1. Fundamentos de la teoría de costes

La **teoría de costes** estudia cómo se traduce la tecnología en términos monetarios: dados un nivel de producción y unos precios de los factores, qué combinación de factores elige una empresa racional y cuánto le cuesta producir.

1.1. Conjunto de elección y premisas

Coste: la economía lo entiende de forma más amplia que la contabilidad. El coste contable recoge los costes explícitos, el implícito es el coste de oportunidad de los recursos propios (el valor de la mejor alternativa a la que se renuncia) y el **coste económico** suma ambos.

Conjunto de elección: es el mismo que el de la teoría de la producción, pero, como se fija el nivel de producción, se restringe al **conjunto de requerimientos de inputs**, las combinaciones de factores que obtienen al menos x, cuya frontera es la isocuanta:

$$V(x) = \{\mathbf{z} \in \mathbb{R}^m_+ \mid f(\mathbf{z}) \geq x\}$$

Supuestos: se pueden resumir en cuatro premisas.

1. *Entorno*: un único producto, en un marco estático y con certidumbre.
2. *Agentes*: la empresa minimiza el coste de cada nivel de producción, condición necesaria para maximizar el beneficio.
3. *Tecnología*: de buen comportamiento, lo que garantiza isocuantas estrictamente convexas y, si los factores son esenciales, soluciones interiores.
4. *Mercado de factores*: la empresa es precio-aceptante, con el vector $\mathbf{p}^z$ dado.

1.2. Gasto en factores e isocostes

Función objetivo: la empresa ordena sus alternativas por el **gasto en factores**, $C(\mathbf{z}) = \mathbf{p}^z \cdot \mathbf{z}$. Es lineal, homogéneo de grado uno en cantidades y en precios, estrictamente creciente en las cantidades, cardinal, medido en unidades monetarias, y observable.

Rectas isocoste: sus curvas de nivel reúnen las combinaciones de igual gasto, $\sum_i p_i^z z_i = C_0$. Con dos factores, su pendiente es el precio relativo, la tasa externa u objetiva a la que el mercado permite sustituir un factor por otro sin alterar el gasto:

$$\left.\frac{dz_2}{dz_1}\right|_{C_0} = -\frac{p_1^z}{p_2^z}$$

Es constante a lo largo de la recta. Un aumento del gasto la desplaza en paralelo hacia fuera, y un cambio del precio relativo la gira en torno a su corte con el eje del factor cuyo precio no varía.

1.3. Tecnología y eficiencia técnica

Restricción: para producir x, la empresa debe elegir una combinación del conjunto de requerimientos, $f(\mathbf{z}) \geq x$. Las isocuantas reúnen las combinaciones **técnicamente eficientes**, y la empresa se sitúa en su tramo decreciente, la región económica, con productividades marginales positivas.

Eficiencia en costes: la eficiencia técnica deja abierta la decisión. En qué isocuanta producir lo decide la teoría de los mercados, y en qué punto de ella, la teoría de costes con los precios de los factores. Al productor le interesan los puntos que usan más del factor relativamente barato, cuyo precio relativo es menor que su productividad marginal relativa. La eficiencia en costes pide una asignación técnicamente eficiente y, además, de coste mínimo.

1.4. Equilibrio y función de costes

La decisión se plantea como un problema primal, la máxima producción con un gasto dado, o como su dual, el mínimo coste de una producción dada. Sus soluciones, relacionadas por dualidad, dan las demandas de factores y la función de costes.

1.4.1. Primal: maximización de la producción

Planteamiento: la empresa maximiza la producción con un gasto dado en factores:

$$\max_{\mathbf{z}} \; x = f(\mathbf{z}) \quad \text{s.a.:} \quad \sum p_i^z z_i = C_0$$

– El lagrangiano es $\mathcal{L} = f(\mathbf{z}) + \lambda(C_0 - \sum p_i^z z_i)$, y, por el teorema de la envolvente, el multiplicador, $\lambda = \partial\mathcal{L}/\partial C_0 = df^*/dC$, es el producto marginal del coste, la producción adicional de un euro más gastado en factores. Si ambos problemas coinciden, $\lambda = 1/\text{Cmg}$.

CPO: la tasa interna a la que la tecnología permite intercambiar los factores, la RMST, iguala la tasa externa, el precio relativo. Reordenando, la productividad marginal por euro es la misma en todos los factores e igual a λ:

$$\text{RMST}_{ji} = \frac{f_i'}{f_j'} = \frac{p_i^z}{p_j^z} \;\Leftrightarrow\; \frac{f_i'}{p_i^z} = \lambda \quad \forall i$$

CSO: la CPO es suficiente si las isocuantas son estrictamente convexas, porque la restricción es lineal.

Solución: las **demandas marshallianas de factores**, $z_i^m = d_i(\mathbf{p}^z, C)$, son homogéneas de grado cero en precios y coste, agotan el presupuesto, $\sum_i p_i^z z_i^m = C$, y son observables. Respecto a su precio tienen pendiente indeterminada, porque recogen el efecto producción y el efecto sustitución. El valor óptimo del primal es la **función indirecta de producción**, $x^* = \phi(\mathbf{p}^z, C)$, cardinal, creciente en el coste y no creciente en los precios, de la que las demandas se obtienen con una identidad análoga a la de Roy:

$$z_i^m = -\frac{\partial\phi(\mathbf{p}^z, C)/\partial p_i^z}{\partial\phi(\mathbf{p}^z, C)/\partial C}$$

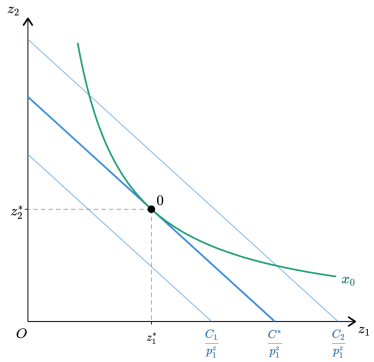
1.4.2. Dual: minimización del coste

Planteamiento: la empresa minimiza el coste de obtener un nivel de producción dado:

$$\min_{\mathbf{z}} \; C = \sum p_i^z z_i \quad \text{s.a.:} \quad f(\mathbf{z}) = x_0$$

CPO y CSO: el multiplicador, $\mu = 1/\lambda$, es el coste marginal. De nuevo, la RMST iguala el precio relativo de los factores, y la isocuanta x_0 es tangente a la isocoste más baja. Cualquier otra combinación técnicamente eficiente sería más cara. La convexidad de las isocuantas garantiza un mínimo. Formalmente:

$$x_0 = f(\mathbf{z}^*), \qquad f_i' = \frac{p_i^z}{\mu} \;\Rightarrow\; \text{RMST}_{ji} = \frac{p_i^z}{p_j^z}$$



Equilibrio de la empresa: tangencia de la isocuanta con la isocoste más baja.

Solución: las **demandas condicionadas de factores**, $z_i^c = d_i^c(\mathbf{p}^z, x_0)$, y, como valor óptimo, la función de costes, $c(\mathbf{p}^z, x_0)$. Las demandas tienen estas propiedades:

– Son homogéneas de grado cero en los precios: una subida equiproporcional no altera la combinación de mínimo coste.
– Su efecto sustitución propio es no positivo, y el cruzado, indeterminado. Al ser condicionadas, no recogen el efecto producción. Con la producción crecen para los factores normales y decrecen para los inferiores.
– Son el gradiente de la función de costes, por el lema de Shephard, y observables, porque la producción que las condiciona es medible, a diferencia de la utilidad en las hicksianas.

1.4.3. Relaciones de dualidad

Los planes óptimos de ambos problemas son idénticos si se intercambia la información, $C_0 = C^*$ o $x_0 = x^*$. Las demandas marshallianas coinciden con las condicionadas evaluadas en la producción máxima, y estas con aquellas evaluadas en el coste mínimo. La función de costes y la indirecta de producción son inversas. En símbolos:

$$d_i^c(\mathbf{p}^z, x_0) = d_i(\mathbf{p}^z, c(\mathbf{p}^z, x_0)), \qquad \phi(\mathbf{p}^z, C) = c^{-1}(\mathbf{p}^z, C)$$

1.4.4. Función de costes y sus curvas

Definición: la **función de costes**, valor del dual, da el coste mínimo de producir x a los precios dados. Análoga a la función de gasto del consumidor, resume toda la información relevante de la tecnología:

$$C^* = c(\mathbf{p}^z, x) = \sum_{i=1}^{m} p_i^z \cdot d_i^c(\mathbf{p}^z, x)$$

Propiedades:

– Es homogénea de grado uno en los precios, estrictamente creciente en la producción y no decreciente en los precios. En símbolos, $c(\lambda\mathbf{p}^z, x) = \lambda \cdot c(\mathbf{p}^z, x)$, y el Cmg y el Cme son también homogéneos de grado uno en precios.
– Es **cóncava en los precios**, es decir, si sube un precio, el coste aumenta menos que con cantidades fijas, por la sustitución. Por eso, la matriz de sustitución es semidefinida negativa y simétrica.
– Por el **lema de Shephard**, sus derivadas respecto a los precios son las demandas condicionadas, $z_i^c = \partial c(\mathbf{p}^z, x)/\partial p_i^z$.
– Por integrabilidad, toda función continua, homogénea de grado uno, no decreciente y cóncava en precios es la función de costes de alguna tecnología. Además, es observable, a diferencia de la función de gasto, que depende de una utilidad no observable.

Coste marginal y medio: el Cmg es la pendiente del coste total, y el Cme, la del rayo desde el origen. El marginal corta al medio en su mínimo.

Economías de escala: son caídas del coste medio al aumentar la escala. Según Marshall, las internas dependen del tamaño de la empresa y dan forma a su curva de coste medio a largo plazo, y las externas dependen del tamaño de la industria y la desplazan. Pueden ser estáticas o dinámicas, y reales o pecuniarias. Las mide la elasticidad-producto del coste, inversa de la elasticidad de escala en la combinación de mínimo coste:

$$\varepsilon_{Cx} = \frac{\text{Cmg}}{\text{Cme}} = \frac{1}{\varepsilon_{x\lambda}}$$

Si es menor que uno, el marginal está por debajo del medio y hay economías de escala; si es mayor, deseconomías. Con precios dados, los rendimientos crecientes equivalen localmente a economías de escala, aunque, si la función no es homotética, puede haber economías de escala sin rendimientos crecientes sobre un radio-vector fijo, porque la empresa cambia la mezcla de factores al crecer.

Precios: la elasticidad del coste respecto al precio de un factor es su participación en el coste, $s_i = p_i^z z_i^c / c$. La **elasticidad de sustitución parcial de Allen** (1938), $\sigma_{ij}^P = c \cdot c_{ij}/(c_i \cdot c_j) = \varepsilon_{z_i p_j}/s_j$, la elasticidad cruzada de la demanda condicionada dividida por la participación de j, mide la respuesta de la demanda de un factor al precio de otro y permite que varíe el empleo de los demás. Es simétrica y clasifica los factores en sustitutivos, si es positiva, y complementarios, si es negativa, caso solo posible con más de dos factores. Con dos, coincide con la de Hicks y es no negativa, $\sigma_{KL}^P = \Delta\,\%(K/L)/\Delta\,\%(w/r) \geq 0$.

En suma, la empresa iguala la RMST al precio relativo de los factores tanto si maximiza la producción como si minimiza el coste, y ambos problemas llevan a la misma combinación. La función de costes reúne la información de la tecnología que importa a la empresa.

2. Teoría neoclásica de los costes

El horizonte temporal determina qué factores pueden alterarse: a corto plazo hay algún factor fijo; a largo, todos son variables, y a muy largo, también la tecnología. La estática comparativa separa los cambios del objetivo de producción, que trazan la senda de expansión, de los de los precios de los factores.

2.1. Corto plazo: proporciones variables

Con al menos un factor fijo, hay **costes fijos**, independientes del volumen y, si son irrecuperables, hundidos, y **costes variables**, que dependen de él.

2.1.1. Análisis gráfico

Costes totales: el fijo es una recta horizontal. El variable tiene, con costes cúbicos, forma de S tumbada e invertida, y crece primero a ritmo decreciente y, desde el punto de inflexión, donde la productividad marginal es máxima, a ritmo creciente. El total es el variable desplazado hacia arriba en el coste fijo.

Costes unitarios:

- El coste fijo medio decrece, porque el coste fijo se diluye con la producción.
- El **coste marginal** es inversamente proporcional a la productividad marginal del factor variable, $\text{Cmg} = w/\text{Pmg}_L$, y tiene forma de U: decrece mientras los rendimientos marginales son crecientes y crece después, pasando por los mínimos del coste variable medio y del coste medio.
- El coste variable medio decrece hasta el **mínimo de explotación**, donde se alcanza el óptimo técnico, la productividad media máxima.
- El coste medio, suma del variable medio y del fijo medio, tiene forma de U, y su mínimo es el **óptimo de explotación**, que, por la dilución del coste fijo, está a la derecha del mínimo de explotación.

2.1.2. Limitaciones a la optimalidad

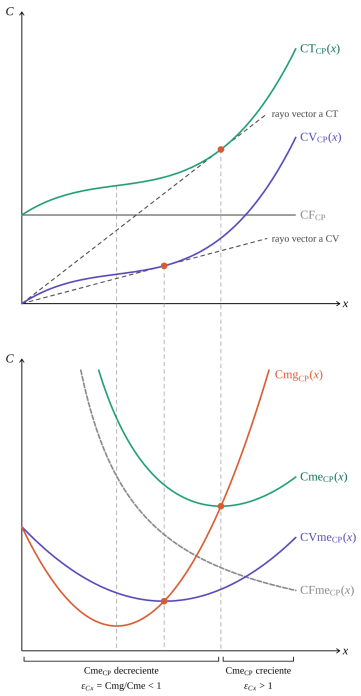
El factor fijo impide cumplir $\text{RMST}_{KL} = w/r$ salvo en la producción para la que la planta es óptima, y la empresa usa combinaciones subóptimas, con exceso o insuficiencia de capital. Los efectos dependen del contrato: con un **arrendamiento**, paga solo las unidades que usa, si el factor es divisible, y mientras no agote la capacidad elige la mezcla como a largo plazo, y con la **propiedad** paga con independencia del uso, de modo que $r\bar{K}$ es un coste fijo.

2.1.3. Estática comparativa

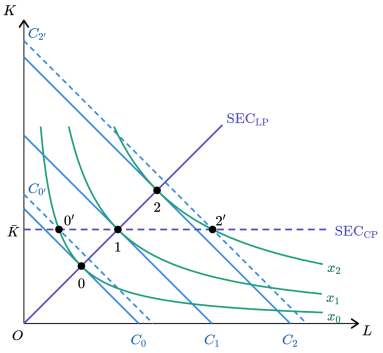
Producción: con el capital fijo, la senda de expansión a corto plazo es la recta $K = \bar{K}$, que solo corta a la de largo plazo donde la planta es óptima. En los demás niveles, el coste a corto plazo supera al de largo. Un cambio del objetivo de producción es un movimiento a lo largo de las curvas. El coste a corto plazo, con $L(x;\bar{K})$ el trabajo que exige la isocuanta de cada nivel, es:

$$\text{CT}_{\text{CP}}(x) = w\,L(x;\bar{K}) + r\bar{K}$$

Precios: los efectos dependen del factor encarecido.



Costes a corto plazo: totales (arriba) y unitarios (abajo).



Factor fijo y condición de tangencia.

- Si suben todos los precios, las curvas se desplazan proporcionalmente y la mezcla no cambia.
- Si sube un **coste fijo** no evitable, como un alquiler o un impuesto de suma fija, suben el coste fijo medio y el medio, pero no el variable medio ni el marginal, de modo que la oferta a corto plazo no cambia y solo cae el beneficio. Un impuesto sobre los beneficios tampoco la altera.
- Si sube un **coste variable**, como el salario o un impuesto específico, suben el variable medio, el medio y el marginal, y la empresa produce menos. Con arrendamiento y capacidad sin agotar, la empresa sustituye trabajo por capital, las curvas suben menos que en proporción y la mezcla cambia según la curvatura de la isocuanta. Con propiedad, el coste variable medio y el marginal se multiplican por w_1/w_0 y la mezcla no cambia.

2.2. Largo plazo: dimensión óptima de planta

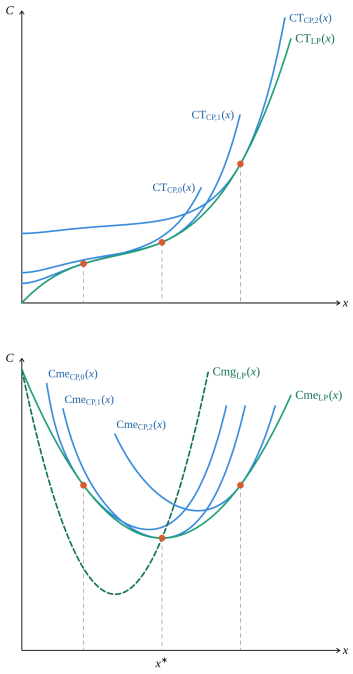
Con todos los factores ajustables, la empresa elige la planta de menor coste. Mientras no ha invertido, los costes a largo plazo son de planificación, *ex ante*, y no hay costes fijos.

2.2.1. Análisis gráfico

Envolventes: como el coste a corto plazo nunca es inferior al de largo, el coste total a largo plazo es la envolvente de los de corto. El coste medio a largo plazo es también la envolvente de los de corto, pero no pasa por sus mínimos salvo en el suyo propio, porque, con su forma de U, cada curva de corto lo toca cuando aún decrece, a la izquierda del mínimo, o cuando ya crece, a la derecha. El marginal a largo plazo no es envolvente, porque el Cmg_{CP} puede ser mayor o menor que él. Se obtiene proyectando en vertical las tangencias entre cada Cme_{CP} y el Cme_{LP}.

Dimensión óptima de planta: se alcanza donde $\text{Cme}_{\text{CP}} = \text{Cmg}_{\text{CP}} = \text{Cmg}_{\text{LP}} = \text{Cme}_{\text{LP}}$, el tamaño que lleva el coste medio a largo plazo a su mínimo, sin economías de escala por aprovechar. Las plantas menores tienen capacidad ociosa, y las mayores trabajan sobrecargadas, con deseconomías de escala. En ella $\varepsilon_{Cx} = 1$,

y la sobrecarga se traduce en horas extra, velocidad antieconómica y averías.



Costes a largo plazo: envolvente y dimensión óptima de planta.

2.2.2. Estática comparativa a largo plazo

Producción: la **senda de expansión del coste** une las tangencias entre isocuantas e isocostes al crecer la producción, con precios relativos constantes. Recorrerla es recorrer la función de costes a largo plazo. Es una recta por el origen si la función es homotética y tiene pendiente negativa si algún factor es regresivo. Un factor es normal si su demanda condicionada crece con la producción e inferior, o regresivo, si decrece, pero no todos pueden ser inferiores, porque el coste mínimo crece con la producción. El factor es normal si $\varepsilon_{z_i x} = (\partial z_i^c/\partial x)(x/z_i^c) > 0$, y con tecnología homotética todas estas elasticidades son iguales a ε_{Cx}. La restricción sobre los inferiores es:

$$\sum_i p_i^z \frac{\partial z_i^c}{\partial x} = \text{Cmg} > 0$$

Precios: si sube el precio de un factor, la isocoste gira y la tangencia se desplaza a lo largo de la isocuanta hacia menos de ese factor, y el coste sube según su participación. Los efectos cruzados son simétricos. Con dos factores, estos son siempre sustitutivos. Un aumento permanente de un coste cuasifijo, independiente del volumen, eleva la curva de coste medio y traslada su mínimo a la derecha; el de un coste variable la eleva con efecto indeterminado sobre su mínimo.

Teorema de Le Châtelier-Samuelson (1947): en la producción para la que la planta es óptima, el coste marginal a corto plazo tiene más pendiente que el de largo, y la oferta a largo plazo es más elástica, porque es tanto más rígida cuantos más factores se mantienen fijos.

2.3. Muy largo plazo: progreso técnico

Idea: las mejoras tecnológicas desplazan las curvas de costes hacia abajo. Con rendimientos constantes y progreso neutral en el sentido de Hicks, $x = A(t)\, f(K, L)$, cada unidad de producto exige $1/A(t)$ unidades de f, y el coste medio cae al ritmo del progreso técnico, $g_{\text{Cme}} = -g_A$:

$$\text{Cme}_{\text{LP}}(t) = \text{Cmg}_{\text{LP}}(t) = \frac{c(w,r)}{A(t)}$$

2.4. Extensiones neoclásicas

Frente a los rendimientos mixtos, los casos polares, con tecnología homogénea de grado h y $c(\mathbf{p}^z, x) = x^{1/h} c(\mathbf{p}^z, 1)$, cambian el perfil del coste a largo plazo, y las economías externas lo trasladan.

Rendimientos constantes: el coste total es una recta por el origen, y el medio y el marginal, una misma recta horizontal, sin economías ni deseconomías de escala, $\varepsilon_{Cx} = 1$.

Rendimientos decrecientes: el coste total es convexo, y el medio y el marginal son crecientes, con el marginal siempre por encima y deseconomías de escala, $\varepsilon_{Cx} > 1$.

Rendimientos crecientes: es el caso que Sraffa (1926) opuso a la forma de U. El coste medio y el marginal decrecen, la regla $p = \text{Cmg}$ genera pérdidas y no se cumple la CSO del problema competitivo, por lo que los rendimientos crecientes son incompatibles con la competencia perfecta y conducen al monopolio natural. El coste total es cóncavo, el medio queda siempre por encima del marginal y $\varepsilon_{Cx} < 1$.

Economías externas (Marshall, 1890): si la oferta de factores no es perfectamente elástica o hay efectos externos tecnológicos, la producción de la industria altera los costes de cada empresa.

- Con **deseconomías externas**, como un factor escaso o la congestión, las curvas suben y la oferta de la industria a largo plazo es creciente.
- Con **economías externas**, como la difusión de conocimientos entre empresas próximas, bajan y la oferta es decreciente. Fundamentan la protección de la industria naciente y la promoción de clústeres.

Permiten una oferta de la industria creciente o decreciente sin renunciar a curvas en U en cada empresa, pero son difíciles de identificar y medir, y Sraffa (1926) las consideró poco frecuentes.

En suma, con rendimientos mixtos hay curvas en U y una dimensión óptima de planta. Pero Sraffa (1926) señaló la incoherencia del esquema con el equilibrio parcial en competencia perfecta: las economías de escala internas solo son compatibles con la competencia imperfecta, en la que la demanda limita el tamaño, respuesta que llegó con Chamberlin (1933) y Robinson (1933).

3. Desarrollos posteriores

Las aportaciones posteriores critican la forma de U y la minimización, relacionan el coste con la empresa y la estructura del mercado, introducen el tiempo y estiman las curvas.

3.1. Críticas a la curva neoclásica

3.1.1. Reserva de capacidad (Stigler, 1939)

Idea: George Stigler cuestiona una dimensión óptima única. La empresa mantiene una **reserva de capacidad** ante imprevistos, como las oscilaciones de la demanda, y por razones técnicas, como las indivisibilidades.

Corto plazo: el coste variable medio tiene forma de bañera, con tres tramos que reflejan la productividad del factor variable.

1. Exceso de capacidad, con productividad media creciente y coste variable medio decreciente. Su elasticidad-producto, $\varepsilon_{\text{CV}x} = \text{Cmg}_{\text{CP}}/\text{CVme}_{\text{CP}}$, es menor que uno.
2. *Reserva de capacidad*: coste variable medio y marginal coinciden y son horizontales, con $\varepsilon_{\text{CV}x} = 1$. Es la zona de uso normal, cuya extensión se fija con el **factor de carga** o coeficiente de aprovechamiento, cociente entre la producción efectiva y la de plena capacidad.
3. Sobrecarga, con el factor fijo plenamente utilizado y coste variable medio creciente. Aquí $\varepsilon_{\text{CV}x} > 1$.

Largo plazo: el coste medio deja de ser la envolvente y pierde su mínimo único. Los costes de producción caen de forma continua, y los administrativos se reducen hasta cierta dimensión y, en las plantas muy grandes, repuntan con mucha lentitud. Si las economías de escala no se agotan, el coste medio baja siempre, en forma de L. Si se agotan, como sostiene Joe Bain, baja hasta la **escala mínima eficiente** y luego es plano, con infinitos puntos de coste mínimo.

Conclusiones: estas formas se parecen más a las curvas observadas, y la reserva de capacidad es una decisión optimizadora, no una ineficiencia. Pero, sin mínimo único, el tamaño de la empresa y el número de empresas quedan indeterminados.

3.1.2. Ineficiencia-X (Leibenstein, 1966)

Idea: Harvey Leibenstein cuestiona que la empresa produzca al mínimo coste. Con información incompleta, contratos incompletos, trabajo penoso y competencia imperfecta, unos mismos *inputs* dan *outputs* distintos. Sin presión de la competencia ni de la supervisión, el esfuerzo cae y se pierden recursos, la **ineficiencia-X**, mayor cuanto más débil es la competencia.

Conclusiones: explica costes distintos con igual tecnología y precios, y añade a la pérdida asignativa del monopolio una pérdida en costes. Se le objeta que, si el control se lleva hasta que su coste marginal iguala su beneficio marginal, no existiría.

3.2. Coste, empresa y estructura de mercado

3.2.1. Coase (1937) y Williamson (1975)

Idea: para Coase, la empresa es una estructura de contratos, y recurrir al mercado tiene **costes de transacción**, que Williamson vincula a la racionalidad limitada, el oportunismo y la especificidad de los activos.

Razonamiento: la dimensión óptima de la empresa se alcanza cuando el coste marginal de organizar internamente una transacción, creciente con el tamaño, iguala el de realizarla en el mercado.

Conclusiones: explica los límites de la empresa y la integración vertical, pero el concepto es amplio y difícil de medir.

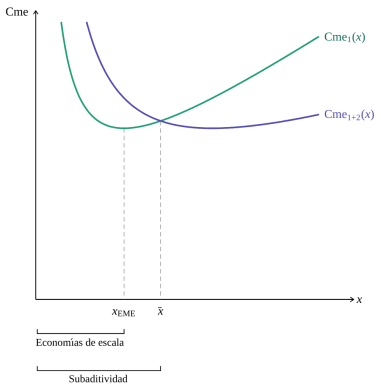
3.2.2. Subaditividad (Baumol, Panzar y Willig, 1982)

Idea: una función de costes es **subaditiva** en x si ningún reparto de x entre varias empresas de igual tecnología es más barato que una sola. Si lo es estrictamente, el mercado es un **monopolio natural**:

$$\text{CT}(x) < \sum_{i=1}^{k} \text{CT}(x_i), \quad \sum_i x_i = x$$

Relación con las economías de escala: en el caso monoproducto, son suficientes, pero no necesarias, y en el multiproducto, ni siquiera suficientes.

- Si el coste medio decrece hasta x, cada parte del reparto tiene un coste medio mayor y la función es subaditiva. En símbolos, $\mathrm{CT}(x) = x \cdot \mathrm{Cme}(x) < \sum_i x_i \cdot \mathrm{Cme}(x_i) = \sum_i \mathrm{CT}(x_i)$.
- Pero puede ser subaditiva aunque el coste medio ya crezca, si la escala mínima eficiente es grande respecto al mercado, porque es una propiedad global, y las economías de escala, local. Gráficamente, el coste medio de dos empresas que se reparten x a partes iguales queda por encima del de una hasta $\bar{x}$, aunque entre x_{EME} y $\bar{x}$ la empresa única esté ya en deseconomías.
- La condición necesaria es alguna no convexidad, como costes fijos o hundidos, indivisibilidades o economías de alcance. Sin ninguna, con costes convexos y $\mathrm{CT}(0) = 0$, la función es superaditiva y la producción se reparte eficientemente entre muchas empresas.



Subaditividad: una empresa es más barata que dos hasta $\bar{x}$.

Conclusiones: define el monopolio natural con una condición exacta, válida para la empresa multiproducto, y da a la regulación un criterio riguroso, pero comprobarla exige conocer el coste de repartos que no se observan.

3.2.3. Economías de alcance (Panzar y Willig, 1977; 1981)

Idea: hay **economías de alcance** cuando producir varios bienes en una empresa cuesta menos que hacerlo en empresas especializadas, y la eficiencia viene de la variedad, no del volumen. Nacen de la dependencia técnica entre productos, de factores indivisibles, de *inputs* públicos internos, como una marca o una red, y de compartir la organización. Su grado mide el ahorro de la producción conjunta:

$$S_C = \frac{\mathrm{CT}(x_1, 0) + \mathrm{CT}(0, x_2) - \mathrm{CT}(x_1, x_2)}{\mathrm{CT}(x_1, x_2)}$$

Ni las economías de escala ni las de alcance implican las otras. Las de alcance son necesarias, pero no suficientes, para la subaditividad multiproducto, y la complementariedad de costes, que aumentar la producción de un bien reduzca el coste marginal del otro, $\partial^2 \mathrm{CT}/\partial x_1 \partial x_2 < 0$, las implica y, con costes fijos comunes y costes marginales propios no crecientes, garantiza la subaditividad. Con $S_C = 0{,}25$, producir por separado costaría un 25 % más.

Conclusiones: explican la diversificación y juzgan la estructura de las industrias reguladas: con S_C alto, separar actividades destruye eficiencia. Pero medirlas exige costes de empresas especializadas que a menudo no existen.

3.2.4. Costes hundidos (Baumol, Panzar y Willig, 1982)

Idea: un coste es **hundido** si no puede recuperarse al abandonar la actividad, como una vía férrea, la publicidad o la I+D, a diferencia del fijo recuperable. Es irrelevante una vez incurrido, pero decisivo antes.

Razonamiento: antes de entrar, es una barrera a la entrada y, para la instalada, de salida. Después, hace creíble el compromiso de permanecer y de luchar. Por eso separa el monopolio natural de la impugnabilidad: sin costes hundidos, la competencia potencial disciplina al monopolio. Con ellos, la regulación recupera su papel.

Conclusiones: precisa qué costes condicionan la entrada y la salida, pero la parte hundida de un coste depende del horizonte y de los mercados de segunda mano, de modo que la distinción es de grado.

3.3. Coste y tiempo

3.3.1. Aprendizaje (Wright, 1936; Henderson, 1968)

Idea: Wright observó en la aeronáutica que las horas por avión caían con la producción acumulada, y Henderson y el Boston Consulting Group lo generalizaron como **curva de experiencia**. Las economías de aprendizaje, o de escala dinámicas, ligan el coste medio a la producción acumulada, con b la elasticidad de aprendizaje:

$$\mathrm{Cme}_t = a \cdot (x_t^{\mathrm{ac}})^{-b}$$

Producir hoy abarata el futuro, por lo que la empresa produce más de lo que igualaría el ingreso y el coste marginales actuales, y el primero en entrar gana una ventaja acumulativa. Explica los precios de penetración y las barreras basadas en la experiencia, el aprendizaje por la práctica de Arrow (1962) en el crecimiento endógeno y el argumento de la industria naciente. Es una regularidad bien documentada, pero empírica: no explica la tasa de aprendizaje y cuesta separarla de las economías de escala estáticas y del progreso técnico. Cada vez que se duplica la producción acumulada, el coste medio cae en la proporción $1 - 2^{-b}$, la tasa de aprendizaje.

3.3.2. Costes de ajuste

Idea: son los costes de alterar la velocidad a la que se incorporan los factores. Eisner y Strotz (1963) introducen los internos, como las mermas de producción al cambiar máquinas, que Lucas (1967) formaliza, y la literatura posterior añade los externos, como las indemnizaciones o la selección de personal.

Forma: la función de costes de ajuste suele suponerse convexa, como $\psi(\Delta z) = \frac{\gamma}{2}(\Delta z)^2$, aunque puede ser asimétrica, lineal o fija. Los factores variables y fijos tradicionales son los casos de costes de ajuste nulos o infinitos, y con costes convexos el ajuste se reparte en varios periodos. Fundamentan la teoría q de la inversión, aunque la evidencia de ajustes a saltos apunta a costes fijos o lineales. La lineal es $\psi(\Delta z) = \gamma \cdot |\Delta z|$, con un ajuste que puede ser inmediato, y la fija, $\psi = \bar{\psi}$ si $\Delta z \neq 0$, como un expediente de regulación de empleo.

3.4. Aplicaciones empíricas

La forma de las curvas de coste es, en último término, una cuestión empírica, de la que dependen el número de empresas que caben en un mercado, la existencia de monopolios naturales y el diseño de la regulación.

3.4.1. La función de costes estimada

Idea: Nerlove (1963) estima una función de costes deducida por dualidad, con los precios de los factores como explicativas, para 145 eléctricas de EE. UU. en 1955. En una industria regulada la producción y los precios son exógenos, y con una Cobb-Douglas el coeficiente de $\ln x$ es $1/h$.

Resultados: con toda la muestra halla rendimientos crecientes, que por grupos de tamaño caen con la escala, de modo que la Cobb-Douglas era demasiado rígida. La translog (Christensen et al., 1973) deja libres los rendimientos y las elasticidades de sustitución, y con ella Christensen y Greene (1976) muestran que en 1970 el grueso de la electricidad se generaba en el tramo plano del coste medio.

3.4.2. Fronteras de costes

Idea: la función estimada supone que todas las empresas minimizan el coste. Las fronteras abandonan ese supuesto, de modo que la función de costes es el mínimo y cada empresa puede quedar por encima. La distancia es la eficiencia económica de Farrell, que mide la ineficiencia-X.

Métodos: son los de la producción, ahora desde el coste. El **SFA** añade a la función de costes el error compuesto, $\ln C_i = \ln c(\mathbf{p}_i^z, x_i) + v_i + u_i$, con la ineficiencia $u_i \geq 0$ sumando, porque encarece. El **DEA** calcula con un programa lineal el coste mínimo de cada empresa a los precios que paga.

Conclusiones: miden cuánto podría ahorrar cada empresa y sirven a la regulación por comparación, pero no siempre ordenan igual a las mismas empresas.

3.4.3. Evidencia y problemas de medición

Evidencia: el coste medio a largo plazo cae con fuerza en escalas pequeñas y luego es casi horizontal, en forma de L más que de U. La escala mínima eficiente suele ser pequeña respecto al mercado, salvo en las industrias de red y de grandes costes fijos.

Cautelas: la **falacia de la regresión** (Friedman, 1955) hace parecer decreciente la curva, porque las empresas que producen por encima de lo normal muestran costes medios bajos. Cada observación es un punto de una curva de corto plazo, y los costes contables omiten el coste de oportunidad del capital propio.