión versaba sobre instrumentos, pero la reaparición de la política industrial, de los aranceles y de las restricciones tecnológicas ha desplazado el desacuerdo hacia los fines.

2. *Geometría variable*: el centro de gravedad se desplaza desde los foros de alcance global, como el G20, hacia acuerdos plurilaterales y bloques regionales, que ganan operatividad a costa de legitimidad.

3. *Ámbitos sin marco consolidado*: la fiscalidad de la economía digital, la gobernanza de la inteligencia artificial y la regulación de los criptoactivos carecen todavía de él.

En todos esos ámbitos, la asimetría entre la dimensión global del fenómeno y el carácter nacional de la autoridad reguladora reproduce el problema que este tema ha examinado.

[15] Las discrepancias surgen en torno a cómo actuar en caso de recesión grave (ortodoxia frente a estímulos económicos).

3.4. Problemas *de facto*: retos y dificultades de la coordinación

Si coordinarse suele beneficiar a quienes participan, ¿por qué no es más frecuente? A los obstáculos analíticos del apartado 2.3 se suman los de orden práctico que se manifiestan en el propio proceso negociador:

1. *Insostenibilidad de la cooperación*: los mandatos de duración limitada acortan el horizonte de los Gobiernos y los vuelven impacientes, de modo que fallan los requisitos que sostienen la cooperación sin coacción (apartado 2.1.3).

2. *Costes de negociación*: la coordinación exige tiempo, recursos técnicos y capital político, y esos costes son ciertos e inmediatos, mientras que los beneficios son inciertos y diferidos. Esos costes aumentan, además, con el número de participantes y con el grado de detalle de los acuerdos (apartado 3.1).

3. *Distribución de las ganancias*: la coordinación define un rango de repartos posibles y no un punto (apartados 2.1.2 y 2.1.4), y cada participante tiene incentivos a retrasar el acuerdo para mejorar su posición, lo que puede impedir que se alcance ninguno.

4. *Problemas políticos*: los Gobiernos responden ante electorados nacionales que no internalizan los beneficios externos de sus políticas, de modo que un compromiso internacional eficiente puede resultar políticamente inviable. A ello se añaden la heterogeneidad de preferencias entre países y la dificultad de comprometer a los sucesores, dado que el Gobierno siguiente puede abandonar los compromisos políticos o denunciar los tratados suscritos por el anterior.

5. *Verificación del cumplimiento*: sin mecanismos fiables de supervisión no cabe detectar la desviación ni, por tanto, activar el castigo. De ahí los procedimientos de evaluación mutua de los foros ya examinados.

Conclusiones

La interdependencia genera efectos desbordamiento que se transmiten por los canales de la renta, monetario, de la inflación y de las expectativas, con un signo que depende del régimen cambiario, del tamaño de las economías y de su rigidez nominal. Esos efectos son condición necesaria, pero no suficiente, para coordinar, pues aun siendo la cooperación Pareto-superior se impone el equilibrio no cooperativo si cada país conserva un incentivo a desviarse y no existe un castigo creíble. De ahí la forma de los foros existentes (G20, OCDE, FSB), con relaciones indefinidas, supervisión continua, presión entre pares y evaluación mutua. La coordinación no es, con todo, siempre preferible. Sus ganancias medidas son moderadas, puede ser contraproducente con inconsistencia dinámica y exige un diagnóstico compartido, de modo que su valor es claro ante bienes públicos globales y discutible en la gestión coyuntural de la demanda.

La coordinación es un problema de teoría de juegos, en el que el dilema del prisionero explica por qué no se alcanza un resultado mejor para todos y los juegos repetidos, cómo puede sostenerse (tema A14). Los canales de la renta y monetario extienden a dos economías grandes el modelo de Mundell-Fleming (tema A32), y el de las expectativas descansa en los equilibrios múltiples de las crisis cambiarias (tema B17) y de la unión monetaria incompleta (tema B16). Por último, la experiencia histórica, de los acuerdos del Plaza y del Louvre a la gestión de la crisis financiera global, muestra que la coordinación avanza en las crisis (tema B21), y el clima es el bien público global en que su necesidad es más clara (tema B31).

El ciclo financiero global ha puesto en cuestión que la flotación aísle de los desbordamientos monetarios. Rey (2013) sostiene que ese ciclo, marcado por la Reserva Federal, reduce el trilema a un dilema. Obstfeld (2015) responde que el tipo de cambio flexible deja a las economías emergentes en mejor posición, aunque no las aísla por completo y eleva el valor de los instrumentos macroprudenciales. Si la tesis de Rey es cierta, gana peso la coordinación, o al menos la gestión de los flujos de capital. Sigue abierto también cómo sostener la cooperación climática. Frente a la lógica del Acuerdo de París (apartado 3.3.1), Nordhaus (2015) propone clubes climáticos que graven con aranceles a quienes no ponen precio al carbono, el castigo creíble que la teoría de juegos exige.