Lecciones de economía general e internacional

Lecciones de economía, vol. 1 · Economía Laboral

Arturo Lorenzo

Edición web 1.0 · noviembre de 2026

Tabla de contenidos

Nota 4

A25. Mercado de trabajo y capital humano 5
- Introducción . . . 5
- 1. Análisis estático: teoría neoclásica de la oferta y demanda de trabajo . . . 7
  - 1.1. Supuestos . . . 7
  - 1.2. Oferta de trabajo . . . 7
  - 1.3. Demanda de trabajo . . . 14
  - 1.4. Equilibrio neoclásico y poder de mercado . . . 20
  - 1.5. Conclusiones . . . 27
- 2. Análisis intertemporal . . . 28
  - 2.1. Supuestos . . . 28
  - 2.2. Oferta de trabajo intertemporal (Lucas y Rapping, 1969) . . . 28
  - 2.3. Demanda de trabajo intertemporal . . . 30
  - 2.4. Evidencia empírica . . . 31
  - 2.5. Conclusiones . . . 31
- 3. Teoría del capital humano . . . 32
  - 3.1. Capital humano como inversión (Becker, 1964) . . . 32
  - 3.2. Otras aportaciones . . . 35
  - 3.3. La función de ingresos de Mincer (1958, 1974) . . . 35
  - 3.4. Evidencia empírica . . . 36
  - 3.5. Conclusiones . . . 37
- Conclusiones . . . 39

A26. Desempleo friccional 40
- Introducción . . . 40
- 1. Búsqueda de empleo y salario de reserva . . . 42
  - 1.1. Contexto y antecedentes . . . 42
  - 1.2. Modelo del salario de reserva (Mortensen, 1970) . . . 44
  - 1.3. Extensiones y evidencia empírica . . . 47
  - 1.4. Conclusiones . . . 50
- 2. Emparejamiento y curva de Beveridge . . . 51
  - 2.1. Supuestos . . . 51
  - 2.2. Modelo . . . 51
  - 2.3. Extensiones y aplicaciones . . . 59
  - 2.4. Conclusiones . . . 61
- 3. Costes de ajuste y dinámica de la demanda de trabajo . . . 62
  - 3.1. Concepto y formas de los costes de ajuste . . . 62
  - 3.2. Demanda de trabajo . . . 62
  - 3.3. Evidencia empírica e implicaciones de política . . . 66
- Conclusiones . . . 68

A27. Determinación de salarios 69
- Introducción . . . 69

1. Contratos implícitos (Baily, 1974; Gordon, 1974; Azariadis, 1975) . . . 71
1.1. Supuestos . . . 71
1.2. Modelo . . . 72
1.3. Conclusiones . . . 73
2. Modelos de negociación colectiva . . . 74
2.1. Fundamentos . . . 74
2.2. Contratos ineficientes: *right to manage* (Nickell y Andrews, 1983) . . . 76
2.3. Contratos eficientes (McDonald y Solow, 1981) . . . 79
2.4. Trabajadores heterogéneos: *insiders-outsiders* (Lindbeck y Snower, 1984-1988) 84
2.5. Conclusiones . . . 88
3. Salarios de eficiencia . . . 89
3.1. Concepto: causas y origen . . . 89
3.2. Información simétrica: condición de Solow (1979) . . . 89
3.3. Información asimétrica: *no-shirking condition* (Shapiro y Stiglitz, 1984) . . . 93
Conclusiones . . . 98
Anexo. Herramientas y propuestas para combatir el desempleo . . . 99

A28. TNP, NAIRU y persistencia **100**
Introducción . . . 100
1. Fundamentos del análisis de la tasa natural de paro (TNP), la NAIRU y la persistencia del desempleo . . . 102
1.1. Debate entre la visión neoclásica y la keynesiana . . . 102
1.2. Síntesis neoclásica . . . 105
1.3. Monetarismo . . . 106
1.4. Neokeynesianos del desequilibrio: la tipología de Malinvaud (1977) . . . 109
1.5. Conclusiones . . . 111
2. Modelos de la tasa de desempleo de equilibrio: expectativas racionales, NAIRU e histéresis . . . 112
2.1. Nueva macroeconomía clásica . . . 112
2.2. Nueva economía keynesiana de primera generación: el modelo de la NAIRU (Layard y Nickell, 1985; Carlin y Soskice, 1990) . . . 114
2.3. Histéresis: la persistencia del desempleo (Blanchard y Summers, 1986) . . . 118
2.4. Conclusiones . . . 121
3. Desarrollos actuales y convergencia del pensamiento económico . . . 122
3.1. El modelo de búsqueda y emparejamiento . . . 122
3.2. La curva de Phillips de la nueva economía keynesiana de segunda generación (NEK2G) . . . 123
3.3. La estimación de la NAIRU y sus límites . . . 126
3.4. La convergencia del pensamiento económico . . . 127
Conclusiones . . . 128

Nota

Este PDF es un extracto de *Lecciones de economía general e internacional* (Lecciones de economía, vol. 1), de Arturo Lorenzo, con el bloque «Economía Laboral» (temas A25-A28). 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.

A25. Mercado de trabajo y capital humano

La teoría neoclásica del mercado de trabajo. Análisis intertemporal de la oferta de trabajo. Teoría del capital humano. Función de ingresos de capital humano y evidencia empírica.

Introducción

La teoría económica del mercado de trabajo se pregunta cómo se relacionan quienes intervienen en él, qué empleo y qué salarios de equilibrio genera esa relación y qué factores institucionales la moldean. El estudio de los salarios nació unido al de la **distribución de la renta**, hasta el punto de que David Ricardo sostenía en sus *Principios de economía política y tributación* (1817) que la gran cuestión de la economía política era cómo se reparte el producto entre terratenientes, capitalistas y trabajadores. Con el tiempo, el análisis se ha desdoblado en una rama microeconómica, que modeliza las decisiones de trabajadores y empresas, y otra macroeconómica, que se ocupa de por qué fluctúa el empleo y de los rasgos estructurales del mercado.

El eje de ese análisis es el debate entre el pensamiento neoclásico, para el que salarios y precios son flexibles y el mercado se vacía, y el keynesiano, para el que la insuficiencia de demanda efectiva y las rigideces pueden impedir el ajuste. Su objeto común es el **desempleo**, es decir, la situación de las personas que, formando parte de la población activa, carecen de empleo y lo buscan activamente. El desempleo puede clasificarse según tres criterios:

1. *Por su causa*: el desempleo *friccional* resulta de la rotación normal del mercado, porque encontrar empleo y cubrir vacantes lleva tiempo; el *estructural*, del desajuste persistente entre las características de la oferta y de la demanda de trabajo (cualificaciones, sectores, territorios) y de las instituciones que impiden el vaciado del mercado; el *estacional*, del patrón intraanual de actividades como la agricultura o el turismo; y el *cíclico*, de las fluctuaciones de la demanda agregada.

2. *Por su duración*: el de *corta duración*, inferior a un año y propio de la rotación friccional, y el de *larga duración*, que deteriora el capital humano y la empleabilidad de quienes lo sufren.

3. *Por su relación con el equilibrio*: la tasa observada se descompone en una tasa *de equilibrio*, que agrupa el desempleo friccional y el estructural y depende de factores reales e institucionales, y un componente *cíclico*, que es el objeto de las políticas de estabilización.

Cada tipo de desempleo tiene una explicación teórica distinta. En la teoría neoclásica, que es la referencia común, el desempleo solo puede ser voluntario o friccional (tema A25). La teoría de la búsqueda y del emparejamiento explica el desempleo friccional (tema A26), y las teorías de la determinación de los salarios, por qué estos no bajan hasta vaciar el mercado, lo que convierte el desempleo involuntario en un resultado de equilibrio (tema A27). Ambas líneas se integran en la tasa de desempleo de equilibrio, desde la tasa natural de Friedman hasta la tasa de paro

que no acelera la inflación (NAIRU), y en la **histéresis**, por la que el desempleo cíclico de hoy puede convertirse en el estructural de mañana (tema A28).

Este tema estudia la teoría neoclásica del mercado de trabajo en tres bloques:

- *Análisis estático*: oferta, demanda y equilibrio neoclásicos, y sus extensiones con poder de mercado (sindicatos, monopsonio y monopolio bilateral).

- *Análisis intertemporal*: la oferta de trabajo como asignación del tiempo entre periodos (modelo de Lucas-Rapping) y la demanda de trabajo con costes de ajuste.

- *Teoría del capital humano*: los modelos de Becker (1964) y Spence (1973) y la función de ingresos de Mincer.

1. Análisis estático: teoría neoclásica de la oferta y demanda de trabajo

Para la teoría neoclásica, el trabajo se intercambia en un mercado parecido al de cualquier otro bien. Su precio, el salario, y su cantidad, el empleo, quedan fijados donde se cruzan la oferta y la demanda.

1.1. Supuestos

Metodología: el análisis es de equilibrio parcial.

Entorno: no hay fallos de mercado (externalidades, bienes públicos, poder de mercado, información imperfecta, costes de transacción ni fricciones de emparejamiento o *matching*). Hay plena flexibilidad de precios y salarios, de modo que el mercado de trabajo se vacía.

Agentes: consumidores-trabajadores racionales, que maximizan su utilidad, y empresas gestionadas por sus propietarios, que maximizan el beneficio. Ambos son numerosos, de modo que ninguno puede influir sobre el salario, es decir, son *salario-aceptantes*.

Tecnología: se aplican los supuestos neoclásicos habituales (producción monoproducto, rendimientos mixtos).

Mercado de bienes: es de competencia perfecta, de modo que las empresas son también precio-aceptantes en el mercado de su producto.

1.2. Oferta de trabajo

El lado de la oferta parte de la elección entre ocio y consumo de cada individuo. De ella se obtiene la oferta de trabajo individual, cuya estática comparativa se estudia antes de agregarla al conjunto del mercado.

1.2.1. La decisión de oferta de trabajo

La oferta de trabajo es una aplicación de la teoría del consumidor con dotaciones (tema A08, Anexo A.2). El individuo consume dos bienes, un **bien compuesto** c (consumo real)[1] y **tiempo de ocio** o, la única alternativa al trabajo remunerado, que, a diferencia del ocio, es un mal[2].

Planteamiento: su función de utilidad es creciente, continua, estrictamente cuasicóncava y diferenciable, $u = u(o, c)$. Está sujeto a una restricción de gasto, $p_c c \leq w(l_0 - o) + M_0$, donde M_0 es la renta no salarial, y a una restricción de tiempo, $l_0 = l + o$, donde l_0 es la dotación de tiempo y l el trabajo. Por tanto:

$$\max_{o,c} \; u(o, c) \quad \text{s.a.: } p_c c = w(l_0 - o) + M_0$$

donde la renta total, $M = w(l_0 - o) + M_0$, suma la renta salarial $w(l_0 - o)$, que depende de cuánto tiempo se dedique al trabajo, y la renta no salarial M_0, que no depende de lo que trabaje y se considera exógena. El lagrangiano es $\mathcal{L} = u(o, c) + \lambda[w(l_0 - o) + M_0 - p_c c]$.

[1] Un bien compuesto es un grupo de bienes cuyos precios varían en la misma proporción, de modo que sus precios relativos se mantienen constantes.

[2] El ocio incluye la producción doméstica. Quien dedica todo su tiempo al ocio es un "no participante". La categoría de desempleado la añade la teoría de búsqueda (tema A26).

CPO: con solución interior y la restricción agotada, a partir del lagrangiano se obtienen tres condiciones de primer orden:

$$p_c c = w(l_0 - o) + M_0 \tag{1}$$

$$\frac{\partial u}{\partial c} = \lambda p_c \tag{2}$$

$$\frac{\partial u}{\partial o} = \lambda w \;\Rightarrow\; \text{RMS}_{co} \equiv \frac{\partial u/\partial o}{\partial u/\partial c} = \frac{w}{p_c} \tag{3}$$

donde (1) establece que el consumidor-trabajador gasta toda su renta; (2) y (3), que la utilidad marginal de cada bien por euro gastado es la misma. El "precio" del ocio es el salario, la renta a la que se renuncia por cada hora no trabajada.

CSO: la cuasiconcavidad estricta de u (curvas de indiferencia ocio-consumo estrictamente convexas) y la linealidad de la restricción garantizan que la tangencia es un máximo único, de modo que las CPO son necesarias y suficientes. Aunque en la solución de esquina no hay tangencia sino desigualdad, la misma convexidad de las preferencias asegura que es el óptimo.

Solución: la función de oferta de trabajo individual presenta dos casos:

▪ *Solución interior*: dividiendo (3) entre (2) se obtiene la *condición de ocio-consumo*. En el óptimo, el consumidor-trabajador adquiere ocio y consumo hasta que su relación marginal de sustitución (RMS, tasa subjetiva de intercambio ocio-consumo) equivale al salario real (tasa objetiva de intercambio ocio-consumo en el mercado o coste de oportunidad del ocio).

▪ *Solución de esquina (salario de reserva)*: si el salario real es suficientemente bajo, el individuo no trabaja, $\text{RMS}_{co}(o = l_0) > w/p_c$. El salario por debajo del cual prefiere no participar es el *salario de reserva* $\underline{w}$, la RMS evaluada en la dotación (todo el tiempo dedicado al ocio), y la condición de participación es $w/p_c > \underline{w}$. El salario de reserva depende de la renta no salarial, del consumo de subsistencia y de las preferencias (valoración relativa del ocio). De la renta no salarial depende directamente si el ocio es normal, porque con más M_0 el consumo es mayor y el ocio más valioso en el margen.

Desarrollo gráfico: la decisión de oferta de trabajo se representa en el plano ocio-consumo (figura A25.1):

1. *Recta de balance*: reúne las combinaciones de ocio y consumo que el individuo puede costear con su renta total. Tiene pendiente $-w/p_c$ y ordenada en el origen $(wl_0 + M_0)/p_c$, la *renta potencial*, y pasa siempre por la dotación $\omega = (l_0, M_0/p_c)$[3].

2. *Curvas de indiferencia*: como u_0, reúnen las combinaciones de ocio y consumo de igual utilidad. Son convexas, porque la utilidad es estrictamente cuasicóncava, y representan más utilidad cuanto más alejadas del origen.

3. *Equilibrio*: en el punto 0, la curva de indiferencia más alejada del origen es tangente a la recta de balance, con ocio o^* y consumo c^*. El trabajo ofrecido, $l^* = l_0 - o^*$, es la distancia horizontal entre el ocio elegido y la dotación.

[3] Equivalentemente, $p_c c + wo \leq wl_0 + M_0$, es decir, la renta potencial se gasta en consumo y en ocio, valorado a su coste de oportunidad w.

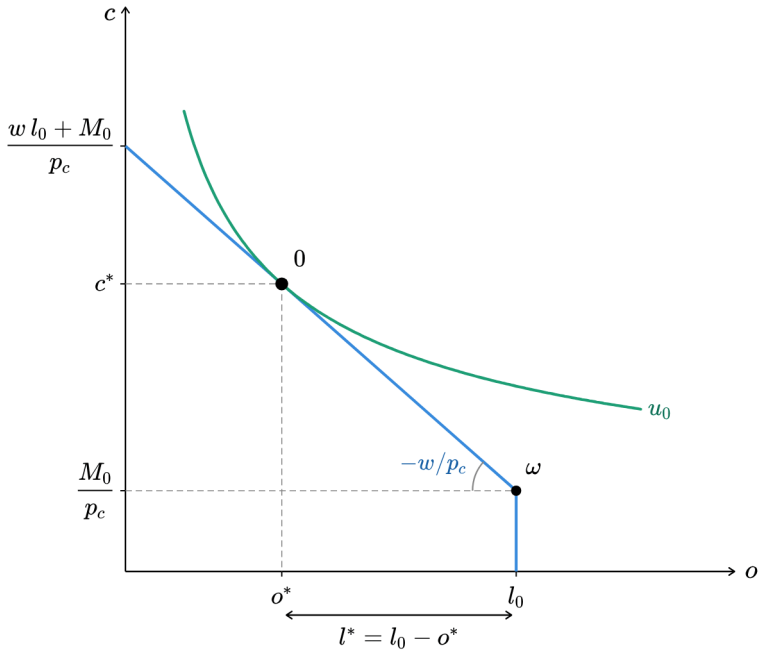


Figura A25.1: Elección ocio-consumo. Se sitúa en la tangencia entre la curva de indiferencia y la restricción presupuestaria, cuya pendiente es el salario real.

1.2.2. La función de oferta de trabajo individual

Definición: del problema primal se obtienen las funciones marshallianas de demanda de consumo, $c^* = c(w, p_c, M_0)$, y de ocio, $o^* = o(w, p_c, M_0)$, y de esta última, a través de la restricción de tiempo, puede obtenerse la *función de oferta de trabajo individual ordinaria*:

$$l^s = l_0 - o^*(w, p_c, M_0) = l^s(w^?, p_c^-, M_0^-)$$

donde los superíndices indican el signo de cada efecto, que se justifica en el punto 3 de las propiedades. Del problema dual (minimizar la renta no salarial necesaria para alcanzar una utilidad dada) se obtiene la *función de oferta de trabajo individual compensada*, $l^c = l^c(w, p_c, u_0)$, que solo recoge el efecto sustitución.

Propiedades:

1. *Continuidad y diferenciabilidad*: son continuas en sus argumentos y diferenciables en las soluciones interiores, lo que permite la estática comparativa. La ordinaria se anula cuando el salario real no supera el de reserva, $\underline{w}$, donde presenta un quiebro.

2. *Transformaciones*: son homogéneas de grado cero, la ordinaria en (w, p_c, M_0) y la compensada en (w, p_c). Por ello, la ordinaria solo depende del salario real, w/p_c, y de la renta no salarial real, M_0/p_c, y la compensada, del salario real y de la utilidad (ausencia de ilusión monetaria).

3. *Inclinación*:

▪ *Ordinaria*: el efecto del salario es de signo indeterminado y el de la renta no salarial es negativo si el ocio es normal. El del precio del consumo es el simétrico del salario, pues ambos operan a través del salario real, aunque p_c reduce también la renta no salarial real, lo que por sí solo eleva la oferta si el ocio es normal. El signo negativo que se asigna a p_c supone, por tanto, que domina el efecto sustitución y que la renta no salarial es reducida.

▪ *Compensada*: es no decreciente en el salario, y creciente en las soluciones interiores, porque solo recoge el efecto sustitución.

4. *Forma*: la teoría no restringe la curvatura de ninguna de las dos. La restricción de segundo orden recae sobre la función de valor del problema dual, la renta no salarial mínima, que es cóncava en el salario y en el precio del consumo y de la que deriva la oferta compensada.

5. *Otras propiedades*:

▪ *Agotamiento de la renta*: en la ordinaria, el consumo agota la renta total, $p_c\, c^* = w\, l^s + M_0$ (condición (1)).

▪ *Dualidad*: la ordinaria coincide con la compensada evaluada en la utilidad máxima que alcanza el individuo, u^*, es decir, $l^s(w, p_c, M_0) = l^c(w, p_c, u^*)$. De esa identidad procede la ecuación de Slutsky con dotaciones.

▪ *Observabilidad*: la ordinaria se estima con datos de horas, salarios y rentas. La compensada, en cambio, no es observable, porque depende de la utilidad, y su elasticidad se obtiene de la ordinaria mediante la ecuación de Slutsky.

1.2.3. Estática comparativa

La estática comparativa de la oferta distingue dos bloques, los efectos de una subida del salario sobre el ocio y las horas ofrecidas (efectos propios) y sobre el consumo (efectos cruzados), y se completa con los de otras variables exógenas.

Efectos propios: el efecto de una subida del salario sobre la oferta de trabajo es ambiguo porque el trabajador es a la vez comprador de ocio (lo consume) y vendedor de tiempo (su dotación se revaloriza). La ecuación de Slutsky con dotaciones (tema A08, Anexo A.2), aplicada a la demanda de ocio, separa esas fuerzas:

$$\underbrace{\frac{\partial o}{\partial w}}_{\text{ET}} = \underbrace{\left.\frac{\partial o}{\partial w}\right|_{u_0}}_{\text{ES}\,\leq\,0} \underbrace{-\,\frac{\partial o}{\partial M}\cdot o}_{\text{ERO}} \underbrace{+\,\frac{\partial o}{\partial M}\cdot l_0}_{\text{ERD}}$$

donde ES es el efecto sustitución (a utilidad constante), ERO el efecto renta ordinario (el de cualquier demanda, en la que el bien se encarece y la renta compra menos) y ERD el efecto renta-dotación (el salario más alto revaloriza las l_0 horas de la dotación). Los dos últimos se suman en el *efecto renta total*, $(\partial o/\partial M)(l_0 - o) = (\partial o/\partial M)\, l$, proporcional a las horas efectivamente vendidas.

▪ *Efecto sustitución*, ES $\leq$ 0 siempre: el salario es el precio del ocio y, al subir, a utilidad constante el individuo sustituye ocio por consumo, con lo que reduce el ocio y aumenta la oferta de trabajo.

- *Efecto renta total*, ≥ 0 si el ocio es normal: como $l = l_0 - o > 0$ para quien trabaja, el ERD domina siempre al ERO y la subida del salario eleva la renta real, de modo que el trabajador, más rico, demanda más ocio y reduce la oferta de trabajo. En cambio, si el ocio fuera inferior, ambos efectos irían en la misma dirección y la oferta sería siempre creciente.

- *Signo total*: la oferta de trabajo es creciente en el salario ($\partial l^s/\partial w > 0$) solo si el efecto sustitución domina al efecto renta total. Si el efecto renta total domina, la curva se vuelve hacia atrás en el plano (l, w). Un bien normal puede comportarse así como un bien Giffen para quien lo vende, ya que el efecto renta actúa a favor del ocio, no en contra.

Gráficamente, en el plano ocio-consumo (figura A25.2), una subida de w hace girar la recta de balance alrededor de la dotación, que sigue siendo alcanzable, y el nuevo equilibrio recoge los tres efectos:

- *ES*: el individuo se desplaza a lo largo de la curva de indiferencia inicial hasta el punto en que la RMS_{co} iguala el nuevo salario real w/p_c (de 0 a 0′) y, como el ocio se ha encarecido, lo reduce.

- *ERO*: si la renta monetaria no cambiase, el ocio más caro reduciría el poder adquisitivo del individuo y, siendo el ocio normal, la demanda de ocio caería también (de 0′ a 0″).

- *ERD*: la dotación de l_0 horas vale ahora más, lo que eleva la renta real por encima del nivel inicial y aumenta la demanda de ocio. Como el individuo vende parte de su dotación, este efecto domina al ERO (de 0″ a 1).

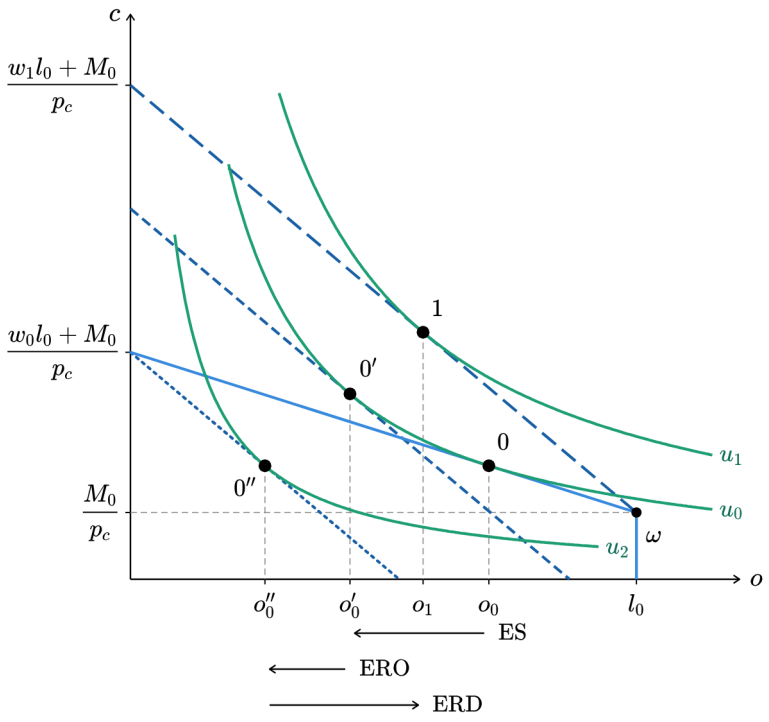


Figura A25.2: Efecto de una subida del salario en el plano ocio-consumo. El cambio se descompone en efecto sustitución ($0 \to 0'$), efecto renta ordinario ($0' \to 0''$) y efecto renta-dotación ($0'' \to 1$).

Proyectado sobre el plano trabajo-salario (figura A25.3), el resultado da la curva de oferta individual, que recoge las horas que el individuo ofrece a cada salario. Si |ES| > |ER|, con ER

el efecto renta total, la curva es creciente, y si $|\text{ES}| < |\text{ER}|$ se vuelve hacia atrás. El punto de giro se alcanza en el salario w', en el que ambos efectos se igualan ($|\text{ES}| = |\text{ER}|$) y la oferta individual alcanza su máximo.

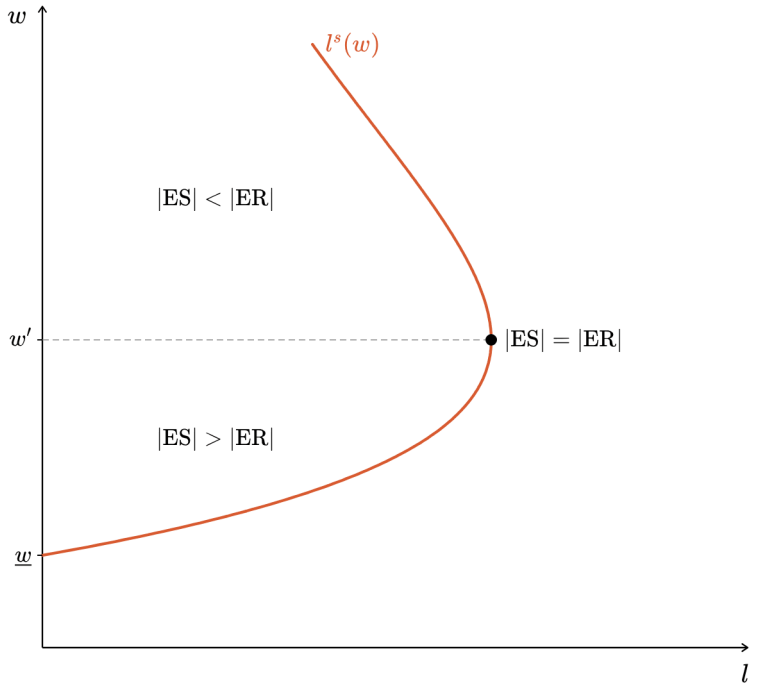


Figura A25.3: Efecto sustitución y efecto renta de una subida salarial sobre la oferta de trabajo. El signo total es ambiguo y explica el tramo regresivo de la oferta de trabajo, que empieza en el salario w', donde $|\text{ES}| = |\text{ER}|$.

Efectos cruzados: el efecto de una subida del salario sobre la demanda del bien de consumo, en cambio, es inequívoco si el consumo es normal. El ES es positivo (el consumo se abarata relativamente respecto al ocio) y el efecto renta total también lo es si el consumo es normal (la dotación se revaloriza). Por ello, el consumo aumenta salvo que sea un bien inferior, pero lo que queda indeterminado es si el mayor consumo se financia trabajando más o solo con el mayor salario por hora.

Otra estática comparativa:

1. *Renta no salarial*: un aumento de M_0 desplaza la recta de balance en paralelo. Si el ocio es normal, aumenta su demanda y se reduce la oferta de trabajo (y sube el salario de reserva), y si es inferior, ocurre lo contrario.

2. *Horas extraordinarias*: una prima salarial que se aplica solo a las horas trabajadas por encima de la jornada, y no a todas, tiene un efecto inequívoco para quien no superaba la jornada. Como no revaloriza la dotación ni las horas ya trabajadas, no genera efecto renta apreciable y actúa como un efecto sustitución puro, de modo que la oferta de trabajo aumenta. De ahí que las empresas la prefieran a una subida general del salario para elevar las horas.

3. *Imposición*: el efecto sobre la oferta de trabajo depende del tipo de impuesto:

- *Suma fija*: un impuesto T_0 desplaza la recta en paralelo y, con ocio normal, aumenta la oferta de trabajo (solo efecto renta).

- *Proporcional sobre el salario*: un tipo τ_w reduce el salario neto a $w(1-\tau_w)$ y tiene los tres efectos anteriores con signo contrario, por lo que su efecto es ambiguo.

Para una misma recaudación, el de suma fija genera siempre más oferta de trabajo, porque no altera el precio relativo del ocio y no produce exceso de gravamen. Por su parte, un impuesto progresivo hace cóncava la recta de balance. Un impuesto proporcional sobre el consumo equivale al proporcional sobre el salario si no hay ahorro ni renta no salarial ($\tau_c = \tau_w/(1-\tau_w)$), pues con $M_0 > 0$ el impuesto sobre el consumo grava también esa renta.

1.2.4. Oferta de trabajo agregada

Función de oferta agregada: bajo trabajo homogéneo e independencia de las ofertas individuales, la oferta de mercado es la suma horizontal de las ofertas individuales (agregación walrasiana), $L^s = \sum_{h=1}^{H} l^{s,h}(w^+, p_c^-, M_0^-)$.

Se acepta generalmente que el tramo relevante es el creciente. El supuesto evita equilibrios múltiples y se justifica porque las subidas de salario atraen a nuevos participantes (*margen extensivo*) aunque los ya empleados reduzcan sus horas (*margen intensivo*).

1.2.5. Evidencia empírica

La elasticidad de la oferta de trabajo es difícil de estimar, por la heterogeneidad de los agentes y porque los datos mezclan el margen intensivo (horas) y el extensivo (participación). Las estimaciones para trabajadores principales son bajas (elasticidad no compensada próxima a cero, compensada positiva pero pequeña) y mayores en el margen extensivo, lo que confirma que los efectos renta y sustitución se compensan en buena medida. La evidencia permite, además, precisar tres aspectos de la teoría:

1. *Cambios en el salario*: para niveles muy bajos de salario la oferta también puede ser decreciente. La curva individual puede tomar así forma de S invertida, con tres tramos: decreciente, creciente y de nuevo decreciente para salarios altos, cuando domina el efecto renta[4].

2. *Cambios en la renta*: la evidencia sobre transferencias no condicionadas (loterías, herencias, experimentos de renta básica) muestra reducciones pequeñas pero significativas de la oferta, coherentes con un ocio normal de baja elasticidad-renta.

3. *Unidades familiares*: la elasticidad de la oferta respecto al salario es baja para los trabajadores principales del hogar y alta para los secundarios, que comparan el salario con el coste de sustituir su producción doméstica. La incorporación de la mujer al mercado de trabajo ha reducido esa diferencia.

[4] Una explicación del tramo inferior es que los trabajadores con salario de reserva ligado a subsidios podrían perderlos si el salario sube modestamente, por lo que reducen su oferta para no exceder el umbral. Otra es la existencia de un consumo de subsistencia que obliga a trabajar más cuanto menor es el salario.

1.2.6. Conclusiones

Aspectos positivos: la oferta de trabajo neoclásica es una aplicación directa de la teoría del consumidor con dotación. El salario real es el precio del ocio, y la descomposición de Slutsky con dotaciones explica que la curva de oferta pueda tener pendiente positiva o negativa según dominen el efecto sustitución o el efecto renta total (ordinario más dotación).

Limitaciones: el modelo es estático y de un solo periodo, ignora las horas fijas por contrato y trata la participación como una decisión de esquina. Además, supone que el individuo elige libremente las horas al salario vigente y que quien no trabaja no quiere hacerlo. Las extensiones intertemporales (sección 2), de capital humano (sección 3) y de búsqueda (tema A26) responden a estas carencias.

1.3. Demanda de trabajo

Del otro lado del mercado, la demanda de trabajo resulta de la maximización del beneficio de la empresa. Tras obtenerla, se estudian su estática comparativa y su agregación al conjunto del mercado, y se contrasta con la evidencia empírica.

1.3.1. La decisión de demanda de trabajo

La demanda de trabajo neoclásica procede de la teoría de la empresa competitiva (tema A16) y de su sistematización por Hicks (1932). A la empresa el trabajo solo le interesa como medio para obtener la producción que vende, y por eso se habla de una demanda *derivada*.

Planteamiento: se obtiene del programa de maximización del beneficio desde el enfoque de la producción (tema A16), para el caso monoproducto, $x = f(\mathbf{z})$, con factores normales, y competencia perfecta en bienes y factores:

$$\max_{\mathbf{z}\geq 0} \ \pi = p \cdot f(\mathbf{z}) - \mathbf{p}^z \cdot \mathbf{z}$$

donde $\mathbf{z} = (z_1, z_2, ...)$ es el vector de cantidades de factores contratadas (variables de decisión, entre ellas el trabajo l), π el beneficio, p el precio del producto, $f(\mathbf{z})$ la función de producción y $\mathbf{p}^z = (p_1^z, p_2^z, ...)$ el vector de precios de los factores (el salario w para el trabajo).

CPO: derivando respecto a cada factor se obtienen dos condiciones:

$$\frac{\partial f/\partial z_i}{\partial f/\partial z_j} = \frac{p_i^z}{p_j^z} \tag{4}$$

$$\text{VPmg}_i \equiv p \cdot \frac{\partial f}{\partial z_i} = p_i^z \tag{5}$$

- *Condición de minimización del coste*: la condición (4) iguala la relación marginal de sustitución técnica (RMST) al precio relativo de los factores, de modo que cualquier cantidad se produce con la combinación de factores más barata.

▪ *Condición de contratación de factores*: la condición (5) establece que cada factor se contrata hasta que el valor de su productividad marginal iguala a su precio. Para el trabajo, $p \cdot \partial f/\partial l = w$, es decir, la empresa contrata hasta que el valor de lo que produce el último trabajador iguala a su salario y, como la empresa es salario-aceptante, la oferta de trabajo a la que se enfrenta es perfectamente elástica al salario de mercado.

CSO: el beneficio ha de ser estrictamente cóncavo en los factores, es decir, la empresa debe operar con rendimientos decrecientes de cada factor y a escala en el tramo relevante[5]. Con un solo factor variable basta $f''_{ll} < 0$, el tramo decreciente de la productividad marginal del trabajo, y bajo ese supuesto las CPO son necesarias y suficientes.

Solución: se obtienen las funciones de demanda de trabajo condicionada e incondicionada.

Desarrollo gráfico: la decisión de demanda de trabajo se representa en el plano trabajo-salario (figura A25.4):

1. *Valor del producto marginal* ($\text{VPmg}_l(l) = p\,\partial f/\partial l$): recoge el ingreso que aporta cada trabajador adicional y, en el tramo relevante, es decreciente por la productividad marginal decreciente.

2. *Valor del producto medio* ($\text{VPme}_l(l)$): recoge el ingreso medio por trabajador. Por debajo del máximo del VPme, el tramo decreciente del VPmg forma la demanda de trabajo, $l^d(w)$, porque a salarios superiores a ese máximo la empresa no cubriría el coste salarial.

3. *Oferta a la que se enfrenta la empresa*: una horizontal en el salario de mercado, w_0, porque la empresa es salario-aceptante.

4. *Equilibrio*: el empleo óptimo, l_0^*, es la intersección de la demanda con esa horizontal (punto 0), donde $\text{VPmg}_l = w_0$.

[5] Formalmente, f ha de ser estrictamente cóncava en ese tramo, con hessiano definido negativo ($f''_{ll} < 0$, $f''_{kk} < 0$ y $f''_{ll}f''_{kk} - (f''_{lk})^2 > 0$).

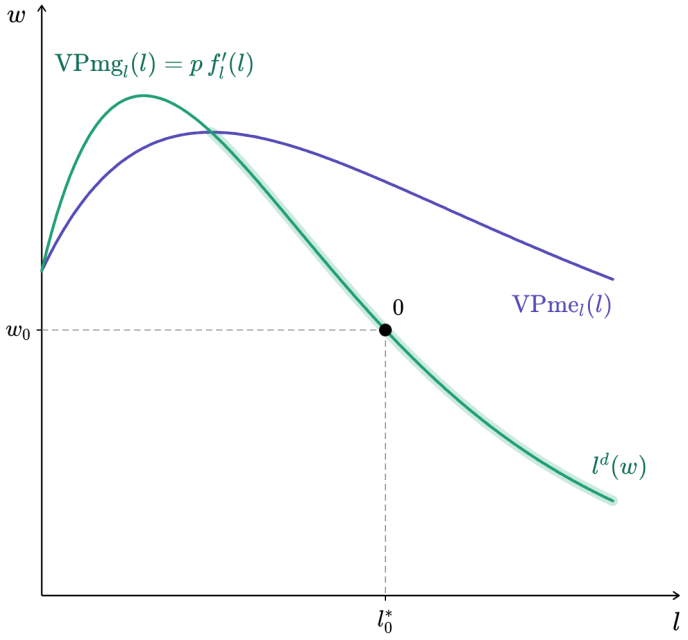


Figura A25.4: Demanda de trabajo a corto plazo. Coincide con el tramo decreciente del valor del producto marginal del trabajo.

1.3.2. La función de demanda de trabajo

Definición: como en la teoría de la producción (tema A12), del problema de la empresa se obtienen dos funciones de demanda de trabajo:

1. *Demanda condicionada*: se obtiene de (4), la minimización del coste para un nivel de producción dado, $l_c^d = l_c^d(w^-, r_K^+, x_0^+)$, donde r_K es el precio del servicio del capital. No depende directamente del precio del producto (solo a través de x_0) y mantiene fija la producción. Su efecto cruzado respecto al precio del otro factor es no negativo con dos factores, que son necesariamente sustitutivos netos (la complementariedad neta solo cabe con tres o más factores).

2. *Demanda incondicionada*: se obtiene de (5), la maximización del beneficio, $l_i^d = l_i^d(w^-, r_K^?, p^+, f_l'^+)$. Sí depende directamente del precio del bien, porque la producción varía con él, y es la demanda relevante para el mercado de trabajo.

Propiedades:

1. *Continuidad y diferenciabilidad*: son continuas y diferenciables en sus argumentos cuando la función de producción es estrictamente cuasicóncava, en la condicionada, y el beneficio es estrictamente cóncavo en los factores, en la incondicionada (CSO). No obstante, a corto plazo, la demanda es discontinua en el salario igual al máximo del valor del producto medio (VPme), porque a salarios superiores el ingreso no cubriría el coste salarial y la empresa no contrataría (cierre).

2. *Transformaciones*: son homogéneas de grado cero en los precios, la incondicionada en (w, r_K, p) y la condicionada en (w, r_K). Por ello, la primera depende de los precios reales de los factores, w/p y r_K/p, y la segunda, de su precio relativo, w/r_K (ausencia de ilusión monetaria).

3. *Inclinación*: el signo negativo respecto al salario es común a ambas, pero la incondicionada añade al efecto sustitución el efecto escala. Sus argumentos y su elasticidad cambian con el horizonte:

- *Corto plazo*: con el trabajo como único factor variable, la demanda de trabajo individual coincide con el tramo decreciente de la curva de VPmg del trabajo situado por debajo del máximo del VPme. Así, $l^d = l^d(w^-, p^+, f_l'^+, \bar{k}^+)$. Depende negativamente del salario, positivamente del precio y de la productividad del trabajo, y positivamente del *stock* de capital. Lo primero se debe a que una caída del salario se compensa con una caída de la productividad marginal para mantener $p\,\partial f/\partial l = w$, lo que exige más trabajadores.

- *Largo plazo*: con todos los factores variables, la demanda ya no coincide con la curva de VPmg, porque esta se desplaza al ajustarse el capital. Confluyen dos efectos que actúan en la misma dirección sobre el trabajo, el efecto sustitución, en la medida en que la tecnología lo permita ($\sigma_{kl} > 0$), y el *efecto escala o producción*. Como a corto plazo el capital está fijo y no cabe la sustitución entre factores, la demanda a largo plazo es más elástica que a corto plazo.

4. *Forma*: la teoría no restringe su curvatura. La restricción de segundo orden recae sobre la función de costes, cóncava en los precios de los factores, en la condicionada, y sobre la función de máximo beneficio, convexa en precios, en la incondicionada.

5. *Otras propiedades*:

- *Envolvente* (*lema de Shephard* y *lema de Hotelling*): la condicionada es la derivada de la función de costes respecto al salario, y la incondicionada es, cambiada de signo, la derivada de la función de máximo beneficio (temas A12, apartado 1.4.4, y A16, apartado 1.3.4).

- *Dualidad*: la incondicionada es la condicionada evaluada en la producción que maximiza el beneficio, $l_i^d(w, r_K, p) = l_c^d(w, r_K, x^*)$, con x^* esa producción. De esa identidad procede la ecuación de Slutsky que se obtiene a continuación.

- *Observabilidad*: se estiman con datos de empleo, salarios y producción.

Gráficamente, en el plano trabajo-salario (figura A25.5), ante una bajada del salario de w_0 a w_1, a corto plazo la empresa se desplaza sobre su curva de VPmg, $\text{VPmg}_l(l;\, k_0)$ (de 0 a 0′). A largo plazo, el mayor empleo eleva la productividad marginal del capital, se contrata más capital y ello desplaza la curva de VPmg del trabajo hacia fuera, hasta $\text{VPmg}_l(l;\, k_1)$ (de 0′ a 1). La demanda de largo plazo, $l^d_{\text{LP}}(w)$, une los puntos 0 y 1 y es, por ello, más plana que la de corto plazo. Por simplicidad, esta figura y las siguientes dibujan las curvas como rectas y, en el caso del VPmg, solo su tramo decreciente relevante, el que forma la demanda en la figura A25.4.

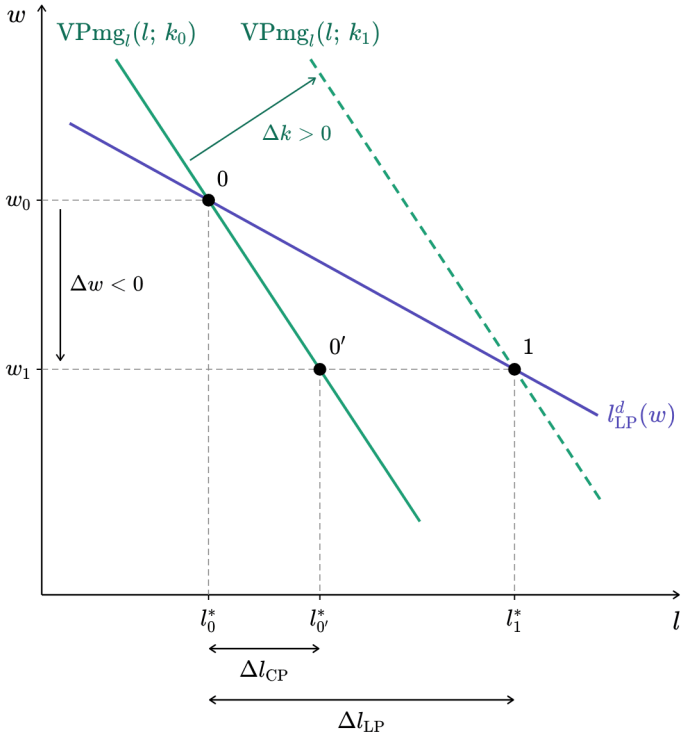


Figura A25.5: Demanda de trabajo a largo plazo. Con capital variable, el efecto sustitución y el efecto escala hacen la demanda más elástica que a corto plazo.

1.3.3. Estática comparativa

Efectos propios: la ecuación de Slutsky para demandas incondicionadas descompone el efecto del salario sobre la demanda de trabajo en un efecto sustitución (a producción constante, la empresa sustituye trabajo por capital, relativamente más barato) y un efecto escala:

$$\frac{\partial l_i^d}{\partial w} = \underbrace{\left.\frac{\partial l_c^d}{\partial w}\right|_{x_0}}_{\mathrm{ES} \leq 0} + \underbrace{\frac{\partial l_c^d}{\partial x} \cdot \frac{\partial x}{\partial w}}_{\mathrm{EP} < 0} < 0$$

donde el ES es no positivo por la concavidad de la función de costes en los precios de los factores y el efecto producción (EP) es negativo porque un factor más caro eleva el coste marginal y reduce la producción. Ambos se mueven en la misma dirección, de modo que una subida del salario reduce inequívocamente la demanda de trabajo.

Efectos cruzados: el efecto del salario sobre la demanda de capital, $\partial k^d / \partial w = \mathrm{ES}^{\geq 0} + \mathrm{EP}^{<0}$, es indeterminado, pues el ES es no negativo (con dos factores, estos son siempre sustitutivos netos) y el EP negativo. No obstante, en las funciones de producción neoclásicas suele predominar el efecto escala y los factores resultan complementarios brutos.

1.3.4. Demanda de trabajo agregada

Función de demanda agregada: cuando cambia el salario, el conjunto de la industria ajusta a la vez su producción, y ese cambio de la producción agregada mueve el precio del producto (una

externalidad pecuniaria). Por eso, a corto y a largo plazo, sumar en horizontal las demandas individuales calculadas a precio constante no basta para obtener la curva del mercado. En su lugar se agregan las *cuasidemandas*, es decir, las demandas de cada empresa una vez incorporado ese efecto sobre el precio, y la curva resultante es menos elástica que aquella suma.

Gráficamente, en el plano trabajo-salario (figura A25.6), se representan la empresa j (panel a) y el mercado (panel b). Si el precio del producto no cambiase, una bajada del salario de w_0 a w_1 llevaría a cada empresa de 0 a $0'$ sobre su curva de VPmg, $\text{VPmg}_l(l_j;\, p_0)$. La suma horizontal de esas respuestas, $\sum_j l_j^d(w; p_0)$, daría el punto $0'$ del mercado. Sin embargo, como todas las empresas contratan más y producen más, el precio cae de p_0 a p_1, la curva de VPmg de cada una gira hacia abajo, hasta $\text{VPmg}_l(l_j;\, p_1)$, y el empleo solo llega a 1. La demanda de mercado $L^d(w)$ une los puntos 0 y 1 y es, por tanto, más inclinada que la suma horizontal.

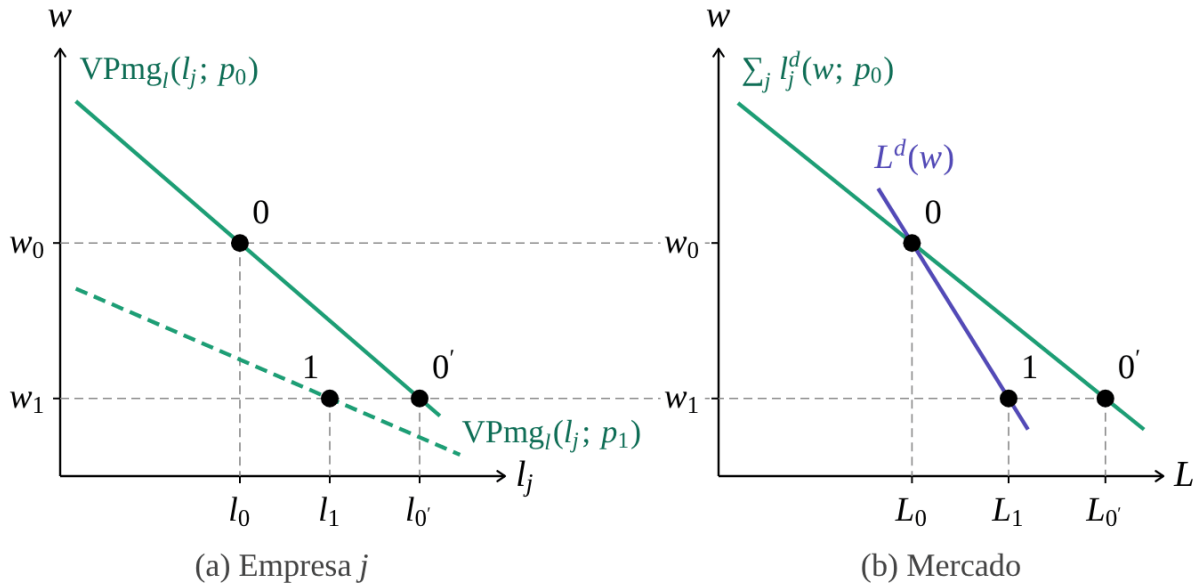


Figura A25.6: Demanda de trabajo de mercado. Ante una bajada del salario de w_0 a w_1, cada empresa pasaría de 0 a $0'$ si el precio del producto no cambiase, pero la caída del precio de p_0 a p_1 hace girar su curva de VPmg y lleva a la empresa a 1 (panel a). La demanda de mercado $L^d(w)$, que une 0 y 1, es más inclinada que la suma horizontal de las demandas individuales a precio constante (panel b).

Elasticidad de la demanda derivada: a largo plazo, en una industria competitiva con rendimientos constantes a escala y oferta de capital perfectamente elástica, la elasticidad de la demanda de trabajo de la industria respecto al salario es

$$\varepsilon_{L^d w} = -\left[(1-s)\,\sigma + s\,|\varepsilon_{Xp}|\right], \qquad s \in (0,1),\ \sigma > 0$$

donde s es la participación del trabajo en el coste total, σ la elasticidad de sustitución entre trabajo y capital y $|\varepsilon_{Xp}|$ la elasticidad-precio de la demanda del producto en valor absoluto.

El primer término recoge el efecto sustitución y el segundo, el efecto escala. De ahí se siguen tres de las reglas de Marshall (1890) y Hicks (1932), según las cuales la demanda de trabajo es más elástica cuanto más fácil es sustituir trabajo por otros factores, cuanto más elástica es la demanda del producto y cuanto mayor es la participación del trabajo en el coste (si $|\varepsilon_{X_p}| > \sigma$). La cuarta, una demanda más elástica cuanto más lo es la oferta de los demás factores, exige relajar el supuesto de oferta de capital perfectamente elástica. Esas reglas explican, por ejemplo, el mayor poder de los sindicatos en industrias con poca sustituibilidad y demanda inelástica (apartado 1.4.2).

1.3.5. Evidencia empírica

La evidencia confirma las predicciones básicas. Así, la demanda de trabajo es decreciente en el salario, con elasticidades de largo plazo claramente mayores en valor absoluto que las de corto plazo. Hamermesh (1993) sitúa las de largo plazo, a producción constante, entre $-0{,}15$ y $-0{,}75$, con $-0{,}3$ como mejor estimación. La sustituibilidad capital-trabajo es, además, mayor para el trabajo poco cualificado, lo que conecta con el cambio técnico sesgado del apartado 3.4.1.

La respuesta de los salarios y del empleo a un aumento de la oferta de trabajo se ha contrastado también con experimentos naturales (apartado 3.4.2). Card (1990) estudia el éxodo del Mariel (1980), que aumentó un 7 % la fuerza de trabajo de Miami en pocos meses, y no encuentra efectos negativos sobre los salarios ni el empleo de los trabajadores poco cualificados residentes.

1.3.6. Conclusiones

Aspectos positivos: la demanda de trabajo se deriva de la maximización del beneficio y hereda las propiedades de las demandas de factores. Es decreciente en el salario real a corto y largo plazo, y en el largo plazo incorpora el efecto sustitución entre factores además del efecto escala.

Limitaciones: supone trabajo homogéneo, ajuste instantáneo del empleo y competencia perfecta en el mercado del producto. Sin embargo, los costes de ajuste (tema A26, sección 3) y el poder de mercado modifican sus conclusiones.

1.4. Equilibrio neoclásico y poder de mercado

Con la oferta y la demanda ya obtenidas, puede caracterizarse el equilibrio competitivo del mercado de trabajo y estudiarse qué ocurre cuando trabajadores o empresas dejan de ser salario-aceptantes.

1.4.1. Equilibrio competitivo

En el mercado, oferta y demanda se cruzan en un único punto, (L^*, w^*). Ese equilibrio es estable y eficiente, y el empleo que fija es el de pleno empleo.

Equilibrio: la existencia y la unicidad se deducen de las pendientes de las curvas en el plano (L, w) (positiva la de oferta, negativa la de demanda). La estabilidad la aporta un salario flexible, que absorbe cualquier exceso de oferta o de demanda. Ninguna empresa tiene poder de mercado, así que todas aceptan el salario dado, y como este resume toda la información sobre productividades y preferencias, mide con exactitud la escasez relativa de trabajo.

Desarrollo gráfico: el equilibrio competitivo se representa en el plano trabajo-salario (figura A25.7):

1. *Oferta de mercado* ($L^s(w)$): recoge las horas que ofrecen los trabajadores a cada salario. Es creciente en el tramo relevante, porque las subidas de salario atraen a nuevos participantes.

2. *Demanda de mercado* ($L^d(w)$): recoge el empleo que desean las empresas a cada salario. Es decreciente, porque el efecto sustitución y el efecto escala actúan en la misma dirección.

3. *Equilibrio*: es su intersección, el punto 0, con salario w^* y empleo L^*. Es estable, pues en los dos casos siguientes el salario vuelve a w^*:

- *Salario superior*: a w', la cantidad ofrecida supera a la demandada y el exceso de oferta presiona el salario a la baja.

- *Salario inferior*: a w'', el exceso de demanda lo presiona al alza.

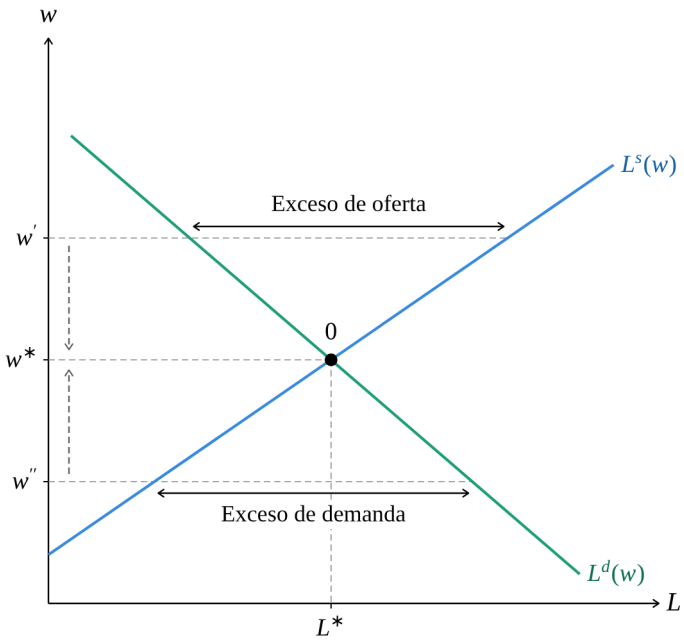


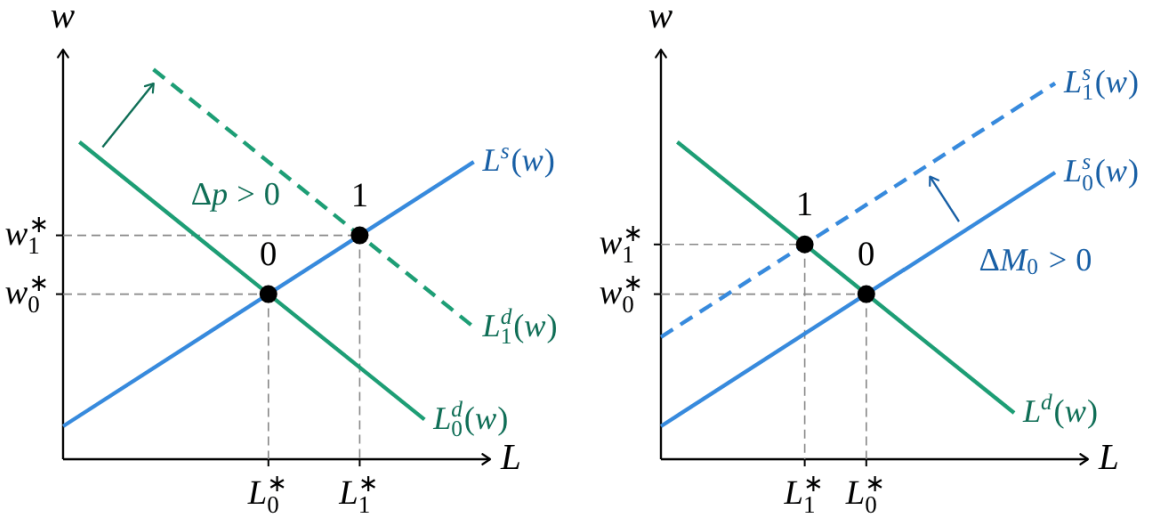
Figura A25.7: Equilibrio competitivo del mercado de trabajo. El cruce de la oferta $L^s(w)$ y la demanda $L^d(w)$ determina el salario w^* y el empleo L^* (punto 0). A un salario superior, w', hay exceso de oferta y, a uno inferior, w'', exceso de demanda, y en ambos casos el salario vuelve a w^*.

Estática comparativa: en el plano trabajo-salario (figura A25.8), los desplazamientos de las curvas alteran el equilibrio:

- *Desplazamiento de la demanda*: una subida del precio del producto o de la productividad del trabajo desplaza la demanda hacia la derecha, de $L_0^d(w)$ a $L_1^d(w)$, y eleva el salario y el empleo de equilibrio (panel a, de 0 a 1).

- *Desplazamiento de la oferta*: un aumento de la renta no salarial, si el ocio es normal, desplaza la oferta hacia la izquierda, de $L_0^s(w)$ a $L_1^s(w)$, de modo que el salario sube y el empleo cae (panel b, de 0 a 1).

- *Papel de las pendientes*: cuanto más rígida es la curva que no se desplaza, más se traduce la perturbación en el salario y menos en el empleo.



(a) Aumento de la demanda

(b) Disminución de la oferta

Figura A25.8: Estática comparativa del equilibrio competitivo. Una subida del precio del producto desplaza la demanda hacia la derecha y eleva el salario y el empleo (panel a). Un aumento de la renta no salarial desplaza la oferta hacia la izquierda, de modo que el salario sube y el empleo cae (panel b).

Bienestar: el equilibrio es eficiente en el sentido de Pareto, pues cumple el primer teorema del bienestar (tema A22). Hay eficiencia técnica (no se desperdician recursos en la producción) y asignativa (el valor del producto marginal del último trabajador contratado iguala a su salario y este a la valoración marginal del ocio). En consecuencia, no existe ninguna reasignación de horas que mejore a alguien sin empeorar a otro.

Sin embargo, la existencia, unicidad y estabilidad del equilibrio no implican que todos trabajen, y se distinguen tres tipos de desempleo:

- *Voluntario*: es el de quienes deciden no participar porque su salario de reserva supera al de mercado, y la teoría neoclásica los considera no participantes, aunque declaren buscar empleo a un salario superior al de mercado.

- *Friccional*: es el desempleo temporal, compatible con el equilibrio, que resulta de la incapacidad de emparejar instantáneamente vacantes y buscadores, y el tema A26 lo endogeniza.

- *Involuntario o clásico*: solo puede existir si alguna rigidez impide la flexibilidad salarial y mantiene el salario por encima del de equilibrio, en cuyo caso el exceso de oferta a w' de la figura A25.7 sería desempleo involuntario.

Como el modelo neoclásico no podía explicar la magnitud y la persistencia del paro de la Gran Depresión, todos los desarrollos posteriores partieron de la teoría neoclásica del mercado de trabajo para modificarla (temas A27 y A28).

Extensiones: el levantamiento del supuesto de salario-aceptación, por uno u otro lado del mercado, se resume en una *condición generalizada de contratación*, en términos de elasticidades:

$$\left(1-\frac{1}{|\varepsilon_{Xp}|}\right)p\cdot\frac{\partial f}{\partial l}=\left(1+\frac{1}{\varepsilon_{lw}}\right)w$$

donde ε_{Xp} es la elasticidad de la demanda del producto y ε_{lw} la de la oferta de trabajo a la que se enfrenta la empresa. Con poder de mercado en cualquiera de los dos lados, el empleo es inferior al competitivo (tema A17). Cabe distinguir tres configuraciones:

- *Competencia perfecta en ambos mercados* ($|\varepsilon_{Xp}|, \varepsilon_{lw}\to\infty$): se recupera $\text{VPmg}_l = w$.
- *Poder de mercado en bienes*: el lado izquierdo es el ingreso del producto marginal, inferior al VPmg (*mark-up*).
- *Poder de mercado en trabajo*: el lado derecho es el gasto marginal, superior al salario (*mark-down*).

1.4.2. Poder de los oferentes de trabajo: sindicatos

Los sindicatos pueden surgir por el **poder compensatorio** (Galbraith, 1952) frente al poder de las grandes empresas.

Planteamiento: el sindicato monopolista elige sobre la curva de demanda de trabajo de la empresa el punto que maximiza su objetivo[6]. Si este es la renta económica de sus afiliados, el excedente del salario sobre el coste de oportunidad del trabajo, su programa es:

$$\max_L\ U = w(L)\cdot L - C(L)$$

donde L es el empleo de los afiliados (variable de decisión), U la renta económica del sindicato, $w(L)$ la demanda inversa de trabajo de la empresa y $C(L)$ el coste de oportunidad de los trabajadores empleados.

CPO:

$$w(L^*)\left(1-\frac{1}{|\varepsilon^d_{Lw}|}\right)=\left.\frac{dC}{dL}\right|_{L^*}=\underline{w}$$

donde $\underline{w}$ es el salario de reserva de los trabajadores. El sindicato iguala el ingreso marginal de la venta de trabajo a su coste marginal, como un monopolista, con un *mark-up* sobre el salario de reserva inverso a la elasticidad de la demanda de trabajo, ε^d_{Lw}.

CSO: el ingreso marginal de la venta de trabajo debe ser decreciente y el coste marginal de oportunidad no decreciente ($C''\geq 0$), como en cualquier monopolio, y bajo ese supuesto la CPO caracteriza el máximo.

Extensiones:

- *Estrategias*: según el objetivo, el sindicato se sitúa en distintos puntos de la demanda de trabajo (figura A25.9, panel (a)):

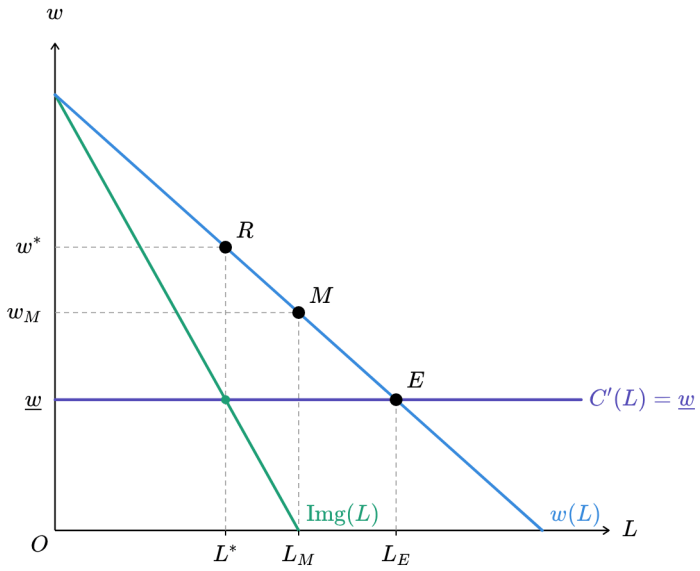
[6] El modelo del *sindicato monopolista* se debe a Dunlop (1944), y el de *contratos eficientes*, a McDonald y Solow (1981).

- *Maximización de la renta económica* (punto R): Img = Cmg, la solución anterior, que genera pérdida de eficiencia y desempleo involuntario.

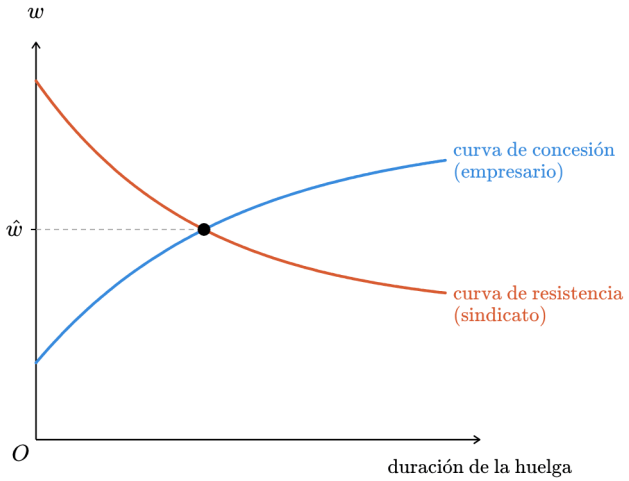
- *Maximización de la masa salarial* (punto M): Img = 0, en el punto de elasticidad unitaria de la demanda de trabajo. El salario es normalmente superior al competitivo (si ese punto queda a la izquierda del cruce oferta-demanda) y, en ese caso, hay desempleo.

- *Maximización del empleo* (punto E): fija el salario de competencia perfecta (no puede obtener mayor empleo sin reducir el salario por debajo del de reserva).

▪ *Negociación*: el equilibrio que finalmente se alcanza depende de cómo se negocie. En el esquema de Hicks (1932), recogido en el panel (b) de la figura A25.9, cuanto más se espera que dure la huelga, menos salario exige el sindicato (curva de resistencia) y más está dispuesto a conceder el empresario (curva de concesión), y el salario queda donde ambas curvas se cortan. Con certidumbre, las dos partes saben cuánto duraría el conflicto, de ahí que pacten ese salario sin llegar a la huelga, que queda como simple amenaza. Los modelos de negociación (derecho a gestionar, contratos eficientes) se desarrollan en el tema A27.



(a) Estrategias del sindicato



(b) Negociación (Hicks, 1932)

Figura A25.9: Objetivos y negociación del sindicato. En el panel (a), sobre la demanda de trabajo, $w(L)$, el sindicato elige R si maximiza la renta económica ($\text{Img}(L) = C'(L) = \underline{w}$), M si maximiza la masa salarial ($\text{Img}(L) = 0$) y E si maximiza el empleo, con el salario competitivo, $\underline{w}$. En el panel (b), la curva de resistencia del sindicato y la curva de concesión del empresario se cortan en el salario pactado, $\hat{w}$.

1.4.3. Poder de los demandantes de trabajo: monopsonio

Con un único demandante de trabajo, la curva de oferta a la que se enfrenta la empresa no es horizontal sino de pendiente positiva en el plano (l, w), porque es la oferta de todo el mercado. Para atraer una unidad más, la empresa debe pagar más a todas las unidades y, por ello, el **gasto marginal** supera al salario:

$$\text{Gmg}_l = w + \frac{\partial w}{\partial l} \cdot l = w\left(1 + \frac{1}{\varepsilon_{lw}}\right) > w$$

donde ε_{lw} es la elasticidad de la oferta de trabajo. La empresa contrata donde $\text{VPmg}_l = \text{Gmg}_l$

(figura A25.10), con menor empleo y menor salario que en competencia perfecta. La brecha entre el VPmg y el salario, el *mark-down*, es la **explotación monopsonística**.

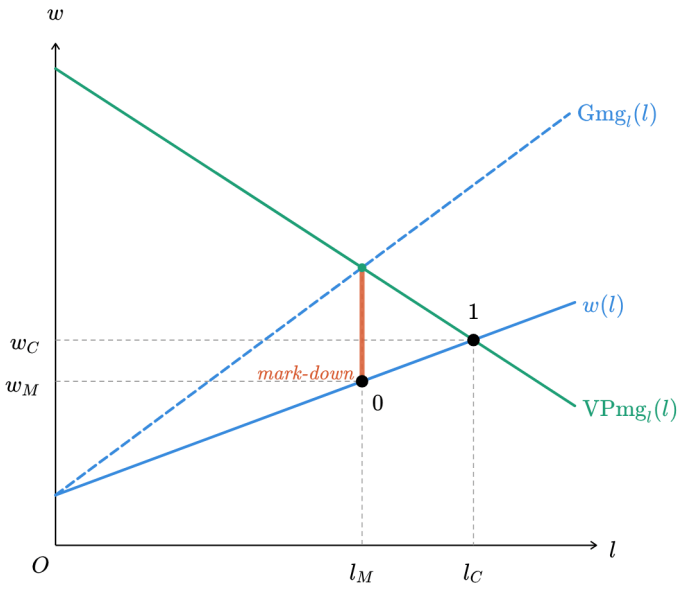


Figura A25.10: Monopsonio en el mercado de trabajo. La empresa contrata l_M, donde $\text{VPmg}_l(l) = \text{Gmg}_l(l)$, y paga el salario que marca la oferta, w_M (punto 0), con menos empleo y menor salario que en competencia, l_C y w_C (punto 1). El segmento vertical en l_M entre $\text{VPmg}_l(l)$ y w_M es el *mark-down* o explotación monopsonística.

Configuraciones de mercado: como la demanda de trabajo es derivada, el equilibrio depende también del poder en el mercado de bienes.

▪ *Competencia perfecta en bienes y monopsonio en trabajo*: el índice de Lerner es nulo, pero la empresa altera el salario con su demanda (salario y empleo inferiores a competencia).

▪ *Monopolio en bienes y monopsonio en trabajo*: se acumulan la explotación monopolística (retribución inferior al VPmg porque el ingreso marginal es inferior al precio) y la monopsonística, de modo que salario y empleo son los más bajos de todas las configuraciones.

▪ *Configuraciones intermedias*: se sitúan entre ambos extremos y, en cualquier caso, cuanto más rígida la demanda de producto, menor el precio pagado al factor.

1.4.4. Monopolio bilateral

Con poder de mercado a ambos lados (un sindicato único frente a una empresa única), el equilibrio no puede determinarse con las herramientas tradicionales. Sin embargo, la teoría acota un rango de salarios posibles, entre el que fijaría el monopsonista y el que fijaría el sindicato, y el punto exacto depende de factores exógenos como la capacidad de negociación de cada parte. Si solo se negocia el salario, el empleo no supera el competitivo (solo lo alcanza si el salario pactado coincide con el competitivo) y, si el salario se fija por encima de este, se produce desempleo involuntario (tema A27).

1.4.5. Evidencia empírica

La evidencia confirma que el poder de mercado es relevante a ambos lados del mercado de trabajo:

1. *Sindicatos*: en EE. UU., los trabajadores sindicados cobran más que los no sindicados de características similares. Esa prima salarial se sitúa en torno al 15 % en 1967-1979 (Lewis, 1986) y al 17-18 % en 1973-2001 (Blanchflower y Bryson, 2004), con una tendencia a la baja en el sector privado.

2. *Monopsonio*: la elasticidad de la oferta de trabajo a la que se enfrenta cada empresa, ε_{lw}, es finita y a menudo reducida (mediana de 1,7 entre las estimaciones que reúnen Sokolova y Sorensen, 2021). Con monopsonio, una elasticidad finita implica salarios inferiores al VPmg. Además, los salarios ofrecidos son menores en los mercados de trabajo locales más concentrados. Así, pasar del percentil 25 al 75 de concentración se asocia a salarios entre un 5 % y un 17 % más bajos (Azar, Marinescu y Steinbaum, 2022).

3. *Salario mínimo*: Card y Krueger (1994) comparan, en un experimento natural (apartado 3.4.2), el empleo en la restauración de Nueva Jersey, que subió el salario mínimo, con el de la vecina Pensilvania, que no lo hizo. No encuentran efectos negativos sobre el empleo, un resultado compatible con el monopsonio.

1.5. Conclusiones

El análisis estático fija el punto de referencia. La oferta de trabajo aplica la teoría del consumidor con dotaciones, y su pendiente es ambigua porque el efecto renta total puede dominar al de sustitución. La demanda, en cambio, cae siempre con el salario, ya que el efecto sustitución y el efecto escala actúan en la misma dirección, y responde más a largo plazo, cuando se ajusta el capital. El equilibrio competitivo es estable y eficiente, y sin rigideces no genera desempleo involuntario.

Sus supuestos (entorno estático, información perfecta, trabajo homogéneo y salario-aceptación) son también su programa de investigación. Cuando trabajadores o empresas dejan de ser salario-aceptantes y solo se fija el salario, el empleo no supera el competitivo, y el desempleo es involuntario si el salario se fija por encima del competitivo. El entorno estático deja sin explicar las fluctuaciones del empleo, que la sección 2 atribuye a la respuesta de la oferta al salario relativo entre periodos. El trabajo homogéneo impide, por su parte, explicar la heterogeneidad de los salarios, objeto de la teoría del capital humano (sección 3).

2. Análisis intertemporal

El análisis intertemporal extiende el modelo estático a varios periodos. Tras fijar sus supuestos, estudia la oferta de trabajo de Lucas y Rapping y la demanda de trabajo con costes de ajuste, y contrasta ambas con la evidencia.

2.1. Supuestos

Metodología: todas las variables se expresan en términos reales, y el modelo es una versión estilizada de Lucas y Rapping (1969), el primero en tratar la oferta de trabajo como una decisión de asignación del tiempo entre periodos.

Entorno: no hay fallos de mercado, y el tiempo es discreto, con horizonte finito T y un tipo de interés real constante y exógeno, r.

Agentes: consumidores-trabajadores sin riqueza inicial (toda la renta es salarial), función de utilidad aditivamente separable en el tiempo[7] con tasa de descuento subjetiva ρ ($\beta = 1/(1+\rho)$), dotación de tiempo normalizada ($o_t + l_t = 1$).

2.2. Oferta de trabajo intertemporal (Lucas y Rapping, 1969)

Origen: Robert Lucas y Leonard Rapping (1969) trasladaron la elección ocio-consumo del análisis estático a un horizonte de varios periodos, y su modelo es el antecedente de la teoría del ciclo real de Kydland y Prescott (tema A42).

Idea: la oferta de trabajo responde al salario relativo entre periodos y no a su nivel, de modo que el individuo concentra su trabajo en los periodos en que está mejor pagado (*sustitución intertemporal del ocio*).

Planteamiento: el consumidor-trabajador maximiza la utilidad intertemporal sujeto a la restricción presupuestaria intertemporal, en la que el valor actual del consumo iguala al valor actual de la renta salarial:

$$\max_{\{c_t, o_t\}} U = \sum_{t=0}^{T} \beta^t u(c_t, o_t) \quad \text{s.a.:} \sum_{t=0}^{T} \frac{c_t}{(1+r)^t} = \sum_{t=0}^{T} \frac{w_t(1-o_t)}{(1+r)^t}, \qquad \beta \in (0,1)$$

donde c_t es el consumo, o_t el ocio, w_t el salario real y $1 - o_t = l_t$ el trabajo del periodo t.

CPO: con solución interior y la restricción agotada, derivando el lagrangiano respecto al consumo y al ocio de cada periodo y eliminando el multiplicador, se obtienen tres condiciones:

$$\text{RMS}_{c_t o_t} = \frac{\partial u/\partial o_t}{\partial u/\partial c_t} = w_t \qquad \text{(intratemporal: idéntica al caso estático)} \tag{6}$$

$$\frac{\partial u}{\partial c_t} = \frac{1+r}{1+\rho} \frac{\partial u}{\partial c_{t+1}} \qquad \text{(ecuación de Euler del consumo)} \tag{7}$$

$$\frac{\partial u/\partial o_t}{\partial u/\partial o_{t+1}} = \frac{1+r}{1+\rho} \cdot \frac{w_t}{w_{t+1}} \qquad \text{(sustitución intertemporal del ocio)} \tag{8}$$

[7] La separabilidad temporal es un supuesto restrictivo, porque excluye la inercia y los hábitos de consumo.

donde (6) reproduce la tangencia del modelo estático en cada periodo; (7) iguala la RMS intertemporal del consumo, $(\partial u/\partial c_t)/[\beta\,\partial u/\partial c_{t+1}]$, al precio relativo del consumo presente, $1+r$; y (8), la novedad del modelo, hace lo propio con el ocio, cuyo precio relativo entre periodos es el salario relativo corregido por el tipo de interés.

CSO: la concavidad estricta de u en cada periodo y la separabilidad temporal hacen estrictamente cóncava la utilidad intertemporal U, de modo que, con una restricción lineal, las condiciones (6)–(8) son necesarias y suficientes.

Solución: se obtiene la función de oferta de trabajo relativa. De (8), con utilidad separable entre consumo y ocio, la oferta de trabajo de un periodo respecto al siguiente es:

$$\frac{l_t^s}{l_{t+1}^s} = f\left[\left(\frac{w_t}{w_{t+1}}\right)^+,\ r^+,\ \rho^-\right]$$

donde depende positivamente del salario relativo (trabajar más cuando el salario es relativamente alto), positivamente del tipo de interés (el ahorro del salario presente rinde más) y negativamente de la tasa de descuento (la impaciencia resta valor al futuro). La *elasticidad de sustitución intertemporal* de la oferta de trabajo, la elasticidad de l_t/l_{t+1} respecto a w_t/w_{t+1}, es el parámetro que determina cuánto responde la oferta a los cambios transitorios del salario.

Estática comparativa: ante perturbaciones, lo decisivo es distinguir entre cambios permanentes y transitorios.

1. *Salario*:

- *Aumento permanente*: eleva w_t y w_{t+1} en la misma proporción, no altera el salario relativo y, por tanto, elimina la sustitución intertemporal. Quedan solo los efectos estáticos de cada periodo (sustitución intratemporal ocio-consumo y renta), de signo ambiguo como en el análisis estático y que en la práctica tienden a compensarse, de modo que el empleo apenas varía.

- *Aumento transitorio*: eleva el salario relativo y, con ello, el individuo concentra su trabajo en el periodo en que está mejor pagado. La oferta aumenta durante la perturbación y cae después.

2. *Tipo de interés real*: un aumento permanente de r hace más atractivo trabajar y ahorrar hoy para consumir mañana, y aumenta la oferta de trabajo relativa presente.

3. *Tasa de descuento*: un aumento permanente de ρ hace que el futuro valga menos y que el individuo trabaje menos hoy (efecto *carpe diem*), lo que reduce la oferta relativa presente.

4. *Perturbaciones tecnológicas*: las permanentes elevan los salarios proporcionalmente en todos los periodos y no alteran la oferta relativa, mientras que las transitorias sí la alteran, y en ello se apoya la teoría del ciclo real.

Conclusiones:

- *Aspectos positivos*: el modelo convirtió la oferta de trabajo en una decisión intertemporal. Es, además, la base de los modelos del ciclo real de Kydland y Prescott y de la calibración como método (tema A42). En ellos, las fluctuaciones del empleo se explican por la sustitución intertemporal del trabajo ante perturbaciones reales, con la conducta siempre óptima y el desempleo siempre voluntario.

▪ *Limitaciones*: son sobre todo empíricas. Para generar los ciclos observados, el modelo exige cambios salariales transitorios grandes y una elasticidad de sustitución intertemporal elevada, cuando las estimaciones microeconómicas la sitúan por debajo de la unidad y los salarios tienen un componente transitorio reducido[8]. Además, sin costes de ajuste, no explica la persistencia del desempleo.

2.3. Demanda de trabajo intertemporal

Del lado de la demanda, la extensión intertemporal descansa en los **costes de ajuste**, que encarecen contratar y despedir (búsqueda, selección, formación, despido). Una versión estilizada de la decisión de contratar permite comparar el caso sin costes de ajuste, en el que la demanda de cada periodo es la estática, con el caso con costes de ajuste, en el que la empresa reparte el ajuste en el tiempo.

Planteamiento: la empresa elige la senda de empleo que maximiza el valor actual de sus beneficios reales, con el capital fijo y un coste cuadrático de variar la plantilla:

$$\max_{\{l_t\}} V_0 = \sum_{t=0}^{T} \frac{1}{(1+r)^t} \left[f(l_t) - w_t l_t - \frac{b}{2}(l_t - l_{t-1})^2 \right]$$

donde l_t es el empleo del periodo t (variable de decisión, con l_{-1} dado), V_0 el valor actual de los beneficios, $f(l_t)$ la función de producción con el capital fijo (el precio del producto se normaliza a la unidad), w_t el salario real, r el tipo de interés real y $b \geq 0$ el parámetro del coste de ajuste. El caso sin costes de ajuste corresponde a $b = 0$.

CPO: derivando respecto al empleo de cada periodo se obtiene una condición de contratación intertemporal:

$$\underbrace{\frac{df}{dl}\Big|_{l_t}}_{\text{Pmg}_l} = w_t + \underbrace{b\,(l_t - l_{t-1}) - \frac{b}{1+r}\,(l_{t+1} - l_t)}_{\text{Cmg}_{\text{ajuste}}} \tag{9}$$

donde (9) iguala la productividad marginal del trabajo al salario más el coste marginal de ajustar la plantilla en t, neto del ahorro descontado en el coste de ajuste de $t+1$. Su lectura depende de los costes de ajuste:

▪ *Sin costes de ajuste* ($b = 0$): la condición (9) se reduce a la estática, $df/dl|_{l_t} = w_t$, en cada periodo. El problema es separable en el tiempo y el empleo salta a su nivel óptimo ante cualquier perturbación.

▪ *Con costes de ajuste* ($b > 0$): la condición (9) enlaza periodos consecutivos, porque contratar hoy encarece el ajuste de hoy y abarata el de mañana. No obstante, cuando el empleo se estabiliza ($l_{t+1} = l_t = l_{t-1}$), (9) recupera la condición estática.

CSO: con $f_l'' < 0$ y $b \geq 0$, el beneficio descontado es estrictamente cóncavo en la senda de empleo, de modo que la condición (9) es suficiente.

Solución: sin costes de ajuste, la demanda de trabajo de cada periodo es la estática, $l_t^d = l^d(w_t)$, y su dinámica se reduce a la de sus determinantes. Con costes de ajuste convexos, en cambio, el empleo de cada periodo depende también del heredado y de los salarios esperados, de modo que

[8] Subidas salariales hacen esperar aumentos futuros, no reducciones, por lo que no incentivan desplazar la oferta al presente.

ante una perturbación permanente se ajusta gradualmente hacia el nuevo nivel estático, tanto más despacio cuanto mayor es b.

Extensiones: si los costes de ajuste son lineales o fijos, la empresa no ajusta el empleo ante perturbaciones pequeñas. Sea cual sea su forma, los costes de ajuste quitan flexibilidad a la empresa, así que el empleo oscila menos que el producto y el desempleo tarda más en corregirse. Sin embargo, el efecto sobre el nivel medio del empleo es en principio ambiguo, porque los costes de ajuste frenan tanto las contrataciones como los despidos (en Bentolila y Bertola, 1990, los costes de despido elevan ligeramente el empleo medio).

> La demanda de trabajo sin costes de ajuste y con costes de ajuste convexos, lineales y fijos, en tiempo continuo, se desarrolla en el tema A26, apartado 3.2.

2.4. Evidencia empírica

La evidencia contrasta la predicción central del modelo de Lucas y Rapping, la respuesta del empleo a los cambios transitorios del salario. Las estimaciones microeconómicas de la elasticidad de sustitución intertemporal de la oferta de trabajo son reducidas, por debajo de la unidad (MaCurdy, 1981; Altonji, 1986), lo que dificulta explicar con ella las fluctuaciones observadas del empleo. Los datos agregados muestran, además, dos regularidades:

- *Corto plazo*: el empleo es marcadamente procíclico y el salario real solo moderadamente, por las rigideces salariales. Por tanto, la mayor parte del ajuste cíclico se produce en cantidades, en contra de lo que predice el modelo con salarios flexibles.

- *Largo plazo*: la tendencia del empleo la marcan la productividad, que eleva los salarios reales con retraso, y la evolución de la participación.

2.5. Conclusiones

El análisis intertemporal añade al estático la dimensión de la asignación del tiempo entre periodos. Su resultado central es que la oferta de trabajo responde al salario relativo y no al nivel, de modo que las perturbaciones transitorias mueven el empleo y las permanentes apenas lo hacen. En ello descansan tanto la teoría del ciclo real como sus críticas.

Del lado de la demanda, la dinámica procede de los costes de ajuste. Sin embargo, oferta y demanda comparten con el modelo estático el supuesto de trabajo homogéneo, que levanta la teoría del capital humano (sección 3).

3. Teoría del capital humano

Además de las teorías anteriores, la economía laboral se apoya en la noción de **capital humano**. La idea tiene precedentes en Smith (el trabajador cualificado como una máquina costosa), pero su formulación moderna, de finales de los años cincuenta y de los sesenta, se debe a la escuela de Chicago.

Sus contribuciones fundacionales son las siguientes:

- *Primer modelo formal* (Mincer, 1958): Jacob Mincer trata la formación como una inversión que el individuo realiza a cambio de salarios futuros más altos. Mincer deriva de ahí la distribución de la renta. Llega a una relación entre años de escolarización y salarios que, con la experiencia, será la ecuación de Mincer.

- *Concepto y programa de investigación* (Schultz, 1961): Theodore Schultz, en su discurso presidencial ante la American Economic Association, convierte la idea de Mincer en una categoría de la contabilidad del crecimiento.

Ello llevó a Schultz a postular que la rápida recuperación de Alemania y Japón tras la Segunda Guerra Mundial, pese a la destrucción de su capital físico, solo se explica por el capital humano de su población. Sostuvo también que buena parte del crecimiento no atribuible a capital y trabajo es rendimiento de la inversión en educación y salud.

- *Teoría microeconómica* (Becker, 1964): Gary Becker trata la educación y la formación como una decisión de inversión, con costes presentes y rendimientos futuros, y distingue la formación general de la específica.

- *Capital humano y desequilibrio* (Nelson y Phelps, 1966; Schultz, 1975): Nelson y Phelps formalizan la idea de que el capital humano acelera la adopción de nuevas tecnologías, de modo que su rendimiento es mayor cuanto más rápido es el cambio técnico. Schultz generaliza esa idea como la capacidad de reasignar recursos ante el desequilibrio. Esta línea enlaza con los modelos de crecimiento endógeno (tema A44).

3.1. Capital humano como inversión (Becker, 1964)

Origen: Gary Becker (*Human Capital: A Theoretical and Empirical Analysis*, 1964) dio forma microeconómica a la noción de Schultz, aplicando el análisis intertemporal a la decisión individual de formarse.

Idea: los apartados anteriores tratan a todos los trabajadores como iguales y les asignan un solo salario de equilibrio. Becker, en cambio, recurre al análisis intertemporal para explicar por qué difieren sus salarios. Educarse o formarse es invertir en capital humano, con un coste presente y un rendimiento futuro, el mayor salario que trae una productividad más alta.

3.1.1. Supuestos

Entorno: información perfecta, dos periodos, mercados de capitales completos y perfectos (tipo de interés r), de modo que se cumple el teorema de separación, según el cual la decisión de inversión es independiente de la de consumo.

Agentes: el agente, además de consumo y ahorro, decide si educarse ($e = 1$) o no ($e = 0$). El modelo hace abstracción de la decisión de ocio.

Tecnología: la educación cuesta C_e, suma de un coste explícito (matrícula y materiales), otro de oportunidad (el salario al que se renuncia) y otro psicológico. A cambio, eleva la productividad, y por eso quien se educa cobra más que quien no, $w_e > w_n$.

3.1.2. Modelo

Planteamiento: el agente elige el consumo de ambos periodos y si educarse o no:

$$\max_{c_1,\, c_2,\, e\in\{0,1\}} u(c_1, c_2) \quad \text{s.a.:} \quad c_1 + \frac{c_2}{1+r} \leq \underbrace{y_0 + \frac{w_n}{1+r}}_{W_n} + e \underbrace{\left(\frac{w_e - w_n}{1+r} - C_e\right)}_{\text{VAN}_e}$$

donde y_0 es la renta del primer periodo, W_n la riqueza sin educación y VAN_e el valor actual neto de educarse, es decir, el diferencial salarial descontado menos el coste. Como e solo entra en la restricción, y lo hace sumando riqueza, se aplica el teorema de separación de Fisher (tema A29).

CPO: el problema se resuelve en dos etapas.

Decisión de inversión: primero, el agente elige e para maximizar su riqueza, con independencia de sus preferencias:

$$e^* = 1 \iff \text{VAN}_e \geq 0 \iff C_e \leq \frac{w_e - w_n}{1+r}$$

Es decir, se educa si el diferencial salarial descontado cubre el coste. La decisión depende directamente del diferencial $w_e - w_n$ e inversamente del tipo de interés (coste de financiación) y del coste de la educación. Se aplica así la regla del VAN de un proyecto de inversión (costes directos e indirectos frente a rendimientos futuros).

Decisión de consumo: después, dada la riqueza $W_n + e^*\text{VAN}_e$, reparte el consumo entre periodos con la condición habitual, $\text{RMS}_{c_2c_1} = (\partial u/\partial c_1)/(\partial u/\partial c_2) = 1 + r$. Si el agente es un padre que decide por su hijo, la misma condición reparte el consumo entre generaciones y la educación se financia con la renta del padre, lo que anticipa el papel de las restricciones de crédito.

CSO: la etapa de inversión es una elección discreta y no requiere condición de segundo orden, sino que basta comparar las dos riquezas. En la de consumo, en cambio, la cuasiconcavidad estricta de u y la linealidad de la restricción garantizan que la tangencia es el máximo.

Equilibrio: la inversión en educación da lugar a trabajadores heterogéneos y a un mercado diferente para cada tipo. Se sigue contratando por productividad marginal, pero ya no hay un único salario de equilibrio.

Desarrollo gráfico: en el plano edad-ingresos, la decisión se representa con los perfiles de ingresos por edad con y sin educación (figura A25.11):

1. *Perfil sin educación*: el individuo percibe w_n desde que empieza a trabajar.

2. *Perfil con educación*: durante los estudios (periodo 1) no percibe salario y paga además el coste directo, de modo que sus ingresos netos son negativos, y durante la vida laboral (periodo 2) percibe $w_e > w_n$.

3. *Áreas de costes y de rendimientos*: el área de costes suma el coste directo y el de oportunidad (el salario w_n al que renuncia mientras estudia), y el área de rendimientos es el diferencial $w_e - w_n$ durante la vida laboral.

4. *Óptimo*: el individuo se educa si el valor actual del área de rendimientos cubre el área de costes, que es la condición $\text{VAN}_e \geq 0$ de la decisión de inversión.

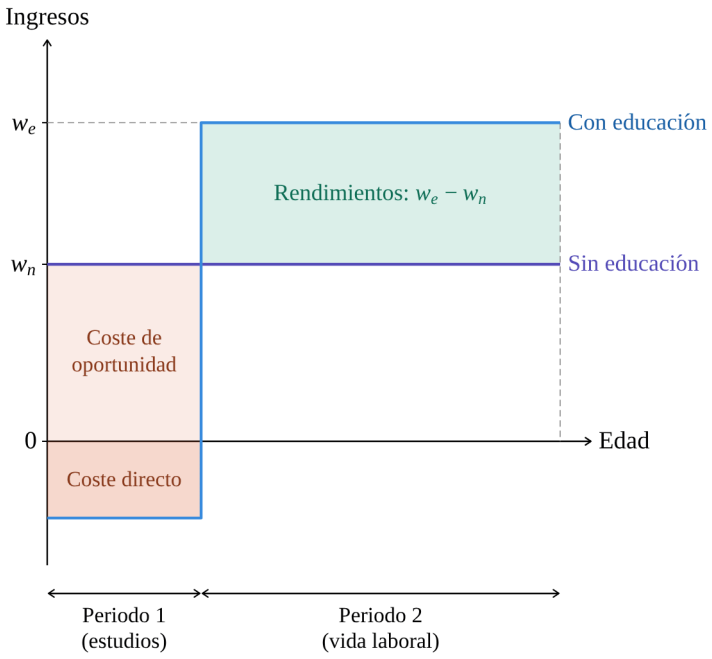


Figura A25.11: Perfiles de ingresos por edad con y sin educación. Durante los estudios, el individuo que se educa soporta el coste directo y el de oportunidad (el salario w_n al que renuncia), y durante la vida laboral obtiene el rendimiento $w_e - w_n$. Se educa si el valor actual de los rendimientos cubre los costes.

Extensiones: Becker separa la formación *general*, útil en cualquier empresa, de la *específica*, valiosa solo en la empresa que la imparte. La general la paga el propio trabajador, porque a ninguna empresa le compensa costear lo que otras pueden aprovechar[9]. En la específica, en cambio, la empresa reparte con el trabajador costes y rendimientos para frenar la rotación.

3.1.3. Conclusiones

Aspectos positivos: el modelo reduce la decisión de formarse a la regla del VAN y ofrece predicciones contrastables (más educación cuanto mayor es el diferencial salarial y menores son el tipo de interés y el coste). Las predicciones han guiado la literatura empírica que se expone en el apartado 3.4. La distinción entre formación general y específica explica, además, quién financia cada una y la menor rotación de los trabajadores con formación específica.

Limitaciones: en su versión básica, el modelo deja fuera las restricciones al crédito, que impiden a las familias sin renta financiar una inversión rentable, y las externalidades del capital humano. Esas dos circunstancias justifican dedicar recursos públicos a la educación, y al omitirlas el modelo no ofrece fundamento para financiar la formación con fondos públicos. La teoría de la señalización discute, además, que la relación entre educación y productividad sea tan directa.

[9] En la práctica, la empresa cofinancia la formación general cuando el mercado de trabajo es imperfecto (Acemoglu y Pischke, 1999).

3.2. Otras aportaciones

Capital humano como señal (Spence, 1973): Michael Spence invierte la causalidad de Becker. En el modelo de Spence, el valor de la educación no está en que eleve la productividad, sino en que, con información asimétrica, sirve al trabajador de *señal* de esa productividad[10].

Como educarse cuesta menos a los más productivos, solo ellos encuentran rentable alcanzar el nivel de educación que las empresas exigen para pagar el salario alto, y el mercado se separa por tipos aunque la educación no añada nada a la productividad.

Si la educación es sobre todo señal, su rendimiento privado supera al social y la financiación pública deja de estar justificada por la productividad (el modelo se desarrolla en el tema A13, apartado 3.4.1).

Crítica marxista (Bowles y Gintis, 1975): Samuel Bowles y Herbert Gintis sostienen que el sistema educativo produce, y el mercado remunera, la capacidad de trabajar en organizaciones capitalistas y obedecer órdenes, y no productividad técnica. Para estos autores, la educación reproduce así la estructura de clases.

Capital humano y crecimiento: la agregación de las decisiones de Becker es la base de los modelos de crecimiento con capital humano (Lucas, 1988; Mankiw, Romer y Weil, 1992), que se desarrollan, respectivamente, en los temas A44 y A43.

3.3. La función de ingresos de Mincer (1958, 1974)

Origen: Jacob Mincer dio a la teoría del capital humano su forma empírica. Su ecuación, que relaciona el logaritmo de los ingresos con los años de educación y de experiencia, se ha convertido en la relación más estimada de la economía laboral.

Idea: si los individuos eligen su educación como una inversión y los mercados de trabajo son competitivos, el salario de cada nivel educativo ha de compensar exactamente el coste de alcanzarlo. De esa condición de equilibrio se deduce la forma de la relación entre educación e ingresos.

Razonamiento: considérese un individuo que decide entre S y $S+1$ años de educación. Se supone que el único coste de un año más es el salario al que renuncia durante ese año (los costes directos son pequeños en comparación) y que la vida laboral es larga. En ese caso, el año adicional es rentable si el aumento proporcional del salario compensa el año de ingresos perdido al tipo de interés r.

En el margen, con individuos idénticos y libre elección, esa compensación es exacta, de modo que $w_{S+1} = w_S(1+r)$ y, tomando logaritmos, $\ln w_S = \ln w_0 + S\ln(1+r) \approx \ln w_0 + rS$. Así, los ingresos crecen exponencialmente con los años de educación, y la pendiente de la relación en logaritmos, r, es a la vez el tipo de interés y la *tasa de rendimiento de la educación*.

Mincer añadió la inversión en formación en el puesto de trabajo, que se concentra al principio de la vida laboral y decrece después, lo que hace que los ingresos crezcan con la experiencia X a un ritmo decreciente. El resultado es la *función de ingresos de Mincer*:

$$\ln w = \beta_0 + \beta_1 S + \beta_2 X + \beta_3 X^2 + \nu, \qquad \beta_1 > 0,\ \beta_2 > 0,\ \beta_3 < 0$$

donde β_0 es el logaritmo del salario de un trabajador sin educación ni experiencia, β_1 la tasa de rendimiento de un año de educación, β_2 y β_3 recogen el perfil cóncavo de los ingresos con la

[10] Spence (1973 y 1974).

experiencia (crecen, alcanzan un máximo y pueden decrecer) y ν las diferencias no observadas de habilidad, suerte y condiciones de trabajo.

Conclusiones:

- *Aspectos positivos*: la ecuación tiene tres lecturas que explican su éxito:

  - *Distribución de la renta*: la dispersión de los ingresos se explica por la de la educación y la experiencia, y el perfil edad-ingresos es la traza de la inversión en capital humano a lo largo de la vida.

  - *Evaluación*: β_1 mide lo que rinde un año de escuela, comparable con el rendimiento del capital físico y entre países.

  - *Regresión*: es estimable con datos individuales, y sus limitaciones (variables omitidas, sesgo de selección, educación tratada como exógena) son el programa de la literatura empírica que se expone a continuación.

- *Limitaciones*: la lectura de β_1 como tasa de rendimiento solo es exacta bajo los supuestos de la derivación (individuos idénticos, coste limitado al salario perdido y vida laboral larga). Sin ellos, β_1 mide solo la diferencia porcentual media del salario asociada a un año más de educación, y la forma lineal en S y cuadrática en X es una aproximación que la evidencia no siempre respalda.

3.4. Evidencia empírica

La evidencia empírica sobre el capital humano abarca las estimaciones de los rendimientos de la educación y la aportación de los experimentos naturales.

3.4.1. Rendimientos de la educación

Estimaciones: los valores de β_1 obtenidos por mínimos cuadrados se sitúan, a igual experiencia, entre el 5 y el 10 % del salario por año de educación en la mayoría de los países.

Psacharopoulos (1994; con Patrinos, 2018) documenta además una regularidad en más de un centenar de países. Los rendimientos sociales son mayores en los países de renta baja, donde la educación es escasa, y, dentro de ellos, mayores para la primaria que para la secundaria y la superior. En los países desarrollados, en cambio, el rendimiento de la educación superior ha crecido desde los años ochenta con el cambio técnico sesgado hacia la cualificación. Ese cambio técnico ha aumentado la dispersión de los salarios pese al aumento de la oferta de titulados (Katz y Murphy, 1992).

En España, las estimaciones se sitúan en el rango bajo del intervalo, entre el 5 y el 8 %, con rendimientos mayores para la educación universitaria que para la secundaria. La *sobreeducación* reduce el rendimiento de los años que exceden los requeridos por el puesto (Duncan y Hoffman, 1981), un fenómeno especialmente estudiado en España.

El sesgo de habilidad: la estimación por mínimos cuadrados puede estar sesgada, porque la habilidad no observada afecta a la vez a la educación elegida y al salario, de modo que parte del rendimiento atribuido a la educación es rendimiento de la habilidad.

Sin embargo, los estudios de gemelos (Ashenfelter y Krueger, 1994), que comparan gemelos idénticos con distinta educación, obtienen rendimientos similares o superiores a los de mínimos cuadrados. El resultado sugiere que el sesgo de habilidad es pequeño y que lo compensa el error de medida de la educación.

La conclusión de la literatura, resumida por Card (1999), es que el rendimiento causal de la educación es real y del orden del 7 al 10 %.

Educación como señal: los intentos de contrastar Becker frente a Spence han buscado si el mercado paga los títulos más que los años. Destacan tres resultados, que indican que la educación es mayoritariamente productiva, con un componente de señal:

- *Efecto piel de oveja* (*sheepskin effect*; Hungerford y Solon, 1987): el año en que se obtiene un título rinde más que los intermedios, lo que es coherente con la señalización.

- *Rendimientos y experiencia*: los rendimientos de la educación no caen con la experiencia (Layard y Psacharopoulos, 1974), cuando una señal pura predeciría que cayeran a medida que la empresa observa la productividad.

- *Autónomos* (Wolpin, 1977): quienes no necesitan señalar a nadie obtienen rendimientos similares a los de los asalariados.

3.4.2. Experimentos naturales (Card, Krueger y Angrist, años noventa)

La aportación de David Card, Alan Krueger y Joshua Angrist (Nobel 2021 a Card, Angrist e Imbens) a la economía laboral fue metodológica, ya que sustituyeron las correlaciones de las regresiones de Mincer por *experimentos naturales*.

Se trata de situaciones en las que una ley, una frontera o el azar asignan el "tratamiento" con independencia de las características de los individuos, lo que permite establecer relaciones de causalidad (tema A01, apartado 3.1.3).

En la economía de la educación, sus resultados fijaron dos regularidades (los obtenidos sobre la inmigración y el salario mínimo se recogen en los apartados 1.3.5 y 1.4.5):

1. *Años de educación e ingresos*: Angrist y Krueger (1991) usan el trimestre de nacimiento y las leyes de escolarización obligatoria como instrumento (los nacidos a principios de año pueden abandonar antes) y Card (1995) la proximidad a una universidad. Sus estimaciones del rendimiento de un año de educación igualan o superan las de mínimos cuadrados (en torno al 13 % en Card, frente al 7 % por mínimos cuadrados).

2. *Recursos escolares*: Card y Krueger (1992) muestran que la calidad de la escuela (ratio alumnos-profesor, duración del curso, salario docente) en el estado de origen afecta al rendimiento de la educación en el mercado de trabajo, comparando cohortes y estados.

3.5. Conclusiones

Aspectos positivos: la función de Mincer convirtió la teoría del capital humano en una hipótesis contrastable que la evidencia de cincuenta años ha confirmado en lo esencial. La educación tiene un rendimiento causal, comparable al del capital físico, que explica una parte sustancial de la desigualdad salarial y de las diferencias de renta entre países y fundamenta la inversión pública en educación cuando existen restricciones de crédito o externalidades (tema A44).

Limitaciones: la evidencia sobre el capital humano tiene cuatro límites principales:

- *Rendimientos privados*: la ecuación de Mincer no mide los rendimientos sociales ni los no monetarios (salud, participación cívica).

- *Habilidad no observada*: la estimación de la ecuación de Mincer exige resolver este problema, y las distintas estrategias no siempre coinciden.

- *Homogeneidad*: la ecuación trata la educación como una cantidad homogénea de años, cuando la calidad y el tipo de estudios importan tanto como su duración.

- *Validez externa*: la evidencia de los experimentos naturales, siendo la más creíble, es local (válida para la población afectada por el experimento) y no siempre generalizable.

Conclusiones

La teoría neoclásica del mercado de trabajo parte de las decisiones individuales. La oferta resulta de la elección entre ocio y consumo, y su pendiente depende del pulso entre el efecto sustitución y el efecto renta total, que puede volver la curva hacia atrás. La demanda, derivada de la maximización del beneficio, es, en cambio, decreciente en el salario y más elástica a largo plazo. De su cruce surge un equilibrio único, estable y eficiente, en el que, sin fricciones, solo cabe el desempleo voluntario. El análisis intertemporal explica las fluctuaciones del empleo por la respuesta de la oferta al salario relativo entre periodos. Por último, la teoría del capital humano explica la heterogeneidad de los salarios como resultado de decisiones de inversión, con un rendimiento causal de la educación que la evidencia confirma (Card, 1999).

La teoría neoclásica del mercado de trabajo aplica la teoría del consumidor con dotaciones, en la que el tiempo es la dotación y el ocio un bien más (tema A08), y la de la empresa competitiva (tema A16). Con un único comprador, el análisis es el del monopsonio (tema A17). El capital humano tiene dos prolongaciones, la señalización, en la que la educación revela la productividad sin aumentarla (tema A13), y el crecimiento endógeno, en el que la acumulación de conocimiento sostiene el crecimiento (temas A43 y A44). Por su parte, las limitaciones del modelo abren los temas siguientes, con las fricciones de búsqueda (tema A26), las rigideces salariales (tema A27) y la tasa de desempleo de equilibrio (tema A28). La sustitución intertemporal de Lucas y Rapping sustenta, además, la teoría del ciclo real (tema A42).

Card (2022) sitúa el poder de las empresas para fijar salarios en el centro de la economía laboral. Por ejemplo, Cengiz, Dube, Lindner y Zipperer (2019) encuentran que las subidas del salario mínimo en EE. UU. apenas han reducido el empleo de bajo salario, un resultado compatible con ese poder. También se discute la medición de las respuestas de la oferta. Chetty, Guren, Manoli y Weber (2011) muestran que las elasticidades de largo plazo de la oferta estimadas con microdatos son coherentes con las agregadas, pero las de sustitución intertemporal son un orden de magnitud menores que las que exigen los modelos del ciclo real. Finalmente, Acemoglu y Autor (2011) explican con un modelo de tareas cómo el cambio técnico puede desplazar a los trabajadores de cualificación media y polarizar el empleo.

A26. Desempleo friccional

Desempleo friccional. La curva de Beveridge, el modelo de búsqueda y emparejamiento de Diamond, Mortensen y Pissarides. Costes de ajuste y dinámica de la demanda de trabajo.

Introducción

> **Introducción común.** Los temas A25 a A28 comparten una introducción, que se desarrolla al comienzo del tema **A25**. Presenta la teoría económica del mercado de trabajo, que estudia el empleo y los salarios de equilibrio y los factores institucionales que los moldean, y la explicación teórica de cada tipo de desempleo.

En la **teoría neoclásica tradicional** (tema A25) todos los trabajadores y todos los puestos son iguales y la información es completa, así que, sin fricciones que frenen el encuentro entre oferta y demanda, el desempleo solo puede ser voluntario. Los datos muestran, en cambio, que conviven personas que buscan trabajo y puestos vacantes. Con información imperfecta aparece un problema de extracción de señales[1]. Así, los agentes han de buscar, y como buscar cuesta, el ajuste deja de ser automático.

En la **teoría de la búsqueda**, McCall (1970) y Mortensen (1970) modelizaron primero la decisión del trabajador, que acepta una oferta solo si el salario supera su **salario de reserva**. Diamond, Mortensen y Pissarides, premiados con el Nobel en 2010, añadieron después las vacantes de las empresas y una **función de emparejamiento** que resume lo difícil que es que unos y otras se encuentren. De ahí surge la **curva de Beveridge**, la relación negativa entre desempleo y vacantes, que permite distinguir una caída de la demanda, que mueve la economía a lo largo de la curva, de un peor emparejamiento, que la desplaza. El modelo explica también por qué el desempleo de equilibrio nunca es nulo, ya que siempre hay trabajadores que cambian de empleo o que se incorporan al mercado. Las fricciones afectan además a las empresas, que soportan **costes de ajuste** al contratar y despedir y adaptan su plantilla con lentitud.

Estos modelos son hoy el marco estándar del mercado de trabajo en la macroeconomía (tema A28) y en la evaluación de políticas. La cuantía y la duración de las prestaciones por desempleo alteran el salario de reserva y la intensidad de la búsqueda, las políticas activas de empleo mejoran el emparejamiento y la protección del empleo frena tanto la destrucción como la creación de puestos. Tras la pandemia, la combinación de desempleo bajo y vacantes muy altas en Estados Unidos (2021-2022) reavivó el debate sobre los desplazamientos de la curva de Beveridge.

El tema estudia el **desempleo friccional**, el que surge porque trabajadores y vacantes tardan en encontrarse cuando la información es imperfecta, y la dinámica de la demanda de trabajo. Para ello, el tema se organiza en tres partes:

[1] Quien ve variar su salario nominal no sabe si la perturbación es agregada, un cambio general de los precios sin consecuencias reales, o específica, un cambio de su salario real que sí debería alterar su oferta. A esa dificultad se la llama problema de extracción de señales.

- *Búsqueda de empleo*: la decisión del trabajador que busca.
- *Emparejamiento*: el modelo de Diamond-Mortensen-Pissarides.
- *Costes de ajuste*: los modelos dinámicos de demanda de trabajo.

1. Búsqueda de empleo y salario de reserva

La teoría de la búsqueda estudia el desempleo friccional desde el lado de la oferta, a partir de cómo deciden los trabajadores, con información imperfecta, qué ofertas aceptar y cuánto tiempo seguir buscando. La sección repasa sus antecedentes y expone el modelo del salario de reserva de Mortensen, sus extensiones y la evidencia empírica.

1.1. Contexto y antecedentes

Antes del modelo del salario de reserva conviene situar el desempleo que la teoría quiere explicar y sus antecedentes: la búsqueda simultánea y la búsqueda secuencial.

1.1.1. La descomposición del desempleo

Para entender las teorías de la búsqueda, que sustentan las de emparejamiento, conviene partir de la descomposición de la tasa de desempleo en dos componentes:

$$u_t = \bar{u}_t + u_t^c$$

donde $\bar{u}_t$ es el **desempleo de equilibrio** (friccional y estructural), no observable, que persiste aunque el mercado esté en su estado estacionario, y u_t^c el **desempleo cíclico**, la desviación respecto a aquel (tema A28).

La teoría neoclásica no puede explicar $\bar{u}$, porque en su equilibrio no queda sin empleo nadie dispuesto a trabajar al salario de mercado. Frente a ella, las teorías de la búsqueda tratan de explicar por qué surge ese desempleo de equilibrio y si es voluntario o involuntario. En un mundo con información imperfecta, encontrar un empleo lleva tiempo y esfuerzo, y el desempleo es, en parte, el tiempo que los trabajadores dedican a buscar.

1.1.2. Búsqueda simultánea (Stigler, 1962)

Origen: la teoría de la búsqueda nace de la economía de la información. George Stigler (1962) aplicó al mercado de trabajo su análisis de la dispersión de precios (1961).

Idea: si buscar es costoso y los salarios están dispersos, el trabajador decide cuántas ofertas consultar igualando el beneficio esperado de una consulta más a su coste, de modo que la dispersión salarial, tomada como dada, es compatible con una búsqueda óptima.

Supuestos:

- *Entorno*: el trabajador conoce la distribución de los salarios que ofrecen las empresas, pero no el salario de cada una.

- *Mercado de factores*: las vacantes son idénticas salvo en el salario. Buscar es costoso (cada consulta tiene un coste c) y la búsqueda es *simultánea*, es decir, el trabajador decide de antemano cuántas empresas n contactar y se queda con la mejor oferta.

Razonamiento: el salario esperado de la mejor de n ofertas, $w^e(n)$, es creciente en n pero a ritmo decreciente[2]. El trabajador elige n hasta que el beneficio marginal esperado de una consulta más iguala a su coste marginal:

$$\frac{\partial w^e}{\partial n} = c$$

donde el lado izquierdo es la ganancia esperada de la consulta n-ésima y c su coste. Por tanto, se consultan más empresas mientras la ganancia esperada de una consulta adicional compense su coste, y no hasta encontrar el salario más alto de la distribución.

Conclusiones:

- *Aspectos positivos*: es el punto de partida de la teoría y explica por qué el mercado laboral no cumple necesariamente la ley del precio único.

- *Limitaciones*: la principal, señalada por McCall, es que la estrategia simultánea no es óptima. Por ejemplo, si la quinta empresa consultada ofrece un salario sorprendentemente alto, tiene poco sentido visitar las $n - 5$ restantes.

1.1.3. Búsqueda secuencial (McCall, 1970)

Origen: la búsqueda secuencial se debe a John McCall, que formuló la regla de parada óptima en 1965 y la aplicó a la búsqueda de empleo en 1970.

Idea: la estrategia óptima es *secuencial*, es decir, contactar empresas de una en una y decidir, tras cada oferta, si aceptarla o seguir buscando.

Supuestos:

- *Entorno*: los de Stigler (distribución salarial conocida, costes de búsqueda).

- *Mercado de factores*: búsqueda secuencial, de modo que el trabajador recibe las ofertas de una en una y decide sobre cada una antes de buscar la siguiente.

Razonamiento: de los supuestos se deduce que el criterio de aceptación es una *regla de parada*, según la cual se acepta la oferta w si y solo si alcanza el *salario de reserva* $\underline{w}$, es decir, si $w \geq \underline{w}$. En caso contrario se rechaza y se sigue buscando. El salario de reserva se fija donde el coste de una búsqueda más iguala a la ganancia esperada de obtener una oferta mejor que $\underline{w}$, de modo que se busca hasta que el salario ofrecido cubre el salario de reserva.

Conclusiones:

- *Aspectos positivos*: amplía la visión de Stigler y hace del salario de reserva el concepto central de la teoría, del que dependen la duración del desempleo y la tasa de salida.

- *Limitaciones*: la principal es que el modelo toma la distribución de salarios como dada, sin preguntarse si las empresas la mantendrían. De hecho, la *paradoja de Diamond* (1971) muestra que, con agentes idénticos, no la mantienen.

[2] La densidad de que una empresa concreta ofrezca el salario w_0 y sea la mejor de n es $H(w_0)^{n-1}H'(w_0)$, el producto de la probabilidad de que las otras $n-1$ ofrezcan menos y la densidad de que ella ofrezca w_0. Como cualquiera de las n puede ser la mejor, el salario esperado de la mejor oferta es $w^e(n) = \int_0^\infty n\, H(w)^{n-1} H'(w)\, w\, dw$, creciente en n a ritmo decreciente.

1.2. Modelo del salario de reserva (Mortensen, 1970)

Origen: Dale Mortensen (1970) dio a la búsqueda secuencial de McCall la forma del modelo del salario de reserva, en un entorno dinámico.

Idea: el desempleado compara en cada instante el valor de seguir buscando con el de aceptar la oferta recibida. El salario de reserva es el que le hace indiferente entre ambos y resume toda su estrategia.

1.2.1. Supuestos

Entorno: el contexto es dinámico, en tiempo continuo, con información incompleta y mercados financieros perfectos y completos (tipo de interés real r estable y exógeno, sin restricciones de liquidez). El trabajador no conoce el salario de cada empresa, pero sí la función de distribución de ofertas $H(w)$, exógena y constante.

Agentes: hay un gran número de oferentes y demandantes racionales, salario-aceptantes. El desempleado busca empleo con intensidad constante y es neutral al riesgo (utilidad instantánea igual al salario y sin desutilidad de la búsqueda)[3].

Mercado de factores:

- *Ofertas*: concretan un salario real constante w mientras dure el contrato. El trabajador recibe las ofertas de una en una, con una tasa de llegada (de Poisson) $\phi > 0$ constante y exógena, en la que se reflejan el perfil del trabajador y las fricciones del mercado.

- *Destrucción*: una vez empleado, el puesto se destruye aleatoriamente a una tasa exógena $\lambda > 0$.

- *Búsqueda*: secuencial y costosa[4], de modo que la disyuntiva no es participar o no, sino seguir buscando o aceptar.

1.2.2. Modelo

> El modelo se desarrolla con detalle para que se entienda de dónde sale cada resultado. Para la exposición basta con enunciar las dos ecuaciones de valor, la condición de indiferencia ($\underline{w} = rU_B^e$), la ecuación implícita del salario de reserva, sin derivarla, la tasa de salida y los signos de la estática comparativa.

Planteamiento: en cada instante el desempleado puede no recibir oferta (y sigue buscando, con valor U_B^e) o recibir una, que acepta si $w \geq \underline{w}$ (obtiene U_T^e) o rechaza si $w < \underline{w}$ (sigue con U_B^e). El modelo se resuelve con dos ecuaciones de valor, análogas a las de un activo cuyo rendimiento por unidad de tiempo es un flujo más una ganancia esperada de capital:

- *Utilidad esperada descontada en situación de búsqueda*: el valor de seguir buscando, U_B^e, cumple la primera ecuación de valor:

$$rU_B^e = z + \phi\left(U_\phi^e - U_B^e\right) \quad \rightarrow \quad U_B^e = \frac{1}{r+\phi}\left[z + \phi U_\phi^e\right] \quad \rightarrow \quad U_B^e = f(z^+, U_\phi^{e+}, r^-, \phi^+)$$

[3] El coste de buscar se recoge en los ingresos netos z y en el tiempo de espera hasta recibir una oferta aceptable.
[4] Sin costes de búsqueda, el trabajador no pararía hasta dar con la oferta más alta posible.

donde z son los ingresos netos durante la búsqueda (subsidios, producción doméstica)[5] y U^e_ϕ el valor esperado de recibir una oferta, $\mathbb{E}[\max(U^e_T(w), U^e_B)]$. En cada instante, el flujo de la búsqueda (rU^e_B) equivale a los ingresos netos más la ganancia esperada de recibir una oferta. Despejando U^e_B, el valor de buscar reúne los ingresos netos y lo que vale recibir una oferta, ponderado por la tasa ϕ a la que llega, y descuenta ambos a la tasa $r + \phi$, con el factor $1/(r + \phi)$. La tasa de descuento suma así al tipo de interés la tasa de llegada de ofertas, porque los ingresos de la búsqueda solo se perciben hasta que llega la siguiente oferta.

▪ *Utilidad esperada descontada de aceptar y trabajar*: el valor de aceptar una oferta con salario w, U^e_T, cumple la segunda:

$$rU^e_T = w + \lambda\,(U^e_B - U^e_T) \quad \rightarrow \quad U^e_T = \frac{1}{r+\lambda}\,[w + \lambda\, U^e_B] \quad \rightarrow \quad U^e_T = f(w^+, U^{e+}_B, r^-, \lambda^-)$$

donde λ es la tasa de destrucción del empleo. El flujo de trabajar equivale al salario menos la pérdida esperada de volver al desempleo. De forma análoga, el valor de trabajar reúne el salario y el valor de volver a buscar, ponderado por la tasa λ a la que se destruye el puesto, y descuenta ambos a la tasa $r + \lambda$, con el factor $1/(r + \lambda)$, porque el salario solo se cobra mientras dura el empleo. Reordenando se obtiene $U^e_T - U^e_B = (w - rU^e_B)/(r + \lambda)$, es decir, la ganancia de aceptar es el exceso del salario sobre el flujo de la búsqueda, descontado a una tasa que incluye la probabilidad de perder el empleo. Las ofertas valen, por tanto, menos cuanto más precario es el empleo.

Equilibrio: el salario de reserva es el que hace al individuo indiferente entre aceptar y seguir buscando:

$$\underline{w} \,\big|\, U^e_T(\underline{w}) = U^e_B \quad \rightarrow \quad \underline{w} = rU^e_B \quad \rightarrow \quad \underline{w}^* = f(z^+, \phi^+, \lambda^-, r^-)$$

donde el salario de reserva es exactamente el flujo de ingresos que la búsqueda reporta. El individuo acepta cualquier oferta superior y rechaza cualquier inferior, y de ahí que el salario de reserva resuma toda su estrategia.

Sustituyendo la ganancia de aceptar, $U^e_T(w) - U^e_B = (w - \underline{w})/(r + \lambda)$, en la ecuación de la búsqueda, y teniendo en cuenta que solo se aceptan las ofertas con $w \geq \underline{w}$, se obtiene la ecuación implícita del salario de reserva:

$$\underline{w} = z + \frac{\phi}{r+\lambda}\int_{\underline{w}}^{\infty} (w - \underline{w})\, dH(w)$$

El salario de reserva iguala así los ingresos durante la búsqueda más el valor de esperar una oferta mejor, que es la ganancia esperada sobre el propio salario de reserva, multiplicada por la tasa de llegada de ofertas y descontada a la tasa $r+\lambda$. Como la integral decrece con $\underline{w}$, derivando implícitamente se obtienen los signos de la estática comparativa, según los cuales $\underline{w}$ crece con z y con ϕ y decrece con λ y con r.

La tasa de salida del desempleo (*hazard rate*) es la probabilidad por unidad de tiempo de recibir una oferta aceptable, $\chi = \phi\,[1 - H(\underline{w})]$, donde $1 - H(\underline{w})$ es la probabilidad de que la oferta supere el salario de reserva. La duración media esperada del desempleo es su inversa, $T_u = 1/\chi = f(z^+, \phi^\pm, \lambda^-, r^-)$.

[5] Incluye también el valor del ocio.

Desarrollo gráfico: la determinación del salario de reserva se representa en el plano salario-valor, con el salario ofrecido en el eje horizontal y el valor descontado en el vertical (figura A26.1):

1. *Valor de aceptar* ($U_T^e(w)$): recoge el valor descontado de aceptar una oferta con salario w y conservar el empleo hasta que se destruya. Es una recta creciente en el salario ofrecido, con pendiente $1/(r+\lambda)$, porque cada unidad de salario se cobra mientras dura el empleo y se descuenta al tipo de interés más la tasa de destrucción.

2. *Valor de seguir buscando* (U_B^e): recoge el valor de rechazar la oferta y esperar la siguiente. Es una horizontal, pues depende de los ingresos durante la búsqueda y de la llegada de ofertas, pero no de la oferta concreta recibida.

3. *Equilibrio*: su intersección, el punto 0, define $\underline{w}$. A su izquierda aceptar vale menos que seguir buscando y el individuo rechaza las ofertas, mientras que a su derecha las acepta.

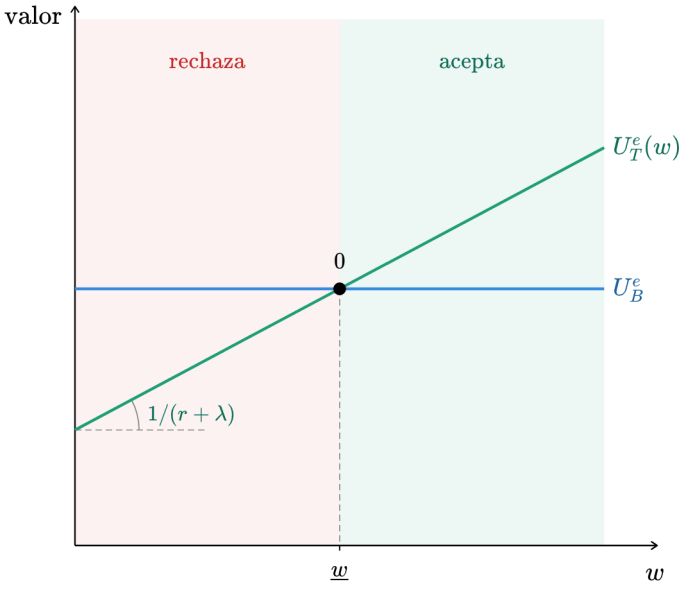


Figura A26.1: Determinación del salario de reserva. Se sitúa en la intersección entre el valor de aceptar una oferta, creciente en el salario, y el valor de seguir buscando, independiente de ella.

Estática comparativa: todo lo que eleva U_B^e (más subsidios, más ofertas) desplaza la horizontal hacia arriba y el salario de reserva hacia la derecha. En cambio, todo lo que reduce el valor de un empleo (mayor destrucción, mayor descuento) aplana la recta $U_T^e(w)$ y reduce también el valor de seguir buscando, pues las ofertas futuras valen menos. Así, la horizontal U_B^e baja y el salario de reserva se desplaza hacia la izquierda. Los efectos de cada variable son los siguientes:

1. *Ingresos netos durante la búsqueda* ($\uparrow z$): sube $\underline{w}$, cae χ, sube T_u. Los desempleados se vuelven más exigentes porque buscar es menos costoso. Por ello, la principal implicación de política es que los subsidios pueden alargar el desempleo friccional.

2. *Tasa de llegada de ofertas* ($\uparrow \phi$): sube $\underline{w}$ (esperar es más barato porque llegan más ofertas) y el efecto sobre χ y T_u es ambiguo, ya que aumentan las oportunidades, pero también la exigencia[6].

3. *Tasa de destrucción* ($\uparrow \lambda$): cae $\underline{w}$, sube χ, cae T_u. Los empleos valen menos porque son más precarios y los trabajadores se vuelven menos exigentes.

4. *Tipo de interés* ($\uparrow r$): los individuos son más impacientes y descuentan más las ofertas futuras, de modo que cae $\underline{w}$, sube χ y se acorta T_u.

1.2.3. Conclusiones

Aspectos positivos: el modelo explica el desempleo friccional como resultado de decisiones racionales con información imperfecta y proporciona las herramientas para analizar los efectos de las prestaciones por desempleo, de la duración del paro y de las políticas de intermediación. Si el trabajador conoce bien $H(w)$, estar parado es para él la mejor opción, pues rechaza voluntariamente toda oferta con $w < \underline{w}$ siguiendo una estrategia de búsqueda óptima.

Limitaciones: el modelo presenta tres:

- *Solo oferta*: no considera la demanda de trabajo ni las vacantes, por lo que no puede decir si todo el desempleo friccional es realmente voluntario.

- *Distribución exógena*: la paradoja de Diamond y la crítica de Rothschild (1973) ponen en cuestión el origen de $H(w)$, que el modelo toma como dada.

- *Efecto desánimo*: no explica por qué algunos individuos dejan de buscar.

1.3. Extensiones y evidencia empírica

La paradoja de Diamond cuestiona el supuesto de una distribución salarial dada, y la teoría posterior lo justifica. El modelo básico se ha extendido, además, para incorporar las características no salariales de las ofertas, el descuento hiperbólico, el esfuerzo de búsqueda, la búsqueda desde el empleo y los servicios públicos de empleo. También cuenta con una amplia contrastación empírica, centrada en la duración del desempleo y en el efecto de las prestaciones.

1.3.1. La paradoja de Diamond (1971)

Origen: Peter Diamond (1971) planteó la paradoja para el mercado de bienes, con consumidores que buscan precios, y la teoría de la búsqueda la trasladó al mercado de trabajo.

Idea: si los trabajadores son idénticos y todos siguen la estrategia de McCall, ninguna empresa tiene incentivo a pagar más que el salario de reserva común, pues cualquier oferta igual a $\underline{w}$ es aceptada.

Razonamiento: basta un coste de búsqueda positivo, por pequeño que sea, para que la distribución de salarios colapse en un único punto, el salario de monopsonio, y desaparezca el motivo

[6] El signo depende de la forma de la distribución de ofertas $H(w)$. Flinn y Heckman (1983) muestran que, si la integral de la función de supervivencia, $\int_w^\infty [1 - H(x)]\,dx$, es log-cóncava (basta con que lo sea la densidad), domina el aumento de las oportunidades, de modo que una mayor tasa de llegada eleva χ y acorta T_u. Van den Berg (1994) debilita las condiciones suficientes obtenidas hasta entonces, que no cumplían las familias de distribuciones de salarios más usadas en la estimación estructural, y extiende el resultado a todas ellas.

para buscar. La dispersión salarial que la búsqueda presupone no es, por tanto, un equilibrio con agentes idénticos.

Extensiones: la teoría posterior respondió por dos vías:

- *Agentes heterogéneos*: la heterogeneidad de los trabajadores (distintos salarios de reserva, como en Albrecht y Axell, 1984) o de las empresas (distinta productividad) sostiene la dispersión.

- *Competencia entre empresas*: si los trabajadores reciben a veces más de una oferta a la vez (Burdett y Judd, 1983) o siguen buscando desde el empleo (Burdett y Mortensen, 1998), las empresas compiten por ellos y la dispersión aparece incluso con agentes idénticos.

Así justificada la dispersión salarial, queda respaldado el supuesto del modelo de Mortensen de una distribución de ofertas dada.

Conclusiones:

- *Aspectos positivos*: muestra que la dispersión salarial, que los modelos de búsqueda parcial toman como dada, no puede suponerse y ha de obtenerse como resultado de equilibrio. Obliga así a modelizar a la vez a trabajadores y empresas, que es el camino de los modelos de equilibrio con búsqueda posteriores.

- *Limitaciones*: el resultado descansa en supuestos extremos (agentes idénticos, empresas que fijan el salario unilateralmente, una sola oferta cada vez y búsqueda solo desde el desempleo) y contradice la dispersión salarial que se observa entre trabajadores similares. Como muestran las extensiones, basta modificar alguno de ellos para recuperarla.

1.3.2. Determinantes no salariales: la utilidad de reserva (Blau, 1991)

Origen: David Blau (1991) señaló que las ofertas difieren también en las horas de trabajo y otras condiciones, y que el buscador las valora conjuntamente.

Idea: el modelo básico supone que el buscador solo valora el salario. Cuando las ofertas difieren en más de una característica, en cambio, el criterio de aceptación pasa a ser una utilidad de reserva.

Planteamiento: con una utilidad $u = w^a h^b$, donde w es el salario (un bien, $a > 0$) y h las horas (un mal, $b < 0$), el individuo compara cada oferta con una *utilidad de reserva* $\underline{u}$ y acepta las que la superan. Tomando logaritmos, la curva de indiferencia de reserva es $\ln w = (\ln \underline{u} - b \ln h)/a$, una recta creciente en el plano $(\ln h, \ln w)$ con pendiente $-b/a > 0$, de modo que el individuo exige más salario para aceptar más horas.

Desarrollo gráfico: la utilidad de reserva se representa en el plano $\ln h$-$\ln w$ (figura A26.2):

1. *Curva de indiferencia de reserva* ($u = \underline{u}$): es una recta creciente, con pendiente $-b/a$, y las ofertas aceptables están por encima de ella.

2. *Salario de reserva* ($\ln \underline{w}$): el criterio del modelo básico sería una horizontal, porque el buscador aceptaría toda oferta con un salario superior a $\underline{w}$ cualesquiera que fueran las horas.

3. *Decisión*: el individuo ya no sigue la regla del salario de reserva. Así, admite un salario inferior a ese umbral si la jornada es corta (punto A) y rechaza uno superior si la jornada es larga (punto B). Los dos criterios se separan más cuanto mayor es la aversión a las horas ($|b|$).

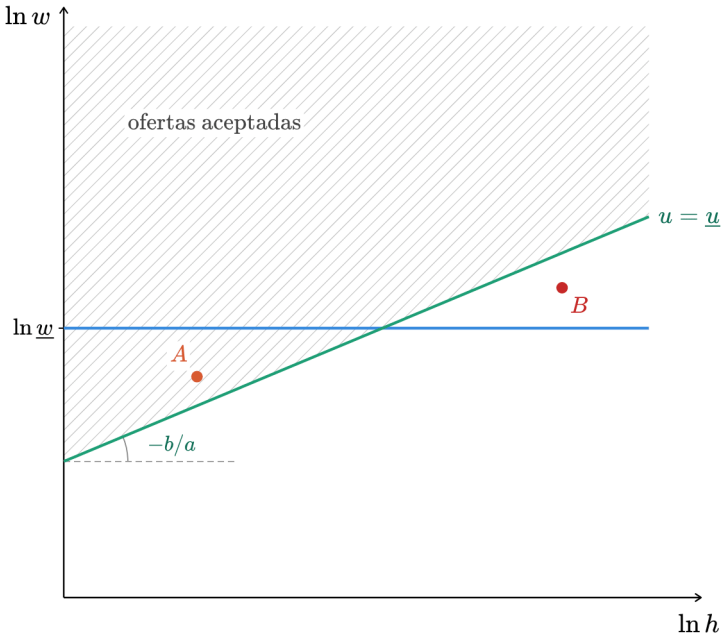


Figura A26.2: Utilidad de reserva (Blau). En el plano (ln h, ln w) las ofertas por encima de la recta de utilidad de reserva son aceptadas, mientras que el criterio del salario de reserva sería una horizontal.

Conclusiones:

▪ *Aspectos positivos*: al introducir dentro de la decisión de aceptar la elección entre ocio y consumo, la extensión tiende un puente entre la búsqueda y la oferta de trabajo del tema A25, apartado 1.2. Da cuenta, además, de que alguien prefiera un puesto con menor sueldo si sus condiciones son mejores (p. ej. una jornada más corta).

▪ *Limitaciones*: como el modelo básico, sigue sin considerar la demanda de trabajo y toma como dada la distribución de las ofertas.

1.3.3. Otras extensiones

1. *Economía conductual* (DellaVigna y Paserman, 2005): con descuento hiperbólico, los buscadores impacientes procrastinan, es decir, buscan con menor intensidad. Aunque también fijan un salario de reserva menor, domina el efecto de la menor intensidad, que alarga el desempleo por una vía distinta de la de los subsidios.

2. *Determinantes del esfuerzo de búsqueda*: la tasa de llegada de ofertas depende del esfuerzo, $\phi = \phi(e)$, que el individuo elige igualando su coste marginal a su beneficio marginal. El esfuerzo depende de la posibilidad de reincorporarse al empleador previo, de las prestaciones por desempleo (que lo reducen además de elevar $\underline{w}$), de las aptitudes de búsqueda y de las restricciones (movilidad, cuidados).

3. *Búsqueda desde el empleo* (*on-the-job search*, Burdett, 1978): los empleados siguen buscando, con un salario de reserva igual al que perciben, lo que explica la movilidad voluntaria y la relación positiva entre salario y antigüedad.

4. *Justificación de los servicios públicos de empleo*: la información sobre vacantes es un bien público y la búsqueda genera externalidades (Pissarides, 1984). La decisión de cada buscador afecta a la probabilidad de encontrar empleo de los demás, de modo que el desempleo de equilibrio puede situarse por encima o por debajo del eficiente (apartado 2.2.3). Esta externalidad enlaza con la teoría del emparejamiento.

1.3.4. Evidencia empírica de la búsqueda

Prestaciones y duración: la predicción central del modelo, que las prestaciones alargan el desempleo, está bien establecida. Además, las salidas del desempleo muestran un pico justo antes de agotarse el derecho a la prestación (Katz y Meyer, 1990). Este pico refleja tanto la aceptación de ofertas antes rechazadas como la salida de la población activa. También en España, Bover, Arellano y Bentolila (2002) encuentran que, a los tres meses de desempleo, la tasa de salida de quienes no cobran prestación duplica la de quienes la cobran, un efecto mayor que el del ciclo económico.

Dependencia de la duración: la tasa de salida cae con el tiempo en desempleo (*dependencia negativa de la duración*), lo que el modelo básico no predice. Se explica por la heterogeneidad no observada (los más empleables salen antes), por la pérdida de capital humano y por el uso de la duración como señal negativa por parte de las empresas. La parte de esta dependencia que no se debe a la heterogeneidad (pérdida de capital humano, estigma) es la base microeconómica de la histéresis del apartado 2.3.2.

Salarios de reserva declarados: las encuestas que preguntan a los parados por su salario de reserva (Krueger y Mueller, 2016) muestran que se sitúa cerca del último salario percibido y que apenas baja con la duración. Ello sugiere que la caída de la tasa de salida con la duración procede menos de la exigencia de los buscadores que de la llegada de ofertas y del esfuerzo de búsqueda.

1.4. Conclusiones

Hasta aquí el análisis se ha centrado en el lado de la oferta. Sin embargo, actualmente el desempleo friccional se entiende como resultado de un proceso natural de doble búsqueda, de empleo por parte de los oferentes y de trabajadores por parte de los demandantes, pues también las empresas buscan. Este desempleo se considera, en general, transitorio, a diferencia del desempleo estructural, más persistente. Para ahondar en estas cuestiones, y en particular para responder si el desempleo friccional es voluntario, hay que ir más allá de los modelos de búsqueda e incorporar las vacantes, como hace la teoría del emparejamiento.

2. Emparejamiento y curva de Beveridge

Origen: Peter Diamond (1982), Dale Mortensen (1982) y Christopher Pissarides (1985), galardonados con el Nobel en 2010, incorporaron al modelo de búsqueda el lado de la demanda (las vacantes) y el proceso de emparejamiento. Mortensen y Pissarides (1994) endogenizaron después la destrucción de empleo, y Pissarides (2000) dio al modelo su forma canónica.

Idea: en un contexto de información imperfecta coexisten puestos vacantes y personas desempleadas, algo imposible en un mercado walrasiano. Por tanto, el mercado de trabajo no se vacía y existe desempleo aun con salarios flexibles. Se presenta una versión estilizada del modelo de Diamond-Mortensen-Pissarides (DMP).

La sección expone los supuestos del modelo, sus ecuaciones, el comportamiento de los agentes, el equilibrio con la curva de Beveridge y la estática comparativa, y sus extensiones y aplicaciones.

2.1. Supuestos

Metodología: entorno de tiempo continuo, donde la población activa se normaliza a uno.

Entorno: la información es incompleta (los agentes conocen el proceso de emparejamiento, pero no pueden coordinarse). Existen costes de búsqueda y de transacción (búsqueda activa, preparación de vacantes)[7]. Los mercados de capitales son completos, con un tipo de interés real r exógeno.

Agentes: existe un elevado número de empresas, cada una con un único puesto, vacante o cubierto, que no se coordinan entre sí. También existe un elevado número de trabajadores neutrales al riesgo, que tampoco se coordinan[8] y no buscan desde el empleo. Una fracción u de ellos está desempleada y $1-u$ empleada. En el modelo, trabajadores y puestos son idénticos, pero la heterogeneidad real de ambos, razón por la cual el emparejamiento lleva tiempo, se resume en la función de emparejamiento.

Mercado de factores: cada puesto cubierto produce un flujo y de un único bien, que es el numerario. Mantener una vacante abierta cuesta c por unidad de tiempo, los puestos se destruyen a una tasa exógena λ y el salario no viene dado por el mercado, sino que se negocia bilateralmente tras el emparejamiento.

2.2. Modelo

El modelo se distingue porque, además de las ecuaciones de comportamiento de los agentes, incluye ecuaciones que describen el funcionamiento del propio mercado de trabajo.

2.2.1. Ecuaciones de funcionamiento del mercado

Tres relaciones describen el funcionamiento del mercado con independencia de las decisiones de los agentes:

Ley de movimiento del desempleo: el modelo describe, en primer lugar, cómo evoluciona el desempleo. La intuición se ve mejor en tiempo discreto, con una ecuación de movimiento del empleo como la siguiente:

[7] En un mundo sin fricciones, a las empresas les resulta indiferente perder empleados (pueden conseguir idénticos al mismo salario sin costes) y a los trabajadores no les preocupa perder su puesto.

[8] Esto implica que una vacante podría recibir muchas solicitudes y otra ninguna.

$$L_t = (1-\lambda)\, L_{t-1} + M_{t-1}$$

donde L_t es el nivel de empleo en el periodo t, λ la tasa de destrucción de empleo, L_{t-1} el nivel de empleo del periodo anterior y M_{t-1} el número de emparejamientos realizados en él.

El empleo de cada periodo es el del periodo anterior, menos los empleos destruidos y más los nuevos emparejamientos realizados en el mercado de trabajo.

En tiempo continuo, que es el marco del modelo, el mismo razonamiento se aplica a la tasa de desempleo. Como la población activa está normalizada a uno, el empleo es $1-u$, y la variación instantánea de la tasa de desempleo es la diferencia entre dos flujos:

$$\dot{u} = \lambda(1-u) - p(\theta)\, u$$

donde $\lambda(1-u)$ es el flujo de empleados que pierden su empleo y $p(\theta)u$ el flujo de desempleados que lo encuentran, a la tasa de salida $p(\theta)$, que depende de la estrechez del mercado, θ (ambas se definen en este mismo apartado). Como es habitual en tiempo continuo, se omite el argumento temporal de las variables.

Función de emparejamiento: constituye el elemento central del modelo, pues relaciona las solicitudes de los desempleados con las vacantes abiertas y determina el número de emparejamientos por unidad de tiempo. Intuitivamente, es una "máquina" que produce un emparejamiento cuando una bola (una solicitud) encesta en una urna (una vacante), y condensa en una expresión todo el proceso de búsqueda mutua y evaluación. Según Pissarides, engloba las fricciones del mercado sin necesidad de hacerlas explícitas.

Formalmente, el número de emparejamientos (*matches*) depende del número de vacantes V y del de desempleados U, $M = M(V,U)$, y es estrictamente creciente en ambos argumentos. Habitualmente, se considera que esta función es de tipo Cobb-Douglas:

$$M = A\, V^{\alpha} U^{1-\alpha}, \qquad A > 0,\ \alpha \in (0,1)$$

donde V son las vacantes, U los desempleados, α la elasticidad de los emparejamientos respecto a las vacantes[9] y A la *eficiencia del emparejamiento*, determinada por factores estructurales (formación y movilidad de los trabajadores, eficacia de los servicios de empleo, desajuste de cualificaciones)[10].

En cuanto a los rendimientos a escala, con rendimientos crecientes se dan efectos de mercado denso (al aumentar las búsquedas el proceso es más eficiente), que pueden generar equilibrios múltiples (Diamond, 1982), y con decrecientes, efectos de congestión. Se suponen, no obstante, rendimientos constantes, los más compatibles con la evidencia, ya que las elasticidades estimadas respecto a vacantes y desempleo suman aproximadamente uno (Petrongolo y Pissarides, 2001).

Grado de estrechez y tasas de transición: la *estrechez del mercado* es la ratio vacantes-desempleados, $\theta \equiv V/U > 0$, y mide lo fácil que es para un desempleado encontrar trabajo:

- *Mercado estrecho*: si $\theta > 1$, las vacantes superan a los desempleados y es fácil encontrar trabajo pero difícil cubrir las vacantes.

[9] Los exponentes α y $1-\alpha$ son elasticidades producto-factor. Si los argumentos se retribuyeran según su productividad marginal, serían sus participaciones en el producto.

[10] La eficiencia del emparejamiento en EE. UU. se degradó notablemente tras la crisis de las *subprime* (Barnichon y Figura, 2015).

- *Mercado holgado*: si $\theta < 1$, lo contrario (p. ej. España tras 2008).

Con rendimientos constantes, las dos tasas de transición dependen solo de θ:

$$q(\theta) \equiv \frac{M}{V} = A\,\theta^{\alpha-1}, \quad q' < 0 \qquad\qquad p(\theta) \equiv \frac{M}{U} = A\,\theta^{\alpha} = \theta\,q(\theta), \quad p' > 0$$

donde $q(\theta)$ es la *tasa de cobertura de vacantes* (la tasa a la que una vacante se cubre por unidad de tiempo) y $p(\theta)$ la *tasa de salida del desempleo* (la tasa a la que un desempleado encuentra empleo).

Un mercado más estrecho favorece a los trabajadores (p sube), pero perjudica a las empresas (q baja). Existe, por tanto, un conflicto de intereses y *externalidades de búsqueda o congestión*, porque cada buscador reduce la probabilidad de éxito de los de su lado y aumenta la del lado contrario. El salario no es, así, el único mecanismo de ajuste, sino que también lo son el número de buscadores y el de puestos.

2.2.2. Comportamiento de los agentes

> Para la exposición no es imprescindible desarrollar las ecuaciones de valor de este apartado, pues basta con enunciar las dos condiciones a las que conducen y pasar al equilibrio. Las dos condiciones son la de participación de los trabajadores ($w \geq \underline{w}$, con $\underline{w}$ creciente en θ) y la de creación de empleo de las empresas ($y = w + (r+\lambda)\,c/q(\theta)$). Las ecuaciones se incluyen, sin embargo, porque de ellas se obtiene la forma cerrada de la curva de salarios.

Trabajadores: condición de participación: como en el modelo de búsqueda del apartado 1.2, el trabajador compara el valor de buscar con el de trabajar, pero ahora la tasa de llegada de ofertas es endógena, $p(\theta)$:

$$rU_B^e = z + p(\theta)\,(U_T^e - U_B^e) \qquad\qquad rU_T^e = w + \lambda\,(U_B^e - U_T^e)$$

donde z son los ingresos durante la búsqueda, U_B^e el valor de buscar y U_T^e el de trabajar. Reordenando la segunda ecuación, la ganancia de aceptar es $U_T^e - U_B^e = (w - rU_B^e)/(r+\lambda)$, y el trabajador participa (acepta el puesto) si esa ganancia es no negativa. Sustituyendo en la primera ecuación se obtiene la *condición de participación*:

$$w \;\geq\; \underline{w} = rU_B^e = z + \frac{p(\theta)}{r+\lambda}\,(w - \underline{w}) \quad \rightarrow \quad \underline{w} = f\,(z^+, \theta^+, \lambda^-, r^-)$$

donde el salario de reserva $\underline{w}$ es el flujo de ingresos que reporta la búsqueda, es decir, los ingresos durante ella más el valor esperado de la ganancia de emparejarse, que crece con la estrechez porque $p(\theta)$ acorta la espera. Para un salario dado, un mercado más estrecho eleva el salario de reserva y la exigencia de los trabajadores, de modo que el salario que exigen es creciente en θ, y esa es la base de la curva de salarios.

Empresas: condición de creación de empleo: cada puesto puede estar vacante (coste c, valor Π_V^e) o cubierto (produce y, paga w, valor Π_C^e), con ecuaciones de valor simétricas a las del trabajador:

$$r\Pi_V^e = -c + q(\theta)\,(\Pi_C^e - \Pi_V^e) \qquad\qquad r\Pi_C^e = y - w + \lambda\,(\Pi_V^e - \Pi_C^e)$$

donde una vacante cuesta c y se cubre a la tasa $q(\theta)$, y un puesto cubierto rinde $y - w$ y se destruye a la tasa λ. Las empresas abren vacantes mientras $\Pi_V^e > 0$. Con libre entrada, por tanto, $\Pi_V^e = 0$ y el valor de un puesto cubierto iguala al coste esperado de cubrirlo, $\Pi_C^e = c/q(\theta)$. Sustituyendo se obtiene la *condición de creación de empleo* (CCE):

$$y = w + (r + \lambda)\,\frac{c}{q(\theta)} \quad \rightarrow \quad \theta = f\left(w^-, y^+, c^-, r^-, \lambda^-\right)$$

donde la empresa crea empleo hasta que el valor de la producción iguala al salario más la capitalización del coste de mantener la vacante, ajustada por la destrucción. Esta condición es la CPO neoclásica con fricciones. A mayor estrechez, más tarda en cubrirse cada vacante y menor es el salario que la empresa puede pagar, por lo que la curva de creación de empleo tiene pendiente negativa en el plano (θ, w).

2.2.3. Equilibrio: desempleo, salario y curva de Beveridge

El concepto de equilibrio es el *estado estacionario*, en el que las tasas de desempleo y de vacantes son constantes porque los flujos de entrada y salida se compensan. El equilibrio estacionario queda determinado por tres condiciones: la negociación del salario, la creación de empleo y el flujo estacionario del desempleo.

Planteamiento: el salario de equilibrio se determina por negociación. El emparejamiento genera una renta para ambas partes, pues ninguna ha de volver a incurrir en costes de búsqueda, y el salario es la regla de reparto de ese excedente. Se modela como una *negociación à la Nash* (tema A14, apartado 2.4.3):

$$w^* = \arg\max_w \left\{ \left(U_T^e - U_B^e\right)^\gamma \left(\Pi_C^e - \Pi_V^e\right)^{1-\gamma} \right\} \quad \Rightarrow \quad U_T^e - U_B^e = \frac{\gamma}{1-\gamma}\left(\Pi_C^e - \Pi_V^e\right)$$

donde $\gamma \in (0, 1)$ es el poder de negociación del trabajador y los dos factores del producto de Nash son las ganancias netas del emparejamiento para el trabajador, $U_T^e - U_B^e$, y para la empresa, $\Pi_C^e - \Pi_V^e$.

CPO: la condición de primer orden reparte el excedente total en proporción $\gamma : (1 - \gamma)$. El resultado es la solución de negociación de Nash, que puede interpretarse como el límite del equilibrio de Nash perfecto en subjuegos del juego de ofertas alternas de Rubinstein (1982) cuando el intervalo entre ofertas tiende a cero (Binmore, Rubinstein y Wolinsky, 1986).

CSO: el producto de Nash es estrictamente log-cóncavo en w, pues ambos excedentes son lineales en el salario, de modo que la CPO define un único salario negociado.

Solución: sustituyendo las ecuaciones de valor y la condición de libre entrada ($\Pi_V^e = 0$, $\Pi_C^e = c/q(\theta)$) se obtiene la *curva de salarios* (CS, *wage curve*) en forma cerrada:

$$w = (1 - \gamma)\, z + \gamma\,(y + c\,\theta)$$

donde z son los ingresos durante la búsqueda y $c\theta$ el coste medio de contratación por desempleado, que las empresas ahorran al cubrir el puesto y que el trabajador captura en proporción a su poder de negociación. Si $\gamma = 0$, el trabajador recibe solo su opción externa, y si $\gamma = 1$, se apropia de toda la producción más el ahorro de búsqueda. En los casos intermedios, el salario es una combinación convexa de ambos.

La curva tiene pendiente positiva en el plano (θ, w) y puede leerse como la oferta de trabajo del modelo, análoga a la neoclásica, pero la analogía termina ahí, pues la curva incorpora el poder de negociación, la opción externa del trabajador y los costes de búsqueda. La tasa de destrucción y el tipo de interés, en cambio, se cancelan en la forma cerrada.

Equilibrio: el modelo se resuelve en dos diagramas:

▪ *Estrechez y salario*: en el plano estrechez-salario (θ, w), la curva de salarios (creciente) y la curva de creación de empleo (decreciente) determinan el grado de estrechez y el salario de equilibrio, θ^* y w^*. Ambos dependen solo de los parámetros (y, z, c, r, λ, γ, A, α) y no de u.

▪ *Desempleo de equilibrio y curva de Beveridge*: determinada la estrechez, el desempleo se obtiene de la ley de movimiento. En estado estacionario ($\dot{u} = 0$) los flujos de entrada y salida del desempleo se igualan:

$$u^* = \frac{\lambda}{\lambda + p(\theta)} = \frac{1}{1 + p(\theta)/\lambda} \quad \rightarrow \quad u^* = f\left(\lambda^+, \theta^-, A^-\right)$$

donde para cada tasa de destrucción y cada grado de estrechez existe una única tasa de desempleo estacionaria, mayor cuanto más se destruye empleo y menor cuanto más estrecho es el mercado y más eficiente el emparejamiento. Escrita en el plano desempleo-vacantes, la misma condición, $\lambda(1-u) = Av^{\alpha}u^{1-\alpha}$, es la *curva de Beveridge* (CB):

$$v = \left[\frac{\lambda(1-u)}{A\,u^{1-\alpha}}\right]^{1/\alpha}$$

donde v y u son las tasas de vacantes y de desempleo. La curva es el lugar geométrico de las combinaciones de ambas compatibles con un desempleo constante.

William Beveridge (1944) observó la relación inversa entre vacantes y desempleo sin modelarla, y Dow y Dicks-Mireaux (1958) la formalizaron como una hipérbola, mientras que el modelo de emparejamiento la deriva de la condición de flujo estacionario. En el plano (u, v), la curva de Beveridge tiene:

- *Pendiente negativa*: en una recesión el equilibrio se desplaza sobre ella hacia la derecha y en una recuperación, hacia la izquierda.
- *Forma convexa respecto al origen*: cada reducción adicional del desempleo exige un aumento cada vez mayor de las vacantes.
- *Posición*: depende de A y λ, y cuanto más próxima al origen está la curva, más eficiente es el emparejamiento o menor la tasa de destrucción.

Es, por ello, una medida de la eficiencia del mercado de trabajo y de la posición cíclica de la economía, y su mera existencia muestra que, pese a haber vacantes, hay desempleo.

En el segundo diagrama, el plano desempleo-vacantes (u, v), la recta de creación de empleo, que parte del origen con pendiente θ^*, corta a la curva de Beveridge y determina u^* y v^*. Ese desempleo es la *tasa de paro friccional* de equilibrio, igual a la tasa natural cuando el salario es el de equilibrio (tema A28, apartado 3.1).

Desarrollo gráfico: la determinación de la estrechez y del salario se representa en el plano estrechez-salario (figura A26.3):

1. *Curva de salarios* (CS): recoge el salario negociado para cada grado de estrechez, $w = (1 - \gamma)\, z + \gamma\, (y + c\, \theta)$. Es creciente, porque un mercado más estrecho mejora la opción externa del trabajador y eleva el ahorro de contratación que este captura. La curva de salarios hace de oferta de trabajo.

2. *Curva de creación de empleo* (CCE): recoge, para cada salario, la estrechez a la que las empresas dejan de abrir vacantes, porque su valor se anula. Es decreciente, porque con un mercado más estrecho cada vacante tarda más en cubrirse y solo compensa abrirla con un salario menor. La curva de creación de empleo hace de demanda de trabajo.

3. *Equilibrio*: su corte, el punto 0, determina θ^* y w^*. Ambos dependen solo de los parámetros y no del desempleo, que se obtiene después en el plano desempleo-vacantes.

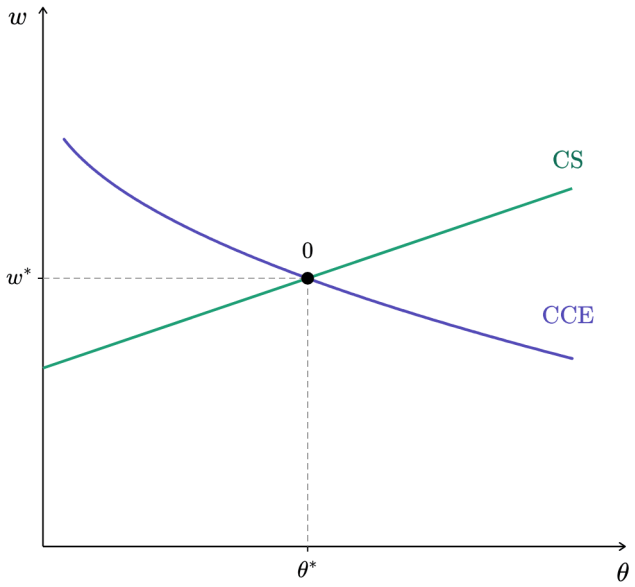


Figura A26.3: Equilibrio del modelo de emparejamiento. La curva de salarios y la condición de creación de empleo determinan la estrechez del mercado y el salario.

La determinación del desempleo y de las vacantes se representa en el plano desempleo-vacantes (figura A26.4):

1. *Curva de Beveridge* (CB): recoge las combinaciones de desempleo y vacantes con las que el flujo de entrada en el desempleo iguala al de salida. Es decreciente, porque con más vacantes los desempleados encuentran empleo antes, y convexa respecto al origen, porque cada reducción adicional del desempleo exige un aumento cada vez mayor de las vacantes.

2. *Recta de creación de empleo* ($v = \theta^* u$): recoge las combinaciones de desempleo y vacantes con la estrechez de equilibrio de la figura A26.3, y por eso parte del origen con pendiente θ^*.

3. *Equilibrio*: su corte, el punto 0, determina u^* y v^*, la tasa de paro friccional de equilibrio y la tasa de vacantes compatibles a la vez con el flujo estacionario y con la estrechez de equilibrio θ^*.

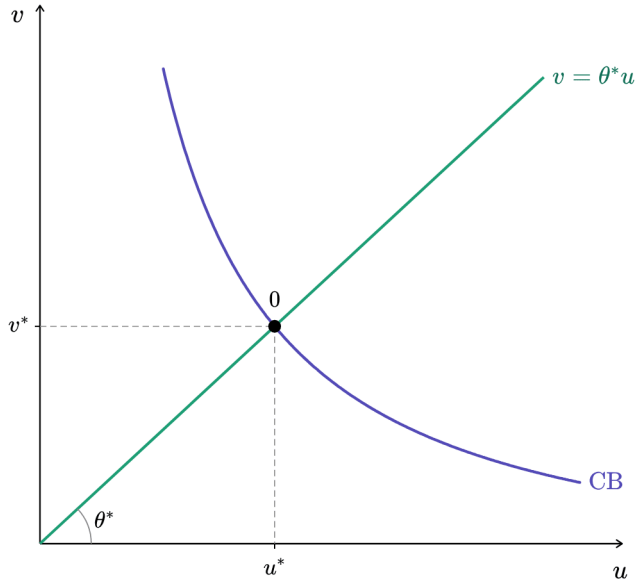


Figura A26.4: Curva de Beveridge (CB). La curva recoge la relación decreciente entre vacantes y desempleo derivada de la condición de flujo estacionario. Su corte con la recta de creación de empleo $v = \theta^* u$ determina u^* y v^* (punto 0).

Bienestar: por las externalidades de búsqueda, el equilibrio descentralizado no es eficiente en general. Lo es solo si el poder de negociación de los trabajadores coincide con la elasticidad de la función de emparejamiento respecto al desempleo, $\gamma = 1 - \alpha$ (condición de Hosios, 1990). Si γ es mayor, se crean pocas vacantes y hay demasiado desempleo, y si es menor, demasiadas. Fuera de ese caso, hay margen para la intervención.

2.2.4. Estática comparativa y dinámica

Estática comparativa: el modelo distingue dos tipos de perturbaciones (Blanchard y Diamond, 1989):

▪ *Perturbaciones coyunturales*: son las que alteran la rentabilidad de crear empleo sin cambiar la eficiencia del emparejamiento ni la tasa de destrucción. Giran la recta de creación de empleo (cambia θ^*) y mueven el equilibrio sobre la curva de Beveridge. Todo lo que hace menos provechoso crear un puesto reduce la estrechez y las vacantes y eleva el desempleo, pero el efecto sobre el salario depende de la curva que se desplace en el plano (θ, w):

- *Subida del tipo de interés r (que descuenta más los rendimientos futuros) o de las cotizaciones*: desplaza hacia abajo la curva de creación de empleo, y el salario cae.
- *Subsidio z*: no entra en la condición de creación de empleo, sino que desplaza hacia arriba la curva de salarios al mejorar la opción externa del trabajador, y el salario sube.
- *Coste de la vacante c*: desplaza hacia abajo la curva de creación de empleo y hacia arriba la de salarios. Aunque gráficamente el efecto parece ambiguo, la forma cerrada muestra que el salario cae, porque la estrechez disminuye más que proporcionalmente, de modo que baja el coste de contratación por desempleado, $c\theta$, que captura el trabajador.

- *Productividad y*: un aumento, al hacer más provechoso crear un puesto, produce lo contrario. Suben la estrechez, las vacantes y el salario, y cae el desempleo.

▪ *Perturbaciones estructurales*: son las que cambian la eficiencia con que vacantes y desempleados se encuentran (A) o la tasa de destrucción (λ), y desplazan la propia curva de Beveridge. Una mejora del emparejamiento (mejores tecnologías de la información, mayor movilidad, políticas activas y agencias de empleo eficaces) desplaza la curva hacia el origen, de CB_0 a CB_1, y, en el plano (θ, w), la curva de creación de empleo hacia arriba. En el plano desempleo-vacantes (figura A26.5), la recta de creación de empleo pasa de $v = \theta_0^* u$ a $v = \theta_1^* u$ y el equilibrio, del punto 0 al 1. Para cada nivel de desempleo se necesitan entonces menos vacantes para mantener el empleo. En cambio, un empeoramiento (menor intensidad de búsqueda por efecto desánimo, desajuste de cualificaciones, desempleo de larga duración, entrada en la población activa de colectivos con menor eficacia de búsqueda) desplaza la curva de Beveridge hacia fuera, con más desempleo para cada nivel de vacantes.

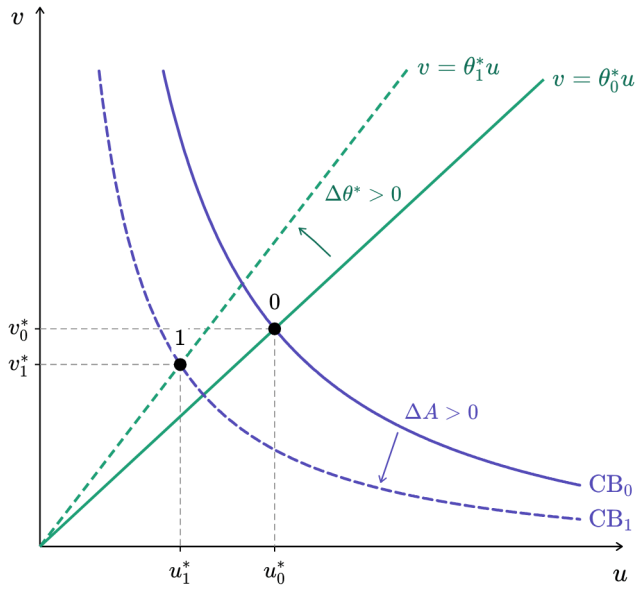


Figura A26.5: Desempleo de equilibrio. La curva de Beveridge y la recta de creación de empleo determinan conjuntamente desempleo y vacantes. Una mejora de la eficiencia del emparejamiento desplaza la curva hacia el origen, eleva la estrechez y lleva el equilibrio de 0 a 1.

Dinámica: las dos variables del modelo se ajustan a velocidades distintas. La estrechez θ es una *variable de salto*. Como abrir o cerrar vacantes no tiene coste de ajuste y θ^* no depende de u, tras una perturbación las vacantes saltan de inmediato a la nueva recta de creación de empleo. Frente a la estrechez, el desempleo u es una *variable de estado*, que solo cambia a través de los flujos de la ley de movimiento, a un ritmo $\lambda + p(\theta)$ que, con los valores habituales, tarda meses en completar el ajuste.

En el plano (u, v), la economía salta verticalmente a la nueva recta de creación de empleo y se desplaza después a lo largo de ella, a medida que u se ajusta, hasta cortar la curva de Beveridge en el nuevo estado estacionario. Esta dinámica genera los *bucles en sentido antihorario* que se observan en los datos. Así, en una recesión las vacantes caen antes de que el desempleo suba, y en la recuperación suben antes de que el desempleo baje.

2.3. Extensiones y aplicaciones

El modelo básico se ha ampliado para endogenizar la destrucción de empleo y explicar la histéresis, y se ha aplicado a la evolución reciente de los mercados de trabajo español y estadounidense.

2.3.1. Destrucción endógena de empleo (Mortensen y Pissarides, 1994)

Origen: en el modelo básico la tasa de destrucción λ es exógena, pero Mortensen y Pissarides la endogenizan en un artículo posterior (1994).

Idea: cada puesto cubierto recibe perturbaciones idiosincrásicas de productividad y la empresa lo destruye cuando la productividad cae por debajo de un umbral, la *productividad de reserva* $\underline{R}$, en el que se anula el excedente conjunto del emparejamiento.

Razonamiento: la productividad de reserva es creciente en la estrechez, pues con un mercado estrecho mejora la opción externa del trabajador, que encuentra antes otro empleo, y el emparejamiento deja de compensar antes. En consecuencia, la *curva de destrucción de empleo* tiene pendiente positiva en el plano $(\theta, \underline{R})$ (figura A26.6) y, con la de creación, de pendiente negativa, determina conjuntamente θ^* y $\underline{R}^*$, y con ellos la tasa de destrucción $\lambda(\underline{R}^*)$ y el desempleo.

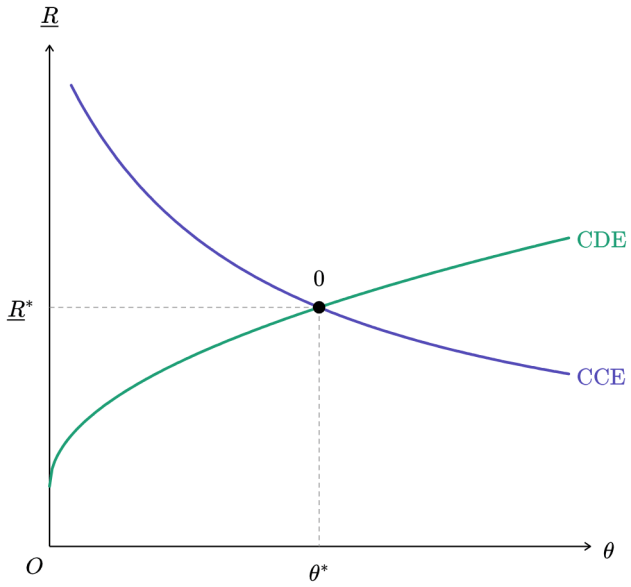


Figura A26.6: Destrucción endógena de empleo (Mortensen y Pissarides, 1994). La curva de destrucción de empleo (CDE), creciente, y la de creación de empleo (CCE), decreciente, determinan en su corte (punto 0) la estrechez θ^* y la productividad de reserva $\underline{R}^*$, y con ella la tasa de destrucción $\lambda(\underline{R}^*)$.

Extensiones: la destrucción de empleo responde al ciclo, pues se concentra en las recesiones, como muestran los datos de flujos de Davis y Haltiwanger (1992), y a las instituciones. Así, los costes de despido reducen $\underline{R}$ y la destrucción, pero también la creación, con efecto ambiguo sobre el desempleo (apartado 3.3), mientras que las prestaciones elevan $\underline{R}$ y la destrucción. El modelo es, por ello, el marco habitual para evaluar la protección del empleo.

Conclusiones:

- *Aspectos positivos*: completa el modelo de emparejamiento al hacer endógenos los dos flujos brutos del mercado de trabajo, de modo que el desempleo responde a las perturbaciones tanto por la creación como por la destrucción de puestos, en línea con la evidencia sobre flujos.

- *Limitaciones*: trata la destrucción como una decisión conjunta y eficiente de la pareja empresa-trabajador, sin distinguir despidos y abandonos voluntarios, y toma como exógenas las perturbaciones de productividad. Además, el efecto de las instituciones sobre el nivel de desempleo es ambiguo, por lo que sus conclusiones de política dependen de la calibración.

2.3.2. Histéresis

Hay **histéresis** cuando el desempleo de equilibrio depende de la trayectoria del desempleo efectivo, de modo que una recesión eleva de forma persistente la tasa natural (tema A28, apartado 2.3). La literatura distingue tres mecanismos:

1. *Salarial o de insiders-outsiders* (Blanchard y Summers, 1986; Lindbeck y Snower, 1988): los empleados que conservan el puesto negocian salarios que solo tienen en cuenta su propio empleo. Por eso, tras una recesión, el salario no baja lo suficiente para recuperar a los despedidos. Este mecanismo opera a través de la curva de salarios (tema A27, apartado 2.4).

2. *De emparejamiento* (Pissarides, 1992): este mecanismo se deriva del propio modelo. El desempleo de larga duración provoca pérdida de capital humano, reconversión y efecto desánimo, que en el modelo se traducen en una caída de la eficiencia del emparejamiento ($\downarrow A$). Por ello, cuando las vacantes vuelven a su nivel, el desempleo es mayor que al inicio, porque la curva de Beveridge se ha desplazado hacia fuera[11].

3. *De capital* (Bean, 1989): la recesión reduce la inversión y el *stock* de capital, y con él la demanda de trabajo compatible con cada salario.

Las tres vías coinciden en que la política de estabilización tiene efectos permanentes, pero difieren en el remedio. La primera pide moderación salarial y reforma de la negociación; la segunda, políticas activas de empleo que restauren la eficiencia del emparejamiento, y la tercera, estímulo a la inversión.

2.3.3. Aplicaciones: los casos español y estadounidense

El caso español: tras la crisis de 2008 la curva de Beveridge se desplazó hacia fuera, lo que indica una pérdida de eficiencia del emparejamiento y no solo una caída de la demanda de trabajo, por cuatro causas documentadas:

1. *Desajuste de cualificaciones*: mano de obra de la construcción sin habilidades para los sectores emergentes.

2. *Menor movilidad geográfica*: el predominio de la vivienda en propiedad frente al alquiler encarece el cambio de residencia (Oswald, 1996).

3. *Histéresis vía desempleo de larga duración*: descapitalización, señal negativa, desmotivación.

4. *Diferencias regionales*: las políticas activas, gestionadas por las comunidades autónomas, difieren entre regiones.

[11] La figura A26.5 representa el caso opuesto, una mejora de la eficiencia ($\Delta A > 0$) que lleva de 0 a 1, de modo que la histéresis equivale a pasar del punto 1 al 0.

El caso estadounidense (2021-2022): tras la pandemia, las combinaciones de desempleo y vacantes de EE. UU. se situaron muy por encima de la curva de Beveridge previa. En 2022, las vacantes alcanzaron máximos de la serie con una tasa de paro similar a la de 2019. Ello abrió un debate entre dos lecturas, y la normalización de 2023 favoreció la segunda:

- *Pérdida persistente de eficiencia del emparejamiento*: exigiría más desempleo para reducir la inflación (Blanchard, Domash y Summers, 2022).

- *Tramo casi vertical de una curva estable*: con un mercado muy estrecho la curva es muy inclinada, de modo que las vacantes podían caer sin apenas aumento del paro (Figura y Waller, 2022).

2.4. Conclusiones

Aspectos positivos: el modelo de emparejamiento es hoy el marco de referencia del análisis del desempleo de equilibrio. Explica la coexistencia de vacantes y parados, dota de microfundamentos a la curva de Beveridge y a la curva de salarios, y permite distinguir las perturbaciones coyunturales de las estructurales, que es lo que la política de empleo necesita. Además, es fácil de contrastar, permite el análisis institucional y ha inspirado las recomendaciones de reforma laboral en España (Manifiesto de los 100, 2009).

Limitaciones: el *puzzle de Shimer* (2005) consiste en que, calibrado con datos de EE. UU. (1951-2003), el modelo predice una volatilidad del desempleo y de las vacantes mucho menor que la observada. En particular, la estrechez simulada tiene una desviación típica diez veces menor que la real, porque el salario negociado absorbe la mayor parte de las perturbaciones de productividad. Las soluciones propuestas son un mayor valor de las actividades de no mercado (Hagedorn y Manovskii, 2008), rigidez salarial (Hall, 2005), costes fijos de emparejamiento (Pissarides, 2009) y heterogeneidad de puestos.

3. Costes de ajuste y dinámica de la demanda de trabajo

Si cambiar la plantilla tiene costes, la empresa tarda en llevarla a su nivel óptimo. Los modelos con costes de ajuste explican así cómo evoluciona el empleo demandado. Del lado de la oferta, la dinámica procede de la sustitución intertemporal del modelo de Lucas-Rapping (tema A25, apartado 2.2), $l_t^s/l_{t+1}^s = f[(w_t/w_{t+1})^+, r^+, \rho^-]$, que no incorpora costes de ajuste, pues el trabajador reasigna sus horas entre periodos sin fricciones. Los costes de ajuste relevantes son, por ello, los de la demanda.

3.1. Concepto y formas de los costes de ajuste

Se llaman **costes de ajuste** los que, además del precio del factor, paga la empresa por cambiar el ritmo al que incorpora factores productivos. Formalmente, $C = C(\dot{L})$ con $C(0) = 0$, $C'(\dot{L}) > 0$ para $\dot{L} > 0$ y $C'(\dot{L}) < 0$ para $\dot{L} < 0$, de modo que el coste crece con la magnitud del ajuste, $|\dot{L}|$[12]. Pueden ser de dos tipos:

- *Internos* (Eisner y Strotz, 1963; Lucas, 1967): pérdidas temporales de eficiencia por cambios en la producción, de carácter implícito contablemente (mermas de producción, formación de los nuevos trabajadores).

- *Externos*: los que nacen de la relación con terceros, de carácter explícito (consultores, costes de contratación, indemnizaciones por despido).

Formas funcionales: determinan la dinámica del empleo (tema A12, apartado 3.3.2):

1. *Convexos*: ajustar es tanto más caro cuanto más rápido se hace ($C'' > 0$). Son *simétricos* si contratar cuesta lo mismo que despedir, *asimétricos* si despedir cuesta más, como imponen las indemnizaciones.

2. *Lineales*: el coste es proporcional al número de trabajadores ajustados ($C'' = 0$), y también pueden ser simétricos o asimétricos.

3. *Fijos*: existe un coste independiente del tamaño del ajuste, como el de un expediente de regulación de empleo.

En todos los casos los costes de ajuste estabilizan el empleo a costa de la eficiencia. La empresa suaviza la variación (costes convexos) o no ajusta ante perturbaciones pequeñas (costes lineales o fijos), de modo que el empleo varía menos que el producto, y tanto menos cuanto mayores son los costes.

3.2. Demanda de trabajo

La dinámica de la demanda de trabajo depende de la forma funcional de los costes de ajuste. Se parte del caso de referencia sin costes y se analizan después los costes convexos, lineales y fijos.

[12] Los factores variable y fijo de la teoría tradicional son casos polares, con costes de ajuste nulos o infinitos, respectivamente.

3.2.1. Sin costes de ajuste

El caso de referencia es la demanda de trabajo neoclásica sin fricciones (tema A25, apartado 1.3), planteada en forma dinámica.

Planteamiento: la empresa maximiza el valor actual de sus beneficios eligiendo la senda de empleo:

$$\max_{\{L_t\}} V(0) = \int_0^\infty e^{-rt}\left[p_t f(L_t) - w_t L_t\right] dt \quad \Rightarrow \quad \frac{\partial \pi_t}{\partial L_t} = 0 \;\Rightarrow\; \underbrace{p_t \left.\frac{df}{dL_t}\right|_{L_t^*}}_{\text{VPmg}_L} = \underbrace{w_t}_{\text{salario}}$$

donde $\{L_t\}$ es la senda de empleo (variable de decisión), $V(0)$ el valor actual de los beneficios, r el tipo de descuento, p_t el precio del producto, $f(L_t)$ la función de producción, w_t el salario y π_t el beneficio de cada instante.

CPO: $p_t\, df/dL|_{L_t^*} = w_t$. Sin costes de ajuste el problema es separable en el tiempo, de modo que la CPO de cada periodo es la estática.

CSO: $f'' < 0$, es decir, el beneficio de cada instante es estrictamente cóncavo en el empleo, por lo que la CPO estática es suficiente.

Solución: en cada instante, la demanda de trabajo es la neoclásica, decreciente en el salario real w_t/p_t, y el empleo salta instantáneamente a su nivel óptimo ante cualquier perturbación.

3.2.2. Costes de ajuste convexos (Holt et al., 1960; Eisner y Strotz, 1963)

Origen: Holt, Modigliani, Muth y Simon (1960) y Eisner y Strotz (1963) introdujeron los costes de ajuste cuadráticos en la decisión dinámica de la empresa.

Idea: si ajustar deprisa es más caro que hacerlo despacio, la empresa reparte el ajuste de la plantilla en el tiempo.

Planteamiento: con costes variables y cuadráticos, $C = (b/2)\dot{L}^2$, el coste marginal de ajuste es creciente y simétrico:

$$\max_{\{L_t\}} V(0) = \int_0^\infty e^{-rt}\left[p_t f(L_t) - w_t L_t - \frac{b}{2}\dot{L}_t^2\right] dt$$
$$\Rightarrow\; \underbrace{p_t \frac{df}{dL_t}}_{\text{VPmg}_L} = w_t + \underbrace{r b \dot{L}_t - b\ddot{L}_t}_{\text{Cmg}_{\text{ajuste}}}$$

donde $\{L_t\}$ es la senda de empleo (variable de decisión), $\dot{L}_t$ y $\ddot{L}_t$ su primera y segunda derivadas temporales, $V(0)$ el valor actual de los beneficios, r el tipo de descuento, p_t el precio del producto, $f(L_t)$ la función de producción, w_t el salario y $b > 0$ el parámetro del coste cuadrático de ajuste $C = (b/2)\dot{L}^2$. El empleo inicial, L_0, está dado.

CPO: la condición de Euler establece que la empresa contrata hasta que el valor de la productividad marginal cubre el salario más el coste marginal de ajustar la plantilla. La cierra la condición de transversalidad, $\lim_{t\to\infty} e^{-rt}\, b\, \dot{L}_t\, L_t = 0$, que descarta las sendas explosivas y deja la que converge al estado estacionario. En el estado estacionario ($\dot{L} = \ddot{L} = 0$) se recupera la condición estática $p\, df/dL|_{L^*} = w$.

CSO: el integrando es estrictamente cóncavo en $(L_t, \dot{L}_t)$, pues $f'' < 0$ y $b > 0$, de modo que la condición de Euler, junto con la de transversalidad, es suficiente (Mangasarian).

Dinámica: el ajuste depende de la naturaleza de la perturbación:

- *Permanente*: el ajuste es gradual y completo hasta el nuevo estado estacionario.

- *Transitoria*: el ajuste es gradual e incompleto, porque no compensa pagar el coste de ajuste para un cambio que va a revertirse.

De ahí que la productividad sea procíclica. Como el empleo va por detrás de la producción, Y/L sube tras un impulso favorable y baja tras uno adverso[13]. Con costes convexos asimétricos (Pfann y Palm, 1993), la dinámica es la misma, pero la velocidad del ajuste depende de la dirección y es más lenta al despedir que al contratar cuando, como imponen las indemnizaciones, despedir cuesta más.

Conclusiones:

- *Aspectos positivos*: explica el ajuste gradual del empleo, distingue la respuesta a perturbaciones permanentes y transitorias y da cuenta de la productividad procíclica. Su forma cuadrática conduce, además, a ecuaciones de ajuste parcial fáciles de estimar.

- *Limitaciones*: la convexidad y la simetría son supuestos de conveniencia analítica que la evidencia microeconométrica no confirma. El modelo tampoco genera los periodos de inacción ni los ajustes a saltos que muestran los datos de empresa, por lo que se ajusta mejor a la dinámica del empleo agregado, que la agregación de ajustes discretos suaviza.

3.2.3. Costes de ajuste lineales (Nickell, 1986)

Origen: Stephen Nickell (1986) analizó el caso de los costes lineales.

Idea: si el coste marginal de ajustar es constante, no hay ganancia en la gradualidad y el ajuste, cuando se produce, es inmediato.

Planteamiento: con costes lineales, $C = c|\dot{L}|$, el programa de la empresa es:

$$\max_{\{L_t\}} V(0) = \int_0^\infty e^{-rt} \left[p_t f(L_t) - w_t L_t - c\,|\dot{L}_t|\right] dt \quad \Rightarrow \quad \underbrace{p_t \left.\frac{df}{dL_t}\right|_{L_t^*}}_{\text{VPmg}_L} = w_t + \underbrace{rc}_{\text{Cmg}_{\text{ajuste}}}$$

donde $\{L_t\}$ es la senda de empleo (variable de decisión), $\dot{L}_t$ su variación por unidad de tiempo, $V(0)$ el valor actual de los beneficios, r el tipo de descuento, p_t el precio del producto, $f(L_t)$ la función de producción, w_t el salario y c el coste unitario de ajuste (constante por trabajador contratado o despedido). El empleo inicial, L_0, está dado.

CPO: $p_t\, df/dL|_{L_t^*} = w_t + rc$ al contratar ($\dot{L}_t > 0$) y $p_t\, df/dL|_{L_t^*} = w_t - rc$ al despedir ($\dot{L}_t < 0$). La productividad marginal cubre el salario más (menos) el flujo equivalente, rc, del coste único de contratar (despedir), y entre ambos valores la empresa no ajusta.

CSO: el integrando es cóncavo en $(L_t, \dot{L}_t)$, pues $f'' < 0$ y $-c|\dot{L}_t|$ es cóncava, de modo que la condición es suficiente. Su no diferenciabilidad en $\dot{L}_t = 0$ genera la banda de inacción.

[13] Si la empresa anticipa una perturbación futura, los costes convexos hacen que esta tenga efectos incluso antes de materializarse.

Desarrollo gráfico: las sendas de ajuste del empleo se representan en el plano tiempo-empleo (figura A26.7), ante una perturbación permanente en t_0 que eleva el empleo desde L_0^*:

1. *Senda con costes lineales*: es un escalón, con salto inmediato a L_1^{lin}, algo inferior al nivel sin costes porque $p\,df/dL = w + rc$.

2. *Senda con costes convexos*: es una curva que se aproxima asintóticamente a L_1^*, donde $p\,df/dL = w$. El ajuste es gradual porque el coste marginal de ajustar, $b\dot{L}$, crece con la velocidad del ajuste.

3. *Estado estacionario*: la senda convexa acaba por encima de la lineal porque con costes cuadráticos el coste marginal de ajuste, $b\dot{L}$, se anula al llegar al estado estacionario y no deja cuña alguna. Con costes lineales, en cambio, el coste unitario c no desaparece por pequeño que sea el ajuste, y su coste de oportunidad, rc, actúa como una cuña permanente sobre el salario.

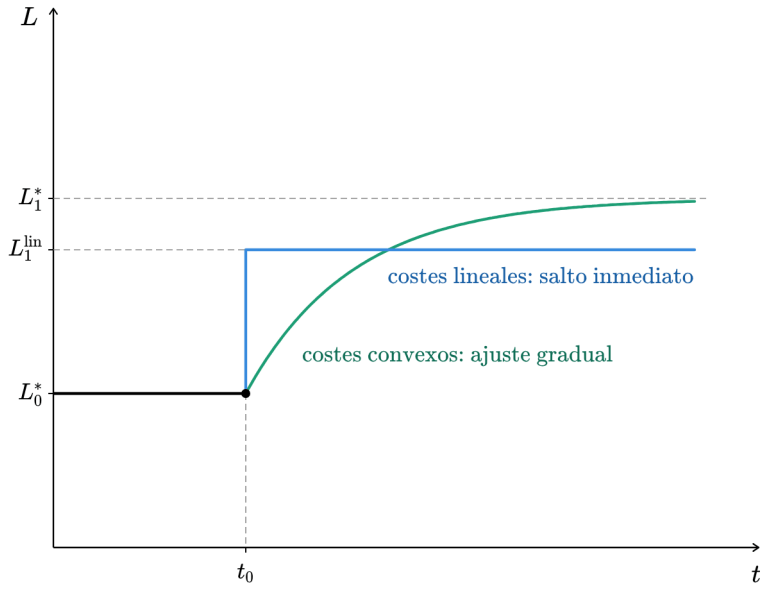


Figura A26.7: Sendas de ajuste del empleo ante una perturbación permanente. Con costes lineales el salto es inmediato (a un nivel final menor, L_1^{lin}), mientras que con costes convexos la aproximación es gradual (hacia L_1^*).

Dinámica: el ajuste depende de la naturaleza de la perturbación:

- *Permanente*: el ajuste es inmediato y de una sola vez, aunque se detiene en L_1^{lin}, algo por debajo del nivel sin costes.

- *Transitoria*: si la desviación del ciclo respecto a la tendencia es pequeña, la empresa no ajusta, pues se encuentra dentro de una banda de inacción alrededor del óptimo, mientras que, si la desviación es grande, el ajuste es inmediato pero incompleto.

La divergencia respecto al nivel sin costes es menor que con costes convexos durante la transición, aunque, a diferencia de estos, persiste a largo plazo una cuña rc. Con costes lineales asimétricos, el ajuste sigue siendo inmediato y la banda de inacción es asimétrica.

Conclusiones:

▪ *Aspectos positivos*: explica la banda de inacción y los ajustes inmediatos, más próximos a lo que muestran los datos de empresa que la gradualidad de los costes convexos. Su versión asimétrica es, además, el marco del análisis de los costes de despido.

▪ *Limitaciones*: un coste proporcional al ajuste no recoge el componente fijo que la evidencia microeconométrica encuentra en los costes de ajuste, por lo que el modelo no explica por sí solo los ajustes grandes e infrecuentes de los datos de empresa. Por ello, el trabajo aplicado lo combina con otras formas de coste.

3.2.4. Costes de ajuste fijos (Hamermesh, 1989)

Origen: Daniel Hamermesh (1989) mostró, con datos mensuales de establecimientos, que el empleo de cada planta se ajusta a saltos, lo que revela un componente fijo en los costes de ajuste.

Idea: los costes de ajuste fijos dan lugar a reglas de ajuste del tipo (S, s), porque con un coste fijo la empresa solo ajusta cuando la desviación respecto a la plantilla óptima compensa pagarlo.

Razonamiento: si ajustar cuesta una cantidad fija k, añadir o quitar un trabajador más no cuesta nada en el margen, así que la empresa, cuando ajusta, lo hace de golpe y del todo. Sin embargo, solo le compensa hacerlo si lo que gana con la plantilla óptima, una vez pagado k, supera lo que obtiene sin tocar la plantilla actual. El resultado es una regla (S, s), en la que el empleo permanece constante dentro de una banda y salta cuando la desviación respecto al óptimo la excede. Se generan así ajustes discretos e infrecuentes en cada empresa que, agregados, producen una dinámica suave.

Conclusiones:

▪ *Aspectos positivos*: reproduce los periodos de inacción y los ajustes grandes e infrecuentes de los datos de empresa, y explica por qué el empleo agregado, que suma decisiones de empresas que ajustan en momentos distintos, se mueve de forma gradual.

▪ *Limitaciones*: la regla (S, s) no conduce a una ecuación de ajuste sencilla, y la dinámica agregada depende de la distribución de las desviaciones entre empresas, por lo que el modelo es más difícil de estimar que el de costes convexos.

3.3. Evidencia empírica e implicaciones de política

Balance empírico: el repaso de Hamermesh y Pfann (1996) constata que la literatura adoptó sobre todo la especificación cuadrática y simétrica. Los estudios microeconométricos con datos de empresa dibujan, sin embargo, costes de ajuste asimétricos, no necesariamente convexos y con un componente fijo de peso, que explican los periodos de inacción y los ajustes a saltos. El análisis admite, además, dos extensiones importantes. Una es el margen intensivo, pues ajustar horas resulta más barato que ajustar trabajadores. La otra es la heterogeneidad de los costes de ajuste entre trabajadores, que hace recaer el ajuste sobre los temporales, como en el mercado dual español.

Costes de despido y empleo (Bentolila y Bertola, 1990): con costes de ajuste lineales y asimétricos, un aumento del coste de despido reduce los despidos en las recesiones más de lo que reduce las contrataciones en las expansiones. La razón es que la empresa descuenta el coste futuro de despedir al contratar, pero lo paga con certeza al despedir. Por tanto, el efecto sobre el empleo medio es pequeño y positivo, mientras que el efecto sobre los flujos es inequívoco, pues el mayor coste de despido reduce la rotación y alarga la duración del desempleo. Hopenhayn y Rogerson

(1993) añaden el coste de bienestar, porque la protección del empleo frena la reasignación de trabajadores de las empresas menos a las más productivas y reduce la productividad agregada.

El mercado dual español (Bentolila y Dolado, 1994): con dos tipos de contrato, indefinidos con altos costes de despido y temporales sin ellos, el ajuste recae sobre los temporales con independencia de su productividad. Los indefinidos, protegidos, presionan por su parte al alza los salarios (*insiders*), y la volatilidad del empleo se concentra en el margen temporal, lo que explica la extraordinaria sensibilidad cíclica del empleo español desde 1984.

Implicaciones de política: la política económica afecta al desempleo no solo a través de las decisiones de búsqueda sino también de la estructura de los costes de rotación. La legislación de protección del empleo estabiliza el empleo a costa de la eficiencia, con sobreempleo en las recesiones y empleo inferior al óptimo en las expansiones.

Conclusiones

El tema ha mostrado que el desempleo friccional es un fenómeno de equilibrio, porque encontrar empleo y cubrir una vacante lleva tiempo cuando la información es imperfecta. Los modelos de búsqueda, de Stigler y McCall a Mortensen, explican la duración del desempleo y el efecto de las prestaciones, aunque toman como dada una distribución de salarios que la paradoja de Diamond pone en cuestión. Después, la teoría del emparejamiento incorpora las vacantes, determina a la vez la estrechez, el salario y el desempleo y da fundamento a la curva de Beveridge. Sin embargo, su equilibrio solo es eficiente si se cumple la condición de Hosios, y el modelo genera menos volatilidad de la observada (el *puzzle* de Shimer). Por último, los costes de ajuste explican por qué el empleo varía menos que el producto y por qué la protección del empleo afecta más a los flujos que al nivel del desempleo.

El desempleo friccional se define por contraste con el mercado sin fricciones, cuyo equilibrio competitivo sirve de referencia (tema A25). El reparto del excedente de cada emparejamiento se resuelve con la solución de negociación de Nash (tema A14), y los costes de ajuste de la plantilla siguen la misma lógica que los del capital en la teoría q de la inversión (tema A34). Frente a la negociación colectiva y los salarios de eficiencia, que atribuyen el desempleo involuntario a salarios superiores al de equilibrio (tema A27), aquí el desempleo procede de las fricciones. Ambas tradiciones confluyen, por último, en la tasa natural de paro y en la histéresis, que explican por qué el desempleo puede persistir tras una perturbación (tema A28).

Tras la pandemia, las combinaciones de desempleo y vacantes de EE. UU. se situaron muy por encima de la curva de Beveridge previa. Blanchard, Domash y Summers (2022) lo interpretaron como una pérdida de eficiencia del emparejamiento, y Figura y Waller (2022), como un movimiento sobre el tramo inclinado de una curva estable. Las propias fricciones generan, además, poder de monopsonio, como en Burdett y Mortensen (1998), y la evidencia reciente lo considera extendido (Manning, 2021). Sigue abierto el diseño de las prestaciones, ya que la ampliación de los subsidios en EE. UU. durante la pandemia redujo poco la tasa de encuentro de empleo (Ganong et al., 2024).

A27. Determinación de salarios

Determinación de salarios: modelos de negociación, salarios de eficiencia y contratos implícitos.

Introducción

> **Introducción común.** Los temas A25 a A28 comparten una introducción, que se desarrolla al comienzo del tema **A25**. Presenta la teoría económica del mercado de trabajo, que estudia el empleo y los salarios de equilibrio y los factores institucionales que los moldean, y la explicación teórica de cada tipo de desempleo.

La teoría neoclásica describe un mercado de trabajo que alcanza un único equilibrio, estable y eficiente. En él, el salario real equivale a la productividad marginal y la oferta casa al instante con la demanda. La flexibilidad del salario lo explica, pues basta con que este se mueva para que el mercado vuelva por sí solo al equilibrio tras cualquier desajuste. Ahora bien, el **desempleo involuntario** persistente, las **rigideces salariales** y las diferencias de salario entre trabajadores similares que se observan en el mercado laboral no encajan con el modelo neoclásico.

Esa discrepancia tiene consecuencias de política económica, porque, si el salario no baja hasta vaciar el mercado por razones que convienen a empresas y trabajadores, el desempleo es involuntario y no se corrige solo con flexibilizar los salarios. Las diferencias entre países apuntan en esa dirección. El paro alto y persistente de la Europa continental de los años ochenta, la *euroesclerosis* (Giersch, 1985), encaja con los modelos de negociación y de *insiders* y *outsiders*, mientras que en mercados con poca afiliación sindical, como el estadounidense, pesan más los salarios de eficiencia.

Estas teorías son también las piezas del modelo de desempleo de equilibrio (tema A28), en el que la negociación o los salarios de eficiencia determinan el salario que se fija por encima del de vaciado. La idea de que pagar más de lo necesario puede ser rentable para la empresa, porque eleva el esfuerzo, la lealtad o la calidad de la plantilla, se apoya tanto en modelos de incentivos (Shapiro y Stiglitz, 1984) como en la lectura sociológica del intercambio de regalos de Akerlof (1982). Los contratos implícitos explican mejor, en cambio, la estabilidad de la renta salarial que el desempleo. De todas estas teorías salen buena parte de las propuestas de reforma del mercado de trabajo, desde la negociación colectiva hasta las políticas activas, que el anexo del tema recoge.

La nueva economía keynesiana desarrolló tres líneas para microfundamentar estas rigideces reales:

- *Contratos implícitos*: conciben el salario como un seguro.
- *Modelos de negociación*: lo conciben como un reparto de rentas.
- *Salarios de eficiencia*: lo consideran un incentivo a la productividad.

Estos modelos explican rigideces reales, no nominales, mientras que las nominales requieren extensiones adicionales. A los trabajadores les preocupa su consumo real (contratos implícitos), su renta real (negociación) o su salario real (eficiencia), de modo que, si cambia la cantidad de dinero, varía el salario nominal, pero no los resultados reales.

1. Contratos implícitos (Baily, 1974; Gordon, 1974; Azariadis, 1975)

Una de las primeras líneas de investigación para explicar la rigidez de los salarios reales parte del reparto del riesgo entre trabajadores y empresas y se apoya, por tanto, en la economía de la incertidumbre (tema A10). Su punto de partida es que los trabajadores no pueden asegurarse en los mercados financieros frente a las fluctuaciones de su renta, de modo que la empresa, con mejor acceso a ellos, puede ofrecerles ese seguro a través del propio contrato de trabajo.

Origen: en la práctica, muchas relaciones laborales duran años, un rasgo que el mercado de trabajo walrasiano no recoge. En vez de recurrir al mercado cada periodo para contratar a quienes necesitan, las empresas retienen a su plantilla durante mucho tiempo. Las teorías de los contratos implícitos buscan explicar por qué, en presencia de ciertos fallos de mercado, los salarios pueden diferir de los de competencia perfecta.

Idea: las relaciones laborales a largo plazo entre trabajadores y empresas hacen que los contratos de trabajo fijen rentas salariales estables (no contingentes al estado de la naturaleza). Los contratos actúan así como seguro ante periodos de demanda de trabajo fluctuante. En consecuencia, la renta salarial se desvincula de las condiciones del mercado en cada periodo, aunque el empleo puede seguir fijándose de forma eficiente. De hecho, en estos modelos, la mejor distribución del riesgo en el largo plazo supone la fijación de una renta salarial fija, menor en los estados favorables que el valor de lo producido por el trabajo. Se habla de *contrato implícito* porque, en la práctica, esas cláusulas no se escriben, y nada fija de forma expresa qué horas y qué salario corresponden a cada realización de la productividad.

1.1. Supuestos

Entorno: información perfecta y precios normalizados (variables reales), de modo que el gasto en consumo real coincide con la renta real total.

Agentes:

- *Trabajadores*: son aversos al riesgo, puesto que no tienen una amplia capacidad para diversificar sus fuentes de renta. Concretamente, toda su renta proviene del trabajo, por lo que el consumo real coincide con la renta real laboral, $c_i = w_i L_i$. Su función de utilidad de Bernoulli es tal que $u' > 0$ y $u'' < 0$, y la desutilidad del trabajo cumple $v' > 0$ y $v'' > 0$. El consumo les reporta utilidad y el trabajo desutilidad, de modo que sus preferencias pueden representarse mediante una función de utilidad esperada de von Neumann y Morgenstern (VNM):

$$u^e = \sum_{i=1}^{n} p_i[u(c_i) - v(L_i)]$$

Los trabajadores cuentan con una utilidad de reserva $\underline{u}$, de manera que su condición de participación será $u^e \geq \underline{u}$. La participación es irrevocable, es decir, una vez que han aceptado un contrato, no hay movilidad laboral.

- *Empresa*: es neutral al riesgo, porque los mercados financieros le permiten diversificarlo mejor que al trabajador. Su tecnología está sujeta a *shocks* de productividad, recogidos en una variable aleatoria no negativa, A_i, que puede tomar n valores distintos, cada uno con una probabilidad p_i de darse. El beneficio esperado de la empresa es:

$$\pi^e = \sum_{i=1}^{n} p_i[A_i f(L_i) - w_i L_i]$$

donde L_i son las horas trabajadas y w_i el salario por hora en el estado i, y $f(L_i)$ la función de producción, con $f' > 0$ y $f'' < 0$.

1.2. Modelo

Planteamiento: la empresa maximiza el beneficio esperado sujeta a la condición de participación del trabajador:

$$\max_{L_i, c_i} \pi^e = \sum p_i[A_i f(L_i) - c_i] \quad \text{s.a.:} \sum p_i[u(c_i) - v(L_i)] \geq \underline{u}$$

donde L_i y c_i son las horas de trabajo y el consumo (igual a la renta laboral, $c_i = w_i L_i$) que el contrato fija en cada estado i (variables de decisión), p_i la probabilidad del estado i, A_i el *shock* de productividad, $f(\cdot)$ la función de producción, $u(\cdot)$ y $v(\cdot)$ la utilidad del consumo y la desutilidad del trabajo del trabajador, $\underline{u}$ su utilidad de reserva y λ (en la CPO) el multiplicador de la condición de participación.

CPO (condiciones de primer orden):

$$\frac{du}{dc_i} = \frac{1}{\lambda} \quad \forall i \tag{1}$$

$$\underbrace{A_i \frac{df}{dL_i}}_{\text{Pmg}_L} = \underbrace{\frac{dv/dL_i}{du/dc_i}}_{\text{RMS}} \tag{2}$$

La condición (1) establece que la utilidad marginal del consumo es constante para todos los estados de naturaleza, por lo que el propio consumo, igual a la renta salarial real $c_i = w_i L_i$, es también constante para todas las realizaciones de la perturbación tecnológica. Es decir, la empresa ofrece un *aseguramiento completo* al trabajador averso al riesgo.

La condición (2) establece que la empresa fija el empleo de cada estado de la naturaleza (cada realización de la perturbación tecnológica) en el nivel que iguala la productividad marginal del trabajo a la relación marginal de sustitución del trabajador, $(dv/dL_i)/(du/dc_i)$. Esta relación equivale a la desutilidad marginal de las horas adicionales expresada en unidades de consumo.

CSO: la concavidad de u y de f y la convexidad de v hacen cóncavo el lagrangiano en (L_i, c_i), de modo que (1) y (2) son necesarias y suficientes.

Solución: el contrato estabiliza la renta laboral, no el salario por hora. Con consumo constante, c^*, el salario por hora, $w_i = c^*/L_i$, se mueve en sentido contrario a las horas, y estas siguen respondiendo a los *shocks* según (2), que es la condición de eficiencia del contrato. Así, el empleo es eficiente en cada estado y no hay desempleo involuntario. Además, como el consumo no varía y en los estados desfavorables se trabajan menos horas, el trabajador está mejor en las recesiones que en los auges.

Extensiones: si solo la empresa observa el *shock*, el contrato eficiente deja de ser compatible con los incentivos, porque la empresa puede declarar el estado que más le conviene. En consecuencia, el contrato óptimo se aparta de la eficiencia y, según los supuestos, genera subempleo en los estados desfavorables (Grossman y Hart, 1981) o sobreempleo (Green y Kahn, 1983).

1.3. Conclusiones

Aspectos positivos: los contratos implícitos explican por qué la renta salarial es estable. El salario cumple en ellos una doble función, pues paga el trabajo y, a la vez, asegura al trabajador frente al riesgo (aquí, el tecnológico). Además, los contratos implícitos responden a la existencia de incertidumbre y de utilización de capital humano específico entre empresas y trabajadores aversos al riesgo. Estos contratos suavizan el consumo, pero hacen que la volatilidad del empleo sea mayor.

Limitaciones: se han formulado cuatro objeciones:

- *Ajuste*: con información simétrica, el empleo es eficiente en cada estado, de modo que el modelo explica la estabilidad de la renta salarial, pero no el desempleo involuntario (Akerlof y Miyazaki, 1980), y el ajuste recae en las horas trabajadas más que en los despidos.

- *Nuevas contrataciones*: no explica por qué la empresa no ofrece salarios menores a quienes contrata.

- *Rigidez nominal*: explica la rigidez de la renta salarial real, pero no puede justificar la rigidez a la baja de los salarios nominales ante perturbaciones negativas de la demanda agregada.

- *Vigencia*: a pesar de su popularidad en los años ochenta, su aplicación en la economía laboral ha estado en declive desde los años noventa. El declive se debe a que las teorías de búsqueda y emparejamiento (modelo de Diamond-Mortensen-Pissarides, tema A26) han superado ampliamente la de los contratos implícitos.

2. Modelos de negociación colectiva

Los modelos de negociación conciben el salario como un reparto de rentas entre la empresa y los trabajadores o el sindicato que los representa. A diferencia del mercado walrasiano, el salario no resulta del cruce de la oferta y la demanda, sino del poder de negociación de cada parte, y el empleo depende de qué variables se negocien. Tras los fundamentos, se exponen el modelo *right to manage*, los contratos eficientes y el modelo de *insiders* y *outsiders*.

2.1. Fundamentos

Antes de los modelos de negociación colectiva se presentan el concepto de negociación y de poder de negociación, la estructura de la negociación según su grado de centralización y las soluciones microeconómicas al problema de negociación. De estas soluciones procede el producto de Nash que utilizan los modelos posteriores.

2.1.1. La teoría de la negociación y el proceso de negociación

La **teoría de la negociación** se ocupa del reparto pactado, entre agentes racionales, de un excedente que solo surge si cooperan. Su aplicación al mercado laboral se dirige sobre todo a la negociación salarial entre trabajadores y empresas. Las dos partes se reparten la renta que genera la relación laboral (lo producido por encima de lo que cada parte obtendría sin acuerdo) mediante la fijación del salario y, en su caso, del empleo. Por ello, el resultado depende en gran medida del **poder de negociación** de cada parte, es decir, de su capacidad para apropiarse de una fracción mayor de ese excedente. Ese poder aumenta con las alternativas de cada parte fuera del acuerdo y con su capacidad para resistir un conflicto.

La negociación laboral se desarrolla, en general, en tres niveles, según negocien el empleador y el trabajador individual, el empleador y un sindicato o la patronal y los sindicatos de un sector o del conjunto de la economía.

2.1.2. Grados de centralización de la negociación colectiva (Calmfors y Driffill, 1988)

Origen: Lars Calmfors y John Driffill (1988) señalan que el grado de centralización o estructura de la negociación laboral colectiva tiene efectos importantes sobre el desempleo.

Idea: los casos extremos (centralizados o descentralizados) producen mejores resultados que los casos intermedios.

Razonamiento: cuando la negociación está muy centralizada, los negociadores son conscientes de los efectos macroeconómicos de sus acuerdos (sobre los precios, el desempleo y los impuestos que financian las prestaciones) y los internalizan, lo que modera sus demandas salariales. Cuando está muy descentralizada, la competencia en el mercado de productos disciplina los salarios, porque una subida en una sola empresa encarece su producto y reduce mucho su empleo. La negociación por sectores, de grado intermedio, combina, en cambio, poder de mercado suficiente para elevar los salarios con una internalización escasa de sus efectos. Por ello, en la muestra de los autores, los países con negociación centralizada (p. ej. Austria o los países escandinavos) o muy descentralizada (p. ej. Japón, Suiza o EE. UU.) presentaban menores niveles de desempleo que los de grado intermedio (p. ej. Bélgica o Países Bajos).

Conclusiones:

- *Aspectos positivos*: la hipótesis explica por qué la relación entre centralización de la negociación y desempleo no es monótona, sino en forma de joroba. La joroba resulta de combinar dos fuerzas, la internalización de los efectos macroeconómicos de los salarios cuando la negociación es centralizada y la disciplina de la competencia cuando está descentralizada. Asimismo, la hipótesis ha orientado el debate sobre el diseño institucional de la negociación colectiva.

- *Limitaciones*: la hipótesis atiende al grado de centralización y omite el de coordinación, que puede lograrse sin centralizar la negociación (Soskice, 1990). Además, la evidencia posterior apenas confirma la forma de joroba, que pierde apoyo empírico desde los años noventa (Cahuc y Zylberberg, 2004).

2.1.3. Fundamentación microeconómica de los procesos de negociación

La fundamentación microeconómica de la negociación ha seguido varios enfoques sucesivos, que van de la indeterminación de Edgeworth a la solución estratégica de Rubinstein.

> La solución de Nash y el modelo de Rubinstein se desarrollan en el tema A14, apartado 2.4.3, y la teoría de la negociación de Hicks, en el tema A25, apartado 1.4.2.

Francis Edgeworth (1881) fue uno de los primeros autores en abordar estas cuestiones. Mostró que el acuerdo debe caer sobre la curva de contratos y ser, por tanto, eficiente en el sentido de Pareto. No obstante, su análisis se queda corto, porque solo asegura que no quedan ganancias por aprovechar, sin decir qué parte se lleva cada negociador, y el resultado queda indeterminado.

Después, John Hicks (1932) incorpora la idea de poder negociador. En su análisis, el salario pactado refleja un pulso entre la presión que el sindicato puede ejercer (p. ej. con la amenaza de huelga) y la resistencia de la empresa. La huelga, sin embargo, no llega a estallar, porque cada parte, con información cierta sobre esas fuerzas, calcula cuánto duraría el conflicto y ambas firman desde el principio el salario que lo evita. El modelo es sencillo, y esa es su ventaja, pero trata el poder de negociación como un dato y no explica de dónde procede.

El enfoque axiomático de John Nash (1950) consiste en fijar las propiedades que debe reunir una solución racional de la negociación. Nash muestra que la negociación entre dos partes tiene una única solución, la que hace máximo el producto de las ganancias netas de ambos agentes, si se respetan estos cuatro axiomas:

1. *Optimalidad paretiana*: el acuerdo no deja sin aprovechar ninguna ganancia mutua.

2. *Invarianza ante transformaciones afines positivas*: la solución no varía si se cambia la función de utilidad por otra equivalente.

3. *Independencia de alternativas irrelevantes*: eliminar opciones no elegidas no altera la solución.

4. *Simetría*: dos jugadores que ocupan posiciones idénticas reciben ganancias iguales.

Si se prescinde del axioma de simetría, la solución recoge el distinto poder negociador de cada jugador. Esta versión se denomina **solución de Nash generalizada**:

$$\max_{(u_1,u_2)\in\mathcal{U}} \Omega = (u_1 - u_1^0)^{\gamma}(u_2 - u_2^0)^{1-\gamma}$$

donde $\mathcal{U}$ es el conjunto de pares de utilidades factibles, u_1 y u_2 son las utilidades que obtienen los dos jugadores con el acuerdo, u_1^0 y u_2^0 las que obtendrían en caso de desacuerdo y $\gamma \in [0,1]$ el poder negociador del jugador 1 (con $\gamma = 1/2$ se recupera la solución simétrica).

Finalmente, Ingemar Stahl (1972) y Ariel Rubinstein (1982) estudian la negociación como un juego dinámico de ofertas y contraofertas, con horizonte finito (Stahl) e infinito (Rubinstein), cuya solución es el equilibrio de Nash perfecto en subjuegos (ENPS). En él, el poder de negociación procede de la paciencia relativa de los jugadores (sus factores de descuento). Cuando el intervalo entre ofertas tiende a cero, el reparto converge a la solución de Nash generalizada, lo que da fundamento estratégico al parámetro γ (Binmore, Rubinstein y Wolinsky, 1986).

2.2. Contratos ineficientes: *right to manage* (Nickell y Andrews, 1983)

> Para el ejercicio oral no hace falta el desarrollo analítico, y basta con exponer las ideas y el gráfico. El desarrollo se recoge para la comprensión conceptual del modelo.

Origen: en la mayoría de los países industrializados, una proporción significativa de los salarios se regula a través de acuerdos colectivos. Por ello, los primeros modelos para representar la negociación colectiva se basan en la existencia de un sindicato monopolista de la oferta de trabajo y se apoyan en las aportaciones de John Dunlop (1944).

Idea: Stephen Nickell y Martyn Andrews (1983) generalizan las aportaciones de Dunlop al considerar que existen distintos grados de monopolización y que la empresa tiene capacidad para despedir o contratar al salario dado. Esa generalización da lugar a los modelos *right to manage*, que reflejan la idea de que, ante el monopolio salarial de los sindicatos, las empresas retienen su "derecho a gestionar" su propia demanda de trabajo.

2.2.1. Supuestos

Entorno:

- *Información*: perfecta.
- *Contexto*: dinámico-secuencial.
- *Variable estratégica*: solo el salario.

La empresa y el sindicato negocian únicamente el salario, y con él ya fijado la empresa elige el nivel de empleo. Si no hay acuerdo, la empresa no obtiene beneficio y el sindicato se queda con el salario de reserva, $\underline{w}$.

Agentes:

- *Sindicato*: agrupa a N^s afiliados iguales y su poder de negociación es $\gamma \in [0,1]$. Su objetivo es una utilidad esperada de tipo VNM, estrictamente creciente en el salario y estrictamente cóncava, pues los afiliados tienen aversión al riesgo:

$$U^e = p \cdot U(w) + (1-p) \cdot U(\underline{w})$$

donde $p = \min\{L/N^s, 1\}$ es la probabilidad de que un afiliado sea contratado, $(1-p)$ la de quedar en paro y $\underline{w}$ el salario de reserva, por debajo del cual el individuo no acepta trabajar. No hay elección de horas, y cada trabajador está empleado o desempleado.

▪ *Empresa*: opera con una función de producción estrictamente creciente ($f' > 0$) y estrictamente cóncava ($f'' < 0$), en la que una tecnología de producción A_θ recoge los *shocks* tecnológicos. Su función de beneficios es $\pi = A_\theta f(L) - wL$, con CPO $A_\theta f'(L) = w$.

2.2.2. Modelo

Planteamiento: el salario negociado es el que maximiza el producto de Nash generalizado de las ganancias netas de ambas partes. La maximización tiene en cuenta que el empleo lo elige después la empresa sobre su curva de demanda y no puede superar la afiliación sindical, y que el salario acordado no puede ser inferior al de reserva:

$$\max_w \ \Omega(w) = \pi(w)^{1-\gamma} \cdot [U^e(w) - U(\underline{w})]^\gamma \quad \text{s.a.:} \begin{cases} L = L^d(w) \leq N^s \\ w \geq \underline{w} \end{cases}$$

donde w es el salario negociado (variable de decisión), $\pi(w) = A_\theta f(L) - wL$ el beneficio de la empresa una vez elegido el empleo $L = L^d(w)$, $U^e(w)$ la utilidad esperada del sindicato, $U(\underline{w})$ la que obtiene en caso de desacuerdo, $\gamma \in [0, 1]$ su poder de negociación y $\underline{w}$ el salario de reserva.

CPO: con $\eta = -\frac{w}{L}\frac{dL^d}{dw}$ la elasticidad-salario de la demanda de trabajo y $U^e(w) - U(\underline{w}) = \frac{L}{N^s}[U(w) - U(\underline{w})]$, la condición de primer orden en logaritmos es:

$$\gamma \left[\frac{dU/dw}{U(w) - U(\underline{w})} - \frac{\eta}{w} \right] = (1 - \gamma)\frac{L}{\pi}$$

El lado izquierdo recoge la ganancia marginal del sindicato (más salario, pero menos empleo) y el derecho, la pérdida marginal de beneficio de la empresa, cada una ponderada por el poder de negociación de la parte correspondiente.

CSO: se supone que el producto de Nash es estrictamente cuasicóncavo en w sobre $[\underline{w}, \infty)$, lo que asegura un único salario negociado interior. La hipótesis se cumple con las formas funcionales habituales (demanda de trabajo isoelástica, utilidad cóncava).

Solución: en el caso general, de la CPO se deduce que el salario pactado sube cuando aumentan el poder negociador del sindicato o el salario de reserva:

$$w^* = w^*(\gamma^+, \underline{w}^+)$$

Como caso límite, con un sindicato neutral al riesgo ($U(w) = w$) y $f(L) = L^\alpha$, la demanda de trabajo es isoelástica, con $\eta = 1/(1 - \alpha)$, y la ratio $wL/\pi = \alpha/(1 - \alpha) = \eta - 1$ es constante. La condición da entonces un margen sobre el salario de reserva que crece con γ y decrece con η:

$$w^* = \underline{w}\left(1 + \frac{\gamma}{\eta - 1}\right)$$

Con $\gamma = 1$ se obtiene la solución de Dunlop (monopolio salarial), $w^* = \frac{\eta}{\eta-1}\underline{w}$, y con $\gamma = 0$, el salario de reserva, $w^* = \underline{w}$, que es el resultado competitivo.

Equilibrio: el resultado depende del reparto del poder negociador γ:

- $\gamma = 0$ (todo el poder en la empresa): el sindicato acepta el salario de reserva $\underline{w}$, la empresa contrata sobre su curva de demanda a ese salario y se logra el equilibrio competitivo. Hay eficiencia asignativa (el resultado es óptimo de Pareto) y productiva (el empleo coincide con el de competencia perfecta).

- $\gamma = 1$ (todo el poder en el sindicato): se alcanza la solución de Dunlop (monopolio salarial). El sindicato decide unilateralmente el salario en la tangencia entre su curva de indiferencia y la curva de demanda, un equilibrio que no es óptimo de Pareto (ineficiencia asignativa y productiva y desempleo involuntario).

- $\gamma \in (0, 1)$: equilibrios intermedios, la solución *right to manage*, con contratos ineficientes en el sentido de Pareto.

Por tanto, la primera implicación de la existencia de poder de negociación en el mercado de trabajo es que los salarios serán superiores a los del equilibrio competitivo y, en consecuencia, existirá desempleo involuntario.

Desarrollo gráfico: el modelo *right to manage* se representa en el plano empleo-salario (figura A27.1):

1. *Curvas de indiferencia del sindicato* (CI_0, CI_1): reúnen los pares de empleo y salario con igual utilidad esperada para el sindicato. Son decrecientes, porque salario y empleo se sustituyen de forma imperfecta, y convexas mientras quedan afiliados sin contratar. Una vez empleados todos ($L = N^s$), cada afiliado trabaja con probabilidad 1 y solo el salario cuenta, de ahí que las curvas se vuelvan horizontales. Lo mismo ocurre en $w = \underline{w}$, donde la utilidad no depende del empleo.

2. *Curvas isobeneficio de la empresa* (π_0, π_1): cada una une los pares de empleo y salario que dan a la empresa un mismo beneficio. Son cóncavas, porque su pendiente, $(A_\theta f'(L) - w)/L$, es positiva a la izquierda de la demanda de trabajo y negativa a su derecha, y representan más beneficio cuanto más cerca están del eje de abscisas.

3. *Curva de demanda de trabajo* ($L^d(w)$): recoge el empleo que la empresa elige para cada salario, el que cumple $A_\theta f'(L) = w$. Une los puntos máximos de las curvas isobeneficio y tiene pendiente negativa (en la figura se representa, por simplicidad, como una recta).

4. *Equilibrio*: se sitúa sobre la curva de demanda. Con $\gamma = 0$ es el punto 0, con salario $\underline{w}$ y empleo competitivo L_C. Con $\gamma = 1$ es el punto 1, con w^* y L^*, donde CI_0 es tangente a la demanda. La zona rayada entre CI_0 y π_0 (lente de mejoras paretianas) recoge los contratos que mejorarían a la vez al sindicato y a la empresa, por lo que el punto 1 no es óptimo de Pareto.

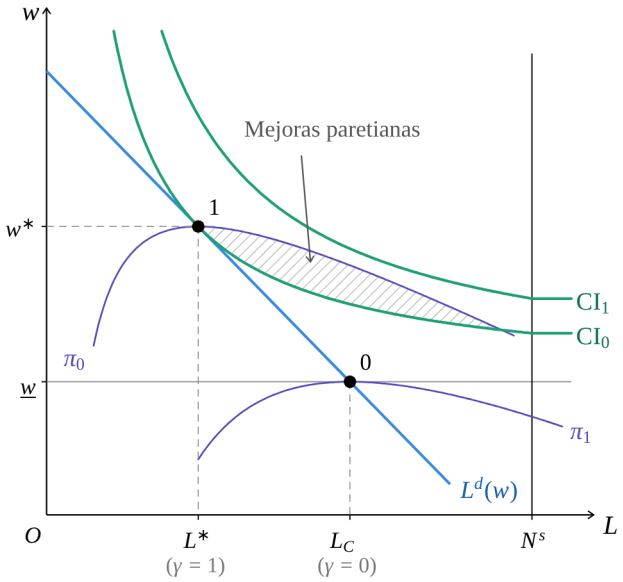


Figura A27.1: Modelo *right to manage*. El sindicato elige el salario sobre la curva de demanda de trabajo de la empresa y el resultado es ineficiente.

Estática comparativa: si la demanda de trabajo es isoelástica (por ejemplo, $f(L) = L^{\alpha}$ con $0 < \alpha < 1$ y la perturbación multiplicativa A_{θ}), un *shock* que la desplace no altera su elasticidad-salario, que es la que determina el salario negociado. Por ello, ante perturbaciones de demanda de trabajo, el salario es rígido y todo el ajuste se realizará vía cantidades, con un nivel de empleo muy volátil.

2.2.3. Conclusiones

Aspectos positivos: el modelo incorpora a la negociación de Nash los factores de los que depende la fuerza del sindicato, y contiene como casos límite el monopolio sindical de Dunlop ($\gamma = 1$) y el equilibrio competitivo ($\gamma = 0$). Su regla salarial es un margen sobre el salario de reserva que crece con el poder sindical y decrece con la elasticidad de la demanda de trabajo. La regla es, además, la base de la curva del salario real negociado (SRN) del modelo de la NAIRU (Layard, Nickell y Jackman, 1991; tema A28, apartado 2.2).

Limitaciones: no obstante, el carácter secuencial de la negociación hace que la solución sea ineficiente siempre que el sindicato tenga algún poder negociador ($\gamma > 0$). La ineficiencia nace de que el acuerdo se sitúa sobre la curva de demanda y deja sin explotar ganancias mutuas que los contratos eficientes sí recogen.

2.3. Contratos eficientes (McDonald y Solow, 1981)

> Para el ejercicio oral no hace falta el desarrollo analítico, y basta con exponer las ideas y el gráfico. El desarrollo se recoge para la comprensión conceptual del modelo.

Origen: Ian McDonald y Robert Solow (1981) suponen que la negociación colectiva abarca también el empleo (y, en su caso, otras condiciones, como el tiempo de trabajo), de modo que ambas partes acuerdan a la vez salario y empleo.

Idea: si salario y empleo se negocian simultáneamente, el acuerdo se sitúa sobre la *curva de contratos*, fuera de la curva de demanda de trabajo, y es eficiente en el sentido de Pareto, aunque el empleo puede diferir del competitivo.

2.3.1. Supuestos

Entorno:

- *Información*: perfecta.
- *Contexto*: estático.
- *Variables estratégicas*: el salario y el empleo en los contratos débiles, y el salario y una prestación para los afiliados desempleados en los fuertes.

Las dos partes negocian a la vez todas las variables del convenio, a diferencia del modelo *right to manage*, en el que la empresa elige después el empleo. En caso de que la negociación fracase, la empresa obtiene beneficio nulo y el sindicato percibe el salario de reserva, $\underline{w}$.

Agentes: se mantiene el sindicato del modelo *right to manage*, con N^s afiliados iguales, poder de negociación $\gamma \in [0,1]$ y utilidad esperada $U^e = p \cdot U(w) + (1-p) \cdot U(\underline{w})$, aunque ahora se admite cualquier perfil de riesgo. También se mantiene la empresa del mismo modelo, con tecnología $A_\theta f(L)$ y beneficio $\pi = A_\theta f(L) - wL$.

2.3.2. Modelo

Planteamiento: de la negociación simultánea se obtiene la curva de contratos, que refleja las asignaciones eficientes en el sentido de Pareto de salario y empleo acordadas por los agentes. Cada uno de sus puntos maximiza el beneficio de la empresa sin que la utilidad esperada del sindicato baje de un nivel dado:

$$\max_{w,\,L} \ \pi = A_\theta f(L) - wL \quad \text{s.a.: } U^e(w,L) \geq \bar{U}$$

donde w y L son el salario y el empleo (variables de decisión), $U^e(w,L)$ la utilidad esperada del sindicato y $\bar{U}$ el nivel de utilidad que se garantiza al sindicato, de modo que al variar $\bar{U}$ se recorre la curva de contratos.

CPO: gráficamente, la curva de contratos es el lugar geométrico de los puntos de tangencia entre las curvas de indiferencia y las curvas isobeneficio. Formalmente, la curva de indiferencia del sindicato tiene pendiente $-\frac{U(w)-U(\underline{w})}{L\,dU/dw}$ y la isobeneficio, $\frac{A_\theta df/dL - w}{L}$, de modo que la curva de contratos cumple:

$$\frac{U(w)-U(\underline{w})}{dU/dw} = w - A_\theta \frac{df}{dL}$$

CSO: las curvas de indiferencia del sindicato son convexas y las isobeneficio, cóncavas, de modo que la tangencia es un máximo.

Solución: la pendiente de la curva de contratos en el plano (L, w) depende del perfil de riesgo de los miembros del sindicato:

- *Aversos al riesgo*: la curva tiene pendiente positiva, porque el sindicato utiliza su poder tanto para cubrir a sus miembros frente al riesgo de desempleo como para subir el salario.

- *Neutrales al riesgo*: la curva es vertical sobre el nivel de empleo competitivo.

- *Propensos al riesgo*: la curva tiene pendiente negativa, porque el sindicato acepta un menor nivel de empleo a cambio de un mayor salario.

Con neutralidad al riesgo, el lado izquierdo de la CPO es $w - \underline{w}$ y la condición se reduce a $A_\theta f'(L) = \underline{w}$, por lo que el empleo es el competitivo para cualquier salario. Con aversión al riesgo, en cambio, la concavidad de U hace el lado izquierdo mayor que $w - \underline{w}$ para $w > \underline{w}$. Así, $A_\theta f'(L) < \underline{w}$ y el empleo supera al competitivo.

En los contratos eficientes fuertes, las prestaciones por desempleo se negocian junto con el salario como segunda variable de elección. Cabe verlas como un subsidio z que la empresa paga a los afiliados sin empleo, que reciben así el salario de reserva más la prestación, $w_u = \underline{w} + z$. Con seguro completo, el pago iguala la utilidad de quien trabaja y la de quien no lo hace ($w = \underline{w}+z$), y el riesgo de desempleo queda cubierto por entero, sea cual sea el perfil del sindicato. A la empresa, sin embargo, ese seguro le cuesta, y su beneficio pasa a ser $\pi = A_\theta f(L) - wL - z(N^s - L)$.

Al estar plenamente asegurados los miembros, al sindicato le es indiferente el nivel de empleo y la negociación deja el empleo al margen como variable de elección. Del lado de la empresa, la CPO es $A_\theta f'(L) = w - z$, porque cada contratación cuesta el salario, pero ahorra el subsidio que habría que pagar a ese afiliado en paro. Con seguro completo, $w - z = \underline{w}$ y la condición queda $A_\theta f'(L) = \underline{w}$. El empleo es, por tanto, el competitivo y no depende de lo que se negocie.

Equilibrio: se sitúa sobre la curva de contratos, en el punto que fija el poder negociador γ, y su eficiencia depende de lo que cubra el convenio:

- *Contratos eficientes débiles*: con $\gamma = 0$, el equilibrio alcanza eficiencia asignativa y productiva. En cambio, con $\gamma > 0$ y un sindicato averso al riesgo, se produce *sobreempleo* (eficiencia asignativa pero no productiva). Dada la dificultad de conjugar ambos tipos de eficiencia, estos contratos son *débilmente eficientes*.

- *Contratos eficientes fuertes*: el empleo es siempre el competitivo, L_C (eficiencia productiva), y el salario se sitúa entre $\underline{w}$ ($\gamma = 0$) y w_1 ($\gamma = 1$). Como el poder negociador solo reparte la renta, sin crear ineficiencias, estos contratos son *fuertemente eficientes*.

Desarrollo gráfico: los contratos eficientes débiles se representan en el plano empleo-salario, con un sindicato averso al riesgo (figura A27.2):

1. *Curvas de indiferencia del sindicato* (CI_0, CI_1) y *curvas isobeneficio* (π_0, π_1, $\pi = 0$): tienen la forma descrita en el modelo *right to manage*.

2. *Curva de demanda de trabajo* ($L^d(w)$): recoge, como en el modelo *right to manage*, el empleo que la empresa elige para cada salario, y une los máximos de las curvas isobeneficio.

3. *Curva de contratos* (CC): une las tangencias entre curvas de indiferencia e isobeneficio y tiene pendiente positiva. Si alcanza el número de afiliados N^s (pleno empleo del sindicato), continúa en vertical sobre $L = N^s$ hasta el beneficio nulo, porque a partir de ese nivel las curvas de indiferencia del sindicato son horizontales.

4. *Equilibrio*: el punto 0 ($\gamma = 0$) es el equilibrio competitivo, con $\underline{w}$ y L_C, y el punto 1 ($\gamma = 1$), con w_1 y $L_1 = N^s$, sobre la vertical de los afiliados y con beneficio nulo, el de mayor sobreempleo.

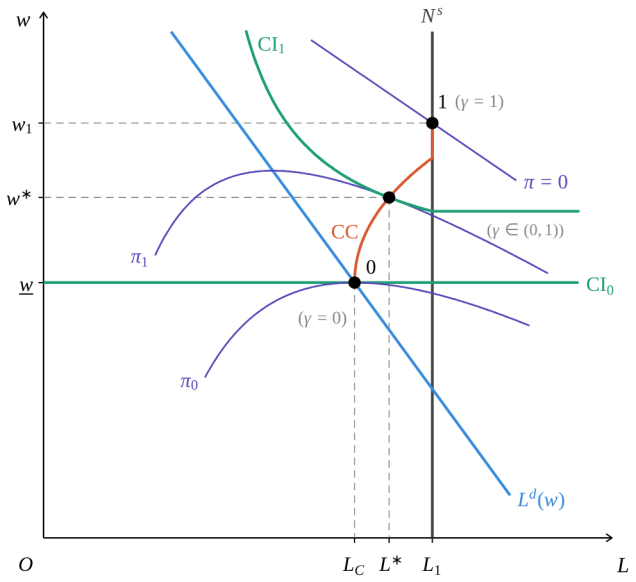


Figura A27.2: Contratos eficientes débiles. La curva de contratos se sitúa a la derecha de la demanda de trabajo, con sobreempleo respecto al modelo *right to manage*.

Los contratos eficientes fuertes, con seguro completo, se representan en el mismo plano (figura A27.3):

1. *Curvas de indiferencia del sindicato* (CI_0, CI_1, CI_2): son horizontales, al margen del perfil de riesgo considerado, puesto que los miembros cobrarán lo mismo independientemente de si trabajan o no.

2. *Curvas isobeneficio* (π_0, π_1, $\pi = 0$): como cada contratación ahorra el subsidio, alcanzan su máximo en el empleo competitivo para cualquier salario, y el lugar de sus máximos deja de ser la curva de demanda de trabajo y pasa a ser una vertical.

3. *Curva de contratos* (CC): une las tangencias entre las curvas de indiferencia, horizontales, y las isobeneficio. Esas tangencias se producen en el máximo de cada isobeneficio. Es siempre vertical sobre el empleo competitivo, porque ese máximo está en L_C para cualquier salario.

4. *Equilibrio*: los acuerdos se sitúan sobre la vertical L_C, entre el punto 0 ($\underline{w}$, $\gamma = 0$) y el punto 1 (w_1, $\gamma = 1$).

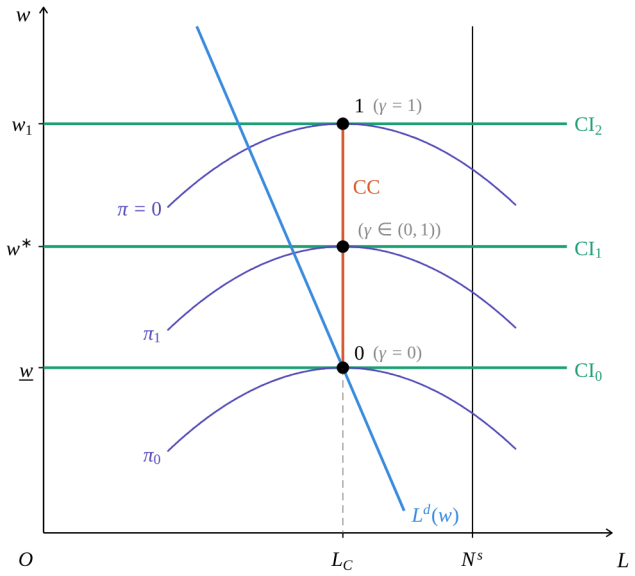


Figura A27.3: Contratos eficientes fuertes. La curva de contratos es vertical y el salario negociado no afecta al empleo.

Estática comparativa: un aumento del poder negociador del sindicato ($\Delta\gamma > 0$) desplaza el acuerdo sobre la curva de contratos, del punto 0 hacia el 1:

- *Contratos eficientes débiles*: con un sindicato averso al riesgo, suben el salario y el empleo, y el sobreempleo es tanto mayor cuanto mayor es γ.

- *Contratos eficientes fuertes*: solo sube el salario, y las rentas se trasladan de la empresa hacia los trabajadores sin que cambie el empleo.

2.3.3. Conclusiones

Aspectos positivos: los contratos eficientes muestran que el poder sindical no genera por sí mismo desempleo. Si se negocia también el empleo, el acuerdo es eficiente en el sentido de Pareto y el empleo puede incluso superar al competitivo (eficiencia débil). Si el convenio asegura además a los afiliados frente al desempleo, el poder sindical solo redistribuye renta entre la empresa y los trabajadores (eficiencia fuerte). En EE. UU., de hecho, además del seguro público de desempleo, algunos convenios negocian prestaciones complementarias (p. ej. en el sector del automóvil), mientras que en Europa las prestaciones por desempleo las fija esencialmente el Estado.

Limitaciones: los convenios rara vez fijan el nivel de empleo, que en la práctica decide la empresa (Oswald, 1993). Un contrato situado fuera de la curva de demanda es difícil de hacer cumplir, porque, una vez pactado el salario, la empresa tiene incentivos a volver a ella (Cahuc, Carcillo y Zylberberg, 2014). La hipótesis de cobertura total de los trabajadores es, además, restrictiva, pues los subsidios negociados no ofrecen coberturas perfectas. Por último, los procesos de negociación suelen dilatarse en el tiempo ante las dificultades para llegar a acuerdos.

2.4. Trabajadores heterogéneos: *insiders-outsiders* (Lindbeck y Snower, 1984-1988)

Origen: los modelos anteriores tratan a todos los afiliados al sindicato como iguales, tanto en preferencias como en influencia sobre el convenio. Sin embargo, en la práctica la posición de cada trabajador en la empresa no es la misma, y tener en cuenta esa diferencia conduce a la teoría de los trabajadores internos y externos. La teoría se inicia con las aportaciones de Assar Lindbeck y Dennis Snower (1984 y 1988).

Idea: las empresas mantienen relaciones con dos grupos de trabajadores:

- *Insiders*: ya trabajan en la empresa, tienen con ella un vínculo específico y el convenio suele recoger sus intereses.

- *Outsiders*: no tienen contrato, pero pueden obtenerlo después de firmado el convenio.

Esta situación da lugar a una asimetría en el poder de negociación en favor de los *insiders* (especialmente cuando existen sindicatos fuertes), que genera rigidez salarial en la medida en que los *insiders* miran por su propio bienestar y no por el de los *outsiders*.

Entre las causas que llevan a esta asimetría de poder se encuentran:

- *Descapitalización*: la falta de atractivo para las empresas de los desempleados de larga duración.

- *Costes de rotación*: los costes de ajuste elevados merman la capacidad de los *outsiders* de competir por puestos cubiertos.

- *Rentas situacionales*: los *insiders* pueden contar con un elevado *know-how* o capital humano específico de la empresa.

- *Cooperación y acoso*: los *insiders* pueden negarse a cooperar con los *outsiders* recién contratados o acosarlos, lo que reduce la productividad de estos y encarece su contratación (Lindbeck y Snower, 1988).

- *Factores institucionales*: la pertenencia a un sindicato o la legislación de protección del empleo de los *insiders*.

2.4.1. Supuestos

Entorno: el modelo es estático, con información perfecta y precios normalizados, de modo que todas las variables son reales. El trabajo es el único factor de producción, y la empresa produce $Af(L)$, con $f' > 0$ y $f'' < 0$, donde A recoge la productividad, que varía con el ciclo, y $L = L_I + L_O$ es el empleo total.

Agentes: la empresa se relaciona con dos tipos de trabajadores. Los *insiders* o trabajadores internos son los L_I que ya emplea y los únicos que influyen en la negociación. Su primer objetivo es conservar el empleo y el segundo, subir su salario real, y fijan además las condiciones de contratación de los *outsiders*, sin interés alguno en crear empleo para ellos. Frente a ellos, los *outsiders* o trabajadores externos son desempleados sin influencia sobre la negociación, cuya oferta es ilimitada al salario w_O, y ningún trabajador acepta un salario inferior al de reserva, $\underline{w}$.

Aunque *insiders* y *outsiders* tienen la misma productividad y son sustitutivos perfectos en la producción, la empresa no puede reemplazar a los primeros por los segundos sin incurrir en

costes de rotación, el de despido, d, y el de contratación, c. Además, los *outsiders* no pueden convertirse en *insiders*, y su salario de entrada depende del convenio de los *insiders*, de modo que los *outsiders* no pueden ofrecerse por un salario inferior para desplazar a los *insiders* (supuesto de salarios dependientes).

2.4.2. Modelo

Planteamiento: el modelo tiene dos etapas. En la primera, los *insiders* fijan su salario, w_I, como el más alto que les permite conservar todos sus puestos, es decir, el que no lleva a la empresa a despedirlos sin sustituirlos ni a reemplazarlos por *outsiders*:

$$\max_{w_I} \; w_I \quad \text{s.a.:} \begin{cases} w_I \leq Af'(L_I) + d \\ w_I \leq w_O + c + d \end{cases}$$

donde w_I es el salario de los *insiders* (variable de decisión), L_I su número, $Af'(L_I)$ su productividad marginal (Pmg_L en la figura), w_O el salario de los *outsiders*, c el coste de contratación y d el de despido. La primera restricción garantiza que a la empresa no le compense despedir al *insider* marginal, porque el ahorro salarial no supera lo que este produce más el coste de despido. La segunda garantiza que tampoco le compense sustituirlo por un *outsider*, que le costaría su salario más los costes de contratación y de despido.

En la segunda etapa, la empresa, dado w_I, decide cuántos *outsiders* contrata:

$$\max_{L_O \geq 0} \; \pi = Af(L_I + L_O) - w_I L_I - (w_O + c)L_O$$

donde L_O es el número de *outsiders* contratados (variable de decisión) y π el beneficio de la empresa.

CPO: como el objetivo de los *insiders* es creciente en w_I, en el óptimo se satura la restricción más exigente, lo que da la condición de fijación salarial:

$$w_I = \min\{A\frac{df}{dL_I} + d, \; w_O + c + d\} \tag{3}$$

Es decir, los *insiders* se apropian de toda la renta que les conceden los costes de rotación. Por su parte, la CPO de la empresa, con $L_O \geq 0$ como restricción, es la regla de contratación:

$$A\frac{df}{d(L_I + L_O)} \leq w_O + c, \quad \text{con igualdad si } L_O > 0 \tag{4}$$

De este modo, la empresa contrata *outsiders* mientras su productividad marginal cubra su salario más el coste de contratación.

CSO: el programa de los *insiders* es lineal, por lo que su solución es la restricción que se satura en (3). El de la empresa es cóncavo en L_O, porque $f'' < 0$, de modo que (4) es suficiente.

Solución: (3) implica, por sí sola, que w_I crece con la productividad A (es procíclico) hasta topar con el techo $w_O + c + d$, por encima del cual a la empresa le compensaría despedir a los *insiders* (coste d) y contratar *outsiders* (coste c). Como los *insiders* están perfectamente asegurados frente al riesgo de despido, fijan el salario más alto que permite (3), tanto mayor cuanto mayores son los costes de rotación, que son la fuente de su poder negociador.

Equilibrio: con el salario que fija (3), la regla (4) determina si la empresa contrata además *outsiders*. El resultado depende del número de *insiders* respecto a L_1, donde $\text{Pmg}_L + d$ alcanza el techo $w_O + c + d$, y a L_2, donde iguala el salario de reserva:

- *Caso A*: por debajo de L_1, el poder de los *insiders* alcanza para exigir el techo salarial, w_O+d+c (salario del *outsider* más coste de despido más coste de contratación). Los *insiders* son tan pocos, y su productividad marginal tan elevada, que la empresa completa la plantilla con *outsiders*.

- *Caso B*: entre L_1 y L_2, los *insiders* ya no alcanzan el techo, pero conservan poder para negociar un salario superior al de reserva, igual a su productividad marginal más el coste de despido ($\text{Pmg}_L + d$). Con ese número de *insiders*, la empresa ni necesita despedir a ninguno ni tiene que incorporar *outsiders*.

- *Caso C*: por encima de L_2, los *insiders* son tantos que la productividad marginal del último trabajador más lo que costaría despedirlo no alcanza el salario de reserva ($\text{Pmg}_L + d < \underline{w}$), y ningún salario aceptable cumple la primera restricción. Conviene entonces a la empresa reducir la plantilla de *insiders* hasta L_2.

Desarrollo gráfico: el modelo de *insiders* y *outsiders* se representa en el plano empleo de los *insiders*-salario (figura A27.4):

1. *Techo salarial* ($w_O + d + c$): es una horizontal, el salario por encima del cual a la empresa le compensa sustituir *insiders* por *outsiders*. No depende del número de *insiders*, porque solo recoge lo que cuesta un *outsider* más los costes de rotación.

2. *Productividad marginal más coste de despido* ($\text{Pmg}_L + d$): recoge el salario por encima del cual a la empresa le compensa despedir al *insider* marginal sin sustituirlo. Es decreciente en el empleo, porque la productividad marginal lo es ($f'' < 0$). Corta al techo salarial en L_1 e iguala el salario de reserva en L_2.

3. *Salario de los insiders* ($w_I(L_I)$): la línea gruesa marca, en cada nivel de empleo, el techo de lo que los *insiders* pueden pedir antes de que a la empresa le convenga despedirlos o reemplazarlos por *outsiders*.

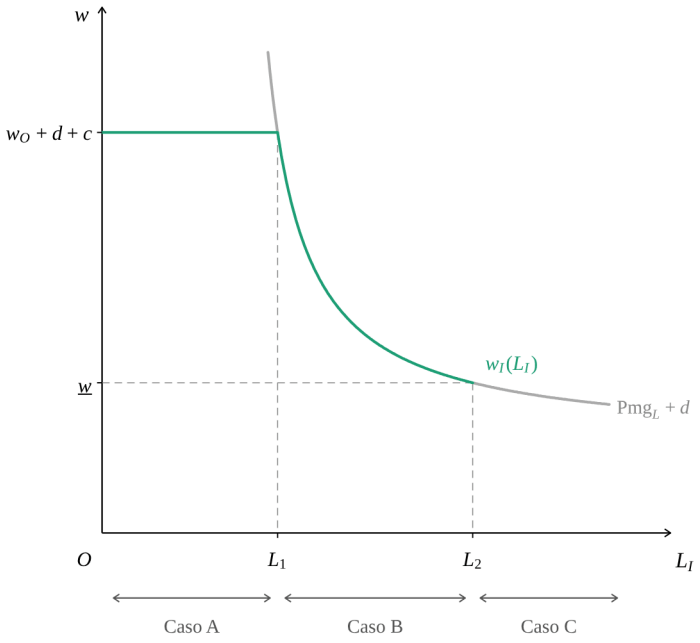


Figura A27.4: Modelo *insiders-outsiders*. Los trabajadores internos fijan el salario que garantiza su propio empleo y dejan fuera a los externos.

Estática comparativa: el carácter anticíclico del salario real que se atribuye al modelo opera a través del número de *insiders*, L_I, pues con $f'' < 0$ una caída de L_I eleva $Af'(L_I)$ en (3):

- *Fases recesivas*: la empresa despide a parte de los *insiders*, y los que se quedan, al ser menos, ganan poder negociador y pueden empujar el salario hacia arriba.

- *Fases de auge*: lo esperable es que la empresa incorpore *outsiders* y que estos pasen a ser *insiders* en el periodo siguiente, algo que el modelo estático no recoge. Con más *insiders*, $f'(L_I)$ es menor y el grupo puede aceptar un salario más bajo a cambio de asegurar sus puestos.

Una interpretación menos extrema de la condición (3) sería que la presencia de *insiders/outsiders* atenúa la respuesta de los costes laborales marginales de las empresas a las fluctuaciones cíclicas[1].

Extensiones:

- *Predicciones*: la marginación de los *outsiders*, salarios anticíclicos bajo ciertas condiciones (por el mecanismo de la estática comparativa) e histéresis, de modo que perturbaciones transitorias pueden tener efectos permanentes sobre el desempleo.

- *Políticas*: las recomendaciones son reducir los costes de rotación, introducir cláusulas de descuelgue y formar a los parados.

[1] Romer (2012, cap. 10).

2.4.3. Conclusiones

Aspectos positivos: el modelo explica tres hechos:

- *Desempleo involuntario*: el salario que imponen estratégicamente los *insiders* es tan alto que a la empresa no le compensa contratar *outsiders*, y quedan sin empleo personas dispuestas a trabajar por el salario de reserva.

- *Persistencia del desempleo (histéresis)*: cuanto más desempleo hay, menos *insiders* quedan y mayor es su poder negociador. Los salarios que imponen suben entonces, los despedidos no se reabsorben y el desempleo se perpetúa.

- *Dualidad del mercado de trabajo*: las condiciones laborales se polarizan, muy favorables para los *insiders* y más precarias para los *outsiders* (temporalidad, salarios más bajos, etc.).

Limitaciones: los supuestos son fuertes:

- *Despido*: si despedir a trabajadores internos no tuviese coste, el convenio buscaría que el producto marginal del trabajo coincidiera con el coste de oportunidad del tiempo de los trabajadores internos.

- *Oferta externa*: si no existiese una oferta ilimitada de trabajadores externos, el salario tendería a aumentar.

- *Salarios dependientes*: es el supuesto más fuerte. Si se levanta, los trabajadores externos se podrían contratar a un salario equivalente al salario vigente en el mercado.

2.5. Conclusiones

En los modelos de negociación, el salario reparte entre empresa y trabajadores una renta que solo surge si cooperan. El reparto depende del poder de cada parte, que la solución de Nash generalizada recoge en el parámetro γ y el modelo de Rubinstein deriva de su paciencia relativa. Influye también la estructura de la negociación, cuyos grados extremos de centralización moderan más los salarios que los intermedios.

El empleo depende, en cambio, de lo que se negocia. Si se pacta solo el salario y la empresa elige después el empleo (*right to manage*), el acuerdo queda sobre la curva de demanda y cualquier poder sindical genera desempleo involuntario. Si se pactan salario y empleo, el acuerdo es eficiente y el poder sindical eleva el empleo o solo redistribuye la renta. Con trabajadores heterogéneos, los costes de rotación dan a los *insiders* un poder que excluye a los *outsiders* y hace persistente el desempleo. En todos estos modelos, el salario elevado procede del poder de los trabajadores, y la sección 3 estudia por qué puede convenir también a la empresa.

3. Salarios de eficiencia

A diferencia de los modelos de negociación, en los de salarios de eficiencia la propia empresa paga un salario elevado porque, además de costes, le reporta beneficios. Tras el concepto y las causas, se exponen dos modelos que difieren en lo que la empresa sabe del esfuerzo, con información simétrica en la condición de Solow y asimétrica en la *no-shirking condition* de Shapiro y Stiglitz.

3.1. Concepto: causas y origen

Los modelos de **salarios de eficiencia** parten de una hipótesis central, la de que pagar más encarece la mano de obra pero también reporta ganancias a la empresa. Existen distintas líneas de argumentación para apoyar esta tesis (Yellen, 1984):

- *Teorías nutricionistas* (Leibenstein, 1957): en países menos desarrollados, un mayor salario puede contribuir a la mejora de la salud de los trabajadores y de su productividad.

- *Teorías sociológicas* (Akerlof, 1982): un salario superior al de mercado eleva la moral y la lealtad a la empresa, y los trabajadores lo corresponden, como en un intercambio de regalos, con un esfuerzo superior a la norma.

- *Teorías de la rotación* (Stiglitz, 1974; Salop, 1979): un salario elevado fomenta la permanencia en la empresa y reduce los costes de rotación.

- *Teorías de la selección adversa* (Weiss, 1980): un salario elevado atrae a trabajadores más capacitados, cuya calidad la empresa no observa.

- *Teorías del riesgo moral* (Shapiro y Stiglitz, 1984): si el esfuerzo no es observable, un salario superior al de vaciado hace costoso el despido y disuade de holgazanear.

Estas teorías son incompatibles con el mercado de trabajo walrasiano, en el que la empresa toma el salario como dado y este vacía el mercado.

3.2. Información simétrica: condición de Solow (1979)

Origen: Robert Solow (1979) formalizó la hipótesis del salario de eficiencia en forma reducida, en la que el esfuerzo es una función del salario que la empresa conoce, sin modelizar por qué responde a él.

Idea: si el esfuerzo depende del salario, la empresa no toma el salario como dado, sino que elige el que minimiza el coste por unidad de trabajo efectivo.

3.2.1. Supuestos

Entorno: el modelo es de corto plazo, con un elevado número de empresas competitivas, información perfecta y precios normalizados. La función de producción tiene en cuenta el trabajo efectivo $e(w) \cdot L$ y no solo el trabajo L.

Agentes: trabajadores y empresas son racionales e idénticos. El esfuerzo no puede fijarse por contrato, pero la empresa conoce cómo responde el esfuerzo al salario, de modo que la información es simétrica. El esfuerzo depende positivamente del salario según una función $e(w)$ que crece más

que proporcionalmente para salarios bajos y crece menos que proporcionalmente para salarios elevados.

3.2.2. Modelo

Planteamiento:

$$\max_{w,L} \ \pi = f[e(w) \cdot L] - wL$$

donde w es el salario y L el empleo (variables de decisión), π el beneficio de la empresa, $e(w)$ la función de esfuerzo, creciente en el salario, $e(w) \cdot L$ el trabajo efectivo y $f(\cdot)$ la función de producción.

CPO:

$$\frac{\partial \pi}{\partial L} = 0 \ \Rightarrow \ \underbrace{\frac{df}{d(eL)}}_{\text{Pmg}_{eL}} = \underbrace{\frac{w}{e}}_{\text{coste por unidad de eficiencia}} \tag{5}$$

$$\frac{\partial \pi}{\partial w} = 0 \ \Rightarrow \ \varepsilon_{ew} = \frac{w}{e} \cdot \frac{\partial e}{\partial w} = 1 \tag{6}$$

La condición (5) indica que la empresa contrata hasta que la productividad marginal del trabajo efectivo iguala su coste, w/e. (6) resulta de combinar la derivada respecto del salario, $df/d(eL)\, de/dw = 1$, con (5), y se conoce como la *condición de Solow*. Establece que, para maximizar beneficios y minimizar el coste por unidad de trabajo efectivo, la elasticidad-salario del esfuerzo es unitaria, independientemente de la situación económica.

CSO: el beneficio es cóncavo en L ($f'' < 0$) y la forma de $e(w)$, convexa para salarios bajos y cóncava para salarios altos, garantiza que el punto de elasticidad unitaria es un mínimo del coste unitario $w/e(w)$, de modo que (5) y (6) son suficientes.

Desarrollo gráfico: la condición de Solow se representa en el plano salario real-esfuerzo (panel superior) y en el plano salario real-coste por unidad de trabajo efectivo (panel inferior) de la figura A27.5:

1. *Función de esfuerzo* ($e(w)$): es primero convexa (el esfuerzo crece a ritmo creciente con el salario) y, superado cierto nivel salarial, pasa a ser cóncava (crece a ritmo decreciente), tramo en el que se alcanza la elasticidad unitaria.

2. *Rayo vector*: parte del origen, y su pendiente es el esfuerzo por unidad de salario, e/w. Por tanto, cuanto más inclinado es el rayo que toca la función de esfuerzo, menos cuesta cada unidad de trabajo efectivo.

3. *Coste por unidad de trabajo efectivo* ($w/e(w)$, panel inferior): es la inversa de la pendiente del rayo vector que pasa por cada punto de la función de esfuerzo. Tiene forma de U, porque decrece mientras la elasticidad del esfuerzo supera la unidad y crece cuando es menor.

4. *Equilibrio*: se sitúa donde la función de esfuerzo es tangente al rayo vector trazado desde el origen con la máxima pendiente ($\varepsilon_{ew} = 1$), con salario w^* y esfuerzo e^*, punto en el que el coste unitario es mínimo, $(w/e)^*$.

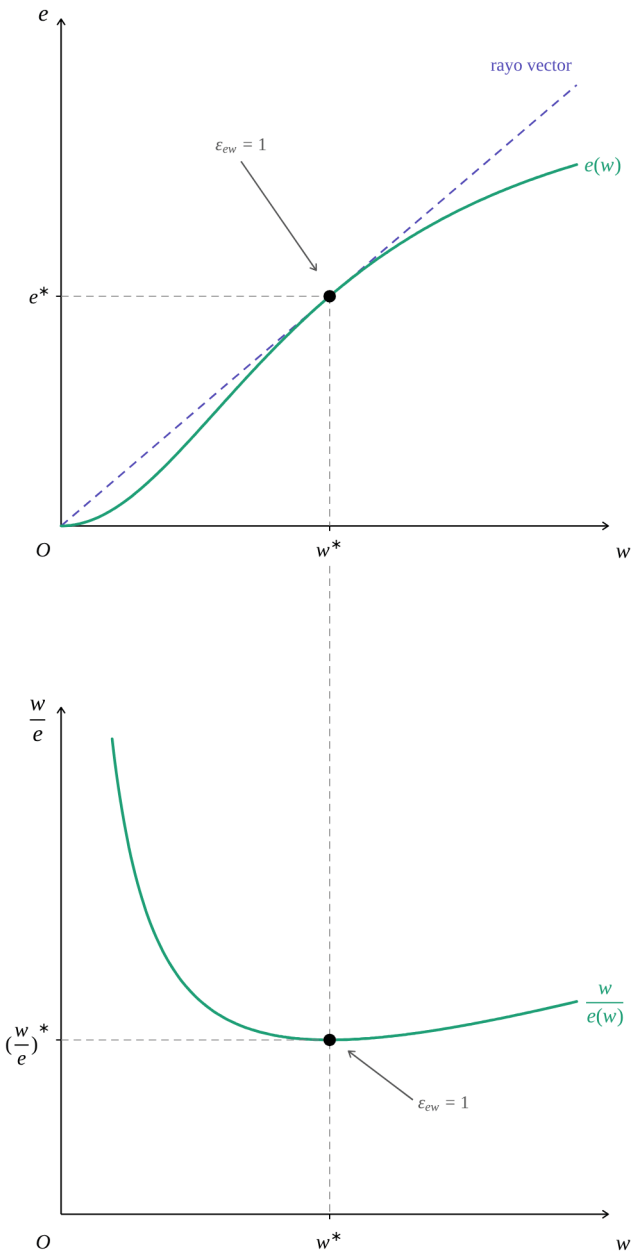


Figura A27.5: Condición de Solow (1979). El salario de eficiencia iguala a la unidad la elasticidad del esfuerzo respecto al salario.

Extensiones: el modelo tiene dos implicaciones:

▪ *Desempleo involuntario*: el salario de eficiencia puede quedar por encima del de vaciado y provocar, en consecuencia, desempleo involuntario. Fijado ese salario, la empresa contrata trabajadores hasta igualar la productividad marginal del trabajo con el salario de eficiencia, al margen de la oferta.

▪ *Rigidez salarial*: frente a una perturbación (*shock*) negativa de demanda, el empresario no querrá rebajar el salario, ya que hacerlo le impediría maximizar beneficios. El salario de eficiencia

lo fija la respuesta del esfuerzo al salario, con independencia de cómo estén la oferta y la demanda. Intuitivamente, recortar el salario haría disminuir el esfuerzo de los trabajadores, y el coste de ello podría superar el ahorro obtenido con la rebaja salarial, por lo que el ajuste se realiza vía cantidades.

Gráficamente, en el plano empleo-salario (figura A27.6), la empresa fija el salario de eficiencia, w^*, por encima del de vaciado, w_C, y contrata sobre la demanda $L_0^d(w)$ hasta el punto 0, con empleo L_0. A ese salario se ofrecen $L^s(w^*)$ trabajadores, y la diferencia es desempleo involuntario. Ante un *shock* negativo ($\Delta A < 0$), la demanda se desplaza a $L_1^d(w)$, el salario no cambia y el empleo cae de L_0 a L_1 (del punto 0 al 1).

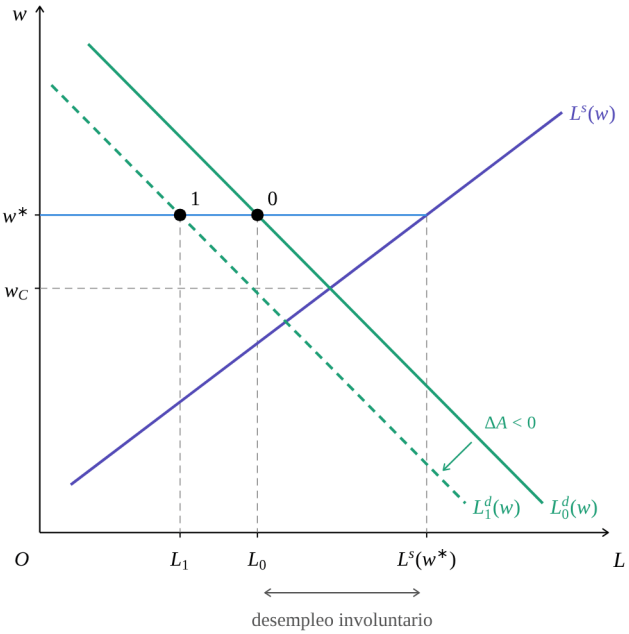


Figura A27.6: Salario de eficiencia y desempleo involuntario. La empresa no reduce el salario aunque exista exceso de oferta de trabajo.

3.2.3. Conclusiones

Aspectos positivos: el modelo permite explicar por qué, a corto plazo, las variaciones de la demanda de trabajo apenas parecen afectar a los salarios reales y se reflejan sobre todo en cambios del nivel de empleo.

Limitaciones: se han formulado cuatro objeciones:

- *Miopía empresarial*: aunque el modelo trata a las empresas como racionales y conscientes de la tentación de rendir poco, no les deja recurrir a otros incentivos para lograr más esfuerzo.

- *Permanencia*: el salario de eficiencia viene fijado por la función de esfuerzo, sin relación con la productividad, así que el salario real quedaría inmóvil indefinidamente, lo que choca con su crecimiento a largo plazo. Para evitarlo, el esfuerzo debe depender también del salario que pagan otras empresas y del paro (Romer, 2012, cap. 10).

- *Oferta laboral*: no se tiene en cuenta.

- *Forma reducida*: la principal limitación es que la relación $e(w)$ se postula sin explicar por qué el esfuerzo responde al salario. El modelo de Shapiro y Stiglitz, en cambio, fundamenta el vínculo entre salario y esfuerzo en la no observabilidad del esfuerzo.

3.3. Información asimétrica: *no-shirking condition* (Shapiro y Stiglitz, 1984)

Origen: Carl Shapiro y Joseph Stiglitz, en *Equilibrium Unemployment as a Worker Discipline Device* (1984), consideran el caso en que la empresa no puede vigilar por completo a sus empleados.

Idea: si el esfuerzo no es observable, el desempleo actúa como mecanismo disciplinario, porque la empresa paga un salario superior al de vaciado para que perder el empleo sea costoso y el trabajador no holgazanee.

3.3.1. Supuestos

Entorno: se mantienen los supuestos del modelo de Solow (trabajo efectivo, precio del producto como numerario y un elevado número de empresas y trabajadores), salvo el de información perfecta.

Agentes: la información es asimétrica, y entre la empresa y los trabajadores se da un problema de riesgo moral por dos motivos:

- *Esfuerzo no observable*: la empresa no ve el esfuerzo, que los trabajadores eligen como discreto, entre $e = 0$ (holgazanear) y $e = e^*$ (esforzarse).

- *Conflicto de intereses*: con más esfuerzo el empresario ingresa más, pero también crece la desutilidad que soporta el trabajador.

3.3.2. Modelo

Planteamiento: la decisión del trabajador determina la oferta de trabajo. En cada momento del tiempo, el trabajador puede encontrarse en tres posibles estados que definen su función de utilidad instantánea:

- *Empleado y esforzándose*: $u = w - e^*$.

- *Empleado y holgazaneando*: $u = w$.

- *Desempleado*: $u = \bar{w}$.

Existen tres posibles transiciones entre estos estados (procesos de Poisson cuyas tasas, probabilidades por unidad de tiempo, no dependen del tiempo en el puesto ni de la causa del desempleo):

- *Salida exógena del empleo* (λ_R): de empleado (se esfuerce o no) a desempleado, por causas ajenas a su conducta.

- *Salida involuntaria del empleo* (λ_D): de empleado holgazán a desempleado, con probabilidad λ_D de ser sorprendido no esforzándose.

- *Salida del desempleo* (a): de desempleado a empleado, con probabilidad a de encontrar un empleo estando parado.

La utilidad vital del individuo se recoge en una función de utilidad intertemporal en tiempo continuo y horizonte infinito, con factor de descuento e^{-rt} y tasa de descuento r.

Solución: las empresas son todas idénticas entre sí y maximizan sus beneficios en función del esfuerzo o no esfuerzo de los trabajadores. Cada empresa elige el nivel de empleo y, frente al riesgo moral, fija un salario lo bastante alto para que a sus trabajadores no les compense holgazanear.

La empresa debe pagar un salario tal que el valor actual de la utilidad esperada de trabajar y esforzarse, $U_T^e(w^*)$, sea al menos igual al de trabajar y holgazanear, U_H^e. En el límite, $U_T^e(w^*) = U_H^e$. Los valores de cada estado, incluido el del desempleado, U_D^e, cumplen ecuaciones de valor de un activo, en las que el rendimiento del estado (r por su valor) es igual a su flujo de utilidad más la ganancia o pérdida esperada de cambiar de estado:

$$\begin{aligned} rU_T^e &= w - e^* + \lambda_R(U_D^e - U_T^e) \\ rU_H^e &= w + (\lambda_R + \lambda_D)(U_D^e - U_H^e) \\ rU_D^e &= \bar{w} + a(U_T^e - U_D^e) \end{aligned}$$

Igualando $U_T^e = U_H^e$ en las dos primeras se obtiene $U_T^e - U_D^e = e^*/\lambda_D$, lo que significa que la renta que el empleo reporta al trabajador que se esfuerza debe igualar la desutilidad del esfuerzo dividida por la probabilidad de ser descubierto. Restando la tercera ecuación de la primera y sustituyendo esa diferencia, el salario óptimo es:

$$w^* = \bar{w} + e^* + \frac{e^*}{\lambda_D}(a + \lambda_R + r)$$

$$w^* = w^*(\bar{w}^+, e^{*+}, a^+, \lambda_R^+, r^+, \lambda_D^-)$$

Concretamente, el salario de eficiencia es:

- *Creciente* con la renta del desempleado ($\bar{w}$), el esfuerzo (e^*), la facilidad para encontrar trabajo (a), la tasa de rotación (λ_R, ya que, cuanto más probable es perder el empleo por causas ajenas a la conducta del trabajador, menos arriesga este si holgazanea) y la tasa de descuento (r, ya que, cuanto más impacientes son los trabajadores, mayor salario deberá ofrecer la empresa).

- *Decreciente* con la probabilidad de sorprender a los holgazanes (λ_D, ya que, cuanto menor es la capacidad de la empresa para detectar la holgazanería, mayor debe ser el salario que disuada de ella).

Además, en el estado estacionario los flujos de entrada y salida del desempleo coinciden ($a(N - L) = \lambda_R L$), de modo que se obtiene la *condición de no escaqueo* (*no-shirking condition*, NSC):

$$w = \bar{w} + e^* + \frac{e^*}{\lambda_D}\left(\frac{L}{N-L}\lambda_R + \lambda_R + r\right)$$

El salario necesario para que los trabajadores no holgazaneen constituye una función creciente del nivel de empleo L. La razón es que, con un empleo elevado, el paro es menor y hay más abandonos de puestos de trabajo, así que a quien está desempleado le resulta más fácil encontrar ocupación. La NSC es asintótica a la oferta de trabajo, porque esfuerzo alto y pleno empleo son incompatibles. Con pleno empleo ($L = N$), los desempleados encuentran trabajo inmediatamente, el despido no tiene coste y ningún salario evita la holgazanería.

Las empresas demandarán trabajo hasta el punto en el que su productividad marginal coincida con el salario, lo que da lugar a una curva de demanda de trabajo convencional con pendiente negativa.

Equilibrio: depende de si la empresa puede supervisar el esfuerzo:

- *Con supervisión perfecta*: el equilibrio se produce en el cruce de la curva de demanda de trabajo y la curva de oferta de trabajo, y es un equilibrio walrasiano de pleno empleo.

- *Con información asimétrica*: el equilibrio se sitúa donde la curva de esfuerzo (NSC) corta a la demanda de trabajo, con desempleo involuntario y un salario superior al de vaciado. El salario de eficiencia resuelve así la asimetría informativa, pero solo como *second best*, porque el desempleo que provoca no produce nada.

Los parados trabajarían con gusto, y esforzándose, por un salario menor, pero no tienen forma de hacer bajar los salarios, porque ninguna empresa los contrataría algo más baratos sabiendo que a ese salario no se esforzarían. Con el salario de vaciado en todas las empresas desaparecería el paro, pero también el incentivo al esfuerzo, ya que el trabajador despedido encontraría enseguida otro puesto igual de pagado (Romer, 2012, cap. 10). El desempleo actúa, así, como mecanismo de disciplina.

Desarrollo gráfico: el modelo de Shapiro y Stiglitz se representa en el plano empleo-salario (figura A27.7):

1. *Curva de demanda de trabajo* ($L^d(w)$): recoge el empleo para el que la productividad marginal iguala al salario y tiene pendiente negativa, porque esa productividad decrece con el empleo.

2. *Curva de esfuerzo o de no escaqueo* (NSC): da el salario de eficiencia mínimo que, según el nivel de empleo, induce a los trabajadores a esforzarse. Crece con el empleo y se aproxima asintóticamente a la oferta de trabajo.

3. *Curva de oferta de trabajo* ($L^s(w)$): perfectamente elástica para un determinado nivel de esfuerzo (al salario $\bar{w}+e^*$, que compensa la renta del desempleado más la desutilidad del esfuerzo) hasta el nivel de pleno empleo, N, en el que se vuelve perfectamente rígida.

4. *Equilibrio*: con información asimétrica, en el punto 0, donde la NSC corta a la demanda, con salario w^* y empleo L^*. La diferencia entre L^* y N es desempleo involuntario. En cambio, con supervisión perfecta, la demanda corta a la oferta en el salario de vaciado, w_C, con pleno empleo.

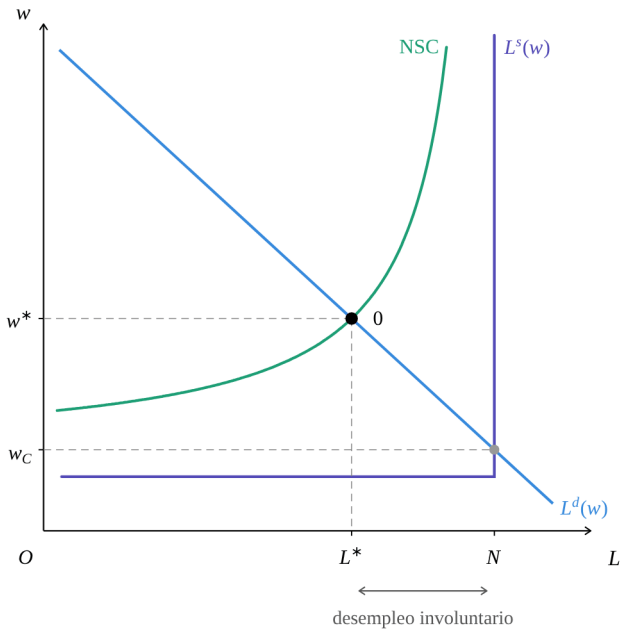


Figura A27.7: Modelo de Shapiro y Stiglitz (1984). La condición de no escaqueo (NSC) corta a la demanda de trabajo por encima de la oferta y genera desempleo de equilibrio como mecanismo disciplinario.

Estática comparativa:

▪ *Shock negativo de demanda*: cae el salario de eficiencia, pero permanece por encima del de vaciado. El ajuste es eminentemente vía cantidades, y su magnitud será tanto mayor cuanto más próximo esté el salario de eficiencia al salario de oferta, $\bar{w} + e^*$, tramo en el que la NSC es más plana. La idea es que la sensibilidad del empleo al salario de eficiencia es mayor cuanto mayor es el desempleo de la economía, porque cuando el desempleo es bajo al trabajador despedido le resulta más fácil encontrar un empleo.

▪ *Cambios en la curva de esfuerzo*: si aumentan la renta del desempleado ($\uparrow \bar{w}$) o la tasa de rotación ($\uparrow \lambda_R$), o sube la tasa de descuento ($\uparrow r$), o disminuye la probabilidad de sorprender a los holgazanes ($\downarrow \lambda_D$), la curva se desplaza hacia arriba. En el nuevo equilibrio hay más desempleo y un salario mayor.

▪ *Caída de la población activa*: con menos activos (N menor), la oferta se traslada hacia la izquierda y la NSC, que ha de seguir siendo asintótica a ella, gana pendiente. El salario sube y el empleo baja. Como $\frac{L}{N-L}\lambda_R + \lambda_R = \lambda_R \frac{N}{N-L}$, la NSC depende solo de la tasa de paro $(N-L)/N$, de modo que un salario de equilibrio más alto exige una tasa de paro menor, por lo que el paro disminuye (tanto la tasa de paro como el número de parados, $N - L$).

▪ *Aumento del esfuerzo requerido*: un e^* mayor eleva tanto la NSC como la oferta con supervisión perfecta y, con la demanda de trabajo sin cambios, el desempleo de equilibrio crece.

3.3.3. Conclusiones

Aspectos positivos: el modelo explica el desempleo involuntario como un equilibrio en el que empresas que maximizan beneficios no bajan el salario aunque haya parados dispuestos a trabajar por menos, porque el paro es el mecanismo que disciplina el esfuerzo. Ofrece además predicciones contrastables, como que el desempleo de equilibrio aumenta con las prestaciones por desempleo y disminuye con la capacidad de supervisión de las empresas.

Limitaciones: supone que la empresa no puede exigir fianzas ni diferir la remuneración con salarios crecientes con la antigüedad, mecanismos que disuadirían la holgazanería sin necesidad de desempleo (Lazear, 1979; Carmichael, 1985). Además, el paro de equilibrio depende de parámetros difíciles de medir, como la probabilidad de detección.

Conclusiones

El enfoque neoclásico postula para el mercado de trabajo un equilibrio único, estable y eficiente que encaja mal con los datos. El tema expone tres fricciones que generan rigideces del salario real y ajuste vía cantidades. La incertidumbre da lugar a los contratos implícitos, en los que la empresa, neutral al riesgo, asegura una renta salarial estable al trabajador averso al riesgo, sin desempleo involuntario. El poder de mercado da lugar a la negociación colectiva, en la que el sindicato o los *insiders* elevan el salario por encima del competitivo, con resultados ineficientes (*right to manage*), eficientes (contratos eficientes) o con histéresis (*insiders* y *outsiders*). Por último, la imposibilidad de fijar el esfuerzo por contrato da lugar a los salarios de eficiencia, que la empresa fija por encima del de vaciado para estimular el esfuerzo, y, si el esfuerzo no es observable, el paro disciplina a los trabajadores.

Este desempleo es clásico, no keynesiano, porque procede de salarios elevados por imperfecciones del mercado. En consecuencia, las políticas más eficaces son las de oferta (abaratar la rotación, revisar la estructura de la negociación, elevar la productividad), salvo cuando hay histéresis, en cuyo caso la demanda también importa. Estos modelos parten del equilibrio competitivo y del monopolio sindical (tema A25) y dan microfundamento a dos piezas de la macroeconomía, las rigideces reales de la nueva economía keynesiana (tema A07) y la curva del salario real negociado que determina la NAIRU (tema A28). Por su parte, las fricciones de búsqueda ofrecen una explicación complementaria del mismo desempleo de equilibrio (tema A26). Los *insiders* y *outsiders* encajan con la *euroesclerosis* de Europa continental (Giersch, 1985), y los salarios de eficiencia, con mercados de baja afiliación sindical como el estadounidense.

En el diseño de la negociación colectiva, la OCDE (2019) concluye que los sistemas coordinados se asocian a más empleo y menor desigualdad salarial que los plenamente descentralizados, lo que matiza la joroba de Calmfors y Driffill. Por su parte, la rigidez a la baja de los salarios nominales y reales está documentada en numerosos países por el International Wage Flexibility Project (Dickens et al., 2007). Finalmente, el poder de mercado de los empleadores (Manning, 2003) ha devuelto al primer plano el monopsonio, con el que una subida moderada del salario mínimo puede no reducir el empleo, como sostienen Card y Krueger (1994, 2000) frente a Neumark y Wascher (2000).

Anexo. Herramientas y propuestas para combatir el desempleo

Las políticas de empleo se articulan en torno a los ejes siguientes:

- *Regulación*: legislación en materia de empleo.

- *Negociación colectiva*: diseño de la negociación y de las instituciones del mercado laboral.

- *Gasto directo*: en políticas pasivas y en políticas activas de empleo. Las pasivas son recursos con función estabilizadora y redistribuidora para absorber *shocks*, como las prestaciones y los subsidios por desempleo. Las activas son recursos destinados a mejorar la empleabilidad de los parados y la creación de puestos de trabajo, como la formación, las subvenciones, las bonificaciones en la contratación o el asesoramiento.

Entre las principales propuestas que se han planteado en el debate económico destacan:

1. *Contrato único*: en países con estructura dual del mercado de trabajo (coexistencia de contratos indefinidos y temporales), una posible solución sería la unificación de las modalidades contractuales. Este tipo de propuestas busca que la carga del ajuste en etapas recesivas no recaiga enteramente en los trabajadores menos protegidos (temporales).

2. *Medidas de capacitación y formación laboral* (p. ej. en ciencia, tecnología, ingeniería y matemáticas): pueden favorecer la movilidad laboral y la productividad de la mano de obra, evitar la descapitalización y mejorar la empleabilidad.

3. *Fondos de capitalización* (p. ej. "mochila austriaca", vigente en Austria desde 2003 y debatida en España, donde no se ha implantado): la empresa acumula parte de la potencial indemnización por despido de forma periódica en una cuenta a nombre del trabajador. La ventaja para la empresa es el fraccionamiento del pago y, por tanto, un abaratamiento del despido en el momento de su ejercicio sin pérdida de cuantía para el trabajador.

4. *Cláusulas de descuelgue*: en presencia de rigideces en el diseño institucional, se pueden favorecer excepciones a los convenios colectivos pactados entre sindicatos y asociaciones empresariales, para dar mayor flexibilidad a la fijación salarial.

5. *Flexiseguridad*: enfoque basado en tres pilares:

- *Flexibilización del despido*: mediante la reducción de la protección al empleo.

- *Protección por desempleo*: mediante el aumento del subsidio al desempleo para reducir los costes económicos y sociales.

- *Políticas activas de empleo*: para garantizar la empleabilidad de los desempleados y evitar el problema de riesgo moral asociado al cobro de las prestaciones.

6. *Salario mínimo interprofesional* (SMI): desde el punto de vista teórico, en un mercado de trabajo competitivo, un SMI superior al salario de equilibrio reduce el empleo, en particular entre los trabajadores menos productivos, aunque con monopsonio un SMI moderado puede elevarlo. Sin embargo, en 138 subidas estatales del salario mínimo en EE. UU. entre 1979 y 2016, el número de empleos de baja remuneración apenas varió en los cinco años siguientes y los salarios de los afectados subieron (Cengiz, Dube, Lindner y Zipperer, 2019). En España, en cambio, tras la fuerte subida del SMI de 2019, el empleo de los trabajadores de menor salario creció menos. El efecto fue mayor entre los de más edad, y las horas y las contrataciones de jóvenes cayeron de forma más acusada (Barceló et al., 2021).

A28. TNP, NAIRU y persistencia

La tasa natural de paro y la NAIRU. La persistencia del desempleo.

Introducción

> **Introducción común.** Los temas A25 a A28 comparten una introducción, que se desarrolla al comienzo del tema **A25**. Presenta la teoría económica del mercado de trabajo, que estudia el empleo y los salarios de equilibrio y los factores institucionales que los moldean, y la explicación teórica de cada tipo de desempleo.

La tasa de desempleo de equilibrio es la referencia con la que se juzgan las políticas macroeconómicas. Friedman (1968) y Phelps (1967) la definieron como la tasa a la que tiende la economía cuando las expectativas se cumplen, la **tasa natural de paro**, y mostraron que la política monetaria solo puede rebajar el desempleo por debajo de ella a costa de una inflación creciente. La versión neokeynesiana, la **NAIRU**, la deriva, en cambio, de la negociación de salarios y la fijación de precios en mercados imperfectos (Layard y Nickell, 1985) y la hace depender de instituciones como las prestaciones o la negociación colectiva.

La **histéresis** (Blanchard y Summers, 1986) añade que el desempleo de equilibrio depende de la historia, porque los parados de larga duración pierden capital humano y vínculos con el mercado de trabajo, y una recesión puede elevar el paro estructural. En la zona del euro y en España, las estimaciones muestran una NAIRU que sube tras cada recesión y baja lentamente después, lo que hace de su medición, siempre imprecisa, una cuestión central para la política económica. De hecho, las estimaciones, con filtros estadísticos o con modelos estructurales, cambian mucho al llegar datos nuevos y obligan a usarlas con cautela al decidir la política monetaria.

La distinción entre el desempleo de equilibrio y el cíclico ordena también las políticas. Las de oferta, que actúan sobre las instituciones del mercado de trabajo, son las únicas que pueden rebajar la tasa de equilibrio de forma duradera, mientras que las de demanda solo mueven el desempleo en torno a ella, salvo que haya histéresis. La nueva síntesis integra hoy la búsqueda y el emparejamiento en la curva de Phillips neokeynesiana, y con ello la tasa natural deja de ser una constante y pasa a depender de las perturbaciones y de las políticas.

El desempleo de equilibrio reúne dos explicaciones del paro que se estudian por separado, las fricciones de búsqueda que originan el desempleo friccional (tema A26) y las rigideces que mantienen el salario por encima del de vaciado (tema A27). Combinadas, determinan una sola tasa de equilibrio, y es la distancia a esa tasa la que enlaza el mercado de trabajo con la inflación (tema A41).

Este tema analiza el desempleo de equilibrio y su persistencia:

- *Fundamentos*: las aportaciones previas a la crítica de Lucas.

- *Desempleo de equilibrio*: sus modelos, de las expectativas racionales a la histéresis.
- *Desarrollos actuales*: la teoría vigente y la convergencia del pensamiento económico.

1. Fundamentos del análisis de la tasa natural de paro (TNP), la NAIRU y la persistencia del desempleo

El análisis de la TNP, la NAIRU y la persistencia del desempleo se apoya en unos fundamentos que arrancan del debate fundacional entre la visión neoclásica y la keynesiana y continúan con la síntesis neoclásica, el monetarismo y los neokeynesianos del desequilibrio.

1.1. Debate entre la visión neoclásica y la keynesiana

El punto de partida del análisis del desempleo es la contraposición entre dos visiones del mercado de trabajo, que difieren en el grado de ajuste del salario ante perturbaciones y, a partir de ahí, en si consideran posible el desempleo involuntario. El debate fija los dos polos entre los que se moverá toda la teoría posterior, es decir, el desempleo como fenómeno de oferta, corregible con flexibilidad, o como fenómeno de demanda, corregible con estabilización.

1.1.1. La visión neoclásica

Origen: la visión neoclásica del mercado de trabajo procede de Marshall y de Pigou (1933), que la construyen sobre la teoría de la oferta y la demanda de trabajo (tema A25).

Idea: el trabajo se compra y se vende igual que cualquier bien. Su precio, el salario real, refleja sin error la escasez relativa de trabajo. Como se ajusta libremente, el mercado se vacía al instante y en equilibrio hay pleno empleo.

Razonamiento: si los agentes optimizan, el equilibrio que resulta es único, estable y eficiente. Para que haya desempleo involuntario algo tiene que impedir que el salario se ajuste (es el *desempleo clásico* de Malinvaud). Por eso la receta son políticas de oferta que supriman esas rigideces, ya que las de demanda no logran mover las variables reales. En suma, sin rigideces el desempleo solo puede ser voluntario o friccional.

Desarrollo gráfico: el mercado de trabajo neoclásico se representa en el plano empleo-salario real (figura A28.1):

1. *Demanda de trabajo* $L^d(w/P)$: es decreciente en el salario real, porque las empresas contratan hasta igualarlo con la productividad marginal del trabajo, que es decreciente.

2. *Oferta de trabajo* $L^s(w/P)$: es creciente en el salario real, porque predomina el efecto sustitución entre ocio y consumo.

3. *Equilibrio*: en el punto 0 se vacía el mercado, con el salario real $(w/P)^*$ y el empleo de pleno empleo L^*.

Con un salario real rígido por encima del de equilibrio, $(w/P)_1$, la oferta L_1^s supera a la demanda L_1^d y la diferencia es desempleo clásico.

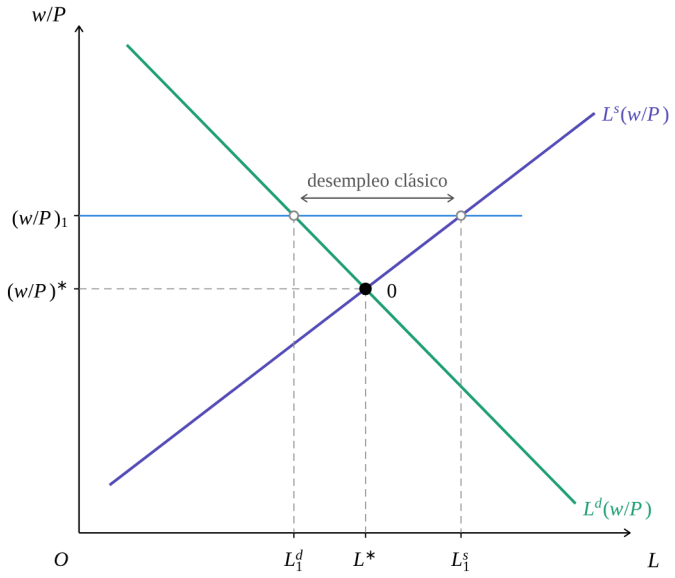


Figura A28.1: Mercado de trabajo neoclásico. Con salario real flexible hay pleno empleo y el desempleo solo puede ser voluntario o friccional.

Conclusiones:

▪ *Aspectos positivos*: ofrece un marco coherente, en el que el salario real y el empleo resultan de las decisiones optimizadoras de empresas y trabajadores, y que sigue siendo la referencia de largo plazo de la teoría posterior. Destaca, además, el papel de las rigideces institucionales en el desempleo y el de las políticas de oferta para reducirlo.

▪ *Limitaciones*: el paradigma no puede explicar el paro masivo y persistente, como puso de manifiesto la Gran Depresión.

1.1.2. La visión keynesiana

Origen: el paradigma neoclásico entra en crisis con la Gran Depresión (1929-1933), cuando la tasa de paro en EE. UU. alcanzó el 24,9 % en 1933 (Lebergott, 1964), una magnitud difícilmente atribuible a decisiones voluntarias de ocio. Por esta razón, John Maynard Keynes (1936) rompe con la idea de que el desempleo es exclusivamente voluntario.

Idea: el desempleo involuntario es posible aunque los salarios sean flexibles, porque el nivel de empleo lo determina la demanda efectiva en el mercado de bienes y no el salario en el mercado de trabajo.

Razonamiento: la teoría keynesiana se ha asociado a menudo con la *ilusión monetaria*, a la que Irving Fisher (1928) dedicó *The Money Illusion*, según la cual los trabajadores confunden salario nominal y salario real. Para Keynes, sin embargo, a los trabajadores les importa cuánto cobran respecto a los demás. Por eso se oponen a que les bajen el salario nominal (rigidez nominal a la baja), ya que esa bajada les haría perder terreno frente a otros grupos. En cambio, aceptan el menor salario real que deja una subida de precios, porque esta afecta a todos por igual y no cambia la posición de nadie.

Como el empleo lo determina la demanda efectiva, y no el salario, una bajada salarial no eleva por sí sola el empleo si no aumenta la demanda agregada. Por tanto, unos precios y salarios flexibles ni bastan para garantizar el pleno empleo ni son imprescindibles para alcanzarlo. Las políticas de demanda, al dar estabilidad a la demanda agregada, se la dan también al mercado de trabajo.

Desarrollo gráfico: el mercado de trabajo keynesiano se representa en el plano empleo-salario real (figura A28.2):

1. *Demanda de trabajo* $L^d(w/P)$: recoge el empleo que las empresas desean para cada salario real. Es decreciente, como en el modelo neoclásico, porque contratan hasta igualar el salario real con la productividad marginal del trabajo, que es decreciente.

2. *Oferta de trabajo* $L^s(w/P)$: recoge el empleo que los trabajadores desean para cada salario real. Es creciente, porque predomina el efecto sustitución, y cortaría a la demanda en el empleo walrasiano L_C, con el salario real $(w/P)_C$.

3. *Demanda efectiva*: es una recta vertical en el empleo $\bar{L}$, el que requiere la producción que las empresas pueden vender en el mercado de bienes.

4. *Equilibrio*: en el punto 0 las empresas están sobre su curva de demanda de trabajo, con el empleo $\bar{L}$ y el salario real $(w/P)_0$.

A ese salario real los trabajadores querrían ofrecer L_0^s, y la diferencia con $\bar{L}$ es desempleo involuntario.

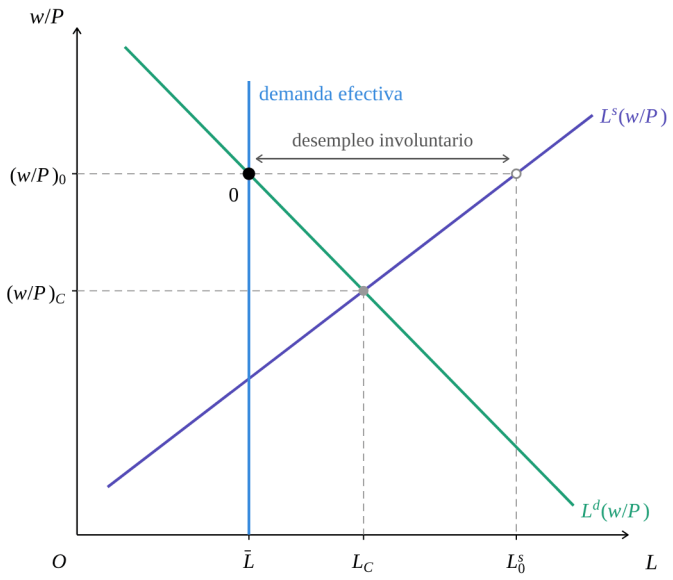


Figura A28.2: Mercado de trabajo keynesiano. El empleo lo fija la demanda efectiva ($\bar{L}$) y, al salario real resultante, aparece desempleo involuntario.

Conclusiones:

▪ *Aspectos positivos*: introduce el desempleo involuntario como resultado de una insuficiencia de demanda efectiva y no de las decisiones de los trabajadores. Con ello da cuenta del paro

masivo que la visión neoclásica no podía explicar y fundamenta las políticas de demanda contra el desempleo.

▪ *Limitaciones*: la resistencia de los salarios nominales a la baja se justifica de manera informal, sin una microfundamentación sólida, que la teoría posterior tuvo que aportar. Además, la visión keynesiana centra el análisis en la demanda y deja en segundo plano los factores de oferta y la inflación, que la síntesis neoclásica y el monetarismo incorporan después.

1.2. Síntesis neoclásica

La síntesis neoclásica busca un terreno común entre ambas visiones. Acepta el equilibrio neoclásico para el largo plazo y trata el desempleo keynesiano como un problema transitorio que las políticas de demanda pueden corregir. En el ámbito de estudio del desempleo, la síntesis neoclásica se apoya en la curva de Phillips y en la ley de Okun.

1.2.1. La curva de Phillips (Phillips, 1958; Lipsey, 1960)

Origen: Alban Phillips (1958) comprobó, con casi un siglo de datos del Reino Unido (1861-1957), que la inflación salarial tendía a ser mayor cuanto menor era la tasa de desempleo. La relación empírica no era lineal, y Richard Lipsey (1960) la dotó después de fundamento teórico.

Idea: existe una relación estable e inversa entre inflación y desempleo, que la política económica puede explotar.

Razonamiento: sobre esta observación empírica, Lipsey deriva una relación teórica con base en tres supuestos:

▪ *Relación demanda de trabajo-inflación salarial*: la inflación salarial crece con el exceso de demanda de trabajo ($\dot{w}/w = \gamma Z_L$).

▪ *Relación demanda de trabajo-tasa de desempleo*: el exceso de demanda decrece con la tasa de desempleo, de forma no lineal (lo que explica la curvatura). La relación linealizada se escribe $Z_L = \nu(u - \bar{u})$, con $\nu < 0$ y $\bar{u}$ la tasa de desempleo (friccional) a la que el exceso de demanda se anula.

▪ *Productividad constante*: como la productividad marginal no varía en el tiempo, la inflación salarial se traslada a la inflación de precios ($\dot{w}/w = \pi$).

Combinando los tres supuestos se obtiene la conocida *curva de Phillips* o curva de Phillips sin expectativas:

$$\pi_t = -\kappa_u(u_t - \bar{u}), \qquad \kappa_u = -\gamma\nu > 0$$

Como Lipsey obtenía primero curvas por industria y después las agregaba, la inflación depende no solo del nivel del desempleo, sino también de su distribución entre industrias.

La curva establece así una disyuntiva (*trade-off*) estable entre inflación y desempleo, y la ley de Okun la traslada del plano desempleo-inflación al plano producción-inflación. Samuelson y Solow (1960) vieron en la curva un *menú de política económica* (véase la figura del tema A41, apartado 2.2.1). Eligiendo un punto de ella se podía hacer un "ajuste fino" del mercado laboral, es decir, comprar un paro bajo al precio de una inflación persistente.

Conclusiones:

▪ *Aspectos positivos*: la curva ofreció la primera relación cuantificable entre los objetivos de inflación y de empleo y fue el instrumento central de la política económica de los años sesenta.

▪ *Limitaciones*: a partir de 1970 se produce el fenómeno de la *estanflación* (elevados desempleo e inflación simultáneos), de especial intensidad en Europa, y la curva de Phillips deja de ser estable empíricamente. Entre las causas señaladas por la literatura figuran los *shocks* del petróleo (1973-1974 y 1979), la caída de la productividad total de los factores, la concentración del poder sindical (especialmente en Europa) y la inestabilidad monetaria posterior a la quiebra de Bretton Woods.

En el plano teórico, además, a la curva le faltaba microfundamentación. La curva solo podía mantenerse en cualquier horizonte si los trabajadores no tenían en cuenta la inflación esperada y sufrían una ilusión monetaria permanente, porque tenían expectativas estáticas, es decir, esperaban que el nivel de precios no cambiara, $\pi_t^e = 0$.

1.2.2. La ley de Okun (1962)

Origen: Arthur Okun (1962) estableció la regularidad empírica que conecta el desempleo con la producción y que permite traducir uno en otra.

Idea: la desviación de la tasa de desempleo respecto a su nivel de equilibrio es proporcional a la brecha entre la producción efectiva y la potencial, $y_t - \bar{y}_t = -\phi\,(u_t - \bar{u})$, donde ϕ es el coeficiente de Okun.

Razonamiento: en su estimación original, el coeficiente valía en torno a tres, de modo que cada punto de desempleo por encima del natural costaba unos tres puntos de producción. El coeficiente es mayor que la participación del trabajo en la producción porque el ajuste del empleo se acompaña de cambios en las horas, en la participación y en la productividad (atesoramiento de trabajo).

En la versión dinámica de la ley, la variación del desempleo depende de la diferencia entre el crecimiento efectivo y el potencial, $\Delta u_t = -\frac{1}{\phi}\,(g_{Y,t} - \bar{g}_Y)$, y por eso el desempleo solo cae si la economía crece por encima de su potencial.

Conclusiones:

▪ *Aspectos positivos*: la ley sigue siendo el puente entre la tasa natural y la brecha de producción (*output gap*) que usa la política de estabilización (tema A42).

▪ *Limitaciones*: es una regularidad empírica sin fundamento teórico propio, y su coeficiente difiere mucho entre países por las características de cada mercado de trabajo. Dentro de la mayoría de los países, sin embargo, la relación ha resultado estable[1], también durante la Gran Recesión (Ball, Leigh y Loungani, 2017). En España el coeficiente es bajo, porque el desempleo responde con mucha intensidad a la producción, lo que refleja la sensibilidad del empleo al ciclo.

1.3. Monetarismo

El monetarismo acepta el marco de la síntesis, pero niega su resultado central, la disyuntiva (*trade-off*) estable, al introducir las expectativas en la curva de Phillips. De ahí nacen sus dos conceptos fundamentales, la tasa natural de paro y la teoría aceleracionista de la inflación.

[1] En EE. UU., por ejemplo, no se rechaza la estabilidad del coeficiente en torno a 1973, pese a los choques de oferta de los setenta (Weber, 1995).

1.3.1. La curva de Phillips ampliada por las expectativas y la tasa natural (Friedman, 1968; Phelps, 1967)

Origen: Milton Friedman (*The Role of Monetary Policy*, 1968) y Edmund Phelps (1967) formularon de forma independiente la crítica a la curva de Phillips y el concepto de tasa natural.

Idea: no existe una disyuntiva permanente entre inflación y desempleo, porque los agentes incorporan sus expectativas de inflación a la negociación salarial. Por tanto, a largo plazo el desempleo vuelve a su *tasa natural*, determinada por factores reales, cualquiera que sea la inflación.

Razonamiento: Friedman parte de que a los agentes les importa el poder adquisitivo de su salario, por lo que negocian teniendo en cuenta los precios que esperan. Para modelizar esas expectativas elige la *hipótesis de las expectativas adaptativas* (HEA), según la cual los agentes corrigen su expectativa anterior en una fracción del error cometido, de modo que solo usan la información contenida en la inflación pasada:

$$\pi^e_{t+1} = \pi^e_t + \lambda(\pi_t - \pi^e_t), \qquad \lambda \in (0, 1]$$

donde π^e_{t+1} es la inflación esperada para el periodo siguiente, π_t la inflación observada en el periodo actual y λ la velocidad de aprendizaje.

Resolviendo recursivamente, $\pi^e_{t+1} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \pi_{t-j}$, es decir, la inflación esperada es una media ponderada de la inflación pasada con pesos geométricamente decrecientes, de modo que, si la inflación sube de forma sostenida, el agente la subestima sistemáticamente. La HEA solo coincide con las expectativas racionales (apartado 2.1.3) en casos particulares. Muth (1960) mostró, por ejemplo, que es óptima cuando la variable es la suma de un paseo aleatorio y un ruido transitorio, con un λ que depende de la varianza relativa de ambos componentes.

Con base en esta idea, los autores monetaristas construyen la *curva de Phillips ampliada por las expectativas*. Como los trabajadores negocian el salario nominal a partir de la inflación que esperan, la inflación efectiva es la esperada más el efecto de la presión del mercado de trabajo,

$$\pi_t = \pi^e_t - \kappa_u(u_t - \bar{u}), \qquad \kappa_u > 0$$

donde $\bar{u}$ es la tasa natural de paro, cuya análoga en el modelo del apartado 2.2, la NAIRU, se escribe $\bar{u} = \alpha/\kappa_u$. La curva de Phillips original es el caso particular $\pi^e_t = 0$.

De la expresión se siguen dos situaciones, según que las expectativas se cumplan o no:

- $\pi_t \neq \pi^e_t$: la inflación observada difiere de la esperada. Si supera a la esperada, los trabajadores han negociado salarios nominales que resultan demasiado bajos en términos reales, las empresas contratan más y el desempleo se sitúa por debajo de la tasa natural, $u_t < \bar{u}$. Si es inferior, ocurre lo contrario. El *trade-off* entre inflación y desempleo existe, pero solo mientras dura el error, por lo que es un fenómeno de corto plazo. Además, cada revisión de las expectativas desplaza la curva, de modo que existen tantas curvas de Phillips de corto plazo como niveles de inflación esperada.

- $\pi_t = \pi^e_t$: la inflación observada y la esperada coinciden. Los salarios reales negociados son los deseados, no hay error que corregir y el desempleo se sitúa en su tasa natural, $u_t = \bar{u}$, cualquiera que sea el nivel de la inflación. La HEA conduce a esta situación de largo plazo conforme los agentes aprenden, y en ella la curva de Phillips es vertical sobre $\bar{u}$ en el plano (u, π). De ahí que no exista *trade-off* que la política pueda explotar de forma permanente.

La tasa natural de paro corresponde, por tanto, al desempleo que hay en el equilibrio de largo plazo, cuando nadie se equivoca en sus previsiones. Abarca el paro friccional y el estructural y depende solo de factores reales e institucionales, entre ellos la eficiencia del emparejamiento, las instituciones laborales y lo competitivos que sean los mercados de bienes.

Desarrollo gráfico: la curva de Phillips ampliada por las expectativas se representa en el plano desempleo-inflación (figura A28.3):

1. *Curvas de Phillips de corto plazo* $\text{CP}_{\text{CP},0}$, $\text{CP}_{\text{CP},1}$ y $\text{CP}_{\text{CP},2}$: cada una recoge las combinaciones de desempleo e inflación posibles con una inflación esperada dada (0 %, 5 % y 10 %). Son decrecientes, porque con las expectativas fijas un desempleo menor eleva los salarios nominales y los precios, y cada revisión al alza de las expectativas desplaza la curva hacia arriba.

2. *Curva de Phillips de largo plazo* CP_{LP}: es vertical sobre la tasa natural $\bar{u}$, porque recoge las situaciones en que las expectativas se cumplen.

3. *Equilibrio*: en el punto 0, sobre CP_{LP}, la inflación observada coincide con la esperada (0 %) y el desempleo está en su tasa natural.

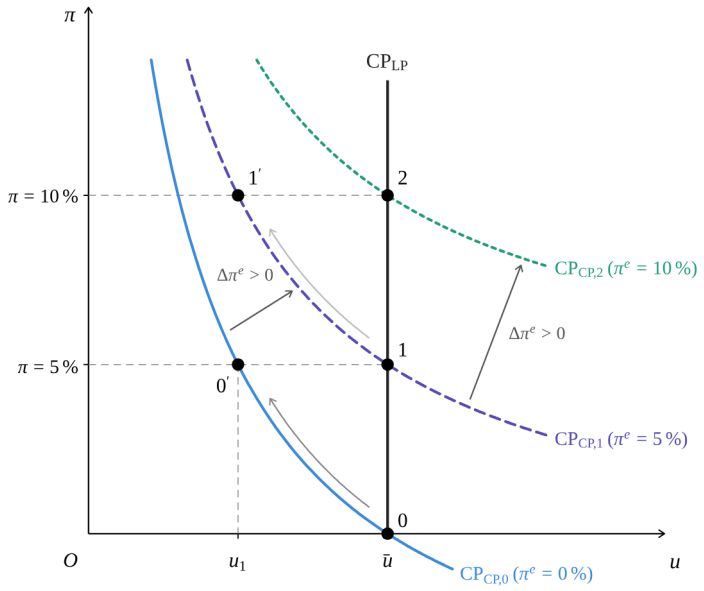


Figura A28.3: Curva de Phillips ampliada por las expectativas. La curva a corto plazo se desplaza y a largo plazo es vertical en la tasa natural.

Dinámica: una expansión que reduce el desempleo a u_1 lleva la economía del punto 0 al $0'$ a lo largo de $\text{CP}_{\text{CP},0}$, con una inflación del 5 %. Al revisarse las expectativas ($\Delta\pi^e > 0$), la curva se desplaza a $\text{CP}_{\text{CP},1}$ y el desempleo vuelve a $\bar{u}$ en el punto 1, con una inflación del 5 % que ya es la esperada. Si la autoridad insiste en mantener u_1, el proceso se repite (de 1 a $1'$ y a 2) con una inflación cada vez mayor.

Conclusiones:

- *Aspectos positivos*: el concepto de tasa natural permanece hasta hoy como guía para la medición del paro cíclico, y la insistencia en las expectativas y los retardos de la política monetaria

transformó la disciplina. Además, la teoría anticipó la ruptura de la curva de Phillips antes de que la estanflación de los años setenta la hiciera evidente.

- *Limitaciones*: la HEA emplea únicamente información pasada, por lo que los agentes cometen errores sistemáticos y predecibles, algo difícil de conciliar con la racionalidad.

1.3.2. La teoría aceleracionista de la inflación (Friedman, 1968; Phelps, 1967)

Origen: Friedman (1968) y Phelps (1967) extraen esta teoría como corolario de la curva de Phillips ampliada por las expectativas.

Idea: el Gobierno solo puede mantener el desempleo por debajo de la tasa natural acelerando la inflación, es decir, explotando permanentemente los errores de predicción de los agentes, lo que exige aumentar continuamente la tasa de crecimiento de la oferta monetaria.

Razonamiento: si la autoridad monetaria fija un objetivo de desempleo inferior a la tasa natural, ha de generar cada periodo una inflación superior a la esperada. Sin embargo, con expectativas adaptativas, la inflación esperada se ajusta a la pasada, de modo que sostener la sorpresa exige una inflación creciente (véase la figura del tema A41, apartado 2.2.2). Si la autoridad deja de acelerar la inflación, la economía regresa a $\bar{u}$ con una inflación más alta. En cambio, si la autoridad persiste, el intento de sostener el empleo por encima de su nivel de equilibrio degenera en una espiral inflacionista.

El corolario de política económica es que solo las políticas de oferta pueden reducir la tasa natural de forma duradera, y que la política monetaria debe renunciar a explotar la curva de Phillips.

Conclusiones:

- *Aspectos positivos*: explica por qué los intentos de mantener el desempleo por debajo de su tasa natural desembocaron en la inflación creciente de los años setenta, y fundamenta la renuncia de la política monetaria a explotar la curva de Phillips.

- *Limitaciones*: el resultado descansa en las expectativas adaptativas. Con expectativas racionales, en cambio, una política sistemática ni siquiera tiene efectos transitorios. Cuando las expectativas están ancladas, como en EE. UU. desde mediados de los años noventa, la curva vuelve a ligar el nivel de la inflación, y no su variación, con el desempleo (Blanchard, 2017, cap. 8).

1.4. Neokeynesianos del desequilibrio: la tipología de Malinvaud (1977)

Finalmente, la escuela del desequilibrio (Clower, 1965; Leijonhufvud, 1968; Barro y Grossman, 1971) reinterpreta a Keynes como una teoría de los fallos de coordinación con precios rígidos y racionamiento.

Origen: Edmond Malinvaud (1977) condensa las aportaciones de Clower y Leijonhufvud sobre los fallos de coordinación y el ajuste dinámico. Bajo la *hipótesis de decisión dual* (los agentes deciden sujetos a las cantidades que efectivamente pueden transar en los demás mercados), con rigideces de precios y racionamiento, identifica tres regímenes de desequilibrio según la situación conjunta de los mercados de trabajo y de bienes. A ellos añade un cuarto caso límite, el subconsumo.

Idea: el desempleo no es un fenómeno único, sino que, según la combinación de excesos de oferta o de demanda en los mercados de trabajo y de bienes, un mismo nivel de desempleo puede tener causas distintas y exigir remedios distintos, incluso opuestos.

Razonamiento: los cuatro regímenes son:

1. *Desempleo clásico*: exceso de oferta de trabajo y exceso de demanda de bienes. El salario real es superior al de vaciado (p. ej. por la acción sindical).

- *Política adecuada*: bajada del salario real, que eleva la oferta de bienes y la demanda de trabajo. Un estímulo de demanda, en cambio, es aquí ineficaz, porque la producción está limitada por la oferta, y solo se precisa si la bajada salarial lleva la economía al régimen de desempleo keynesiano.

2. *Desempleo keynesiano*: exceso de oferta de trabajo y exceso de oferta de bienes. El salario real puede ser inferior, igual o superior al de vaciado, porque el problema no está en el precio del trabajo, sino en la demanda efectiva.

- *Política adecuada*: una subida limitada del salario real o, sobre todo, la eliminación de la restricción de demanda mediante el impulso fiscal.

3. *Inflación reprimida*: los controles administrativos impiden que precios y salarios suban hasta su nivel de equilibrio, así que tanto en bienes como en trabajo hay exceso de demanda. El desequilibrio aflora entonces como escasez y racionamiento, no como inflación. La inflación reprimida es el diagnóstico clásico de las economías planificadas y de las economías de guerra con controles.

- *Política adecuada*: liberalizar precios y salarios para que absorban el exceso de demanda, acompañada de una contracción de la demanda agregada (fiscal o monetaria) que evite que la liberalización se traduzca en una inflación abierta.

4. *Subconsumo*: exceso de demanda de trabajo y exceso de oferta de bienes, un caso lógicamente posible pero infrecuente en la práctica. Por eso, su interés es sobre todo teórico, como el régimen que completa la tipología.

- *Política adecuada*: elevar el salario real, que estimula el consumo (y absorbe así el exceso de oferta de bienes) y modera la demanda de trabajo de las empresas (y reduce así su exceso).

Gráficamente, con la oferta monetaria y el gasto público dados, los regímenes se reparten el plano salario nominal-nivel de precios, (w, P), en torno al equilibrio walrasiano, W (figura A28.4). Un nivel de precios alto reduce los saldos reales y la demanda de bienes (desempleo keynesiano), un salario alto respecto a los precios contrae la oferta de las empresas (desempleo clásico), y salarios y precios bajos generan exceso de demanda en ambos mercados (inflación reprimida). En cambio, el subconsumo solo aparece sobre la frontera entre el desempleo keynesiano y la inflación reprimida, de ahí que sea un caso límite.

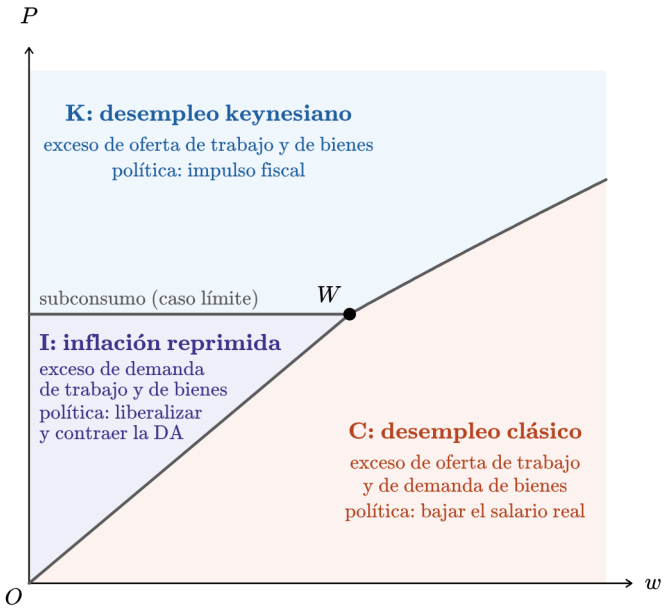


Figura A28.4: Regímenes de desequilibrio de Malinvaud (1977). Con la oferta monetaria dada, las fronteras que parten del equilibrio walrasiano, W, separan el desempleo keynesiano (K, precios altos), el desempleo clásico (C, salario real superior al de vaciado) y la inflación reprimida (I, salarios y precios bajos). El subconsumo es el caso límite de la frontera entre K e I.

Conclusiones:

- *Aspectos positivos*: la caracterización de los regímenes permite discriminar la política óptima según el tipo de desequilibrio, frente a las recetas universales.

- *Limitaciones*: son metodológicas, pues las rigideces de precios se postulan sin microfundamentar y la identificación empírica del régimen vigente resulta difícil, lo que explica el declive del enfoque tras la revolución de las expectativas racionales.

1.5. Conclusiones

Los fundamentos parten de dos diagnósticos opuestos. Para la visión neoclásica, el desempleo involuntario exige una rigidez que impida el ajuste del salario real, y se corrige con políticas de oferta. Para Keynes, el empleo lo fija la demanda efectiva, y el desempleo se corrige estabilizándola. La síntesis neoclásica concilia ambas posturas con la curva de Phillips, un menú estable de inflación y desempleo que la ley de Okun traslada a la producción.

El monetarismo niega ese menú al introducir las expectativas. Cuando los agentes corrigen sus errores, el desempleo vuelve a la tasa natural, y mantenerlo por debajo exige acelerar la inflación. Malinvaud muestra, por último, que un mismo desempleo puede responder a regímenes distintos y exigir remedios opuestos. Quedan abiertos dos problemas, los errores sistemáticos que implican las expectativas adaptativas y las rigideces de precios que el enfoque del desequilibrio postula sin microfundamentar. La sección 2 los aborda con las expectativas racionales y con el modelo de la NAIRU, que deriva el desempleo de equilibrio de la negociación salarial y de la fijación de precios.

2. Modelos de la tasa de desempleo de equilibrio: expectativas racionales, NAIRU e histéresis

La sección anterior deja establecido que existe una tasa de desempleo de equilibrio hacia la que la economía vuelve cuando las expectativas se cumplen, pero la tasa natural de Friedman es un concepto más que un modelo. Es decir, no dice cómo se forman las expectativas cuando los agentes no se limitan a mirar el pasado, de qué depende esa tasa, ni por qué el desempleo europeo no volvió a su nivel anterior tras cada recesión.

Los tres modelos de esta sección responden a esas tres preguntas:

- *Nueva macroeconomía clásica*: sustituye las expectativas adaptativas por las racionales y lleva la lógica de Friedman a su límite, de modo que, sin errores sistemáticos, la curva de Phillips es vertical también a corto plazo.

- *Modelo de la NAIRU*: de la primera generación neokeynesiana, abandona el mercado walrasiano y deriva la tasa de equilibrio de la negociación salarial y de la fijación de precios en mercados imperfectos, con lo que la hace depender de instituciones observables.

- *Histéresis*: muestra que esa tasa puede no ser un ancla sino una variable que sigue al propio desempleo, lo que devuelve a la política de demanda un papel a largo plazo que los dos modelos anteriores le niegan.

2.1. Nueva macroeconomía clásica

La nueva macroeconomía clásica (NMC) aborda el desempleo de equilibrio con tres aportaciones: la crítica de Lucas, la sustitución intertemporal en la oferta de trabajo y la curva de Phillips con expectativas racionales.

2.1.1. Crítica de Lucas (1976)

Origen: Robert Lucas (*Econometric Policy Evaluation: A Critique*, 1976) cuestionó el uso de los modelos de la síntesis para evaluar políticas.

Idea: Lucas sostiene que los modelos econométricos de la síntesis no pueden usarse para evaluar políticas alternativas, porque sus parámetros estimados dependen del régimen de política vigente. Al cambiar el régimen, cambian las reglas de decisión de los agentes y, con ellas, los parámetros estimados.

Razonamiento: la crítica exige microfundamentar los modelos y modelizar explícitamente las expectativas, que es el programa de la nueva macroeconomía clásica y, después, de la nueva economía keynesiana. Aplicada al mercado de trabajo, implica que la curva de Phillips estimada bajo un régimen no sirve para predecir bajo otro.

Conclusiones:

- *Aspectos positivos*: explica por qué la curva de Phillips estimada en los años sesenta dejó de funcionar cuando la política trató de explotarla, y su exigencia de modelos con parámetros estructurales, invariantes a la política, es hoy un estándar de la macroeconomía aplicada.

▪ *Limitaciones*: su relevancia empírica es discutida, porque los contrastes encuentran pocos casos en que los parámetros de los modelos econométricos tradicionales cambien de forma significativa con la política (Ericsson e Irons, 1995). Además, la invariancia de los parámetros de preferencias y tecnología es un supuesto difícil de contrastar.

2.1.2. Oferta de trabajo y sustitución intertemporal (Lucas y Rapping, 1969)

Origen: Robert Lucas y Leonard Rapping (1969) fueron los primeros en explicar las fluctuaciones del empleo dentro de un mercado de trabajo que se vacía continuamente.

Idea: los trabajadores eligen cuánto trabajar comparando el salario real actual con el esperado en el futuro, de modo que las fluctuaciones del empleo son respuestas voluntarias de sustitución intertemporal del ocio ante salarios percibidos como transitoriamente altos o bajos (tema A25, apartado 2.2).

Razonamiento: combinada con la información imperfecta sobre los precios agregados, la sustitución intertemporal permite a la NMC generar fluctuaciones del empleo sin abandonar el vaciado continuo de los mercados. Por ello, el desempleo cíclico es ocio elegido, y solo la tasa natural es desempleo en sentido propio.

Conclusiones:

▪ *Aspectos positivos*: ofrece la primera explicación de las fluctuaciones del empleo con agentes optimizadores y mercados que se vacían, y su mecanismo de sustitución intertemporal pasa a los modelos del ciclo real (tema A42, apartado 2.2).

▪ *Limitaciones*: el mecanismo exige una elasticidad intertemporal de la oferta de trabajo elevada, que la evidencia microeconométrica no respalda (Altonji, 1986). Además, resulta difícil interpretar el aumento del desempleo en las recesiones como ocio elegido.

2.1.3. Curva de Phillips con expectativas racionales (Lucas, 1972; Sargent y Wallace, 1975)

Origen: Robert Lucas (1972) derivó la curva de Phillips de un modelo con expectativas racionales e información imperfecta, y Thomas Sargent y Neil Wallace (1975) extrajeron su implicación para la política económica.

Idea: la NMC deja atrás las expectativas adaptativas y supone que son racionales (*hipótesis de las expectativas racionales*, HER). Al prever, los agentes aprovechan con eficiencia toda la información a su alcance, incluido cómo funciona la economía y qué política se ha anunciado, y por eso solo se equivocan al azar y sin sesgo. Desaparece así lo que en Friedman daba lugar al *trade-off* de corto plazo, unos agentes que erraban siempre en la misma dirección.

Razonamiento: en la parábola de las islas, un productor ve subir el precio de lo que vende sin conocer el nivel general de precios. Producir más solo le interesa si la subida es de precio relativo, no si todos los precios han subido, y ha de averiguar qué caso se da. Si atribuye parte de la subida a un cambio relativo, aumenta la producción y el empleo. Al agregar y aplicar la ley de Okun, el desempleo se desvía de su tasa natural solo en la medida en que el nivel de precios (en logaritmos, p_t) difiera del esperado (p_t^e):

$$u_t = \bar{u} - \frac{\zeta}{\phi}(p_t - p_t^e) + \varepsilon_t, \qquad \zeta, \phi > 0$$

donde ζ mide la respuesta de la producción a la sorpresa de precios, ϕ es el coeficiente de Okun y ε_t una perturbación aleatoria.

Con expectativas racionales, $p_t - p_t^e$ es un error impredecible con la información disponible. Por eso, el desempleo baja de su tasa natural únicamente cuando hay sorpresas inflacionarias, y vuelve a ella en cuanto se disipan.

Así, una política monetaria sistemática es anticipada e incorporada en p_t^e, y no tiene efecto real ni siquiera a corto plazo. Por tanto, la curva es vertical también a corto plazo, salvo por las sorpresas.

Conclusiones:

- *Aspectos positivos*: la HER obliga a modelizar explícitamente las expectativas y a microfundamentar los modelos, y su implicación de política, que la estabilización sistemática es inútil, fijó los términos del debate posterior.

- *Limitaciones*: para que el modelo genere fluctuaciones observables, los agentes han de confundir cambios de precios relativos con cambios generales durante periodos largos, lo que es poco creíble cuando el nivel de precios se publica mensualmente. Por su parte, el mecanismo de sustitución intertemporal en que se apoya exige una elasticidad de la oferta de trabajo poco realista. Por todo lo anterior, la propia NMC evolucionó hacia los modelos del ciclo real (tema A42), y la nueva economía keynesiana conservó la HER pero abandonó el vaciado continuo.

2.2. Nueva economía keynesiana de primera generación: el modelo de la NAIRU (Layard y Nickell, 1985; Carlin y Soskice, 1990)

Origen: Richard Layard y Stephen Nickell (1985) y, en su versión de manual, Wendy Carlin y David Soskice (1990) formularon el modelo de salarios y precios del que resulta la NAIRU. Layard, Nickell y Jackman (1991) lo sistematizan y lo contrastan con datos de la OCDE.

Idea: la nueva economía keynesiana acepta de la NMC las expectativas racionales y la exigencia de microfundamentos, pero rechaza su punto de partida, porque el mercado de trabajo no es walrasiano. Los salarios se negocian y los precios los fijan empresas con poder de mercado, de modo que el desempleo de equilibrio no es el que vacía el mercado sino el que hace compatibles las aspiraciones de trabajadores y empresas sobre el reparto de la renta.

2.2.1. Supuestos

Entorno: la economía es cerrada y los mercados de trabajo y de bienes son imperfectos.

Agentes: sindicatos y empresas, ambos con poder de mercado y sin ilusión monetaria, de modo que negocian en términos reales esperados y lo relevante es el salario real y el margen real, no sus valores nominales.

- *Sindicatos*: en el mercado de trabajo, un sindicato (o los trabajadores en negociación colectiva) fija el salario real que aspira a obtener, y esa aspiración crece con la presión del mercado, es decir, decrece con el desempleo.

- *Empresas*: en el mercado de bienes, las empresas fijan el precio como un margen sobre el coste marginal, y ese margen determina el salario real que están dispuestas a pagar.

Tecnología: el trabajo es el único factor variable y opera con rendimientos constantes, $Y = A_L \cdot L$, de modo que la productividad marginal coincide con la media y es constante. El coste marginal es, por tanto, el salario dividido por la productividad.

2.2.2. Modelo

Planteamiento: el equilibrio resulta de dos relaciones, la que fija el salario real que los trabajadores pretenden y la que fija el que las empresas conceden. El desempleo es la variable que reconcilia ambas pretensiones. Si el salario real al que aspiran los trabajadores supera al que las empresas pagan, la diferencia se traduce en una inflación superior a la esperada.

Salario real negociado (SRN): al inicio de cada periodo, los trabajadores negocian un salario real esperado, w/P^e, que depende negativamente del desempleo (a mayor desempleo, menor poder negociador) y positivamente de un conjunto de factores institucionales z (prestaciones por desempleo, salario mínimo, cotizaciones):

$$\frac{w}{P^e} = f(u^-, z^+) \tag{1}$$

Salario real pagado (SRP): las empresas fijan un margen constante $\mu = 1/(\sigma - 1) > 0$ sobre el coste marginal, que por los rendimientos constantes coincide con el coste laboral unitario:

$$P = (1 + \mu)\,\frac{w}{A_L} \tag{2}$$

$$\frac{w}{P} = \frac{A_L}{1 + \mu} = (1 - I_L)\,A_L$$

donde σ es la elasticidad de sustitución entre variedades (la elasticidad-precio de la demanda de cada empresa) e $I_L = 1/\sigma$ el índice de Lerner. El SRP es horizontal en el plano empleo-salario real, pues no depende del nivel de empleo, sino solo de la productividad y del grado de competencia[2]. El margen puede ser, no obstante, anticíclico, porque la colusión es más difícil de sostener en los auges (Rotemberg y Saloner, 1986) o porque la elasticidad-precio de la demanda es procíclica. En ese caso, el SRP tendría pendiente positiva.

Equilibrio (la NAIRU): exige que las aspiraciones recogidas en (1) y (2) sean compatibles, SRN = SRP, lo que implica el cumplimiento de las expectativas ($P^e = P$):

$$f(\bar{u}, z) = \frac{A_L}{1 + \mu}$$

La tasa de desempleo que resuelve esta ecuación es la *NAIRU* (*Non-Accelerating Inflation Rate of Unemployment*), es decir, la tasa a la que las aspiraciones de trabajadores y empresas sobre el reparto de la renta resultan mutuamente compatibles. Por eso, a esa tasa la inflación no se acelera. La NAIRU es única y estable en ausencia de histéresis, y es también compatible con desempleo involuntario, pues el salario negociado se sitúa por encima del de vaciado. Además, la NAIRU engloba el desempleo estructural y el friccional.

Conviene distinguirla de la tasa natural de Friedman. Ambas son tasas de equilibrio a largo plazo determinadas por factores reales, pero la tasa natural es un concepto de vaciado del mercado, es

[2] En un marco de competencia imperfecta existe consenso en que los precios no responden mucho ante cambios en la demanda (Carlin y Soskice, 2006, p. 140).

decir, el nivel que generaría el equilibrio general walrasiano. Friedman, no obstante, incorporaba ya en ese equilibrio las imperfecciones del mercado y los costes de información y movilidad. La NAIRU, en cambio, es un concepto de compatibilidad entre las aspiraciones de agentes con poder de mercado. Por ello, la NAIRU hace depender explícitamente el desempleo de equilibrio de instituciones (sindicatos, prestaciones, margen empresarial) que en el modelo competitivo no existen. En la práctica, los dos términos se usan como sinónimos para designar la tasa a la que la inflación es estable.

Desarrollo gráfico: la determinación de la NAIRU se representa en el plano empleo-salario real (figura A28.5):

1. *Salario real negociado* (SRN): es creciente en el empleo, porque un menor desempleo refuerza el poder negociador de los trabajadores, y se hace muy inclinado al acercarse a la población activa.

2. *Salario real pagado* (SRP): es horizontal en $A_L/(1+\mu)$, porque con rendimientos constantes el salario real que conceden las empresas no depende del empleo.

3. *Población activa* (PA): marca el empleo máximo y es una vertical, porque no depende del salario real. La distancia horizontal entre el empleo y esta recta mide el desempleo.

4. *Equilibrio*: en el punto 0 se cortan el SRN y el SRP, con el empleo L^*. La distancia entre L^* y la población activa es el desempleo de la NAIRU, $\bar{u}$.

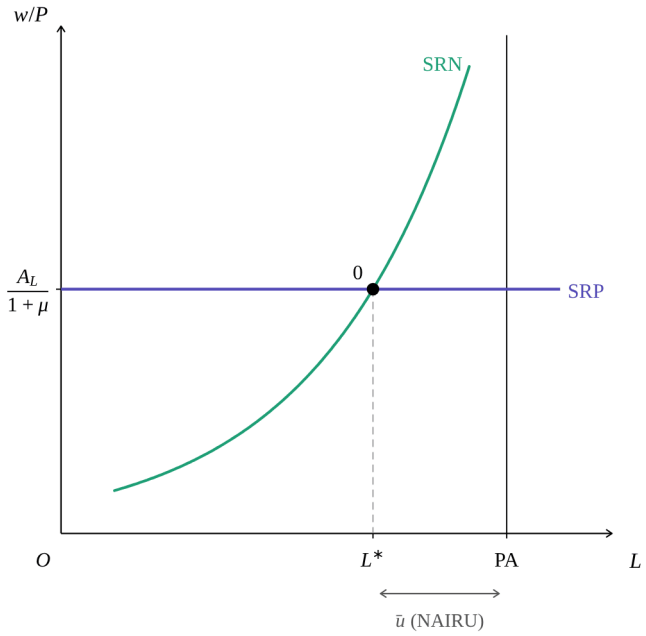


Figura A28.5: Determinación de la NAIRU. El salario real negociado (SRN) se iguala al salario real pagado (SRP) en el empleo L^*. La distancia hasta la población activa (PA) es el desempleo correspondiente a la NAIRU, $\bar{u}$.

Estática comparativa: se distinguen las políticas de demanda y las de oferta:

▪ *Políticas de demanda*: una expansión de demanda reduce transitoriamente el desempleo por debajo de la NAIRU. El menor desempleo refuerza el poder negociador de los trabajadores y eleva el salario negociado por encima del pagado (SRN > SRP). Las empresas trasladan entonces los mayores salarios a los precios y se desata una espiral salarios-precios que devuelve la economía a la NAIRU con una inflación superior. Las políticas de demanda solo tienen, pues, efectos transitorios, y a costa de inflación.

▪ *Políticas de oferta*: son las únicas que alteran la NAIRU, porque desplazan una de las dos curvas, y las óptimas son las que eliminan las rigideces de ambos mercados. Sus efectos son los siguientes:

- *Factores institucionales*: un aumento de z (prestaciones, salario mínimo) desplaza el SRN hacia arriba y eleva la NAIRU.
- *Competencia*: un aumento de μ (menor competencia en el mercado de bienes) reduce el SRP y eleva la NAIRU.
- *Productividad*: un aumento de A_L (formación) desplaza el SRP hacia arriba y reduce la NAIRU. Sin embargo, si las aspiraciones salariales acaban indexándose a la productividad, el efecto es solo transitorio.
- *Fiscalidad*: las rebajas de impuestos y cotizaciones reconcilian las aspiraciones de ambas partes a un menor nivel de desempleo.

Extensiones:

▪ *Curva de Phillips de la primera generación neokeynesiana*: sin imponer el cumplimiento de las expectativas, y suponiendo por concreción $f(u,z) = z\,e^{-\kappa_u u}$, la combinación de (1) y (2) da una curva de Phillips con fundamento en la puja distributiva entre sindicatos y empresas:

$$\pi_t = \pi_t^e + \alpha - \kappa_u u_t = \pi_t^e - \kappa_u(u_t - \bar{u}), \qquad \kappa_u > 0$$
$$\bar{u} = \frac{\alpha}{\kappa_u}, \qquad \alpha = \ln(1+\mu) + \ln z - \ln A_L$$

Se trata de la curva ampliada por las expectativas del apartado 1.3.1, con la diferencia de que la NAIRU ya no es un dato sino una función de los parámetros estructurales. Así, la NAIRU crece con el margen μ y con los factores institucionales z, y decrece con la productividad A_L.

▪ *Evidencia sobre los determinantes*: los estudios de panel para la OCDE que siguen este enfoque (Nickell, 1997) asocian una NAIRU más alta a las prestaciones generosas de duración indefinida sin exigencia de búsqueda activa. También la asocian a la sindicación elevada sin coordinación de la negociación y a una cuña fiscal elevada sobre el trabajo. La protección del empleo, en cambio, reduce a la vez las entradas y las salidas del desempleo y afecta más a su duración que a su nivel.

Blanchard y Wolfers (2000) muestran que ni las perturbaciones ni las instituciones explican por sí solas la subida del paro europeo, pues muchas instituciones ya existían cuando el paro era bajo. Lo explica su interacción, ya que unas mismas perturbaciones adversas (desaceleración de la productividad, tipos de interés reales altos) elevaron más y de forma más duradera el desempleo en los países con instituciones más rígidas.

2.2.3. Conclusiones

Aspectos positivos: el modelo fundamenta el desempleo de equilibrio en la estructura del mercado de trabajo y del de bienes (poder de negociación, márgenes, generosidad de las prestaciones), lo que lo hace más rico que la tasa natural friedmaniana y permite analizar las políticas estructurales. Además, da a la curva de Phillips ampliada un fundamento alternativo al de la NMC, basado en la puja distributiva en mercados imperfectos y no en la información imperfecta sobre los precios.

Limitaciones: la NAIRU no es observable y sus estimaciones son inestables (apartado 3.3). A ello se suma que el modelo es estático y no explica por qué el desempleo europeo se mantuvo alto tras las perturbaciones de los años setenta, lo que motiva la hipótesis de histéresis.

2.3. Histéresis: la persistencia del desempleo (Blanchard y Summers, 1986)

Origen: Olivier Blanchard y Lawrence Summers (*Hysteresis and the European Unemployment Problem*, 1986) tomaron de la física el término histéresis para describir la persistencia del desempleo europeo.

Idea: la subida del desempleo europeo en los años setenta y ochenta reveló que la NAIRU estimada parecía evolucionar a la par que el desempleo observado, en lugar de actuar como atractor estable. La *histéresis* designa precisamente esa dependencia, pues considera que la tasa de desempleo de equilibrio depende de la evolución pasada del desempleo efectivo. Se trata, por tanto, de un fenómeno de medio plazo. Según el desempleo vuelva muy despacio a su nivel de partida o no vuelva, la histéresis es parcial (persistencia) o pura.

2.3.1. Razonamiento: los mecanismos

Blanchard y Summers identifican tres mecanismos por los que el desempleo pasado eleva el desempleo de equilibrio:

1. *Reducciones en la acumulación de capital*: las perturbaciones negativas reducen el empleo y la inversión, y la caída del *stock* de capital provoca disminuciones adicionales de la demanda de trabajo.

2. *Pérdida de capital humano*: el desempleo prolongado deprecia las cualificaciones y genera desgaste psicológico (efecto desánimo), lo que dificulta la reinserción (Phelps, 1972). Esta pérdida es uno de los mecanismos con mayor respaldo empírico, aunque Blanchard y Summers (1986) consideraron más prometedor el de *insiders-outsiders*.

3. *Dinámica* insiders-outsiders: los *insiders*, que concentran el poder negociador, fijan el salario atendiendo solo a su propio empleo, de modo que en las recesiones no se ajusta a la baja lo suficiente para reabsorber a los despedidos[3].

[3] En su interpretación menos extrema (Romer, 2012, cap. 10), la dinámica *insiders-outsiders* solo mitiga la sensibilidad cíclica de los costes marginales del trabajo.

2.3.2. Desarrollo en el modelo de la NAIRU

Analíticamente, la histéresis consiste en hacer depender la NAIRU del desempleo pasado. La forma más simple es una media ponderada de la NAIRU estructural y del desempleo del periodo anterior:

$$\bar{u}_t = (1-h)\,\bar{u}^* + h\,u_{t-1}, \qquad h \in [0,1]$$

donde $\bar{u}^*$ es la NAIRU estructural de largo plazo y h el grado de histéresis. Se dan así tres casos:

- *Sin histéresis*: con $h = 0$ la NAIRU es constante, como en el modelo de Layard y Nickell.

- *Histéresis parcial*: con $0 < h < 1$ hay persistencia, pues tras una perturbación la NAIRU se desplaza y vuelve solo gradualmente a $\bar{u}^*$. La NAIRU de largo plazo es única, aunque a corto o medio plazo puede haber varias.

- *Histéresis pura* (formulación original de Blanchard y Summers): con $h = 1$, la NAIRU es simplemente el desempleo del periodo anterior y el desempleo sigue un paseo aleatorio, sin nivel de largo plazo al que volver. En esta forma extrema de persistencia deja de existir una NAIRU única, y las políticas de demanda vuelven a tener efectos a largo plazo.

Sustituyendo en la curva de Phillips, $\pi_t = \pi_t^e - \kappa_u(u_t - \bar{u}_t)$, la inflación deja de depender de la distancia a una tasa fija y pasa a depender, en el caso puro, de la variación del desempleo, $\pi_t = \pi_t^e - \kappa_u\,\Delta u_t$.

Gráficamente, en el plano empleo-salario real del modelo de Layard y Nickell (figura A28.6), el SRN recoge el salario real que negocian los trabajadores para cada nivel de empleo y es creciente, porque un menor desempleo refuerza su poder negociador. El SRP, horizontal, es el que pagan las empresas con su margen.

La histéresis se representa mediante un desplazamiento de la curva SRN. Tras una contracción de demanda, el desempleo aumenta ($\Delta u_{t-1} > 0$) y parte de los desempleados deja de presionar sobre la negociación. Los *insiders* restantes, con mayor poder negociador efectivo, sostienen sus salarios. La curva se desplaza hacia la izquierda, de SRN_0 a SRN_1, y, con el SRP sin cambios, el equilibrio pasa del punto 0 al 1, con menor empleo ($L_1^* < L_0^*$) y una NAIRU superior ($\bar{u}_1 > \bar{u}_0$). Por tanto, cuando la coyuntura mejora, el desempleo ya no retorna a su nivel inicial.

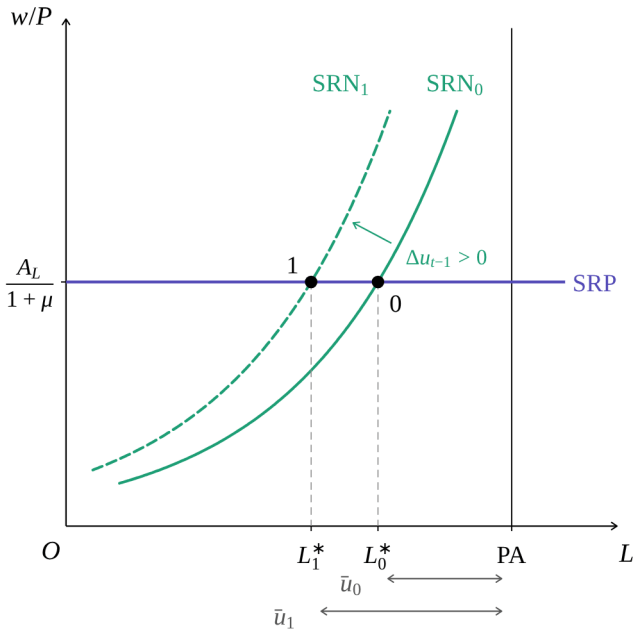


Figura A28.6: Histéresis. Un *shock* transitorio desplaza de forma permanente la curva SRN y eleva la NAIRU.

El corolario de política es que las políticas activas de empleo (formación, intermediación, recualificación de parados de larga duración) pueden favorecer el retorno del desempleo a su nivel inicial, porque frenan la pérdida de capital humano (segundo de los mecanismos anteriores), que el apartado 3.1 reinterpreta como una caída de la eficiencia del emparejamiento. La dinámica *insiders-outsiders* se desarrolla en el tema A27, apartado 2.4.

2.3.3. Conclusiones

Aspectos positivos: la histéresis explica lo que la NAIRU estática no podía, la persistencia del desempleo europeo tras las perturbaciones de los años setenta y ochenta, y devuelve a la política de demanda un papel a largo plazo. Además, sus tres mecanismos han recibido apoyo empírico, y el caso español es una de sus ilustraciones más citadas.

Limitaciones: la literatura posterior considera que la histéresis pura es una hipótesis demasiado extrema, por tres razones:

- *Outsiders*: los desempleados sí pueden afectar, siquiera parcialmente, al resultado de la negociación.

- *Insiders*: afrontan a su vez una probabilidad positiva de caer en el desempleo, lo que modera sus demandas.

- *Empresas*: pueden recurrir a la amenaza creíble de despido y sustitución.

En consecuencia, la dinámica *insiders-outsiders* genera una persistencia elevada ($0 < h < 1$), pero no histéresis pura, y el desempleo acaba volviendo a una NAIRU de largo plazo, aunque lentamente.

2.4. Conclusiones

Los tres modelos convierten la tasa natural en una teoría del desempleo de equilibrio. En primer lugar, la nueva macroeconomía clásica la combina con expectativas racionales, de modo que solo las sorpresas apartan el desempleo de esa tasa y una política sistemática carece de efectos reales incluso a corto plazo. En segundo lugar, el modelo de la NAIRU sitúa el equilibrio donde el salario real que negocian los trabajadores coincide con el que pagan las empresas. La tasa depende entonces de instituciones (prestaciones, sindicatos, margen empresarial), y solo las políticas de oferta la alteran. Finalmente, la histéresis suma a esos determinantes el desempleo pasado y devuelve a la política de demanda efectos duraderos, aunque la literatura considera más plausible una persistencia elevada que la histéresis pura.

Quedan por precisar los determinantes que resume el vector z y la forma de medir la tasa. La sección 3 lo hace con el modelo de búsqueda y emparejamiento, que da contenido a esos factores, y con los métodos de estimación de la NAIRU y sus límites.

3. Desarrollos actuales y convergencia del pensamiento económico

La teoría actual del desempleo de equilibrio incorpora las fricciones de búsqueda y la curva de Phillips neokeynesiana y afronta el problema práctico de estimar la NAIRU. La sección termina con la convergencia del pensamiento económico.

3.1. El modelo de búsqueda y emparejamiento

Los desarrollos actuales parten del modelo de búsqueda y emparejamiento, que microfundamenta el desempleo de equilibrio como resultado de un proceso costoso de encuentro entre trabajadores y vacantes. Su aportación a este tema es triple.

- *Nuevos parámetros*: en primer lugar, ofrece una teoría de la tasa de desempleo de equilibrio en la que esta depende de la eficiencia del emparejamiento, de la tasa de destrucción de empleo, del poder negociador de los trabajadores y del coste de mantener una vacante. Ello da contenido a los factores z del modelo de la NAIRU.

- *Naturaleza de las perturbaciones*: en segundo lugar, distingue con precisión las perturbaciones coyunturales, que mueven el equilibrio sobre la curva de Beveridge, de las estructurales, que la desplazan y alteran la tasa de equilibrio. Dota así de un instrumento empírico a la distinción entre paro cíclico y estructural.

- *Persistencia y eficiencia*: en tercer lugar, explica la persistencia como pérdida de eficiencia del emparejamiento tras las recesiones (pérdida de capital humano, desajuste). Ello convierte a la histéresis en un fenómeno del proceso de búsqueda y no solo de la negociación salarial. El modelo ha permitido también integrar las fricciones de búsqueda en los modelos del ciclo[4]. Asimismo, hace de las políticas activas de empleo el instrumento para desplazar la curva de Beveridge hacia el origen.

Gráficamente, en el plano desempleo-vacantes (figura A28.7), el equilibrio resulta de dos curvas:

1. *Curva de Beveridge* (CB_0): recoge las combinaciones de desempleo y vacantes en que las entradas en el desempleo y las salidas se compensan. Es decreciente, porque con más vacantes los desempleados encuentran trabajo antes, y convexa respecto al origen, porque cada reducción adicional del desempleo exige un aumento cada vez mayor de las vacantes.

2. *Recta de creación de empleo* ($v = \theta_0^* u$): recoge las combinaciones de desempleo y vacantes con la estrechez de equilibrio. Parte del origen con pendiente igual a la estrechez del mercado θ_0^*, determinada por la condición de creación de empleo de las empresas.

3. *Equilibrio*: en el punto 0 se cortan ambas, con la tasa de desempleo de equilibrio $\bar{u}_0$ y las vacantes v_0.

Una caída de la eficiencia del emparejamiento ($\Delta A < 0$), una perturbación estructural, desplaza la curva hacia fuera, de CB_0 a CB_1, y reduce la estrechez, de modo que la recta gira a $v = \theta_1^* u$, menos inclinada. El equilibrio pasa del punto 0 al 1, con una tasa de desempleo de equilibrio mayor ($\bar{u}_1 > \bar{u}_0$).

[4] Merz (1995) y Andolfatto (1996) fueron los primeros en integrar fricciones de búsqueda en modelos del ciclo real.

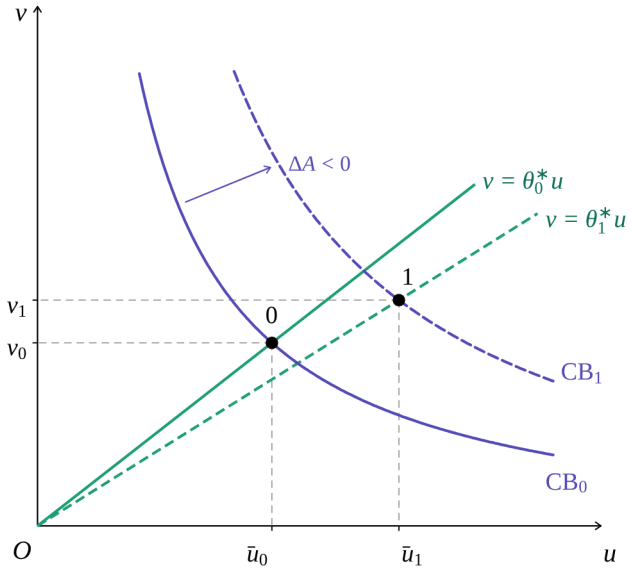


Figura A28.7: Curva de Beveridge (CB) en el modelo de búsqueda y emparejamiento. Una caída de la eficiencia del emparejamiento ($\Delta A < 0$) desplaza la curva hacia fuera y hace menos inclinada la recta de creación de empleo ($v = \theta^* u$), y el equilibrio pasa de 0 a 1, con una tasa de desempleo de equilibrio mayor ($\bar{u}_1 > \bar{u}_0$).

El modelo de Diamond, Mortensen y Pissarides se desarrolla en el tema A26, apartado 2. Aquí se recoge su aportación a la teoría de la tasa de desempleo de equilibrio.

3.2. La curva de Phillips de la nueva economía keynesiana de segunda generación (NEK2G)

Origen: la nueva síntesis neoclásica (tema A07) deriva la curva de Phillips de la fijación escalonada de precios a la Calvo (1983) con expectativas racionales. La versión híbrida se debe a Galí y Gertler (1999) y la salarial, a Erceg, Henderson y Levin (2000), que Galí (2011) reformula en términos de desempleo.

Idea: los modelos de la NEK2G, o de la nueva síntesis neoclásica (NSNC), derivan una curva de Phillips plenamente microfundamentada y prospectiva. En ella, la inflación de hoy depende de la inflación esperada y de la brecha de producción o, de forma equivalente, del desempleo.

3.2.1. Supuestos

Entorno: las empresas deciden bajo incertidumbre en un modelo dinámico, en tiempo discreto y con horizonte infinito ($T \to \infty$), lo que, junto con la rigidez de precios a la Calvo, hace que las decisiones de precios de hoy dependan de la inflación esperada en el futuro. Los precios se fijan a la Calvo, de manera que cada periodo solo una fracción $1 - \xi$ de las empresas, seleccionada aleatoriamente, puede reajustar su precio. El nivel agregado de precios evoluciona entonces como una media ponderada del precio pasado y del precio óptimo de reajuste, $P_t = \left[\xi P_{t-1}^{1-\sigma} + (1-\xi)(P_t^*)^{1-\sigma}\right]^{1/(1-\sigma)}$, lo que introduce la rigidez nominal agregada.

Agentes: hogares idénticos con vida infinita, que eligen consumo, trabajo y tenencia de bonos, y un continuo de empresas, cada una productora de una variedad, que fijan su precio. Todos forman sus expectativas según la HER.

Mercado de bienes: hay competencia monopolística en el mercado de bienes, con un continuo de variedades.

3.2.2. Modelo

Planteamiento: el modelo parte de dos programas de optimización intertemporal, el de los hogares y el de las empresas:

- *Hogares*: maximizan la utilidad esperada intertemporal sujeta a su restricción presupuestaria:

$$\max_{c,\ell,b} U^e = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \, u(c_t, 1-\ell_t) \right], \qquad \beta \in (0,1)$$
$$\text{s.a.:} \begin{cases} c_t + (b_t - b_{t-1}) \leq w_t \ell_t + r_t b_{t-1} \\ b_{-1} \text{ dado} \\ \lim_{T\to\infty} \mathbb{E}_0 \left[b_T / R_T \right] \geq 0 \end{cases}$$

donde c_t es un índice de consumo que agrega todas las variedades, ℓ_t el trabajo, b_t la tenencia de bonos, w_t el salario real, r_t el tipo de interés real de los bonos, $u(\cdot)$ la utilidad instantánea y $\beta \in (0,1)$ el factor de descuento. Los bonos iniciales, b_{-1}, están dados, y la condición de no-Ponzi, con $R_T \equiv \prod_{s=1}^{T}(1+r_s)$, impide endeudarse indefinidamente.

- *Empresas*: cada productor de variedad maximiza el valor esperado descontado de sus beneficios:

$$\max_{P_{i,t}} V^e = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \, \Pi(P_{i,t}, Y_{i,t}, C_t) \right], \qquad \beta \in (0,1)$$

$$\text{s.a.:} \begin{cases} Y_{i,t} = f(\ell_{i,t}) & \text{(tecnología)} \\ Y_{i,t} = \left(\dfrac{P_{i,t}}{P_t} \right)^{-\sigma} C_t & \text{(demanda de la variedad)} \\ P_{i,t} = P_{i,t-1} \text{ con probabilidad } \xi, \qquad \xi \in [0,1) & \text{(restricción de Calvo)} \end{cases}$$

donde $P_{i,t}$ es el precio de la variedad i (la variable de decisión), P_t el nivel general de precios, $Y_{i,t}$ la producción de la variedad, C_t el consumo agregado, $\Pi(\cdot)$ el beneficio del periodo, $f(\ell_{i,t})$ la tecnología, σ la elasticidad de la demanda de la variedad, ξ la probabilidad de no poder reajustar el precio y β el factor de descuento.

CPO:

- *Hogares*: derivando su lagrangiano respecto a c_t, ℓ_t y b_t, se obtienen dos relaciones:

- *Ecuación de Euler del consumo*: $\partial u / \partial c_t = \beta \mathbb{E}_t[(\partial u / \partial c_{t+1})(1 + r_{t+1})]$, que iguala la utilidad marginal del consumo presente con la utilidad marginal esperada del consumo futuro, descontada y capitalizada, y que loglinealizada da lugar a la *IS dinámica*.

- *Condición intratemporal de oferta de trabajo*: $\mathrm{RMS}_{c\,(1-\ell)} = (\partial u/\partial(1-\ell_t))/(\partial u/\partial c_t) = w_t$, que iguala la relación marginal de sustitución consumo-ocio al salario real.

▪ *Empresas*: de la condición de primer orden se obtiene el precio óptimo de reajuste, $P_t^* = P_t^*(\sigma, \mathrm{Cmg}, \mathrm{Cmg}^e)$, que depende de tres elementos:

- *Elasticidad de la demanda*: σ determina el margen deseado sobre el coste marginal, $1+\mu = \sigma/(\sigma-1)$.

- *Coste marginal corriente*: Cmg es el coste marginal nominal de hoy, sobre el que la empresa aplicaría el margen si pudiera ajustar el precio cada periodo.

- *Costes marginales esperados*: Cmg^e recoge la senda esperada de los costes marginales futuros, $\mathbb{E}_t[\mathrm{Cmg}_{t+k}]$ para $k = 1, 2, ...$, ponderados por $(\beta\xi)^k$, es decir, descontados y multiplicados por la probabilidad ξ^k de que el precio fijado hoy siga vigente en $t+k$.

En resumen, la empresa sabe que no podrá modificar el precio durante un tiempo, así que no aplica su margen solo al coste actual sino a una media ponderada del actual y de los que prevé. De ahí que la inflación dependa de las expectativas.

CSO: las condiciones de primer orden de ambos programas caracterizan un máximo, porque la utilidad del hogar es estrictamente cóncava en consumo y ocio y el valor esperado de la empresa es estrictamente cuasicóncavo en su precio[5].

Equilibrio (curva de Phillips neokeynesiana): agregando los precios óptimos con la estructura de Calvo y loglinealizando en torno al estado estacionario, se obtiene la curva de Phillips de la NEK2G, que generalmente adopta dos formas equivalentes:

1. *Inflación como función del coste marginal*: $\pi_t = \beta\mathbb{E}_t[\pi_{t+1}] + \kappa_{\mathrm{cm}}\,\widehat{\mathrm{cmg}}_t$, donde $\widehat{\mathrm{cmg}}_t$ es la desviación del coste marginal real respecto de su nivel de equilibrio.

2. *Inflación como función de la brecha de producción*: dado que el coste marginal se mueve con ella, la forma habitual es:

$$\pi_t = \beta\mathbb{E}_t[\pi_{t+1}] + \kappa\,\tilde{y}_t + \varepsilon_t^{\pi}, \qquad \kappa > 0,\ \beta \in (0,1)$$

donde $\tilde{y}_t$ es la brecha de producción y ε_t^{π} un *shock* de oferta (de costes). Aplicando la ley de Okun ($\tilde{y}_t \approx -\phi(u_t - \bar{u}_t)$), se obtiene la versión en términos de desempleo:

$$\pi_t = \beta\mathbb{E}_t[\pi_{t+1}] - \kappa_u\,(u_t - \bar{u}_t) + \varepsilon_t^{\pi}, \qquad \kappa_u > 0,\ \beta \in (0,1)$$

La curva se diferencia de las curvas de Phillips anteriores en dos aspectos:

▪ *Forward-looking*: la curva es prospectiva, porque la inflación de hoy depende de la inflación esperada mañana, no de la pasada.

▪ *Microfundamentación*: los parámetros de la curva son funciones de los parámetros profundos (p. ej. κ depende de la frecuencia de ajuste de precios). Por eso la curva es robusta a la crítica de Lucas, siempre que esa frecuencia no varíe con el régimen de política.

[5] En el programa del hogar, la restricción es lineal y se exige además la condición de transversalidad sobre los bonos. En el de la empresa, la cuasiconcavidad (y la concavidad en un entorno del óptimo) se debe a que la demanda de la variedad es isoelástica con $\sigma > 1$ y el coste marginal es no decreciente.

Estática comparativa: se distinguen los desplazamientos sobre la curva (variaciones de la brecha de producción o del desempleo cíclico) de los desplazamientos de la curva (cambios en la inflación esperada y *shocks* de oferta ε_t^π).

Extensiones: Galí (2011) revitaliza la curva de Phillips salarial original de Phillips, pero microfundamentada y con expectativas:

$$\pi_t^w = \beta\, \mathbb{E}_t[\pi_{t+1}^w] + \kappa_w\, (\bar{u}_t - u_t)$$

donde π_t^w es la inflación salarial y $\kappa_w > 0$ su sensibilidad a la brecha de desempleo. La importancia de esta aportación es doble:

- *Teórica*: muestra que el modelo de la NEK2G con fijación escalonada de salarios a la Calvo implica una curva de Phillips salarial en términos de desempleo, y cierra así el círculo iniciado por Phillips (1958).

- *Empírica*: la ecuación se ajusta razonablemente bien al comovimiento de la inflación salarial y el desempleo en EE. UU. durante las últimas décadas, incluso con una tasa natural constante.

> La curva de Phillips neokeynesiana se desarrolla en el tema A07, apartados 1.4.1 y 2.2.3 (versión salarial), y su papel en la teoría de la inflación, en el tema A41, apartado 2.4.2.

3.2.3. Conclusiones

Aspectos positivos: la curva de Phillips neokeynesiana deriva la relación inflación-actividad de la fijación escalonada de precios con expectativas racionales, lo que le da microfundamentos y la hace prospectiva. La inflación depende así de la brecha de producción actual y de las esperadas en el futuro. En ausencia de *shocks* de costes ($\varepsilon_t^\pi = 0$), el modelo básico presenta también la "divina coincidencia" (Blanchard y Galí, 2007), por la que estabilizar la inflación equivale a estabilizar la brecha de producción, de modo que no existe conflicto entre ambos objetivos.

Limitaciones: en su versión pura, la curva no genera la persistencia de la inflación observada en los datos, lo que obliga a versiones híbridas con un componente retrospectivo (Galí y Gertler, 1999). Además, su estimación empírica es sensible a la medida de la brecha y de los costes marginales.

3.3. La estimación de la NAIRU y sus límites

Para la política monetaria, la NAIRU es la referencia que separa el paro cíclico, que la demanda puede reducir, del estructural, que no. Por ello, los bancos centrales y los organismos internacionales la estiman de forma sistemática. Los métodos son tres:

- *Estructurales*: la derivan de un modelo de determinación de salarios y precios con las instituciones como variables explicativas.

- *Estadísticos*: la extraen como componente tendencial del desempleo observado mediante filtros.

- *Semiestructurales*: estiman una curva de Phillips con NAIRU variable en el tiempo (Gordon, 1997; Staiger, Stock y Watson, 1997).

La lección de esa literatura es la incertidumbre. De hecho, los intervalos de confianza de la NAIRU estimada abarcan varios puntos porcentuales, sus revisiones son grandes y sus estimaciones en tiempo real difieren mucho de las posteriores. Ello limita su uso como guía de la política monetaria y explica que los bancos centrales la traten como una referencia entre otras y no como un objetivo. En la zona del euro y en España, además, las estimaciones muestran una NAIRU que sube tras cada recesión y baja lentamente después, lo que es coherente con la histéresis. Con todo, ese patrón refleja en parte que los métodos estadísticos y semiestructurales tienden, por construcción, a seguir al desempleo observado.

3.4. La convergencia del pensamiento económico

El recorrido del tema muestra una convergencia. De la nueva macroeconomía clásica la teoría actual conserva las expectativas racionales, la microfundamentación y la idea de que la tasa de desempleo de equilibrio la determinan factores reales. De la nueva economía keynesiana conserva las rigideces nominales y reales que hacen que la demanda agregada afecte al empleo a corto plazo y que el desempleo de equilibrio sea el resultado de mercados imperfectos y no de un mercado walrasiano. El modelo de la NAIRU y la curva de Phillips neokeynesiana son el punto de encuentro, pues combinan una tasa de equilibrio con microfundamentos y una relación inflación-desempleo de corto plazo. La política monetaria no puede explotar esa relación de forma sistemática, pero encuentra en ella margen para estabilizar.

Conclusiones

El análisis del desempleo de equilibrio ha evolucionado desde la dicotomía neoclásico-keynesiana hasta una amplia convergencia. La curva de Phillips, inestable en su formulación original, se reformula al incorporar las expectativas, adaptativas (Friedman y Phelps) o racionales (Lucas), y de ahí nace la tasa natural de paro. El modelo de Layard y Nickell la reinterpreta como NAIRU, la tasa que hace compatibles las aspiraciones de sindicatos y empresas, y la hace depender de instituciones observables. La histéresis completa el cuadro, porque la pérdida de capital humano, la menor acumulación de capital y la dinámica *insiders-outsiders* hacen que esa tasa siga en parte al desempleo observado. En consecuencia, las políticas de oferta reducen esa tasa de forma duradera, pero las de demanda recuperan efectos duraderos cuando hay histéresis, y permanentes solo si esta es pura.

El desempleo de equilibrio reúne los hilos del bloque de economía laboral. La oferta y la demanda de trabajo fijan la referencia sin fricciones (tema A25), las fricciones de búsqueda y emparejamiento dan contenido a los determinantes de la tasa natural (tema A26) y las teorías de la rigidez salarial explican la curva de salarios del modelo de la NAIRU (tema A27). A su vez, la curva de Phillips enlaza el mercado de trabajo con la teoría de la inflación (tema A41) y con la política monetaria del modelo de tres ecuaciones (tema A07). Los desarrollos recientes (Blanchard y Galí, 2010; Galí, 2011) integran las fricciones de búsqueda en el modelo neokeynesiano y recuperan la curva de Phillips salarial, de modo que el desempleo vuelve a ser una variable del modelo y no un residuo.

La Gran Recesión ha reabierto el debate sobre la histéresis. Yagan (2019) muestra que las áreas de EE. UU. más expuestas a la crisis de 2007-2009 tenían todavía en 2015 tasas de empleo menores, y Blanchard, Cerutti y Summers (2015) documentan que muchas recesiones van seguidas de una producción duraderamente inferior a la tendencia previa. También sigue abierta la discusión sobre la pendiente de la curva de Phillips. Hazell, Herreño, Nakamura y Steinsson (2022) estiman una curva muy plana y atribuyen la estabilidad de la inflación al anclaje de las expectativas. En cambio, a raíz de la inflación pospandemia, Benigno y Eggertsson (2023) defienden una curva no lineal, más inclinada cuando el cociente entre vacantes y desempleo es elevado.