Universidad de Oviedo

Universidad de León

INTRODUCCIÓN AL ANÁLISIS EMPÍRICO DE LA PRODUCCIÓN

Antonio Álvarez Pinilla
Universidad de Oviedo
Departamento de Economía

Carlos Arias Sampedro
Universidad de León
Departamento de Economía

Luis Orea Sánchez
Universidad de Oviedo
Departamento de Economía

MAYO 2003

ÍNDICE

PREFACIO

El trabajo econométrico aplicado es una tarea complicada en la que no es suficiente dominar los conocimientos que se adquieren en un libro de texto de econometría. El origen de este libro tiene lugar en las múltiples preguntas que han asaltado a los autores durante dos décadas de actividad en el campo de la microeconometría aplicada. En una gran parte de los casos, la respuesta a esas preguntas no pudo encontrarse en un libro de econometría y hubo que buscarla en artículos, en múltiples consultas a pacientes colegas o usando el conocido método de “prueba y error”. De esa experiencia surge este libro cuyo objetivo es servir de puente entre la econometría teórica y los trabajos aplicados en economía de la producción.

Debe quedar claro que no hemos intentado hacer un libro de recetas, que contenga todos los ingredientes y las proporciones necesarias para cada "plato". Al contrario, el libro sólo pretende hacer más fácil el camino a los estudiantes interesados en el campo de la microeconomía aplicada. Al mismo tiempo, hemos pretendido hacer lo más atractiva posible la lectura del libro (en la medida que la lectura de la econometría aplicada puede ser atractiva). Por este motivo, todos los capítulos están salpicados de referencias históricas acerca de cuándo se descubrió qué o de quién opina esto o lo otro, con el fin de que el lector se sumerja dentro del ámbito de la econometría aplicada.

La conjunción de teoría y práctica es la aportación mas importante del libro, que sin duda tiene antecedentes en otros libros como los de Berndt (1993), Thomas (1993), Chambers (1988), o en otros textos de econometría aplicada ya mas lejanos, como los de Desai (1976) y Bridge (1971).

El libro va dirigido a estudiantes que quieren hacer investigación aplicada en economía de la producción. El libro puede usarse como texto en un curso avanzado de microeconomía aplicada, que de hecho fue la idea inicial bajo la que fue concebido, o también como manual auxiliar de un curso introductorio de econometría.

Por ultimo, el libro es limitado en su alcance ya que no pretende ser un compendio exhaustivo de todo el conocimiento en el área de la economía de la producción aplicada. En primer lugar, el enfoque es paramétrico y econométrico, quedando sin cubrir tanto la amplia literatura del análisis no paramétrico de la producción como la del análisis paramétrico por medio de técnicas de programación matemática. Asimismo, son muchos los temas de econometría que no tienen cabida en el texto.

El capítulo de agradecimientos es muy amplio, dado que son muchas las personas que han ayudado a la elaboración de este libro. En primer lugar, es necesario dar las gracias a unas cuantas generaciones de alumnos de la asignatura de Economía Industrial II de la Universidad de Oviedo, los cuales han contribuido a corregir sucesivas versiones de apuntes de clase. Asimismo, estamos en deuda con aquellos colegas que han leído el documento:.

Por último, es necesario reconocer que este libro no es fruto exclusivo del esfuerzo personal. En este sentido, un capítulo especial lo merecen aquellas personas que nos han enseñado intuición económica y, fundamentalmente, a combinar rigor con sencillez.

NOMENCLATURA

Subíndices

i	: subíndice para observaciones (i=1...N)
j,k	: subíndices para inputs variables (j,k=1...J)
h	: subíndice para inputs fijos (h=1...H)
l,m	: subíndice para outputs (l,m=1...M)
t	: tiempo (t=1...T)

Variables

y_m	: outputs
x_j	: inputs variables
z_h	: inputs fijos
p_m	: precios de los outputs
w_j	: precios de los inputs variables
r_h	: precios de los inputs fijos
C	: costes
π	: beneficios
S_j	: participaciones en los costes (“cost shares”)
e_j	: elasticidades-output
e	: rendimientos a escala o elasticidad de escala
ε_{Cy}	: elasticidad de tamaño
u,v	: perturbaciones aleatorias

Funciones

$f(\odot)$	: función de producción
$C(\odot)$	: función de costes
$\pi(\odot)$	: función de beneficios

CAPÍTULO 1

EL ANÁLISIS ECONOMÉTRICO APLICADO

En este capítulo se introduce al lector a los principales aspectos del análisis econométrico aplicado. Después de un breve repaso histórico sobre el origen y evolución de la econometría, se discute con cierta amplitud las diferencias entre los modelos económicos y los modelos econométricos. Por último, se hacen algunas reflexiones sobre la organización del trabajo empírico y sobre el uso de ordenadores.

1.1. ¿Qué es la econometría?

Schumpeter (1934) tiene en su conocida obra *Historia del Análisis Económico* un capítulo titulado "Los económetras y Turgot" en el que remonta el origen de la econometría a algunos trabajos realizados principalmente por economistas del siglo XVII. Al principio del capítulo dice: "Los individuos ... en este capítulo ... tienen en común... el espíritu del análisis numérico. Todos ellos han sido económetras. Su obra ilustra realmente a la perfección qué es la econometría y qué intentan hacer los económetras".[1]

A pesar de la afirmación de Schumpeter y aunque Engel y algunos autores neoclásicos pueden considerarse como precursores, la econometría, tal y como se concibe hoy en día, es una disciplina relativamente joven. Los primeros trabajos de naturaleza verdaderamente econométrica aparecieron en el primer tercio del siglo XX, destacando entre otros los de Moore (1914), Working (1927), Cobb y Douglas (1928), Schultz (1928) y Waugh (1928).[2]

La fundación de la Econometric Society en 1930 y la publicación de la revista *Econometrica* en 1933 fueron dos grandes hitos en la historia de la econometría. Frisch,[3] primer editor de la revista, expone en el primer editorial cuál es su principal objetivo: "...promover estudios dirigidos a la unificación de las aproximaciones teórico-cuantitativa y empírico-cuantitativa a los problemas económicos y que estén imbuidos de un razonamiento constructivo y riguroso, similar al que ha venido a dominar en las ciencias naturales".

Más adelante, Frisch pasa a explicar en qué consiste la econometría: "...la econometría no es lo mismo que la estadística económica. Tampoco es idéntica a lo que llamamos teoría económica general, aunque una parte considerable de esta teoría tiene un carácter cuantitativo. Tampoco debería la econometría ser tomada como sinónimo de la aplicación de las matemáticas a la economía. La experiencia ha demostrado que cada uno de esos tres puntos de vista, el de la estadística, el de la teoría económica y el de las

[1] Schumpeter estudia en ese capítulo, entre otros, la obra de Petty, Cantillon y Quesnay.

[2] Los primeros trabajos econométricos no contaban con el cuerpo teórico del que se dispone hoy en día. De hecho, muchos de los aspectos más sencillos de los modelos econométricos fueron forjándose poco a poco. Una exposición del desarrollo histórico de los modelos formales en econometría puede verse en el libro de Morgan (1990).

[3] Frisch recibió en 1969 el primer Premio Nobel de Economía (junto con Jan Tinbergen) por su importante contribución al desarrollo de la econometría.

matemáticas, es una condición necesaria pero no suficiente por sí misma para el entendimiento real de las relaciones cuantitativas en la vida económica moderna. Es la unificación de las tres, la que es poderosa. Es esta unificación la que constituye la econometría".[4]

Aunque esta idea persiste, la noción que se tiene de la econometría no es muy clara, como se puede comprobar leyendo las definiciones que proporcionan casi todos los libros de texto. Éstas varían desde algunas ciertamente complejas, hasta otras relativamente sencillas, como la de Intriligator *et al.* (1996): "Econometría es la rama de la economía relacionada con la estimación empírica de relaciones económicas".[5]

También es importante destacar el trabajo realizado por la *Cowles Commission* de la Universidad de Chicago, a la que se debe una buena parte de los avances iniciales de la Econometría.[6]

Dado que el objetivo de la econometría es estimar modelos económicos, parece conveniente explicar qué se entiende por un modelo económico y cuáles son sus características distintivas. Este tema se aborda en la siguiente sección.

1.2. Los modelos económicos

La realidad económica es compleja, por lo que los economistas tienden a representarla por medio de modelos, es decir, haciendo abstracción de aquellos elementos de la realidad que no son esenciales para entender el fenómeno en cuestión. Es evidente que abusar de la abstracción puede hacer que el modelo se aleje excesivamente de la realidad, pero, por otra parte, si el modelo no simplifica suficientemente lo complejo de la realidad puede suceder que sea imposible llegar a la mínima comprensión del fenómeno que se quiere estudiar.

Los modelos económicos que se emplean en econometría tienen dos características: son modelos causales y están expresados en forma matemática. El

[4] El primer editorial de *Econometrica* y un buen número de los primeros trabajos econométricos han sido recopilados por Darnell (1994) en su obra *The History of Econometrics*.

[5] La falta de consenso sobre la definición de la econometría ha dado lugar a un artículo sobre el tema (Tintner, 1953).

objetivo fundamental de estos modelos es representar una estructura determinada, es decir una relación estable entre una serie de variables. Las variables se suelen clasificar en dos grandes grupos, la dependiente o endógena y las independientes o exógenas. La distinción, a nivel teórico, entre variables endógenas y exógenas es una característica esencial del proceso de simplificación que implica la modelización. Los modelos tratan de explicar una parte de una realidad compleja que se representa mediante variables endógenas, cuyo valor explica el propio modelo, y variables exógenas que afectan al fenómeno analizado y cuya determinación no se explica en el modelo. El modelo permite analizar cómo cambian las variables endógenas cuando cambian las variables exógenas.

En un contexto experimental, esta distinción entre variables endógenas y exógenas sería suficiente para cuantificar las relaciones entre las variables. El fenómeno de interés se estudia en un marco que elimina el efecto de cualquier otro factor que no sean las variables exógenas consideradas. El experimento se repite para diversos valores de las variables exógenas, de modo que es posible cuantificar cómo se modifican las variables endógenas ante cambios en las variables exógenas.

En un contexto no experimental es necesario asegurarse de que la variación en la variable endógena se debe exclusivamente a un cambio en la variable exógena y que no se trata de un caso en que las dos hayan cambiado simultáneamente por un factor ajeno a la relación que se trata de analizar. Por ejemplo, es fácil imaginar que aumentar el agua de regadío en una tierra aumenta la producción. Sin embargo, durante una sequía es probable que simultáneamente la producción disminuya y que se utilice más agua de riego. En estas circunstancias, si se analizan esos datos, se encuentra una relación inversa entre agua de riego y producción. El problema econométrico consiste en que la variable independiente (agua de riego) está correlacionada con la perturbación aleatoria (si no se incluye el clima como un regresor). En estas circunstancias el método de mínimos cuadrados ordinarios no proporciona estimaciones insesgadas de los parámetros poblacionales.

Si se tienen dos variables, x e y, siendo x la variable independiente e y la dependiente, su relación se puede expresar usando el concepto matemático de función como:

---

[6] La página web de la Cowles Foundation es http://cowles.econ.yale.edu/

$$y = f(x) \qquad (1.1)$$

donde, f representa la relación funcional entre x e y.

El modelo (1.1) refleja que variaciones en *x* son causa de variaciones en *y*. De hecho, tal como está escrito el modelo, son la única causa de variaciones en *y*. Asimismo, el modelo no indica si existe causalidad en la otra dirección, aunque en principio se supone que la causalidad va en una sola dirección. El concepto de causalidad no está claro muchas veces y hay que distinguirlo claramente del de correlación. Hay tres causas por las que dos variables pueden estar correlacionadas: (i) una es causa de la otra, (ii) las dos están causadas por una tercera variable, y (iii) la relación se debe al azar. El modelo (1.1) describe una situación del primer tipo, es decir, *x* es causa de *y*.

La determinación de la causalidad suele ser de tipo teórico, más que empírico, aunque en el contexto de series temporales, Granger ha desarrollado un test basado en la idea de que el futuro no puede explicar el pasado, lo que permite identificar estadísticamente la dirección de causalidad. Es decir, una variable es causa de otra en el sentido de Granger si la predicción de valores actuales de *y* puede mejorarse usando valores retrasados de *x*. El test empírico consiste en hacer una regresión de los y en valores pasados, actuales y futuros de x. Si los coeficientes de los valores futuros no son significativos y los otros sí, se puede decir que *x* causa *y*.[7]

En muchas ocasiones la complejidad de la relaciones económicas requiere más de una ecuación para representarlas adecuadamente. De hecho, una por cada variable endógena que haya en la estructura. Estos modelos se conocen como modelos de ecuaciones simultáneas, debido a que se supone que las variables endógenas se determinan simultáneamente. Un ejemplo es el modelo de oferta y demanda en el que las variables endógenas son el precio y la cantidad de equilibrio del mercado. Parece lógico suponer que ambas se determinan al mismo tiempo por la interacción de la oferta y la demanda.

1.3. Los modelos econométricos

[7] Los desarrollos más recientes en el tema del análisis empírico de la causalidad pueden verse en Granger (1988).

En econometría, al igual que en economía, el objetivo es explicar el comportamiento de una variable en función de otras. Por eso, el punto de partida de la econometría es el modelo económico. La diferencia está en que la econometría pretende cuantificar la relación entre las variables económicas.

Desde un punto de vista empírico, el concepto que subyace a todo modelo econométrico es el de variación. Es decir, el objetivo de un modelo econométrico es explicar la variación que presenta una variable, llamada dependiente, por medio de la variación de otras variables que se llaman independientes. Esto implica que el poder explicativo de una variable depende de lo mucho o poco que varíe y de la relación que tenga su patrón de variación con el de la variable dependiente. De aquí se puede deducir, por ejemplo, que una constante no tiene poder explicativo para explicar la variación de la variable dependiente.[8]

El modelo económico, tal como está escrito en (1.1), establece que variaciones en x son la única causa que produce variaciones en y. Esa relación es determinística ya que para cada valor de x existe un único valor de y. Por tanto, los valores de y quedan unívocamente determinados una vez que se conocen x y $f(\cdot)$.

Las relaciones determinísticas son comunes en las ciencias físicas. Un ejemplo sencillo es la ley de la gravedad. En efecto, si se deja caer una piedra de cualquier masa desde una altura de 4,9 metros tarda en llegar al suelo 1 segundo. Este resultado puede repetirse tantas veces como se quiera si se controlan las condiciones experimentales, es decir, controlando todos los factores que pueden afectar al resultado del experimento. En este ejemplo, es necesario controlar la resistencia que el aire ofrece ante la caída del objeto, ya que estamos interesados en analizar los efectos de la fuerza gravitatoria y no las características aerodinámicas del objeto lanzado.

En el análisis empírico de las relaciones económicas resulta imposible controlar multitud de factores que afectan al fenómeno analizado pero que no son esenciales

[8] Una posible confusión puede surgir porque el término independiente de una regresión, que es muchas veces significativo, es en realidad el coeficiente de una variable que sólo toma el valor 1, es decir, de una constante. Sin embargo, hay que tener en cuenta que lo que en realidad explica el término independiente es un nivel (una media) pero no sirve para explicar variabilidad.

para explicar dicho fenómeno. En términos prácticos, esto significa que es posible observar varios valores de la variable dependiente para un mismo valor de las variables independientes. Por este motivo, las variables económicas se modelizan como variables aleatorias, cuyo valor no se conoce con certeza sino con una determinada probabilidad. Una relación estocástica entre dos variables se puede representar del siguiente modo:

$$y = f(x) + u \tag{1.2}$$

donde *u* es una variable aleatoria, a la que se denomina perturbación aleatoria.[9]

En principio, *u* puede seguir cualquier tipo de distribución de probabilidad y su presencia implica que para cada valor de *x* existe una distribución de valores de *y*. El modelo (1.2) es la suma de una parte determinística, f(x), y de una parte aleatoria, *u*.[10] Por tanto, *y* también es una variable aleatoria. De hecho, las propiedades aleatorias de *y* vienen dadas exclusivamente por *u*, por lo que *y* seguirá la misma distribución que *u*. En general, se supone que los factores no controlables son, en cierto modo, independientes del fenómeno estudiado. La idea es que la variación en *y* se debe a un componente que es fijo (determinístico) y a otro componente que no es predecible (aleatorio). La econometría tiene por objeto buscar (estimar) la parte determinística de los modelos económicos.

La especificación de la relación entre *x* e *y* requiere escoger una forma funcional para f(·). Si, por simplicidad, se supone que es lineal, la ecuación (1.2) se convierte en:

$$y_i = \beta_0 + \beta_1 x_i + u_i \tag{1.3}$$

donde β_0, β_1 son parámetros que caracterizan la relación entre *x* e *y.* El subíndice *i* indica que la relación se cumple para cada observación.

Los modelos económicos tienen normalmente un fin explicativo, es decir, el objetivo del investigador suele ser explicar algún fenómeno de tipo económico.

[9] Los adjetivos estocástico y aleatorio se usarán indistintamente.

[10] Sin embargo, esto no es siempre así. Hay modelos donde la variación aleatoria constituye la totalidad de la variación de la variable dependiente. Un ejemplo es el conocido paseo aleatorio (*random walk*) en el que el presente es igual al pasado más un componente de error. Es decir, el modelo se puede escribir como $y_t=y_{t-1} + u_t$.

Alternativamente, los modelos econométricos pueden usarse con fines predictivos, es decir, para predecir el valor de la variable dependiente en períodos futuros.

Este libro se centra en la especificación y estimación de modelos explicativos. En este contexto, la capacidad de predicción del modelo es importante ya que es la prueba de la validez explicativa del modelo propuesto. Sin embargo, en ocasiones, es posible lograr predicciones más precisas mediante sofisticadas búsquedas de correlaciones que a través del análisis del fenómeno. Esta segunda noción de predicción no entra en los objetivos del presente trabajo.

1.3.1. Ejemplo: una función de producción agraria

El concepto de relación estocástica puede entenderse mejor con un ejemplo tomado de la agricultura. Si se quiere estudiar el efecto de la dosis de fertilizante (x) sobre la producción de un cultivo (y), la forma habitual de hacerlo es mediante un ensayo experimental en el que se siembran varias parcelas aplicando distintas dosis de fertilizante a cada una y repitiendo el experimento varias veces. En la Tabla 1.1 se recogen los datos de uno de esos experimentos cuyo objetivo era estudiar la respuesta de la hierba (medida en kg de materia seca por hectárea) al abono nitrogenado. Las seis dosis experimentales se repitieron nueve veces.[11]

Estos datos se representan de forma gráfica en la Figura 1.1, en la que se observa que, para cada dosis de fertilizante, la producción no toma un único valor (como sería en el caso determinístico) sino que hay toda una "distribución" de valores de *y* para cada valor de *x*, lo que refleja gráficamente la idea estadística de distribución condicionada.[12]

Tabla 1.1. *Producción de hierba (kg MS/ha) según dosis de nitrógeno*

	Dosis de nitrógeno (kg/ha)					
Repetición	0	20	40	60	80	100
1	1567	1338	2128	2345	2481	2470

[11] El experimento fue llevado a cabo en el Centro de Experimentación Agraria de Villaviciosa (Asturias) en el año 1987.

[12] Un ejemplo similar con datos no experimentales es el de la relación entre el salario y los años de formación (ver Goldberger, 1998; p. 2).

2	1166	1233	1368	1997	2245	2095
3	765	1311	1744	2445	2361	2561
4	859	1003	1484	2616	2748	2283
5	1159	1369	2347	3412	1912	2291
6	752	1407	2439	2226	2995	2565
7	901	1116	1645	2325	2194	1780
8	809	1443	1691	2001	1929	1839
9	992	2114	2207	2671	2166	1905

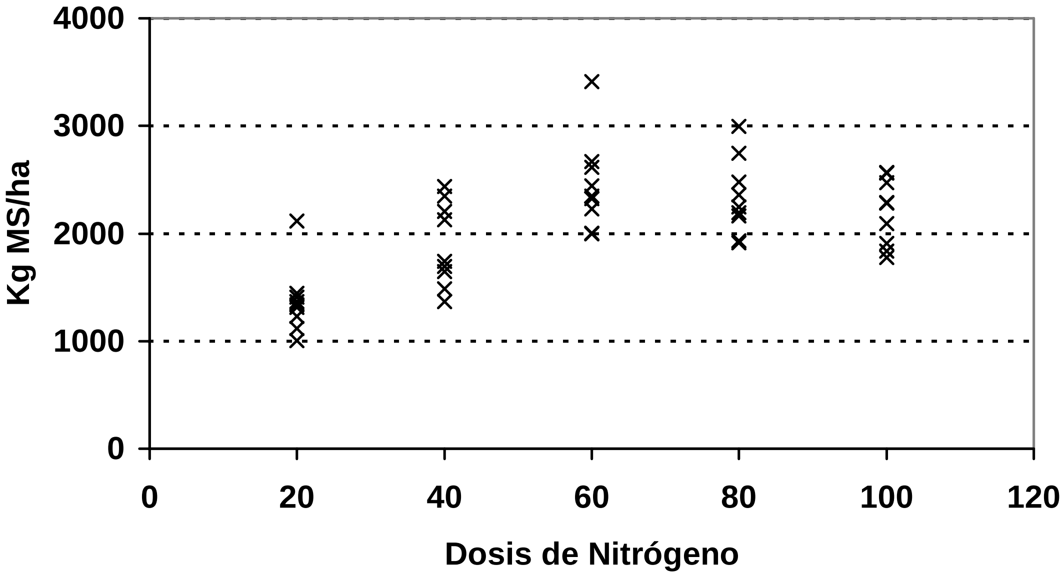

Figura 1.1. *Distribución de y condicionada en x*

Este ejemplo permite entender el significado de que la variable independiente es no estocástica o, como se suele decir algunas veces, 'las X son fijas en muestras repetidas'. En efecto, en la primera repetición de 6 observaciones las dosis de nitrógeno son 0, 20, 40, 60, 80 y 100, pero las siguientes repeticiones tienen los mismos valores, es decir, la misma X.[13]

1.3.2. La interpretación de los parámetros del modelo

Aunque el fin de la econometría es encontrar buenas estimaciones de los parámetros del modelo, el objetivo del investigador es de tipo económico. Es decir, el interés del investigador suele ser estimar conceptos económicos, tales como características de un proceso productivo (productividades marginales, elasticidad de escala,...) o de una función de demanda (elasticidades precio,...).

En este sentido, las características de interés pueden estar parametrizadas en el modelo o ser una combinación de los parámetros. En el ejemplo de la función de producción de la hierba, la productividad marginal del nitrógeno viene dada por la derivada de la función con respecto al único input. En el caso de que se considere una especificación lineal del modelo, como en (1.3), la productividad marginal del nitrógeno es β_1. En este caso, el concepto económico de interés coincide con un parámetro del modelo. Sin embargo, una productividad marginal constante contradice el supuesto que subyace en cualquier modelo de producción: la ley de los rendimientos decrecientes. Por ese motivo puede ser más conveniente emplear una función cuadrática, en vez de una especificación lineal:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i \tag{1.4}$$

En este modelo, la productividad marginal del nitrógeno (PMaX) se convierte en:

$$PMaX_i = \beta_1 + 2\beta_2 x_i \tag{1.5}$$

Ahora, el concepto de interés (PMaX) no coincide con un parámetro, sino que es una función de dos parámetros y de los datos, de tal forma que si β_2 es negativa, se cumple la ley de los rendimientos decrecientes. Por tanto, la PMaX no es constante sino que depende de la dosis empleada (existe, por tanto, una productividad marginal para cada valor de X).

Aunque en el ejemplo anterior β_0 puede interpretarse como la producción media esperada de una parcela sin abono, es importante destacar que, en general, el termino independiente no debe interpretarse como el valor que alcanza *y* cuando *X*=0. Esto se debe a que normalmente las observaciones que se tiene para las X no suelen estar cercanas a cero, por lo que el término independiente es simplemente el resultado de que las funciones se proyectan en el espacio cortando los ejes en algún punto. Por tanto, las funciones deben interpretarse en el rango relevante descrito por los datos. Como señalan

[13] En ese sentido dice Leamer (1983): “A los económetras les gustaría proyectar la imagen de investigadores agrarios que dividen una granja en una serie de parcelas más pequeñas de tierra y que seleccionan los niveles de fertilizante para ser usados en cada parcela”.

Rao y Miller (1971): "Cuando todas las variables independientes son cero, la observación no pertenece a la subpoblación bajo investigación y la ecuación de regresión no tiene una interpretación valida".[14]

Una cuestión que suele cruzar la mente de muchos estudiantes es por qué se supone en la especificación del modelo que los parámetros son constantes, es decir, que son iguales para todos los individuos. En principio, parece lógico admitir que cuanto menos restrictivo sea un modelo mejor podrá representar la realidad. En ese sentido, el modelo puede flexibilizarse haciendo que los parámetros varíen individualmente, por lo que (1.3) podría escribirse de la siguiente forma:

$$y_i = \beta_{0i} + \beta_{1i} x_i + u_i \tag{1.6}$$

El problema es que el modelo (1.6) no tiene ningún interés empírico ya que no proporciona ninguna información. La razón es que la parte determinística no establece estructura alguna sobre la relación entre las variables sino que simplemente indica que cada observación es ella misma. De hecho, en ese modelo la perturbación aleatoria es redundante, ya que la parte determinística del modelo "explica" perfectamente la nube de puntos. Adicionalmente, puede decirse que (1.6) no es estimable puesto que con N observaciones se tienen que estimar 2N parámetros, lo que es imposible.

Desde un punto de vista más filosófico, si no existe algún tipo de estructura común del fenómeno analizado la investigación científica no existiría. El astrónomo y divulgador científico Carl Sagan explica magistralmente esta idea al comentar que un grano de sal contiene un número mayor de partículas que neuronas tiene el cerebro humano. De ahí que la posibilidad de entender algo tan trivial como un grano de sal radica en que consta de un número muy elevado de repeticiones de una estructura química relativamente sencilla (Sagan, 1994).

Sin embargo, en algunas circunstancias la constancia de los parámetros del modelo no parece un supuesto razonable. Existen dos ejemplos importantes en que el supuesto de constancia de los parámetros puede ser ciertamente contraproducente. En primer lugar, cuando se trabaja con series temporales relativamente largas es de esperar que algunos parámetros cambien con el tiempo, lo que se conoce como cambio

[14] El lector avanzado encontrará importantes excepciones a esta regla. La más evidente es una variable binaria que toma los valores 1 y 0.

estructural. En segundo lugar, se encuentra lo que se conoce como la crítica de Lucas (Lucas, 1976) que aparece cuando se intenta usar los modelos econométricos con fines de política económica. Lucas argumenta que si se simula una variación de una variable exógena (tipo de interés,...) y los individuos pueden anticipar ese cambio, reaccionarán, de modo que los parámetros del modelo no pueden considerarse constantes. Estos dos problemas requieren un esfuerzo de modelización y estimación especial que será discutido más adelante en este libro.

Una vez caracterizado el modelo, permanece la pregunta de cuáles son los valores poblacionales de β_0 y β_1. Esta pregunta es imposible de responder ya que nunca se van a conocer los valores poblacionales. Por tanto, se necesita un procedimiento que permita aproximar su valor de alguna manera. Este procedimiento se conoce como estimación de los parámetros. Para ello se necesita un método de estimación y unos datos. Como la población no se dispone nunca de ella, los datos van a ser un subconjunto de la población que se denomina muestra. En principio se supondrá que la muestra es aleatoria. Este término quiere decir que, de algún modo, es necesaria la independencia entre las observaciones de la muestra. El ejemplo más típico es que cuando se trate de averiguar las intenciones de voto en una consulta electoral no debe preguntarse sólo a las personas que salen y entran de la sede de un determinado partido político ya que las respuestas (las observaciones) estarán todas relacionadas y no contienen información suficiente para conocer una característica de la población. De hecho, los encuestadores buscan la independencia de las observaciones situándose en diferentes lugares de la ciudad. Un modo de buscar la independencia es hacer algún tipo de sorteo para elegir las posiciones de los encuestadores. Este último punto puede explicar el uso del adjetivo de aleatorio.[15]

Un caso particular que merece cierta reflexión ocurre cuando la población tiene un número tan pequeño de individuos que parece ocioso hacer una distinción entre población y muestra y usar técnicas estadísticas. Un ejemplo ilustrativo puede ser los bancos de un país. Por un lado parece atractivo pensar que se tiene acceso fácil a la población y que los resultados que se obtengan se refieren directamente a la población. Sin embargo, si se piensa un poco más nos damos cuenta que cualquier

---

[15] En los libros de texto se distingue entre muestreo aleatorio y muestreo aleatorio estratificado. En el primer caso, las observaciones son extraídas al azar de la población. En el segundo caso, las observaciones son obtenidas al azar en una subpoblación definida por ciertas características de las variables explicativas.

aspecto que tomemos (rentabilidad de los bancos) pudo haber sido distinto con cierta probabilidad. Es decir, que el conjunto de individuos que observamos parecen tener propiedades similares a una muestra. La solución a este enigma consiste en reconocer que las poblaciones son un objeto abstracto que creamos para el análisis científico. En este sentido, no existen fuera de nuestro análisis y son, por tanto, inobservables. Por tanto, existe una población ideal de bancos de los cuáles observamos la muestra que las circunstancias han hecho que sea accesible en un determinado periodo. Un punto clave de este razonamiento es que la muestra que se observa no es aleatoria sino determinada por las circunstancias económicas que rodean a la banca.

1.4. La econometría aplicada

La utilización de la econometría teórica (la de los libros de texto) en el terreno de la economía aplicada no es fácil. El trabajo empírico en economía requiere una combinación adecuada de teoría económica, análisis de datos, econometría teórica y manejo de ordenadores. En esta combinación, la econometría no entra en forma aditiva sino que interacciona con todas las demás. Por esta razón, el éxito del trabajo empírico requiere, además de un adecuado conocimiento de los distintos elementos que lo integran, de una buena dosis de experiencia.

Quizás sea esta la razón por la que, en comparación con los libros de econometría teórica, no son muchos los libros sobre econometría aplicada que se pueden encontrar. Los mas conocidos probablemente sean los de Bridge (1971), Desai (1982) y Berndt (1991). En el campo específico de la economía de la producción, destaca el texto de Chambers (1988) y alguna recopilación de trabajos empíricos como la de Brown (1967). También algunos libros generales de econometría dedican alguna parte a estudiar aplicaciones econométricas. Entre éstos cabe reseñar los libros de Thomas (1993) e Intriligator *et al.* (1996).

Las dificultades que conlleva el trabajo econométrico aplicado, así como la ligereza con la que a veces se lleva a cabo, ha suscitado un buen número de críticas hacia la disciplina. Así, por ejemplo, Orcutt decía que “Aplicar la econometría es como intentar aprender las leyes de la electricidad escuchando la radio”. En realidad, la mayor parte de estas críticas no van dirigidas contra la econometría sino contra una buena parte del trabajo econométrico aplicado. Efectivamente, es muy fácil hacer trabajo econométrico: sólo se necesitan datos y un programa para estimar un modelo. Sin embargo, como se irá viendo a lo largo de este libro, el "buen" trabajo econométrico aplicado dista mucho de ser algo sencillo.[16]

Edward Leamer es uno de los económetras de reconocido prestigio que más ha llamado la atención sobre la separación entre econometría teórica y aplicada. En un famoso libro, Leamer (1978) dice: "Nos dividimos confortablemente entre una orden sacerdotal de estadísticos teóricos, por un lado, y una legión de inveterados pecadores analistas de datos, por otra. Los sacerdotes tienen poder para crear listas de pecados y son adorados por el talento que muestran. Los pecadores no se supone que tienen que evitar los pecados; sólo necesitan confesar sus errores abiertamente".

Uno de los aspectos de la econometría aplicada más criticado por muchos económetras es la avidez por una gran mayoría de los principiantes en este campo (y de otros no tan principiantes) de buscar un alto R^2 o estadísticos significativos. Goldberger ha bautizado esta forma de "validación" de los modelos como "R^2 fishing" (la pesca del R^2) o "the kitchen sink approach" (la aproximación del fregadero) en el sentido de que para lograr altos R^2 se introduce un elevado número de variables explicativas con el mismo criterio que se echan los platos sucios al fregadero, es decir, hasta llenarlo (R^2=1).

Los practicantes de la pesca del R^2 suelen "torturar" el modelo con una amplia variedad de trucos tales como añadir términos cuadráticos, cambiar el deflactor, agregar variables, etc. hasta que el proceso de estimación produce el resultado deseado (t's significativas). Esta forma de actuar, conocida como "data mining" se traduce en un grave problema consistente en que la probabilidad de cometer un error de tipo I es mayor que la supuesta en los test de inferencia estadística (Lovell, 1983). Esto se debe a que la cantidad de pruebas realizadas para encontrar estadísticos significativos hace que se puedan encontrar por casualidad en vez de por haber encontrado una "buena" especificación.

---

[16] Hendry (1980) compara la econometría con la alquimia ya que se pueden obtener resultados inverosímiles, tales como el que él expone en su artículo, consistente en explicar la tasa de inflación del Reino Unido en función de la pluviometría. Evidentemente, la econometría tiene más de alquimia que de ciencia cuando se aplica mal. Sin embargo, puede surgir una luz de esperanza si se piensa que los logros de la química actual deben algo a la alquimia de siglos anteriores.

Sin embargo, en muchas ocasiones la teoría económica no define con precisión todas las variables que deben aparecer en un modelo empírico. En estos casos la práctica común es estimar varios modelos con diferentes subconjuntos de variables hasta elegir el modelo que el investigador considera "mejor". Esta práctica es problemática ya que los modelos estimados con diferentes subconjuntos de variables dan lugar a diferentes resultados. En concreto, los parámetros pueden cambiar de signo, por lo que es difícil saber el efecto real de una variable explicativa en la variable explicada.[17]

A continuación se expone un ejemplo empírico en el que se intenta aclarar que el R^2 es un "ídolo" falso, al que no se debe adorar. El razonamiento que suelen hacer los principiantes es que si el R^2 es bajo, el modelo no es bueno por lo que hay que buscar alguna alternativa que permita obtener un R^2 mayor. Este es un grave error que conviene aclarar con un ejemplo. Para ello se va a usar el anterior ejemplo de la respuesta a la fertilización nitrogenada. En concreto, se ha estimado la función de producción cuadrática (1.4). Los resultados pueden verse en la Tabla 1.2.

Tabla 1.2. *Estimación de la función de producción cuadrática*

Parámetro	Coeficiente	Desv. Estándar	t-ratio
β_0	894.483	108.60	8.23
β_1	36.309	5.10	7.10
β_2	-0.228	0.04	-4.64
N = 54		R^2 = 68%	

Como se puede ver, las estimaciones de todos los parámetros son significativas, el signo de los coeficientes confirma la forma de U invertida que mostraba la figura de los datos y el R^2 es del 68%. En principio, parece que el investigador podría estar satisfecho con esta estimación. Sin embargo, en agronomía

[17] Este problema aparece en la literatura empírica de crecimiento. La teoría económica no identifica de forma precisa las variables que definen el estado estacionario. Ante este problema, Sala-i-Martín (1997) utiliza un método riguroso para conocer la probabilidad de que el efecto de una determinada variable se mantenga en diferentes especificaciones del modelo. En el caso en que la probabilidad sea alta se considera que esa variable debe aparecer en la especificación final del modelo.

es bastante frecuente realizar esta estimación usando sólo las medias para cada dosis. Es decir, las nueve repeticiones que hay para cada dosis se sustituyen por la media aritmética de las mismas, pasando de tener 54 observaciones a tener sólo 6. Los resultados de estimar en las medias pueden verse en la Tabla 1.3.

Tabla 1.3. *Estimación de la función de producción cuadrática en medias*

Parámetro	Coeficiente	Desv. estándar	t-ratio
β_0	894.483	151.83	5.89
β_1	36.309	7.14	5.08
β_2	-0.228	0.06	-3.32
N = 6		R^2 = 95%	

Como se puede comprobar las estimaciones obtenidas de los tres parámetros del modelo son las mismas usando todas las observaciones o las medias para cada dosis. El resto de los resultados son diferentes. Las desviaciones estándar de los coeficientes son mayores cuando se estima con las medias, lo cual es una consecuencia de que se emplean menos observaciones para estimar, por lo que la precisión que se consigue es menor. Pero el hecho diferencial más destacado es la elevada diferencia en el R^2 a favor del modelo estimado con las medias. La razón por la que se obtiene un R^2 mayor es porque al tomar las medias de las producciones, se está reduciendo el efecto del componente aleatorio de la producción.[18]

El hecho de que el R^2 sea mayor, hace que muchos investigadores hagan públicos los resultados obtenidos sólo con las medias en vez de con todos los datos. Dado que las estimaciones de los parámetros son idénticas en ambos modelos, las productividades marginales y los valores óptimos que se obtienen son iguales. Sin embargo, el investigador está falseando la bondad del ajuste de su modelo puesto que está diciendo que la parte determinística de su modelo explica un 95% de la variabilidad de la variable dependiente, cuando no es verdad, ya que explica una proporción bastante menor (68%).

[18] El lector más avanzado reconocerá en este ejemplo el hecho de que el método de mínimos cuadrados lo que hace es ajustar la línea que mejor pasa por las medias de las distribuciones condicionadas de Y en X.

Esperamos que este ejemplo haya servido para aclarar que el principal criterio para juzgar un modelo no puede ser su R^2. Un bajo R^2 puede ser ciertamente una indicación de una mala especificación. Así, en el ejemplo anterior, si en vez de estimar una función cuadrática, la cual se ajusta claramente a la estructura de los datos, se hubiera estimado una función lineal, se habría obtenido un R^2 más bajo, indicando que la función estimada no se ajusta bien a la nube de puntos. Sin embargo, no siempre que el R^2 sea bajo puede interpretarse como que la especificación del modelo no es adecuada, puesto que hay fenómenos económicos en los que existe una variabilidad intrínseca muy elevada. En estos casos, la parte aleatoria va a tener un papel importante en la explicación de la variable dependiente, lo que quedará reflejado en un "verdadero" bajo R^2.

Esta discusión sobre valores altos o bajos de modelos no relacionados no debe confundirse con la comparación de R^2 entre modelos anidados. Por ejemplo, cuando se añade una variable a un modelo el R^2 no puede disminuir. En este caso, el aumento del R^2 puede ser usado para contrastar estadísticamente la hipótesis nula de que el parámetro de la variable añadida es cero contra la alternativa de que es distinto de cero.

1.5. La organización del trabajo empírico

Como se ha dicho anteriormente, el trabajo econométrico no es sencillo y requiere una buena dosis de experiencia. Algunos libros de econometría dedican un capítulo a explicar cómo se debe desarrollar un proyecto econométrico (Johnson *et al.*, 1987; Griffiths, *et al.*, 1993; Intriligator *et al.*, 1996). A continuación se dan una serie de pistas para guiar a los que se inician en el mundo de la econometría aplicada. Para ello distinguiremos entre los elementos que debe tener un buen trabajo empírico y cómo se debe estructurar el mismo.

1.5.1. Los elementos de un buen trabajo empírico

Los principales elementos que debe tener todo buen trabajo econométrico son:

1) Una buena pregunta

Una buena pregunta es la que hace referencia a un tema interesante y además no tiene una respuesta obvia. Probablemente no sea una pregunta técnica sino económica.

Algunos ejemplos en el campo de la economía de la producción pueden ser: ¿son más eficientes las empresas grandes que las pequeñas?, ¿reaccionan los productores igual ante subidas que ante bajadas en el precio del producto?, ¿ha contribuido el progreso técnico a la sustitución de trabajo por capital?, etc...

En este punto el estudiante puede preguntarse cómo se encuentra una buena pregunta. Berndt (1991; p.) dice que Blinder siempre comenta con sus estudiantes que lo más perjudicial que le enseñaron en econometría en el MIT fue que las hipótesis no se obtienen de los datos. En ese sentido Blinder aconseja mirar los datos, hacer gráficos, mirar a las medias y los momentos, buscar observaciones "extrañas,... En resumen, hay que aprender a conocer tus datos.

2) Buenos datos

Los datos deben permitir responder a la pregunta planteada anteriormente. En algunos trabajos empíricos se contesta parcialmente (o no se contesta) a la pregunta de interés planteada debido a que los datos empleados no contienen la información necesaria. A veces, esta justificación no es aceptable ya que, aunque la fuente de información primaria puede no tener todas las variables necesarias, eso no quiere decir que no se puedan obtener mediante una encuesta auxiliar.

Aunque casi siempre se ha relegado los datos como algo marginal en el análisis econométrico, lo cierto es que constituyen un apartado fundamental. Una popular expresión inglesa "garbage in, garbage out", es decir, "basura dentro, basura fuera", indica que si lo que se introduce en el modelo no es bueno, lo que se va a obtener tampoco lo puede ser.

La importancia de la calidad de los datos pasa desapercibida algunas veces, sobre todo por jóvenes investigadores, que suelen poner más énfasis en usar un método de estimación sofisticado que en la búsqueda y el tratamiento correcto de los datos. A este respecto, un prestigioso económetra, Dennis Aigner (1988) dice: "...encuentro un poco desilusionante que sólo recientemente me haya dado cuenta de lo realmente dependientes que somos de datos generados por otros y con fines que no se corresponden necesariamente con el nuestro".

3) Un modelo

El modelo debe permitir expresar la pregunta. A ser posible, es preferible contar con un modelo que tenga algún tipo de comportamiento económico, como maximización de beneficios o minimización de costes.

La elección del modelo es, probablemente, el principal problema al que se enfrenta el investigador aplicado. Aunque en principio la teoría económica debe ser la principal guía para la especificación del modelo, lo cierto es que en la mayor parte de los casos la teoría suele ser demasiado general, por lo que la aplicación al caso concreto de estudio requiere de modificaciones que permitan incorporar todas las especificidades necesarias. Como dice Pesaran (1988): "Casi todas las teorías económicas son incompletas como explicación de fenómenos observados. A menudo incluyen variables inobservables, operan con cláusulas *ceteris paribus* y procesos de ajuste no especificados. El económetra se enfrenta entonces a la poco envidiable tarea de la especificación antes de que pueda empezar a confrontar la teoría con los datos".

El problema de la especificación suele resolverse en la investigación aplicada empleando un modelo ya desarrollado en la literatura. Sin embargo, el joven investigador que se enfrenta al análisis de un problema económico no debe abandonar la idea de poder construir un modelo específico para el problema en cuestión. Esta no es una tarea fácil y algunos reconocidos economistas han querido compartir su experiencia en este tema, con la publicación de sus consejos sobre cómo modelizar (Varian, 1997).[19]

Una característica común de la mayoría de los documentos dedicados a comentar cómo modelizar un problema, es la insistencia en que, al menos al principio, el modelo debe ser *sencillo*.[20]

4) Un método de estimación adecuado

Dado que el objetivo de los trabajos aplicados depende en última instancia de la calidad de las estimaciones obtenidas, es importante que el método de estimación empleado esté en consonancia con las características del modelo a estimar, con el fin de que los estimadores tengan las propiedades deseables.

---

[19] Este trabajo puede consultarse en la página web de Hal Varian: http://info.berkeley.edu/~hal/

[20] Uno de los máximos exponentes de esta opinión es Paul Krugman. Puede consultarse el documento "How I work" en su página web: http://web.mit.edu/krugman/www/

Como se verá en el capítulo 5, la mayor parte de las relaciones económicas de interés junto con la naturaleza de los datos hacen que mínimos cuadrados ordinarios sea un método de estimación que raramente produce estimadores con propiedades deseables, es decir, insesgados y eficientes.

1.5.2. La estructura de un buen trabajo empírico

Aunque no todos los trabajos empíricos son iguales, existe una homogeneidad en casi todos ellos en lo relativo a la organización de la presentación. A continuación se dan algunos consejos sobre cómo estructurar un trabajo de investigación empírico.[21]

Capítulo 1: Introducción

La Introducción es una parte fundamental de un trabajo empírico. Entre otras cosas debe tener el siguiente contenido:

- El problema económico que se pretende analizar.
- La importancia que tiene el análisis del problema.
- Antecedentes de análisis del problema en cuestión.
- Un párrafo, que se podría denominar párrafo "pero", en el que se explica que a pesar de los análisis previos citados en el punto anterior, existe un hueco no tratado o tratado de diferente forma que justifica la realización del trabajo.
- Un objetivo específico del trabajo, que, a poder ser, debería expresarse en términos de una (o más) hipótesis.
- Un esquema de la estructura del trabajo.

Capítulo 2: El modelo teórico

En este capítulo se debe resumir la teoría económica relevante para el análisis del problema que se trata. Este apartado debe contener:

- Los supuestos que subyacen al análisis.
- La descripción del modelo que se pretende emplear.

Capítulo 3: Los datos[22]

---

[21] Dado que el destino de la mayoría de los trabajos empíricos es su publicación en una revista científica, una interesante guía de cómo publicar trabajos puede verse en Hamermesh (1992).

Este apartado puede estar incorporado en el siguiente capítulo o constituir una sección independiente si la base de datos tiene una complejidad grande. Algunos aspectos que deben comentarse en este apartado son:

- Explicar la fuente de donde se han tomado los datos. En caso de que se haya hecho una encuesta, se debe explicar con claridad todos los aspectos de ejecución de la misma.
- Comentar los problemas de medición de las variables. Es normal que algunas de las variables del modelo teórico no sean observables, por lo que se suele proceder a usar "proxies" de las mismas. En ese caso se debe hacer una reflexión sobre si los problemas de observabilidad ponen en entredicho los resultados del análisis.
- En muchas ocasiones el número de observaciones con las que se hace el análisis empírico es diferente del número original de observaciones. La razón es que se han eliminado algunas observaciones. Los motivos más habituales son que para algunas observaciones no existen datos de todas las variables relevantes o que algunas observaciones presentan datos "sospechosos" de no ser verdaderos. Todos estos procedimientos de selección deben ser claramente especificados. En cualquier caso, es importante señalar que si la muestra aleatoria, estos procesos pueden dar lugar a la pérdida de esta propiedad en la muestra resultante.
- Es conveniente presentar una tabla en la que figuren las estadísticas descriptivas de los datos.
- En general, es aconsejable intentar ver la información que proporcionan los datos previamente a la estimación econométrica. Este tipo de técnicas suelen denominarse como "Análisis Exploratorio de Datos" (*Exploratory Data Analysis*). Una buena exposición de estos métodos puede verse en Hartwig y Dearing (1979). Una herramienta muy útil pero que desgraciadamente no se suele ver mucho en la práctica econométrica habitual es el uso de gráficos.[23]

---

[22] En algunos trabajos, la sección explicativa de los datos se encuentra después de la especificación del modelo empírico. Sin embargo, en algunos casos, la estructura de los datos puede condicionar la especificación del modelo y el método de estimación. Por ejemplo, no es lo mismo estimar una función de producción si se tienen datos de corte transversal que si se dispone de datos de panel.

[23] Dado que los gráficos tienen dos dimensiones, puede parecer que su utilidad para el análisis de relaciones con más de una variable independiente es limitado. Kennedy (1998) defiende su empleo: "Quizás el *software* econométrico debería tener algún modo de impedir que se corran regresiones hasta que los datos hayan sido examinados".

Capítulo 4: El modelo empírico

Esta es probablemente la sección más importante de un trabajo aplicado. En ella, aparte de revisar otros trabajos empíricos similares, se espera encontrar:

- Una especificación empírica clara del modelo a estimar, en la que hay que hacer referencia a la parte aleatoria del modelo.
- El método de estimación a emplear, con algunos comentarios acerca de cómo se van a tratar posibles problemas de estimación, tales como heterocedasticidad, autocorrelación, etc.

Capítulo 5: Estimación y resultados

En esta sección se presenta la estimación del modelo y también los resultados de la investigación, es decir, el contraste de la hipótesis del trabajo. Algunos consejos específicos son:

- Presentar los resultados de la estimación en una tabla ordenada, procurando buscar claridad en la presentación. Por ejemplo, debe evitarse copiar los resultados de la salida del ordenador y escribir números con 8 decimales.
- Debe contestarse con claridad a la(s) pregunta(s) que se haya formulado como objetivo de la investigación. El trabajo no acaba con la estimación del modelo, eso es sólo un paso intermedio para llegar a poder contrastar la idea de partida.
- Se deben "explotar" los resultados empíricos al máximo. Es decir, aparte de fijarse en el (los) parámetro(s) de interés para el contraste de la hipótesis nula, deben comentarse otros resultados obtenidos, como el signo y magnitud de otros parámetros.

Capítulo 6: Conclusiones

Todo trabajo tiene que tener unas conclusiones. Aunque esto parece evidente, hay dos aspectos que conviene destacar:

- Las conclusiones no son lo mismo que un resumen del trabajo. Muchas veces, en el apartado final, se espera encontrar alguna conclusión, ya sea metodológica (ventaja de un modelo o de un método de estimación sobre otro) o de tipo económico (recomendaciones de política económica), pero sólo se encuentra un resumen.

- Por el contrario, no se deben sacar conclusiones que no estén directamente obtenidas del modelo analizado.

Consejos globales

Aunque los puntos anteriores ofrecen una buena pauta para guiar el trabajo empírico, conviene dar dos consejos finales que afectan al trabajo en su totalidad:

- El trabajo ha de parecer importante. Al final del trabajo conviene dar una lectura al mismo. Si la impresión que se tiene es que no parece haberse realizado una contribución al estado de la cuestión del problema analizado, quizás convenga en ese momento replantearse el análisis y pensar qué le falta al trabajo para hacer esa contribución.
- El trabajo ha de parecer “honrado”. Esto quiere decir que el trabajo ha de presentarse con naturalidad contando las cosas como se hicieron, sin esconder ningún problema y destacando todos los detalles importantes del análisis realizado. Una característica que debe tener todo trabajo “honrado” es la de ser replicable. La replicabilidad de un trabajo implica que si otro investigador tuviese acceso a los mismos datos, podría llegar a reproducir los mismos resultados.[24]

1.6. La econometría y el ordenador

Como se dijo anteriormente, el trabajo econométrico de estimación requiere el uso de algún tipo de software especializado. Hoy en día son muchos los programas econométricos que hay en el mercado. Probablemente, el estudiante que quiera empezar a dar sus primeros pasos por el mundo de la econometría aplicada deba empezar a utilizar alguno de los más fáciles de usar como Limdep o Econometric Views. Sin embargo, aunque estos programas permiten estimar la gran mayoría de los modelos, en muchos casos es necesario recurrir a programas más flexibles, que permiten programar rutinas de estimación novedosas o poco comunes. El más popular hoy en día es Gauss.

[24] Aunque, en principio, parece que esta característica debería estar presente en todos los trabajos empíricos, la realidad demuestra que no es así (véase Dewald *et al.*, 1986). La importancia científica de la replicabilidad hace que algunas revistas de prestigio exijan a los autores de los trabajos aceptados para publicación que acompañen su base de datos y que ésta esté disponible para cualquier investigador que la solicite.

Pero no debe confundirse el manejo de los programas con el trabajo econométrico con el ordenador, que dista mucho de ser algo sencillo. Uno de los principios que debe guiar a todo buen investigador es el de intentar minimizar la posibilidad de cometer errores y, cuando se trabaja con ordenadores, esa posibilidad aparece con más frecuencia de lo que se suele pensar.

En ese sentido, parece conveniente dar algunos consejos de tipo práctico de cara a la forma de trabajar con los ordenadores:

1) *Introducir los datos originales en un fichero*

El tipo de fichero más conveniente es probablemente una hoja de cálculo, ya que ese formato puede ser leído por casi todos los programas estadísticos. Después se debe hacer una copia para trabajar y no volver a usar el fichero original más. Este es un aspecto importante que permitirá evitar que por cualquier tipo de accidente involuntario, se destruyan o modifiquen total o parcialmente los datos originales.

2) *Comprobar que los datos se han leído correctamente*

Una vez leídos los datos en el programa econométrico, es inexcusable comprobar que el programa ha leído los datos correctamente. En muchas ocasiones, especialmente cuando los datos están en ficheros de texto, los programas cometen errores de lectura por diversos motivos (no encuentran un fin de línea, se pueden confundir si los datos están separados por tabuladores, etc.). Por ello, no debe confiarse uno si al ejecutar una orden de lectura el programa no da ningún error, porque puede haber leído algo, pero mal. Para ello, si la base de datos es muy grande, lo que hace prácticamente inviable listarlos, es muy conveniente pedir las estadísticas descriptivas de todas las variables. Mirando especialmente los máximos y los mínimos de cada variable se puede detectar, no solamente errores de lectura, sino también algún valor anómalo de los datos. Otro aspecto importante es comprobar que el número de observaciones leídas se corresponde con de las originales.

3) *Transformar las variables usando el programa econométrico*

Todas las modificaciones de variables (tomar logaritmos, deflactar, etc.) deben hacerse dentro del programa econométrico. Si son muchas las transformaciones, es conveniente hacer un programa que contenga todas las modificaciones y que cree un nuevo fichero de datos con las variables originales y las transformadas. De esta manera, cuando se haga el análisis aplicado no habrá que repetir las transformaciones cada vez que se empiece a trabajar.

4) Crear un fichero de texto (con un editor o procesador de textos) en el que se vayan recogiendo todos los detalles relativos a la evolución de la investigación. Por ejemplo, en caso de haber varios ficheros de datos, lo que contiene cada uno, el significado de las distintas variables, sus unidades de medida, la fuente de donde provienen, etc.

5) Los programas deben estar bien documentados

Los ficheros que contengan las instrucciones del programa econométrico deben tener comentarios que permitan a otra persona entender lo que se está haciendo. Este es un aspecto muy importante del proceso de estimación que tiene varias ventajas. En primer lugar, permite reducir el número de errores en caso de que el programa sea largo y haya distintas alternativas de estimación. En segundo lugar, permite recordar con claridad lo que se está haciendo en caso de que el investigador abandone el trabajo durante un cierto período y lo recupere después. Por último, permite enviar el programa a colegas, para prestar o solicitar ayuda o para iniciar una colaboración científica.

6) Hacer copias de seguridad

Hacer <u>diariamente</u> una copia de seguridad de todos los ficheros (datos y programas). Aunque este consejo parece obvio, es frecuente encontrar casos en los que muchos investigadores se lamentan de los daños ocasionados por un colapso del disco duro, etc.

Bibliografía

Aigner, D., (1988), "On Econometric Methodology and the Search for Causal Laws", *The Economic Record*, 64, 323-325.

Berndt, E., (1991), *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley.

Bridge, J.L., (1971), *Applied Econometrics*, North Holland.

Brown, M. (ed.), (1967), *The Theory and Empirical Analysis of Production*, National Bureau of Economic Research.

Chambers, R., (1988), *Applied Production Analysis*, Cambridge University Press.

Cobb, C.W. y P.H. Douglas, (1928) "A Theory of Production", *American Economic Review*, 18(1), 139-72.

Darnell, A., (1994), *The History of Econometrics*, vols. I y II, Edward Elgar.

Desai, M., (1982), *Applied Econometrics*, Phillip Allen.

Dewald, W.G., J. Thursby y R. Anderson, (1986), "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project", *American Economic Review*, 76(4), 587-603.

Goldberger, A.S., (1998), *Introductory Econometrics*, Harvard University Press.

Granger, C.W.J., (1988), "Some Recent Developments in a Concept of Causality", *J. of Econometrics*, 199-211.

Griffiths, W.E., R.C. Hill y G.G. Judge, (1993), *Learning and Practicing Econometrics*, John Wiley.

Hamermesh, D.S., (1992), "The Young Economist's Guide to Professional Etiquette", *J. of Economic Perspectives*, 6(1), 169-179.

Hartwig, F. y B.E. Dearing, (1979), *Exploratory Data Analysis*, Sage.

Hendry, D. (1980), "Econometrics - Alchemy or Science?", *Economica*, 47, 387-406.

Intriligator, M.D., R. Bodkin y C. Hsiao, (1996), *Econometric Models, Techniques and Applications*, 2nd edition, Prentice-Hall.

Johnson, A.C., M.B. Johnson y R.C. Buse, (1987), *Econometrics: Basic and Applied*, Macmillan.

Kennedy, P., (1998), *A Guide to Econometrics*, Fourth edition, The MIT Press.

Leamer, E.E., (1978), *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Ney York, John Wiley.

Leamer, E.E., (1983), "Let´s Take the Con out of Econometrics", *American Economic Review*, 73(1), 31-43.

Lovell, M.C., (1983), "Data Mining", *Review of Economics and Statistics*, 65, 1-12.

Lucas, R., (1976), "Econometric Policy Evaluation: A Critique", *Carnegie Rochester Conferences on Public Policy*, 1, 19-46.

Moore, H.L., (1914), *Economic Cycles. Their Law and Cause*, New York: Macmillan.

Morgan, M., (1990), *The History of Econometric Ideas*, Cambridge University Press.

Pesaran, M.H., (1988), “The Role of Theory in Applied Econometrics”, *The Economic Record*, 64, 336-9.

Rao, P. y R. Miller, (1971), *Applied Econometrics*, Wadsworth.

Sagan, C., (1994), *El cerebro de Broca*, Editorial Crítica, Barcelona.

Sala-i-Martin, X., (1997), “I Just Run Two Million Regresions”, *American Economic Review*, 87(2), 178-183.

Schultz, H., (1928), *Statistical Laws of Demand of Supply with Special Application to Sugar*, University of Chicago Press.

Thomas, R.L., (1993), *Introductory Econometrics. Theory and Applications*, Longman.

Tintner, G., (1953), “The Definition of Econometrics”, *Econometrica*, 21, 31-40.

Varian, H., (1997), “How to Build an Economic Model in Your Spare Time”, en M. Szenberg, (ed.) *Passion and Craft: Economists at Work*, University of Michigan Press.

Waugh, F., (1928), “Quality Factors Influencing Vegetable Prices”, *J. of Farm Economics*, 19, 185-196.

Working, H., (1927), “What Do Statistical Demand Curves Show?”, *Quarterly J. of Economics*, 41, 212-215.

CAPÍTULO 2

EL ANÁLISIS PRIMAL DE LA PRODUCCION

En este capítulo se describen distintas formas de representar la tecnología y su aplicación para realizar estudios empíricos en el campo de la producción. El capítulo empieza con la definición de la tecnología y las distintas formas de representarla. El siguiente apartado se dedica a analizar las principales características de interés económico de una tecnología, como son las productividades marginales, los rendimientos a escala y las posibilidades de sustitución entre factores. A continuación, se analizan ciertas propiedades de la tecnología que permiten simplificar el análisis empírico. Entre ellas se pueden citar propiedades como la homogeneidad, homoteticidad y la separabilidad. Finalmente, se hace una breve introducción al análisis dinámico de la producción.

2.1. Análisis de la producción

La teoría de la producción estudia las decisiones relativas a la producción de uno o más bienes (outputs) y la utilización de factores productivos (inputs) a nivel microeconómico, esto es, desde el punto de vista de las unidades de producción básicas, las empresas, o de toda una industria o zona geográfica. Las aplicaciones de esta teoría son muy diversas y de gran relevancia. Por un lado, el análisis de la producción permite conocer los cambios en las decisiones productivas de los empresarios ante cambios en la tecnología o en las condiciones de mercado en las que se desenvuelven las empresas. En este sentido, la teoría de la producción permite evaluar los efectos de diversas medidas de política económica, como el establecimiento de cuotas a la producción, de salarios mínimos o medidas de regulación medioambiental. Por otra parte, los análisis de la productividad, la eficiencia o el cambio técnico son claves para entender cuestiones tan relevantes como el desempleo, la inversión, el crecimiento económico, las desigualdades regionales o el papel que juegan las infraestructuras públicas en el desarrollo regional. Finalmente, algunos temas de actualidad como el agotamiento de los recursos naturales, la deforestación o la degradación ambiental también pueden estudiarse dentro del marco de un proceso productivo.

El análisis teórico de la producción se basa en el uso de modelos esquemáticos de las actividades productivas.[25] Básicamente, se usa una representación de la tecnología, una descripción de los mercados de inputs y outputs y se supone un determinado comportamiento optimizador para los productores. Con la intención de mostrar la relevancia de la teoría económica para el análisis empírico, este capítulo repasa cuestiones básicas de la teoría de la producción y extiende algunos resultados teóricos que facilitan el análisis cuantitativo de los procesos productivos.

2.2. La tecnología

[25] En este punto, es importante recordar que un modelo es una representación simplificada de la realidad. Los fenómenos económicos suelen presentar una gran complejidad. La simplificación, es decir, la elección de las características claves, es a la vez la esencia del proceso de análisis y la fuente de numerosas críticas al análisis económico. Ante esta controversia, los autores más prestigiosos optan por un enfoque eminentemente práctico en el que los modelos se juzgan por su capacidad de ayudar a entender la economía y no por el grado de realismo de sus supuestos (véase la clásica referencia de Friedman (1953)).

La tecnología juega un papel fundamental en el análisis económico ya que es una de las restricciones a las que se enfrenta el productor de cara a maximizar sus beneficios o minimizar costes. Por tanto, se trata de un elemento clave a la hora de explicar las decisiones de producción.

La tecnología es la relación técnica entre inputs usados y outputs producidos, la cual se puede definir en sentido estricto o amplio. De acuerdo con la primera concepción, la tecnología representa únicamente el conjunto de procesos de producción técnicamente viables y disponibles para las empresas en un momento del tiempo. Desde esta perspectiva, la tecnología se define en términos "ingenieriles". Por otra parte, en el análisis económico, la tecnología se define en términos más amplios incluyendo, además de las relaciones puramente técnicas, la estructura jerárquica de las empresas, el sistema de incentivos, etc. que de una forma u otra determinan también la capacidad de transformación de unos inputs en outputs.

Es evidente que la tecnología de empresas de distintos sectores no se parecen. Así, por ejemplo, una fábrica de zapatos, un bar, una empresa siderúrgica y un invernadero agrícola tienen aparentemente poco en común. Esto podría llevar a concluir que debe existir una representación distinta de la tecnología para cada actividad, que recoja en cada caso las peculiaridades existentes en ese sector productivo. Sin embargo, los economistas han optado por ver lo que es esencial en todas las tecnologías: se usan inputs para producir outputs. Esto constituye un claro ejemplo de modelización económica, que permite, haciendo abstracción de lo que no es sustancial al problema, representar lo verdaderamente importante.

La representación de la tecnología consiste en la definición matemática de la relación entre inputs y outputs. Existen varias formas de representar la tecnología en las que el investigador puede apoyarse para realizar un estudio empírico. Entre ellas se encuentran algunas ampliamente conocidas (como la función de producción) y otras que no lo son tanto (la función de distancia). Al finalizar el apartado siguiente, el lector debería ser consciente del arsenal de posibilidades del que se dispone para el estudio empírico, y de que el enfoque adecuado puede ser distinto en cada caso, dependiendo de las condiciones del estudio.

Las representaciones tecnológicas que se describen a continuación se pueden definir o clasificar en paramétricas y no paramétricas. El enfoque paramétrico especifica la relación entre inputs y outputs mediante una forma funcional concreta. Por su parte, el enfoque no-paramétrico define la relación en términos más generales, sin especificar una forma funcional. Desde un punto de vista práctico, el enfoque paramétrico permite usar herramientas comunes del cálculo (derivadas, condiciones de optimización) en el análisis teórico y, en el ámbito empírico, permite el uso de técnicas econométricas convencionales. El inconveniente es que cualquier forma funcional que se elija para representar el proceso productivo implica una simplificación de la realidad y, por tanto, algún resultado del análisis estará relacionado con la forma funcional elegida y no con las características esenciales del fenómeno analizado (a este respecto véase el Capítulo 4 de este volumen). Por otra parte, el análisis no-paramétrico presenta como principal ventaja su grado de generalidad. Sus desventajas son la imposibilidad de usar las herramientas comunes del cálculo citadas previamente y, desde el punto de vista empírico, requiere el uso de métodos de análisis de desarrollo reciente y uso menos generalizado.[26]

2.2.1. El conjunto de requerimiento de inputs

El conjunto de requerimiento de inputs se define como el conjunto de todos aquellos vectores de inputs $x \in R_+^n$ que permiten obtener, al menos, el vector de outputs $y \in R_+^m$. Dicho conjunto se va a denotar por L(y) y se puede expresar matemáticamente como:

$$L(y) = \{ x : (y,x) \text{ sea factible} \} \tag{2. 1}$$

Por ejemplo, en una compañía de aviación el vector x puede tener dos componentes (x_1,x_2) donde x_1 representa el número de trabajadores y x_2 el número de aviones. Un caso más realista consistiría en añadir un tercer componente x_3 que mida las instalaciones y maquinaria con la que cuentan. Por su parte, el vector y puede tener dos componentes (y_1,y_2) donde y_1 representa el número de pasajeros e y_2 la carga que se transporta. Es fácil ver que con los mismos factores se puede transportar más pasajeros y menos carga o viceversa. Sólo se trata de configurar los aviones de

[26] Uno de los trabajos precursores en esta línea de investigación es el de Hanoch y Rothschild (1972) en el que se desarrollan tests sobre la estructura de la tecnología y sobre el objetivo de los productores que no requieren ninguna forma funcional. Otros trabajos influyentes son el de Varian (1984) y el de Chavas y Cox (1988).

modo un poco distinto. Sin embargo, alguien podría decir que los pasajeros se suben solos al avión y la carga necesita personal que lo haga. Desde el punto de vista de los factores, es posible reducir el número de aviones necesarios si se tiene más personal. Al tener más trabajadores, se pueden tener más turnos que permitan usar los aviones continuamente. También se pueden cargar, descargar y poner en estado de vuelo más rápido si se tiene más personal con lo que se necesitan menos aviones para hacer el mismo servicio a los clientes.

En la Figura 2.1 se representa el conjunto de requerimiento de inputs para el caso de un único output y dos inputs. El área sombreada nos indica la cantidad de inputs que permiten obtener (con la tecnología disponible) al menos el nivel de producción y. Así, a modo de ejemplo, si el nivel de producción viene determinado por la isocuanta

$$x_1 = \frac{y^{1/\alpha_1}}{x_2^{\alpha_2/\alpha_1}} \tag{2. 2}$$

el conjunto de requerimientos de inputs vendría dado por

$$L(y) = \{ (x_1, x_2) : x_1^{\alpha_1} x_2^{\alpha_2} \geq y \} \tag{2. 3}$$

El signo mayor o igual en (2.3) indica que L(y) incluye combinaciones de inputs técnicamente ineficientes, lo que sucede cuando existe alguna alternativa que permite: a) producir más output usando la misma cantidad de inputs; o b) producir la misma cantidad de output usando menos inputs. Así, mientras que la combinación A de la Figura 2.1 es técnicamente eficiente, el punto B es una combinación ineficiente.

Esta forma de representar la tecnología es, en principio, apropiada para analizar empresas cuyo nivel de producción viene determinado exógenamente (un ejemplo ilustrativo son los hospitales) puesto que el conjunto de requerimientos de inputs se define dado un nivel de producción. Para las empresas que tienen un mayor grado de control sobre los outputs que sobre los inputs, es más adecuado una descripción de la tecnología definida dado el nivel de inputs. Dicha representación se denomina conjunto de posibilidades de producción.

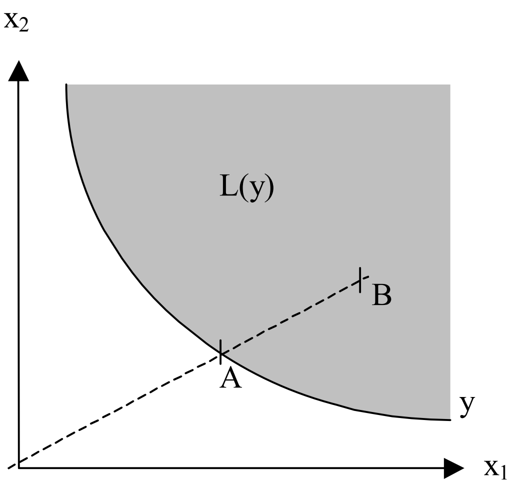

Figura 2.1. *Conjunto de requerimiento de inputs*

2.2.2. El conjunto de posibilidades de producción

El conjunto de posibilidades de producción se define como el conjunto P(x) de todos los vectores de outputs $y \in R_+^m$ que se pueden producir usando el vector de inputs $x \in R_+^n$. Es decir:

$$P(x) = \{ y : x \text{ puede producir } y \} \tag{2. 4}$$

Así, dadas las cantidades de inputs x_1 y x_2 para la tecnología (2.3), el conjunto de posibilidades de producción vendría dado por

$$P(x_1, x_2) = \{ y : y \leq x_1^{\alpha_1} x_2^{\alpha_2} \} \tag{2. 5}$$

En la Figura 2.2 se representa el conjunto posibilidades de producción para el caso de dos outputs y un único input. El área sombreada indica la cantidad de outputs que se puede obtener con la utilización de *x* unidades del input. Al igual que el conjunto de requerimiento de inputs, P(x) incluye tanto combinaciones de outputs técnicamente eficientes (punto C) como combinaciones ineficientes (punto D).

Si bien el conjunto de requerimiento de inputs y el conjunto de posibilidades de producción permiten mantener un alto grado de intuición sobre la relación entre inputs y outputs, presentan serias dificultades para ser estimados con técnicas econométricas sencillas. Intuitivamente, el problema se encuentra en el signo de desigualdad que aparece en la definición de ambas formas de representar la tecnología y que los modelos econométricos tradicionales están diseñados para estimar relaciones de igualdad.[27] Por ello, desde el punto de vista econométrico, es preferible utilizar representaciones de la tecnología que se apoyen en relaciones de igualdad, como las que se presentan a continuación.

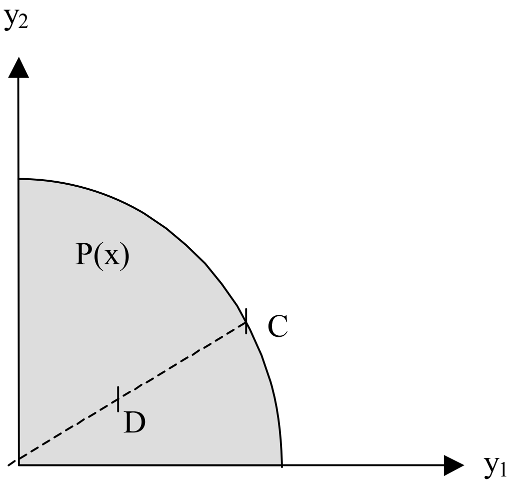

Figura 2.2. *Conjunto de posibilidades de producción*

2.2.3. La función de producción

La función de producción, que es el elemento central del análisis tradicional de la producción, relaciona los inputs usados con el *máximo* output que es posible obtener con dichos inputs. En el caso de un único output, la función de producción se suele representar como:

$$y = f(x) \tag{2. 6}$$

donde *y* representa la producción máxima que es posible obtener a partir de un vector de inputs x. A diferencia de las formas anteriores, la función de producción sólo representa las posibilidades tecnológicas que son técnicamente eficientes, excluyendo aquéllas combinaciones ineficientes.[28] Es decir:

$$f(x) = \max \{ y : x \in L(y) \} \quad ; \quad y \in R_+ \tag{2. 7}$$

[27] Sin embargo, el conjunto de posibilidades de producción o de requerimientos de inputs son idóneos cuando se utilizan técnicas de programación matemáticas como es el conocido Análisis Envolvente de Datos (DEA) que se apoya en relaciones de desigualdad.

De acuerdo con esta definición, la función de producción se puede ver como la "frontera" del conjunto de requerimientos de inputs L(y), lo que en términos gráficos equivale a la isocuanta de la Figura 2.1. Dada la tecnología definida por (2.3), la función de producción vendría dada por la expresión

$$y = x_1^{\alpha_1} x_2^{\alpha_2} \tag{2. 8}$$

Esta función es un caso particular de la función de producción que Cobb y Douglas propusieron en su famoso artículo de 1928. Para el caso de dos inputs la función de producción Cobb-Douglas (CD, en adelante) puede escribirse como:

$$y = Ax_1^{\alpha_1} x_2^{\alpha_2} \tag{2. 9}$$

El parámetro A puede interpretarse como una medida de la eficiencia técnica de la empresa (esto es, $A=A_i$ donde el subíndice *i* denota a la i-ésima empresa) puesto que para cualquier combinación de inputs, el output aumenta con A. Si las observaciones sobre las que se define la función son de series temporales y corresponden a algún sector determinado, entonces A se puede interpretar como un indicador del estado de la tecnología (en este caso, A se puede escribir como A(t) donde *t* es la variable tiempo) y su variación temporal como el efecto del cambio técnico sobre la producción.

El origen de la función CD tiene que ver con una observación empírica que hizo el economista Paul Douglas, de la Universidad de Amherst. Douglas encontró que las rentas del trabajo constituían una proporción casi constante en el tiempo del valor del producto total de los Estados Unidos. En su interés por caracterizar esa relación, pidió la ayuda de su colega, el matemático Cobb, y su colaboración tuvo como resultado la función expresada en (2.9). Desde su aparición, esta función ha sido la más usada en el análisis empírico de la producción, aunque hoy en día se prefiere el uso de formas más flexibles.[29]

---

[28] Cabe destacar que el concepto de función de producción implícitamente incorpora la existencia de un gestor eficiente. Esto implica que todas las empresas están situadas sobre la función de producción. Sin embargo, en la realidad se observa que hay empresas que comparten la misma tecnología y utilizan la misma cantidad de inputs pero producen distinto nivel de output, lo que sugiere que hay empresas que no emplean la tecnología correctamente. El análisis de la ineficiencia técnica, que es una importante área de investigación en economía de la producción, se desarrolla en el capítulo 6.

[29] Las formas funcionales flexibles se estudian con detalle en el Capítulo 4.

Uno de los problemas más importantes que se suele achacar a la función de producción es que no permite la representación de tecnologías con varios outputs. Sin embargo, eso no es cierto puesto que existen funciones de producción multiproducto, que expresan la producción de un output en función de los inputs y del resto outputs. [30] Así, en su pionero texto de econometría, Klein (1953) propuso estimar una función de producción del tipo:

$$y_1 = y_2^{-\beta_2} x_1^{\alpha_1} x_2^{\alpha_2} \qquad (2.\ 10)$$

donde el signo menos delante de β_2 indica que la tasa de sustitución entre los dos outputs es negativa, como la existente en la Figura 2.2. Como la función (2.10) representa la máxima cantidad que se puede producir del primer output, dada la cantidad del segundo y dadas unas cantidades de inputs, dicha función de producción se puede interpretar como la “frontera” del conjunto de posibilidades de producción P(x).[31]

Aunque Klein ya estimó una función de producción con dos outputs, la inmensa mayoría de los trabajos empíricos de producción con tecnologías multiproducto han empleado funciones de costes o beneficios. La utilización de estas funciones exige, a diferencia de lo que ocurre con una función de producción multiproducto, disponer de información sobre precios, que los precios sean exógenos (es decir, los productores sean precio-aceptantes) así como suponer comportamiento maximizador (minimizador) de beneficios (costes). Tales supuestos no se cumplen siempre, especialmente en empresas públicas o reguladas.

2.2.4. La función de distancia

Otra representación primal de una tecnología multiproducto, que en los últimos años ha sido ampliamente utilizada, es la función de distancia desarrollada originalmente por Shephard (1953). La popularidad de la función de distancia se debe a que no presenta ninguno de los problemas que presenta la función de producción

[30] Las funciones de producción multiproducto se desarrollan en Beattie y Taylor (1993).

[31] Klein observó, sin embargo, que la curvatura de la función de producción multiproducto (2.10) podría ser problemática (esto es, que la frontera del conjunto de posibilidades de producción fuera convexa en lugar de cóncava) si se requiere algún tipo de optimización económica. Dicho problema posteriormente se ha tratado de solucionar mediante la utilización de otras formas funcionales, que sin embargo imponían fuertes restricciones sobre la tecnología, como son la separabilidad entre outputs e inputs o la de elasticidades de sustitución y de transformación constantes (a este respecto véase la sección siguiente de este capítulo).

multiproducto y a que permite tratar convenientemente la existencia de productores ineficientes.

La idea que hay detrás de una función de distancia es, como su nombre indica, medir la distancia que separa una empresa (si es ineficiente) de la "frontera" tecnológica. La medida de esa distancia puede hacerse, al menos, siguiendo dos direcciones, lo que da lugar a dos tipos de funciones de distancia: orientadas a los inputs o a los outputs.

Así una función de distancia orientada a los outputs (D_O) trata medir la distancia que separa una empresa de la "frontera" del conjunto de posibilidades de producción, comparando el nivel de producción de dicha empresa con el de una empresa eficiente, que, al estar localizada sobre la "frontera, utiliza la misma cantidad de inputs pero obtiene un mayor nivel de producción. En este sentido, se puede definir la función de distancia orientada a los outputs como la mayor expansión equiproporcional que se puede hacer de un vector de outputs sin aumentar la utilización actual de inputs. Matemáticamente:

$$D_O(y,x) = \min\{\lambda : y/\lambda \in P(x)\} \qquad (2.\ 11)$$

Por definición, si la empresa está produciendo en su frontera de posibilidades de producción, la función de distancia tomará el valor 1. Sin embargo, si la empresa está en el interior del conjunto de posibilidades de producción, $D_O(y,x)$ será menor que 1. Al dividir el vector de outputs de dicha empresa por $D_O(y,x)$, proyecta esa observación ineficiente sobre la frontera del conjunto de posibilidades de producción, haciéndola eficiente,[32] lo que en términos de la Figura 2.2, es equivalente a mover la observación D al punto C de la frontera.

Por su parte, en la función de distancia orientada a los inputs (D_I) se compara la cantidad de inputs empleada por una empresa con el de una eficiente, que al estar está situada sobre la isocuanta, produce lo mismo pero utilizando menos inputs. En este sentido, se puede definir la función de distancia orientada a los inputs como la máxima reducción equiproporcional de un vector de inputs dado un vector de outputs. Es decir, se busca en qué proporción se puede reducir el uso de todos inputs sin que por ello disminuya ninguno de los outputs producidos. Matemáticamente:

$$D_I(y,x) = \max\{\delta : x/\delta \in L(y)\} \qquad (2.\ 12)$$

Por tanto, $D_I(y,x)$ será siempre mayor que la unidad si la empresa es ineficiente e igual a 1 si es eficiente (esto es, si está situada sobre la isocuanta relevante). Al igual que la anterior función de distancia, si se divide el vector de inputs de una empresa ineficiente por $D_I(y,x)$ se convierte dicha observación en eficiente, trasladándola a la frontera del conjunto de requerimientos de inputs, lo que en términos de la Figura 2.1 es equivalente a mover la observación B (ineficiente) al punto A (eficiente).

Las definiciones anteriores se centran en la "distancia" de un determinado vector de output o inputs a un conjunto de referencia. En ese sentido, puede ser difícil ver que la función de distancia constituye realmente una descripción o representación de la tecnología. Sin embargo, indirectamente la función de distancia define los outputs que es posible producir a partir de un determinado vector de inputs (y/λ) o los inputs necesarios para producir un vector de outputs (x/δ). En esencia, estas definiciones son una descripción de la tecnología, lo que se puede entender mejor con un ejemplo.

Hemos dicho que la función de producción multiproducto (2.10) se puede interpretar como la "frontera" del conjunto de posibilidades de producción P(x), esto es, incluye únicamente observaciones eficientes. Por lo tanto, si se divide un vector de outputs cualquiera por D_O, se ha de cumplir:

$$\frac{y_1}{D_O}\left(\frac{y_2}{D_O}\right)^{\beta_2} = x_1^{\alpha_1} x_2^{\alpha_2} \qquad (2.\ 13)$$

lo que, despejando D_O, se obtiene la expresión de la función de distancia orientada al output que representa la *misma* tecnología que la función de producción (2.10):

$$D_O(y_1, y_2, x_1, x_2) = \frac{y_1^{\frac{1}{1+\beta_2}}\, y_2^{\frac{\beta_2}{1+\beta_2}}}{x_1^{\frac{\alpha_1}{1+\beta_2}}\, x_2^{\frac{\alpha_2}{1+\beta_2}}} \qquad (2.\ 14)$$

Nótese que los exponentes que acompañan a los outputs suman la unidad, lo cual indica que la función de distancia (2.14) y, en general, todas las funciones de

[32] Esto es, $D_O(y/D_O,x)\equiv 1$.

distancia orientadas a los outputs son, por definición, homogéneas de grado uno en outputs.[33]

La estimación de la función de distancia (2.14) exige realizar algunos supuestos sobre el grado de ineficiencia. Si se supone que todas las empresas son eficientes (como cuando se estima (2.10) por mínimos cuadrados ordinarios), los parámetros de la expresión (2.14) se pueden estimar igualando D_O a la unidad, e imponiendo homogeneidad de grado uno en outputs.

En definitiva, existe un amplio abanico de posibilidades para representar una tecnología determinada. La elección de una u otra dependerá de factores como las peculiaridades de la tecnología a representar, la disponibilidad de datos o el objetivo del análisis empírico a realizar. Dadas las limitaciones de espacio existentes, en este texto se ha optado por la simplicidad y la intuición de la función de producción, por lo que a lo largo de este capítulo nos centraremos en dicha función como forma de representar la tecnología.

2.2.5. Axiomas de la tecnología

Antes de pasar a analizar las características más relevantes de la tecnología, es preciso señalar que siempre que se realiza un análisis aplicado de la producción hay una serie de hipótesis mantenidas que no son sometidas a contraste sino que se suponen ciertas. Estos axiomas son simplificaciones plausibles de la tecnología que se hacen porque se consideran compatibles con el comportamiento observado a los productores en el mundo real y que se traducen en propiedades que deben cumplir las distintas representaciones de la tecnología.

En concreto, para que una función de producción sea considerada de buen comportamiento debe cumplir las siguientes propiedades: monotonicidad, cuasiconcavidad, continuidad, diferenciabilidad con producción nula en ausencia de inputs (es decir, $f(0)=0$).[34] Mientras que la ultima propiedad es sumamente intuitiva, el resto merecen alguna aclaración. La propiedad de monotonicidad refleja el hecho de que las empresas no contratan factores productivos (costosos) a no ser que su

[33] De forma análoga, todas las funciones de distancia orientadas a los inputs son homogéneas de grado uno en inputs.

[34] Ver Diewert (1982; p. 537-38).

aportación al proceso productivo sea positivo. Por otra parte, no es común observar un proceso productivo con dos inputs en el que se use exclusivamente uno de los inputs. En este sentido, la convexidad de la isocuantas hacia el origen (o, lo que es lo mismo, la cuasiconcavidad de la función de producción) garantiza que no se den soluciones de esquina en los procesos productivos.[35] Adicionalmente, el análisis neoclásico suele suponer a menudo que la función de producción es continua y diferenciable, lo cual no es estrictamente necesario pero sí muy conveniente ya que permite usar el cálculo diferencial para obtener los resultados más importantes de la teoría económica de la producción.

Un punto adicional de interés sobre la tecnología, que muchas veces pasa desapercibido en los análisis tanto teóricos como empíricos, es el supuesto implícito de que la tecnología es exógena a los productores. Efectivamente, la inmensa mayoría de los análisis de producción se supone que existe un stock de conocimiento tecnológico que está disponible para los productores. Esto quiere decir que los productores no toman decisiones sobre la tecnología y que, por tanto, ésta es exógena.[36]

2.3. Características de la tecnología

El objetivo principal del análisis empírico de la producción es la medición de la información económica que caracteriza el comportamiento de los agentes económicos. Puesto que en este caso se está estudiando la tecnología primal, la información relevante desde el punto de vista económico se puede resumir en tres características: las productividades marginales, los rendimientos a escala y las posibilidades de sustitución entre inputs.

2.3.1. Productividades marginales

La productividad marginal de un factor es una característica de la tecnología a corto plazo, ya que se define como la cantidad en la que varía la producción cuando

[35] Una función cuasicóncava es una transformación monótona de una función cóncava. Las funciones cóncavas aseguran que se usan todos los inputs en el proceso productivo pero implican necesariamente la existencia de productos marginales decrecientes.

se aumenta la utilización del factor en una unidad adicional (manteniendo constante la cantidad empleada de los demás factores). Se puede ver que se trata de una tasa de cambio, por lo que en términos matemáticos se puede definir la productividad marginal del factor j como:

$$PMaX_j = \frac{\Delta y}{\Delta x_j} \tag{2. 15}$$

Cuando el incremento en el input tiende a cero, la productividad marginal de un factor es igual a la derivada parcial de la función de producción con respecto a ese factor.

$$PMaX_j = \frac{\partial y}{\partial x_j} \tag{2. 16}$$

Se dice que una tecnología es monótona cuando las productividades marginales son siempre positivas. Esto significa que incrementos en las cantidades utilizadas de input no pueden hacer que disminuya la producción.

Además, un supuesto comúnmente utilizado es que la producción aumente a una tasa decreciente, lo que implica que el producto marginal sea decreciente. Esto se debe a la ley de los rendimientos decrecientes, que dice que si se utilizan cantidades adicionales de un factor variable en presencia de (al menos) un factor fijo, los incrementos de producción a partir de un determinado punto serán cada vez más pequeños, pudiendo llegar a ser negativos. Esta ley parece indicar que las funciones de producción tienen un máximo, por lo se violaría el axioma de la monotonicidad de la función de producción expuesto en la sección anterior.

Sin embargo, no se debe confundir la posibilidad teórica que tienen todas las tecnologías de llegar a usar inputs en una cantidad tal que su producto marginal sea negativo, con la posibilidad de que se observen esas situaciones en la realidad. Bajo el supuesto de maximización de beneficios, los productores nunca usarán cantidades de un input hasta el punto de llegar a reducir su producción. Por tanto, si los resultados empíricos arrojan ese resultado, el investigador debe repasar las

---

[36] La tecnología exógena es, simplemente, un supuesto simplificador conveniente cuando el interés del análisis no se centra en la adopción de tecnología sino en su uso. Sin embargo, existen análisis empíricos en los que el centro de interés es el propio proceso de adopción de la tecnología, lo que da lugar que la tecnología pueda considerarse endógena (Mundlak, 1988).

estimaciones (o más probablemente buscar algún error de especificación) en vez de apresurarse a divulgar ese “descubrimiento”.

La comparación de las productividades marginales de los inputs se hace difícil ya que vienen expresadas en unidades de output por unidad de input. Por eso se suele emplear una medida adimensional de los cambios en la producción atribuibles a cambios en la utilización de un determinado input: la elasticidad. La elasticidad del output respecto al input *j* mide el porcentaje en que aumenta el output cuando dicho input aumenta en un uno por ciento. Es decir:

$$e_j = \frac{\partial f(x)}{\partial x_j} \frac{x_j}{f(x)} \tag{2. 17}$$

Aunque el producto marginal es, como se verá en epígrafes posteriores, un concepto muy relevante, no es una magnitud directamente observable ya que su calculo requiere disponer de una representación de la tecnología. Sin embargo, el concepto de producto medio, que se define como el cociente entre el output y la cantidad empleada de un determinado input, sí es directamente observable:

$$PMeX_j = \frac{y}{x_j} \tag{2. 18}$$

El producto medio es, por su facilidad de cálculo, una característica de la tecnología frecuentemente empleada en argumentos informales sobre economía de la producción. Diferenciando la expresión (2.18) con respecto al input *j* se obtiene la relación que existe entre ambos tipos de productividad:

$$PMaX_j = PMeX_j + x_j \frac{\partial PMeX_j}{\partial x_j} \tag{2. 19}$$

El producto marginal es mayor que el producto medio cuando el producto medio es creciente ($\partial PMeX_j / \partial x_j > 0$), menor que el producto medio cuando este es decreciente ($\partial PMeX_j / \partial x_j < 0$), y coinciden en el punto máximo del producto medio ($\partial PMeX_j / \partial x_j = 0$). Desde un punto de vista económico, si unidades adicionales de un input variable aportan más al producto que la media de las anteriores (PMa>PMe), el producto medio tiene que crecer. Si las unidades adicionales de input variable aportan menos al producto que la media de las anteriores (PMa<PMe) el producto medio tiene que decrecer.

En el caso de la función Cobb-Douglas los parámetros son las elasticidades de producción. En efecto, para el caso del input 1 en la ecuación (2.9) se tiene:

$$e_1 = \frac{\partial y}{\partial x_1}\frac{x_1}{y} = \alpha_1 A x_1^{\alpha_1 - 1} x_2^{\alpha_2}\frac{x_1}{y} = \alpha_1$$

Se suele suponer que el signo de los coeficientes es positivo y menor que 1. Por tanto, en esta forma funcional, las productividades marginales (PMa) van siempre por debajo de la función de producto medio (PMe), y el cociente entre la PMa y la PMe de cada factor es constante. Efectivamente, para el caso del factor X_1 se tiene:

$$PMaX_1 = \alpha_1 \frac{y}{x_1} = \alpha_1 PMeX_1 \quad \Rightarrow \frac{PMaX_1}{PMeX_1} = \alpha_1 \tag{2. 20}$$

2.3.2. Optimo físico y óptimo económico

Una aplicación del concepto de productividad marginal es la diferencia entre el óptimo físico y el óptimo económico. Por óptimo físico se entiende la cantidad del input *j* para la que la función de producción tiene un máximo. Esa cantidad (X_j^{max}) será aquélla para la que la pendiente de la función de producción sea cero. Es decir, la condición del óptimo físico es que $PMaX_j=0$.

Por su parte, el óptimo económico, que se representa como X_j^*, es la cantidad del input *j* que hace que el beneficio sea máximo. El beneficio (Π) de una empresa es igual a la diferencia entre ingresos y costes. Si las empresas son precio aceptantes (esto es, los precios son exógenos para la empresa), tanto los ingresos como los costes son función de los inputs utilizados. En este caso, el problema de la empresa puede verse como qué cantidades de inputs debe utilizar para maximizar el beneficio, es decir:

$$\max_{x_1, x_2, \ldots, x_n} \Pi(x_1,\ldots,x_n) = P_y f(x_1,\ldots,x_n) - \sum_{j=1}^{n} w_j x_j \tag{2. 21}$$

donde P_y es el precio del output y w_j es el precio del input j. Las condiciones de primer orden para la maximización del beneficio son:

$$\frac{\partial \Pi}{\partial x_j} = P_y \frac{\partial f}{\partial x_j} - w_j = 0 \quad \Rightarrow \quad P_y f_j = w_j \tag{2. 22}$$

donde f_j es la productividad marginal del input *j*. Es decir, las condiciones anteriores indican que la empresa debe usar cantidades adicionales de un factor variable hasta que el valor del producto marginal del factor sea igual a su precio. Es decir:

$$VPMaX_j = w_j \quad \Rightarrow \quad PMaX_j = \frac{w_j}{P_y} \quad \forall j \qquad (2.\ 23)$$

Resolviendo esta ecuación para X_j, se obtiene la cantidad del input *j* para la que se alcanza el óptimo económico, X_j*. Como se puede ver en (2.23), el óptimo físico y el óptimo económico coincidirán en el caso de que el factor productivo sea gratis (w_j=0). En caso contrario, cuando el precio real del input sea positivo, el óptimo físico siempre será mayor que el económico debido a que el producto marginal es decreciente (ver Figura 2.3).

La gran relevancia de la ley de los rendimientos decrecientes no es tenida en cuenta por la "sabiduría convencional" de ciertos sectores. Por ejemplo, en el sector ganadero es frecuente considerar como bueno que las explotaciones intenten conseguir una producción *por* vaca cada vez mayor. Este es un claro error, como indica la Figura 2.3, en la que la producción puede ser los litros de leche y el input, la cantidad de vacas. Obsérvese que obtener la máxima producción por vaca es en realidad maximizar el producto medio de cada vaca. Por lo tanto, es beneficioso usar unidades adicionales del factor variable. Este proceso aumenta el beneficio hasta el máximo (producto marginal igual a precio del factor variable) pero reduce el producto medio.

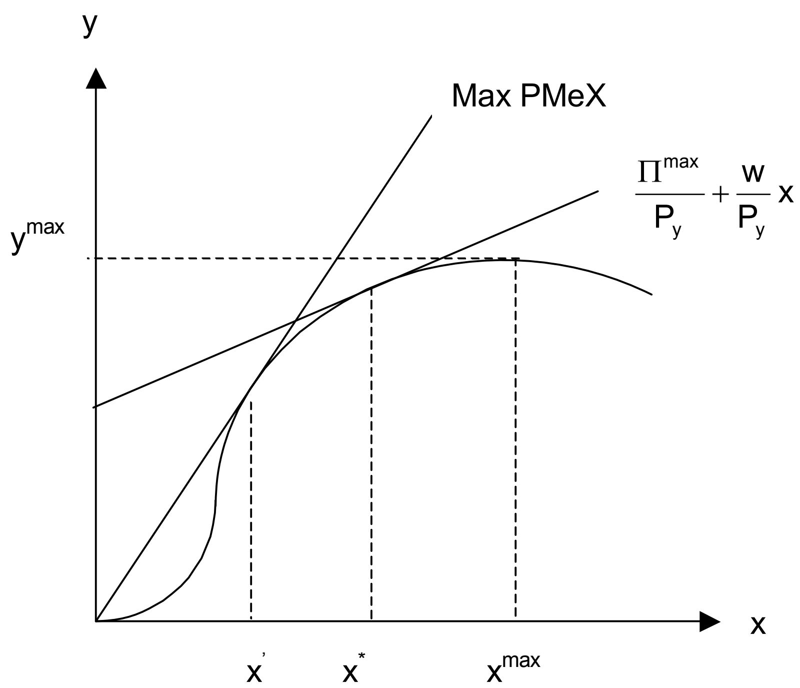

Figura 2.3. *Diferencia entre óptimo físico y económico*

El ejemplo de producción lechera anterior es parecido a la cuestión general de si tiene sentido para una empresa maximizar el producto medio de un factor. Como se puede ver en la Figura 2.3 la cantidad de factor para la que el producto medio es máximo es menor que el óptimo económico. Por tanto, parece claro que la maximización del producto medio *per se* no es un buen objetivo para las empresas. Como dice el propio Stigler (1961; p. 47): "...[es] una proposición básica de la economía que uno nunca debería mirar a los productos medios, sólo a los productos marginales". Y un poco más adelante: "Hasta lo que yo sé, ni un solo principio teórico de importancia puede hacerse sobre productos medios de los factores".

Para hacernos un idea de lo diferentes que pueden ser el óptimo económico y el óptimo físico, a continuación se calcula la diferencia entre ambas cantidades para el caso del ejemplo de la fertilización de la hierba mostrado en el Capítulo 1. La producción de la hierba en función del nitrógeno (N) puede especificarse como una función de producción cuadrática:

$$y = \beta_0 + \beta_1 N + \beta_2 N^2 \qquad (2.\ 24)$$

La función del producto marginal del nitrógeno se obtiene derivando la anterior función de producción con respecto a la cantidad del input:

$$PMaN = \frac{\partial y}{\partial N} = \beta_1 + 2\beta_2 N \tag{2. 25}$$

Igualando a cero la anterior expresión se obtiene que el óptimo físico será:

$$N_{max} = -\frac{\beta_1}{2\beta_2} \tag{2. 26}$$

mientras que el óptimo económico será:

$$N^* = \frac{\frac{W}{P_y} - \beta_1}{2\beta_2} \tag{2. 27}$$

La estimación por mínimos cuadrados de la ecuación (2.24) con los datos de la Tabla 1.1 permite obtener las siguientes estimaciones de los parámetros de la función:

$$\hat{\beta}_0 = 894 \quad \hat{\beta}_1 = 36.31 \quad \hat{\beta}_2 = -0.228$$

Si el precio de la hierba es 3 céntimos de Euro por kilogramo y el del nitrógeno 30 céntimos/kg, el óptimo físico es N^{max} =79.6 kg/ha y el óptimo económico es N*=57.7 kg/ha.

2.3.3. Rendimientos a escala

El concepto de rendimientos a escala hace referencia a cómo varía la producción cuando aumentan todos los inputs en la misma proporción, es decir, cuando se aumenta la escala de la empresa. Supóngase que se está utilizando un vector de inputs *x* para producir una cierta cantidad de output *y*, decidiendo multiplicar (escalar) todos los inputs por una cantidad determinada λ>0. Se dice que la tecnología presenta respectivamente rendimientos decrecientes, constantes y crecientes a escala si se cumple:

$$\begin{aligned} f(\lambda x) &< \lambda f(x) \quad , \ \forall \lambda \geq 0 \\ f(\lambda x) &= \lambda f(x) \quad , \ \forall \lambda \geq 0 \\ f(\lambda x) &> \lambda f(x) \quad , \ \forall \lambda \geq 0 \end{aligned} \tag{2. 28}$$

En el primer caso, la duplicación de inputs en un proceso productivo da lugar a un output menor que llevar a cabo dos procesos productivos separados. En el segundo caso, el resultado de duplicar los inputs es equivalente a llevar a cabo dos procesos productivos separados. En el último caso, la acumulación de inputs en un proceso productivo es beneficiosa, ya que el producto de un proceso productivo con el doble de inputs es superior al de dos procesos productivos separados.

Un modo intuitivo de entender este concepto es preguntarse que debería hacer una multinacional presente en varios países o regiones que se encontrase con cada una de las tres tecnologías descritas. En el caso de los rendimientos decrecientes a escala está claro que no tiene ventajas productivas el centralizar los recursos en una única planta y transportar el output a los lugares de consumo. Es más razonable, producir en cada uno de los países y ahorrarse los costes de transporte. Por otra parte, esta decisión es óptima si los costes de coordinación no son tan elevados que superan los beneficios de la producción descentralizada en los diferentes países. En el caso de los rendimientos constantes a escala parece que desde el punto de vista de la producción es indiferente producir en un solo país o en varios. En este caso, la respuesta vendrá dada por la relación entre los costes de coordinación (favorecen la decisión de producción centralizada) y los costes de transporte del output (favorecen la decisión de producción descentralizada). Por último, en el caso de rendimientos crecientes a escala la decisión razonable es producir en un solo país a menos que los costes de transporte sean tan elevados que superen las ventajas productivas que se obtienen de concentrar la producción en un único proceso productivo.

Una cuestión importante es qué tipo de tecnologías da lugar a diferentes rendimientos a escala. En principio, siempre que sea posible escalar el proceso, van a existir al menos rendimientos constantes a escala. Esto es así porque si se puede replicar exactamente un proceso, por ejemplo, construyendo una fábrica igual al lado de la primera, cabe esperar que después de duplicar todos los inputs se duplique también el output.

Los rendimientos decrecientes a escala se explican por la imposibilidad de duplicar tanto la cantidad como la calidad de todos los factores. Respecto a la cantidad, cuándo una empresa duplica su tamaño ¿se duplica al mismo tiempo la capacidad de los gestores? Así, es probable que en una pequeña ganadería, el ganadero conozca con bastante exactitud cuántas y qué vacas están en celo, cuestión sumamente difícil cuando se disponen de cientos de vacas. En esta circunstancia se observarían rendimientos decrecientes a escala. El mismo resultado se observaría si, por ejemplo, en el análisis Ricardiano[37] se puede duplicar una explotación agraria en términos de superficie cultivada, capital y trabajo. Sin embargo, la nueva explotación estará situada sobre un terreno de distinta fertilidad. En el razonamiento Ricardiano, el terreno más fértil ha sido cultivado en primer lugar, por lo que la segunda explotación agraria produce menos que la primera y el output no llegará a duplicarse.

Por otra parte, no está claro que haya muchas actividades productivas que puedan proporcionar rendimientos crecientes a escala. Samuelson y Nordhaus analizan en su libro de texto el ejemplo de la tecnología de un oleoducto, consistente básicamente en una tubería. Si se duplica el diámetro de la tubería, el volumen que puede circular por unidad de tiempo (output) se ve aumentado por 4, por lo que existen rendimientos crecientes a escala.

Un ejemplo de moda es el de las redes de transmisión de datos.[38] El objetivo es conectar varias ciudades que tienen la misma población y están situadas a la misma distancia cada una de la siguiente. Por simplicidad, hacemos que la distancia entre cada par de ciudades consecutivas sea 1. Con una unidad de fibra óptica se pueden unir dos ciudades A y B y realizar una conexión. Con una segunda unidad de fibra óptica se puede unir la ciudad B con la C. En este caso, las conexiones posibles no son sólo A-B y B-C, sino también A-C. Es decir, duplicando la cantidad de fibra óptica (el input) se triplica el output (número de conexiones posibles). En general, el número de conexiones posibles cuando se usan N unidades de fibra óptica viene dado por:

$$f(N) = \frac{N(N-1)}{2} \qquad (2.\ 29)$$

donde, f(N) representa el número de conexiones posibles. La función de producción en (2.29) presenta rendimientos crecientes a escala.

---

[37] La cita erudita apropiada es Ricardo (1959), Una aproximación intuitiva a esta cuestión puede encontrarse en Heilbroner (1985). Para un tratamiento más formal puede consultarse Blaug (1988). Sin embargo, esta cuestión suele tratarse razonablemente en cualquier libro de texto de introducción a la economía.

[38] Este tipo de análisis se relaciona actualmente con la denominada “nueva economía”. Sin embargo, un fenómeno semejante ocurrió hace más de un siglo con la instalación de las líneas telegráficas (Krugman, 1999).

Es importante señalar que la "maldición" Ricardiana de los rendimientos decrecientes a escala puede aparecer en este ejemplo con cierta facilidad. En un modelo más realista las ciudades tendrían distintos tamaños de población y estarían situadas a distintas distancias. En ese sentido, unir una nueva ciudad puede dar lugar a una conexión menos valiosa (por el número de ciudadanos que puede usarlos). En este caso, las empresas que crean la red tienen un incentivo para unir las ciudades más grandes y más cercanas. Ese incentivo, se reduce a medida de las ciudades que se unen son más pequeñas y están más alejadas.

Los primeros estudios que estimaron funciones de producción Cobb-Douglas encontraron que los rendimientos a escala eran casi siempre igual a 1. El motivo es que bajo el supuesto de competencia perfecta, las condiciones de maximización del beneficio son:

$$\frac{\partial \pi}{\partial x_j} = p\alpha_j \frac{y}{x_j} = w_j \quad \Rightarrow \quad \alpha_j = \frac{w_j x_j}{py} \tag{2. 30}$$

Es decir, los coeficientes de la CD son las proporciones de las rentas de cada input en el valor del output total. Por tanto, no es de extrañar que las primeras estimaciones de funciones de producción CD usando como inputs capital y trabajo obtuvieran estimaciones de los coeficientes muy cercanas a las "shares" observadas para cada input (0.75 para el trabajo y 0.25 para el capital, aproximadamente).

Los diferentes tipos de rendimientos a escala que se acaban de definir son globales. Sin embargo puede ocurrir que una tecnología presente rendimientos crecientes a escala en algunos intervalos de producción y decrecientes en otros. Por lo tanto, en ocasiones resulta útil emplear una medida local de los rendimientos de escala. Esta medida es la elasticidad de escala, que se define como el cambio porcentual que se produce en el output debido a un cambio de un uno por ciento en todos los inputs. Matemáticamente:

$$e(x) = \left.\frac{\partial \ln f(\lambda x)}{\partial \ln \lambda}\right|_{\lambda=1} \tag{2. 31}$$

Hay que destacar que la anterior expresión se evalúa en λ=1, lo cual quiere decir que se está considerando la escala actual de operaciones. Cuando la elasticidad de escala es mayor, igual o menor que uno se dice que la tecnología presenta rendimientos crecientes, constantes o decrecientes a escala *localmente*.

Puesto que los rendimientos a escala hacen referencia a cambios en la producción provocados por variaciones en la cantidad empleada de *todos* los factores, la elasticidad de escala se puede interpretar como el efecto acumulado de aumentar (o disminuir) cada uno de los inputs involucrados en el proceso de producción. En este sentido, la elasticidad de escala se puede expresar como una *suma* de elasticidades parciales, esto es:

$$e(x) = \frac{\partial \ln f(\lambda x)}{\partial \ln \lambda} = \frac{\partial f(\lambda x)}{\partial \lambda} \frac{\lambda}{f(\lambda x)} = \frac{\lambda}{f(\lambda x)} \sum_{j=1}^{k} \frac{\partial f(x)}{\partial \lambda x_j} \frac{\partial \lambda x_j}{\partial \lambda} \tag{2. 32}$$

Si se evalúa esta expresión en λ=1, se obtiene:

$$e(x) = \sum_{j=1}^{k} \frac{x_j}{f(x)} \frac{\partial f(x)}{\partial x_j} = \sum_{j=1}^{k} e_j(x) \tag{2. 33}$$

donde $e_j(x)$ es la elasticidad-output del input j. En la práctica, para calcular la elasticidad de escala en un punto, lo único que hay que hacer es por tanto sumar las elasticidades de cada input, evaluadas en dicho punto.

2.3.4. La sustitución entre inputs

La capacidad de sustitución entre inputs es una característica clave de cualquier tecnología.[39] El productor puede reducir el impacto de una subida en el precio de los inputs o aprovechar la caída del precio de un input si es capaz de sustituir unos inputs por otros en el proceso productivo. Esto se debe a que la función de producción permite que haya varios vectores de inputs capaces de producir al menos un nivel de output determinado, por lo que para producir un nivel de output determinado existe la posibilidad de sustituir unos inputs por otros.

De lo anterior se deduce que la sustitución entre inputs se realiza a lo largo de una isocuanta, por lo que la capacidad de sustitución entre inputs de una tecnología queda reflejada en la forma de la isocuanta. Parece, por lo tanto, que en procesos productivos con dos inputs se podría medir la capacidad de sustitución por la pendiente de la isocuanta, que mide el coste de oportunidad de un input en términos de otro.

[39] En una tecnología multiproducto se podría analizar la sustitución entre outputs.

Dado que el valor de la pendiente depende de las unidades de medida, se han desarrollado medidas adimensionales de las posibilidades de sustitución entre un par de inputs denominadas *elasticidades de sustitución*. Este concepto fue introducido por Hicks (1932) como una medida adimensional de la curvatura de una isocuanta en el caso de una tecnología con dos factores productivos y puede expresarse como:

$$\sigma = \frac{\partial \ln(x_2/x_1)}{\partial \ln(f_1/f_2)} \tag{2. 34}$$

La elasticidad de sustitución puede definirse, por tanto, como el porcentaje en que cambia el cociente entre dos inputs cuando la relación marginal de sustitución técnica cambia en un uno por ciento. Bajo el supuesto de minimización de costes, la relación marginal de sustitución técnica entre dos inputs se iguala al cociente de sus precios, por lo que la elasticidad de sustitución definida en (2.34) puede expresarse como el porcentaje en que cambia la proporción entre dos inputs cuando el cociente de sus precios cambia en un uno por ciento:

$$\sigma = \frac{\partial \ln(x_2/x_1)}{\partial \ln(w_1/w_2)} \tag{2. 35}$$

La generalización del concepto de elasticidad de sustitución para tecnologías con más de dos factores es bastante complicado. Por ese motivo, se han desarrollado elasticidades de sustitución *parciales*, es decir, entre un par de inputs. La más conocida es la elasticidad de sustitución de Allen, que es una generalización de la expresión (2.34) para el caso de varios inputs y puede expresarse como:

$$\sigma_{ij} = \frac{\sum x_k f_k}{x_i x_j} \frac{F_{ij}}{F} \tag{2. 36}$$

donde F_{ij} es el cofactor de f_{ij} en el hessiano de la función de producción y F es el determinante del hessiano de la función de producción. Para el caso de dos inputs, las expresiones (2.34) y (2.36) coinciden. Binswanger (1974) demuestra que la expresión (2.36) puede escribirse como:

$$\sigma_{ij} = \frac{e_{ij}}{s_j} \tag{2. 37}$$

donde e_{ij} representa la elasticidad del input i con respecto al precio del input j y s_j representa la proporción que representa el coste del factor j en el coste total.

En la función de producción Cobb-Douglas la elasticidad de sustitución de Allen es siempre igual a 1. Esto es fácil de comprobar a partir de la fórmula para la relación marginal de sustitución técnica de X_1 por X_2 ($RMST_{12}$):

$$RMST_{12} = \frac{PMaX_1}{PMaX_2} = \frac{\alpha_1 x_2}{\alpha_2 x_1} \quad \Rightarrow \quad \ln RMST_{12} = \ln\frac{\alpha_1}{\alpha_2} + \ln\frac{x_2}{x_1} \qquad (2.\,38)$$

Usando la fórmula para la elasticidad de sustitución de Allen en (2.34) se tiene que:

$$\sigma = \frac{\partial \ln(x_2 / x_1)}{\partial \ln RMST_{12}} = 1 \qquad (2.\,39)$$

Como un intento de superar las rigideces de la Cobb-Douglas, y, en particular, que la elasticidad de sustitución sea siempre igual a 1, Arrow, Chenery, Minhas y Solow (1961) introdujeron la denominada función CES (Constant Elasticity of Substitution, es decir, Elasticidad de Sustitución Constante). Dicha función se puede expresar como sigue:

$$y = \gamma(\delta x_1^{-\theta} + (1-\delta)x_2^{-\theta})^{-\frac{\rho}{\theta}} \qquad (2.\,40)$$

donde γ es un parámetro de eficiencia o de status tecnológico, δ es un parámetro de distribución, ρ recoge los rendimientos a escala y θ está relacionado con la elasticidad de sustitución. En efecto, la elasticidad de sustitución de Allen (σ) es constante, aunque no necesariamente igual a 1:

$$\sigma = (1+\theta)^{-1} \qquad (2.\,41)$$

La función CES se convierte en la CD cuando θ=0 (y, por tanto, la elasticidad de sustitución es igual a 1). Asimismo, si θ=-1 y ρ=1 la función se convierte en lineal.

Algunos autores han cuestionado la relevancia de la elasticidad de sustitución de Allen como medida de sustitución entre inputs. Blackorby y Russell (1989) demuestran que con una tecnología multifactor la elasticidad de sustitución de Allen no es una medida que indique las posibilidades de sustitución entre inputs. En cierto modo, la elasticidad de sustitución de Allen es una generalización matemática de una medida de sustitución que es razonable para dos inputs, pero que pierde su significado económico en un contexto de más de dos inputs. En su lugar, recomiendan el uso de la elasticidad de sustitución de Morishima, que se puede definir como sigue:

$$M_{ij} = \left.\frac{\partial \ln\left(x_j/x_i\right)}{\partial \ln\left(w_i/w_j\right)}\right|_{dw_i=0} \tag{2. 42}$$

o, alternativamente, como:

$$M_{ij} = \frac{f_j x_j}{\sum_k f_k x_k}\left(\sigma_{ij} - \sigma_{jj}\right) \tag{2. 43}$$

La elasticidad de sustitución de Morishima, al contrario que la de Allen, no es simétrica, es decir, $M_{ij} \neq M_{ji}$. Asimismo, dos factores complementarios según la elasticidad de Allen pueden ser sustitutivos según la de Morishima. Blackorby y Russell muestran en su trabajo que esta medida de sustitución proporciona resultados mucho más razonables en un contexto multifactorial que la elasticidad de sustitución de Allen.

2.4. Características "útiles" de la tecnología

En el apartado anterior se han analizado las características de la tecnología que son relevantes desde un punto de vista económico. A continuación, se presentan algunas características de la tecnología que, si bien no son tan relevantes para el estudio del comportamiento de las empresas, su importancia radica en que facilitan notablemente el trabajo empírico. Entre ellas podemos destacar la homogeneidad, la homoteticidad y la separabilidad entre factores de producción.

2.4.1. Homogeneidad

Una función de producción es homogénea de grado k si al multiplicar todos los inputs por una misma constante λ, la producción queda multiplicada por λ^k. Matemáticamente:

$$f\left(\lambda x_1, \lambda x_2, \ldots, \lambda x_n\right) = \lambda^k f\left(x_1, x_2, \ldots, x_n\right) \tag{2. 44}$$

donde λ es cualquier numero real. Si la función de producción es homogénea de grado k, aplicando el teorema de Euler se obtiene el siguiente e interesante resultado:

$$\sum_j \frac{\partial f}{\partial x_j} x_j = ky \tag{2. 45}$$

o, lo que es lo mismo,

$$k = \sum_j \frac{\partial f(\cdot)}{\partial x_j}\frac{x_j}{f(\cdot)} = \sum_j e_j(x) = e(x) \tag{2. 46}$$

Es decir, la suma de las elasticidades output de todos los inputs es igual al grado de homogeneidad de la función. Por tanto, existe una relación directa entre homogeneidad y rendimientos a escala. Si la función de producción es homogénea de grado k, la función tendrá rendimientos a escala crecientes, constantes o decrecientes, dependiendo de si k es mayor, igual o menor que 1.

La utilidad empírica de este resultado es doble. Por una parte, dado que el grado de homogeneidad es constante, una forma de “imponer” un tipo de rendimientos a escala en todo el dominio de datos, es estimar una función homogénea. Por otra parte, el número de parámetros a estimar se reduce si se cumple algún tipo de homogeneidad. Para verlo, supongamos que estamos interesados en estimar la siguiente función de producción:

$$y = \alpha_1 x + \alpha_2 x^2 \tag{2. 47}$$

Si se multiplica la cantidad de input empleada por λ, la producción resultante, que se puede denominar $y(\lambda)$, sería igual a:

$$y(\lambda) = \lambda\left[y + \alpha_2 x^2(\lambda - 1)\right] \tag{2. 48}$$

La ecuación (2.48) indica que el nivel de producción aumentará en la misma proporción que el input (esto es, $y(\lambda)=\lambda\cdot y$) si $\alpha_2=0$. Por lo tanto, la homogeneidad de grado uno de la función (2.47) se cumplirá cuando este parámetro sea nulo. Es inmediato deducir, entonces, que si la función de producción (2.47) fuera homogénea de grado uno, no sería necesario estimar el parámetro α_2.

Es fácil comprobar que la función Cobb-Douglas es homogénea y que el grado de homogeneidad coincide con la suma de los exponentes de los inputs.

$$f(\lambda x_1, \lambda x_2) = \lambda^{\alpha_1+\alpha_2} x_1^{\alpha_1} x_2^{\alpha_2} = \lambda^{\alpha_1+\alpha_2} y \tag{2. 49}$$

Por tanto, los rendimientos a escala son constantes a lo largo de toda la función e iguales a la suma de los exponentes de los factores. Si se supone que la función es homogénea de grado 1, en este caso $\alpha_1+\alpha_2=1$, y la función Cobb-Douglas puede escribirse como:

$$\ln(y/x_1) = \alpha_0 + \alpha_2 \ln(x_2/x_1) \qquad (2.\ 50)$$

Este es un buen ejemplo para comprender que si la función de producción es homogénea, es posible reducir el número de parámetros a estimar. En efecto, si la función Cobb-Douglas es homogénea sólo es necesario estimar dos parámetros en vez de tres, ya que ahora α_1=1-α_2.

Una propiedad importante de una función homogénea de grado k, es que sus primeras derivadas son homogéneas de grado k-1. Por tanto, si una función de producción es homogénea de grado uno, las productividades marginales de los factores van a ser homogéneas de grado cero. Es decir, si se duplica la cantidad de factores, la productividad marginal no cambia. Una consecuencia inmediata de este resultado es que la relación marginal de sustitución técnica, que se define como el cociente de productividades marginales, tampoco cambia con la escala de producción. Esta propiedad es conveniente para asegurar que algunas decisiones de la empresa son independientes de la escala de las operaciones.

2.4.2. Homoteticidad

Una función de producción homotética es una transformación monótona creciente de una función de producción homogénea. Por tanto, las funciones de producción homogéneas son homotéticas, aunque lo contrario no tiene por que ser necesariamente cierto. Una función de producción homotética se puede escribir como:

$$y = g[f\ (x_1, x_2, ..., x_n)] \qquad (2.\ 51)$$

donde f(·) es una función homogénea de algún grado arbitrario y g[·] es una transformación monótona.

Las funciones de producción homotéticas son una generalización de las funciones homogéneas que preservan las convenientes propiedades de las últimas. En concreto, en la familia de funciones de producción homotéticas la pendiente de las isocuantas no cambia a lo largo de cualquier radio vector que sale del origen. Por lo tanto, el mapa completo de isocuantas puede conocerse a partir de una sola, ya que las restantes son expansiones o contracciones a escala de la misma. Sin embargo, las funciones homogéneas presentan el mismo nivel de rendimientos a escala para cualquier nivel de producción. Esta propiedad puede ser demasiado restrictiva ya que no es de esperar que se puedan obtener indefinidamente los mismos resultados con independencia de la expansión de la actividad que se realice. Una adecuada transformación monótona de la función homogénea puede hacer que los rendimientos a escala varíen con el nivel de producción.

La función Cobb-Douglas es homotética. Una sencilla forma de comprobarlo es que la pendiente de las isocuantas sólo depende de la proporción de factores, es decir, del radio vector en el que nos movemos.

$$\frac{dx_2}{dx_1} = -\frac{\alpha_1}{\alpha_2}\frac{x_2}{x_1} \qquad (2.\ 52)$$

Las funciones homotéticas, por lo tanto, son idóneas para modelizar procesos productivos con rendimientos a escala que cambian dependiendo del nivel de producción, ya que se consigue este objetivo con un número de parámetros a estimar muy reducido.

2.4.3. Separabilidad

La condición de separabilidad se refiere a la estructura interna de la función de producción. Se dice que una función es separable en un subconjunto de inputs si las tasas marginales de sustitución entre los inputs que pertenecen a dicho subconjunto son independientes de las cantidades empleadas de otros inputs. En concreto, dada la función de producción

$$y = f(x_1, x_2, x_3) \qquad (2.\ 53)$$

los dos últimos inputs (x_2 y x_3) son separables del primer input (x_1) si se cumple que:

$$\frac{\partial(f_2 / f_3)}{\partial x_1} = 0 \qquad (2.\ 54)$$

El significado de esta expresión es que la pendiente de una isocuanta entre los inputs x_2 y x_3 no depende de lo que ocurra con el primer input.. Esto implica que las decisiones óptimas respecto a los dos últimos factores productivos no se ven afectadas por las decisiones tomadas respecto al primer input, x_1. Por tanto, si la función de producción es separable en algún grupo de factores, las intensidades relativas en la utilización de dichos factores pueden ser optimizadas (y analizadas) separadamente del resto de factores productivos (Berndt y Christensen, 1973).

La separabilidad de un conjunto de inputs implica la existencia de un subproceso que se puede analizar independientemente. En términos matemáticos, la propiedad de separabilidad implica la existencia de una función que permite agregar todos los inputs involucrados de un subproceso y sustituir dichos inputs por un input "agregado".[40] En otras palabras, la separabilidad permite realizar agregaciones de inputs sin que ello suponga la introducción de restricciones *ad hoc* en la modelización de la tecnología. Así, por ejemplo, si la función de producción (2.53) viene dada por la expresión:

$$y = (x_1 + x_2 x_3)^2 \tag{2. 55}$$

se puede comprobar fácilmente que la condición de separabilidad (2.54) únicamente se cumple para el par de inputs x_2 y x_3. Esto significa que se puede expresar la función de producción (2.55) sin pérdida de generalidad como sigue:

$$y = (x_1 + z)^2 \quad , \quad z = g(x_2, x_3) = x_2 x_3 \tag{2. 56}$$

La agregación es conveniente desde el punto de vista teórico como un elemento de simplificación. Desde el punto de vista empírico es importante porque permite reducir la dimensión del problema o estimar modelos cuando no se dispone de datos desagregados.

No es frecuente encontrar ejemplos didácticos de separabilidad en producción por lo que vamos a intentar construir uno a continuación. Una sofisticada empresa de servicios financieros abre un servicio de cafetería que sirve encargos en los puestos de trabajo. El hecho de no tener que salir del edificio o no tener que caminar hasta la cafetería hace que el producto de la empresa aumente. Este incremento de producción de la empresa financiera es independiente de cómo se organiza la cafetería. No importa si tienen muchas máquinas y pocos empleados o viceversa. Lo único que importa es qué nivel de servicio se ofrece a los empleados y cuánto tiempo ahorran. En este sentido, los inputs del servicio de cafetería serían separables del resto de los inputs de la empresa.

[40] Otro problema distinto es encontrar la fórmula más adecuada para agregar un número determinado de inputs. A este respecto, Ritcher (1966) demuestra que bajo condiciones bastante generales el mejor índice viene dado por el índice Divisia o, su versión discreta, el índice de Törnqvist.

Se puede generalizar la propiedad de separabilidad para empresas que produce más de un output. En este caso, la separabilidad entre outputs e inputs implica que se pueden agregar todos los outputs en un único output “agregado”, de tal forma que la tecnología de una empresa que produce dos outputs con dos inputs se puede escribir como:

$$g(y_1, y_2) = f(x_1, x_2) \qquad (2.\ 57)$$

donde $g(\cdot)$ se puede interpretar como una medida agregada de todos los outputs. En este caso, el análisis de empresas multiproducto se puede abordar agregando todos los outputs en un único valor (utilizando algún índice de cantidades en el que se pondere adecuadamente cada uno de los outputs, normalmente sus precios o sus participaciones el los ingresos totales) y estimar una función de producción para un único output.[41]

Obsérvese que la tecnología representada por la función de producción multiproducto (2.10) o la función de distancia (2.14) es separable entre outputs e inputs. El supuesto de separabilidad, en este caso, implica esencialmente que uno puede cambiar la combinación de inputs sin afectar a la pendiente de la curva de posibilidades de producción. Como ponen de manifiesto Coelli y Perelman (2000), este supuesto puede ser bastante restrictivo, por ejemplo, en el caso del transporte ferroviario: “cabe esperar que sea más fácil sustituir transporte de mercancías por transporte de pasajeros cuando la proporción entre trabajo y capital sea reducida [es decir, cuando existen muchos vagones de carga] que cuando sea elevada”.

2.5. Análisis dinámico de la producción

El análisis de la producción expuesto hasta aquí es estático en el sentido de que el tiempo no tiene una consideración específica. Se supone implícitamente que todos los períodos de tiempo son independientes, esto es, las decisiones realizadas en el pasado no afectan a los resultados del presente. Sin embargo, la realidad parece ser muy distinta puesto que la reacción de las empresas ante cambios en las variables exógenas no es instantánea en muchos casos, sino que tiene que transcurrir un tiempo hasta que la empresa se ajusta hasta su equilibrio a largo plazo.

[41] Esta alternativa, sugerida por Mundlak, (1963), presenta sin embargo serias dificultades.

El principal motivo por el que los ajustes no son instantáneos se debe a ciertas características de la tecnología y/o de los mercados. A este tipo de restricciones, que hacen que el ajuste de la dotación de algunos factores de producción presente determinados costes al margen del precio a pagar por las unidades de factor, se les conoce como costes de ajuste.

Las circunstancias que dan lugar a la aparición de costes de ajuste son muy diversas. Las principales están asociadas a características de la tecnología, tales como entrenamiento de trabajadores, aprendizaje en el uso de nuevo equipo, etc. Estos son los denominados costes de ajuste internos. Se trata de procesos necesarios para la incorporación de nuevas unidades de factor que originan reducciones respecto a la cantidad de output que la empresa puede producir si la totalidad de la dotación de factores cuasi-fijos[42] ha sido incorporada con anterioridad. También existen costes de ajuste externos, cuyo origen viene dado por causas externas a la empresa, tales como imperfecciones en los mercados financieros, costes de transporte, costes de búsqueda, etc. Generalmente, este tipo de costes se traduce en un desembolso monetario que la empresa debe realizar para conseguir ajustar su dotación de factores.

En presencia de costes de ajuste internos, la función de producción de la empresa, que se supone dos veces continuamente diferenciable, viene dada por:

$$y = f(x, k, \dot{k}) \qquad (2.58)$$

donde *y* es el output *x* es el vector de factores variables, *k* es la dotación de factores cuasi-fijos que se utiliza en un determinado momento (que, por simplicidad, supondremos que es únicamente el capital), y $\dot{k}$ representa la inversión realizada durante el período en los factores cuasi-fijos (esto es, la diferencia entre el stock de capital del período presente y el del anterior).

La existencia de costes de ajuste de tipo interno asociados al proceso de variación del stock de factores cuasi-fijos se modeliza con la inclusión de $\dot{k}$ como argumento de la función de producción.[43] En la Figura 2.4 se representa la producción en función de la inversión (tanto positiva como negativa) en factores cuasi-fijos,

---

[42] Se entiende por factores cuasi-fijos aquellos que presentan costes de ajuste y que, por tanto, originan la ralentización del proceso de ajuste hacia la situación de equilibrio a largo plazo.

[43] Dos de los trabajos pioneros en este área son Treadway (1970) y Mortensen (1973).

manteniendo constante la dotación de factores cuasi-fijos (k). Esto justifica la forma de la función, ya que dado un determinado stock de k, la producción será menor cuantas más unidades se hayan incorporado al proceso productivo en el período actual.

En otras palabras, en presencia de costes de ajuste internos, una empresa que haya añadido unidades de algún factor cuasi-fijo en el período actual no es capaz de producir tanto como otra empresa en la que todas sus unidades hayan sido incorporadas en períodos anteriores. La cantidad de output perdida como consecuencia del ajuste en la dotación de factores depende de la rapidez con la que se realice ese proceso de ajuste.

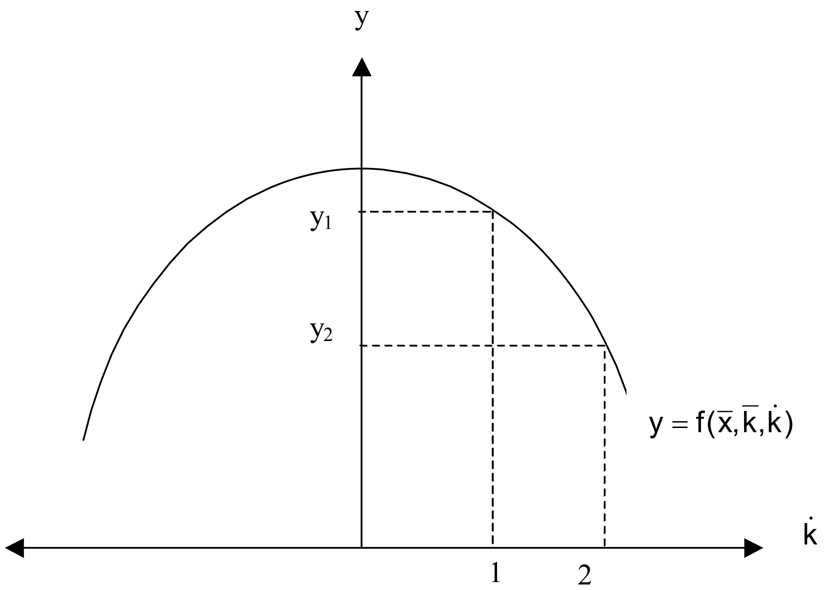

Figura 2.4. *Efecto sobre la producción de la inversión en factores cuasi-fijos*

El empresario, cuando planifica su contratación de factores, debe considerar cuál es la forma más rentable de ajustar la dotación de factores de la empresa. Un ajuste rápido permite aumentar la corriente de ingresos, pero puede elevar el coste del proceso de inversión. En este contexto, el agente tratará de determinar cuál es la "velocidad" de ajuste que permite maximizar la corriente descontada de beneficios, teniendo en cuenta el efecto de la "velocidad" tanto sobre la corriente de ingresos como sobre la corriente de costes. En este contexto, el objetivo del empresario será determinar la evolución temporal de la dotación de factores que hace máxima la corriente descontada de beneficios obtenida por la empresa. Esta es la diferencia fundamental con un modelo estático donde el objetivo de la empresa es la maximización de beneficios en cada período, ya que cada uno de los períodos es independiente de los demás.

Bibliografía

Arrow, K., H. Chenery, B. Minhas y R. Solow, (1961), "Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics*, 29, 155-73.

Beattie, B. y C.R. Taylor, (1993), *The Economics of Production*, Wiley, New York.

Berndt, E. y L. Christensen, (1973), "The Translog Function and the Substitution of Equipment, Structures and Labour in U.S. Manufacturing 1929-1968", *J. of Econometrics*, 1, 81-114.

Binswanger, H.P., (1974), "A Cost Function Approach to the Measurement of Factor Demand Elasticities of Substitution", *American J. of Agricultural Economics*, 56, 377-386.

Blackorby, C. y R. Russell, (1989), "Will the Real Elasticity of Substitution Please Stand up? (A Comparison of the Allen/Uzawa and Morishima Elasticities)", *American Economic Review*, 79, 82-88.

Blaug, M., (1988), *Teoría económica en retrospectiva*, Fondo de Cultura Económica.

Chavas, J.P. y T. Cox, (1988), "A Nonparametric Analysisis of Agricultural Technology", *American J. of Agricultural Economics*, 70(2), 303-310.

Cobb, C. y P. Douglas, (1928) "A Theory of Production", *American Economic Review*, 18(1), 139-72.

Coelli, T. y S. Perelman (2000), "Medición de la eficiencia técnica en contextos multiproducto", en A. Álvarez (Coordinador), *La medición de la eficiencia y la productividad*, Pirámide, Cap. 6, 113-138.

Diewert, W., (1982), "Duality Approaches to Microeconomic Theory", en K. Arrow y M. Intriligator (eds.), *Handbook of Mathematical Economics*, Vol. II, North-Holland.

Friedman, M., (1953), "The Methodology of Positive Economics", en *Essays in Positive Economics*, University of Chicago Press.

Hanoch, G. y M. Rothschild, (1972), "Testing the Assumptions of Production Theory: A Nonparametric Approach", *J. of Political Economy*, 80, 265-275.

Heilbroner, R., (1985), *Vida y doctrina de los grandes economistas*, Orbis.

Hicks, J., (1932), *A Theory of Wages,*

Klein, L., (1953), *A Textbook of Econometrics*, Row Peterson, New-York.

Krugman, P. (1999), "Networks and Increasing returns: a Cautionary tale", http://web.mit.edu/krugman/www/#slate.

Mortensen, D.T., (1973), "Generalized Costs of Adjustment and Dynamic Factor Demand Theory", *Econometrica*, 48, 1755-1762.

Mundlak, Y., (1963), "Specification and Estimation of Multiproduct Production Functions", *J. of Farm Economics*, 45, 433-443.

Mundlak, Y., (1988), "Endogenous Technology and the Measurement of Productivity", en S. Capaldo y J. Antle (eds.), *Agricultural Productivity. Measurement and Explanation*, Resources for the Future.

Ricardo, D., (1959), *Principios de Economía Política y Tributación*, Fondo de Cultura de Económica, México.

Richter, M.K., (1966): “Invariance Axioms and Economic Indexes”, *Econometrica*, 34, 739-755.

Shephard, R., (1953), *Cost and Production Functions*, Princeton University Press.

Stigler, G., (1961), “Economic Problems in Measuring Changes in Productivity”, en *Output, Input, and Productivity Measurement*, National Bureau of Economic Research, Princeton University Press.

Treadway, A.W., (1970), Adjustment Costs and Variable Inputs in the Theory of the Competitive Firm”, *J. of Economic Theory*, 2, 329-347.

Varian, H., (1984), “The Nonparametric Approach to Production Analysis”, *Econometrica*, 52, 579-598.

CAPÍTULO 3

EL ANALISIS DUAL DE LA PRODUCCION

En este capítulo se estudia la aproximación dual al análisis de la producción. En primer lugar se describen las diferencias entre la aproximación primal y la dual. A continuación se desarrolla la función de costes, sus propiedades teóricas y la dualidad con la función de producción.

3.1. Diferencias entre el análisis primal y el dual

El análisis empírico *primal*, estudiado en el capítulo anterior, se centra en la representación de la tecnología, que es uno de los elementos centrales en el análisis de las decisiones de producción. Puesto que las decisiones de los individuos se modelizan suponiendo un comportamiento optimizador sujeto a la restricción que implica la tecnología,[44] el modelo empírico genera predicciones cuantitativas contrastables sobre las decisiones de producción observadas. Por su parte, el análisis empírico *dual* analiza directamente los objetivos de los individuos (por ejemplo, maximización de beneficios o minimización de costes) y sus decisiones (por ejemplo, demanda de inputs y oferta de outputs). Este enfoque constituye un camino alternativo para modelizar las decisiones de producción observadas.

La existencia de dos vías de análisis tiene importantes ventajas. En ocasiones, problemas intratables desde el primal son sencillos o incluso triviales en el análisis dual (Varian, 1992; p. 81). Por ejemplo, en el enfoque primal, el análisis del efecto del precio de un factor en su demanda se efectúa a través del estudio de las soluciones de un enrevesado sistema de ecuaciones obtenido bajo el supuesto de minimización de costes. Puesto que este supuesto está ya incorporado en el análisis dual, ese efecto se puede deducir directamente de la función de costes.[45]

Desde el punto de vista empírico, el análisis dual supone una ampliación de las posibilidades que tiene un investigador para realizar un análisis empírico de la producción. En este sentido, la dualidad se puede ver como un camino alternativo cuando otro camino es dificultoso o está cortado. Por ejemplo, la estimación de una función de producción requiere la existencia de datos detallados sobre los inputs y outputs del proceso productivo. Puede darse el caso de que exista información sobre outputs producidos pero no sobre los inputs usados. En este caso, el camino de la función de producción está cerrado pero, si existen datos sobre los precios de los factores, la función de costes proporciona una oportunidad de analizar empíricamente el fenómeno de interés. Otro caso posible, es que no existan datos sobre la cantidad

[44] La tecnología implica una restricción en el sentido de que no todos los outputs pueden ser producidos con unos determinados inputs.

producida por una empresa. En este caso, parece que incluso el camino de la función de costes está cortado (dado que mide el coste de obtener un determinado nivel de producción). Sin embargo, si existen precios de los inputs y de los outputs la función de beneficio proporciona una oportunidad para el análisis empírico.

El concepto de dualidad se usa casi coloquialmente para referirse al hecho de que un mismo fenómeno puede estudiarse desde dos puntos de vista alternativos. En términos más rigurosos, la teoría de la dualidad contiene un conjunto de teoremas que aseguran la coherencia interna de los dos caminos de análisis. En otras palabras, la dualidad analiza las propiedades de las funciones de costes, beneficios, ingresos, etc.[46] que aseguran que son el resultado de un proceso de optimización a partir de una tecnología caracterizada por unas mínimas propiedades de regularidad.

3.2. La función de costes

En primer lugar, es necesario recordar la importante distinción que existe entre dos términos relacionados pero claramente distintos: coste y función de costes. Por coste se entiende la suma de los gastos en cada uno de los inputs; es decir, es una variable. Su expresión matemática es:

$$C = \sum_{j} w_j x_j \qquad (3.\ 1)$$

donde *C* es el coste, *w* son los precios de los inputs, *x* son las cantidades empleadas de los inputs y el subíndice *j* denota inputs.

Por su parte, la función de costes, C(y,w), es una relación entre el coste y un conjunto de variables explicativas que se determinan por un proceso de optimización. De hecho, la función de costes indica el mínimo coste de producir cada nivel de output, dados

[45] Paris y Caputo (1995) desgranan las ventajas e inconvenientes del análisis de la producción a través de la función de costes. En general, afirman que algunos problemas que se atribuyen a este enfoque son en realidad dificultades del tema que se analiza y no del enfoque. En este artículo, se analizan en detalle dos problemas económicos tratables en un enfoque dual pero que, hasta la fecha, no han podido ser analizados desde el primal.

[46] A estas funciones se les pone el apellido de duales ya que relacionan el valor óptimo de la función objetivo con los parámetros de la propia función objetivo y las restricciones. Así, por ejemplo, la función de costes relaciona el coste con los precios de los inputs (parámetros de la función objetivo) y con el nivel de output (un parámetro de la restricción tecnológica).

unos precios de los factores. Esta función se obtiene minimizando el coste de producir el output deseado, sujeto a la restricción que impone la tecnología. Es decir,

$$\begin{aligned} C(y,w) = \quad & \min_{x} \sum_{j} w_j x_j \\ \text{s.a.} \quad & y = f(x) \end{aligned} \tag{3. 2}$$

En el programa (3.2) se escogen las cantidades óptimas de los factores productivos, dados la tecnología y los precios de los inputs.

El lagrangiano correspondiente a ese programa de minimización es:

$$L = \sum w_j x_j + \lambda(y - f(x)) \tag{3. 3}$$

donde λ es el multiplicador de Lagrange.

Las condiciones de primer orden para la minimización de (3.3) pueden expresarse como:

$$\begin{aligned} & w_j - \lambda f_j(x) = 0 \quad \forall j \\ & y - f(x) = 0 \end{aligned} \tag{3. 4}$$

Las condiciones de primer orden para un proceso productivo con k inputs definen un sistema de k+1 ecuaciones con k+1 incógnitas (k inputs y λ). La solución a ese sistema de ecuaciones son las cantidades óptimas de los inputs, x(y,w) y el multiplicador de Lagrange evaluado en el óptimo. Las condiciones de segundo orden se pueden expresar en términos de un hessiano orlado (Silberberg, 2000; p.183). Desde un punto de vista económico, las condiciones de segundo orden implican la cuasiconcavidad de la función de producción.[47]

Las cantidades óptimas de inputs x(y,w) que se deben usar para minimizar el coste de producción cambian con los precios de los inputs y con la cantidad de output. A la relación funcional entre las cantidades óptimas de inputs, sus precios y las cantidades de output se la conoce como función de demanda de inputs y se suele escribir como:

[47] Una función (de producción) cuasicóncava tiene curvas de nivel (isocuantas) convexas hacia el origen. La cuasiconcavidad es una propiedad menos restrictiva que la concavidad. La concavidad implica necesariamente rendimientos decrecientes a escala mientras que esto no es necesario en una función cuasicóncava.

$$x_j = x_j(y, w) \tag{3. 5}$$

Sustituyendo las cantidades óptimas de factores en la expresión del coste, se obtiene:

$$C = \sum_j w_j x_j(y, w) = C(y, w) \tag{3. 6}$$

donde C(y,w) indica mínimo coste.[48]

Por tanto, la ecuación (3.6) establece una relación funcional entre el coste de producción y los precios de los inputs y la cantidad de output. Este resultado es sumamente importante desde el punto de vista empírico, ya que *limita* el número de variables a incluir en la estimación de una función de costes: sólo hay que emplear datos de precios de inputs y cantidades producidas.[49] Adicionalmente, la teoría económica proporciona una serie de propiedades que la relación funcional citada debe poseer para poder ser considerada una función de costes.

La discusión precedente contiene una gran parte de los elementos que explica la popularidad de la función de costes en el análisis empírico. En primer lugar, bajo el supuesto de minimización de costes, los inputs usados son cantidades elegidas por los productores y, por tanto, están afectadas por elementos aleatorios parecidos a los que afectan la transformación de inputs en outputs. Por tanto, existe un cierto escepticismo sobre la exogeneidad de los inputs a la hora de estimar una función de producción (ver Griliches y Mairesse, 1997). La función de costes permite analizar el proceso productivo sin sufrir este problema ya que las variables explicativas son los precios de los inputs (no las cantidades), que al determinarse en mercados perfectamente competitivos se pueden considerar exógenos. En segundo lugar, la función de costes permite analizar las características de la tecnología empleada por las empresas minimizadoras de costes de la misma forma que con una función de producción (véase el último apartado de esta sección). En tercer lugar, la función de costes permite modelizar con cierta facilidad procesos productivos multiproducto. De

[48] En la ecuación (3.6) se supone que el coste observado C es igual al coste mínimo C(y,w), por lo que estamos asumiendo que las combinaciones de inputs seleccionadas por las empresas son técnica y asignativamente eficientes.

[49] Pueden existir razones empíricas que justifiquen la inclusión de otras variables si, existen diferencias en la calidad en los inputs y outputs, cambios en la tecnología, etc.

hecho, todo el análisis previo se mantiene en el caso de que *y* represente un vector de outputs en vez de un número real.

3.2.1. Propiedades de la función de costes

A continuación se estudian las propiedades de la función de costes. Cada propiedad va acompañada de un breve comentario. La demostración rigurosa de estas propiedades puede consultarse en manuales generales de teoría económica como Varian (1992), Silberberg (2000) o en la monografía sobre economía de la producción de Chambers (1988).

La función de costes presenta las siguientes propiedades: 1) es monótona creciente en los precios de los inputs; 2) es monótona creciente en el nivel de output; 3) es homogénea de grado 1 en los precios de los inputs; y 4) es cóncava en los precios de los inputs.

La propiedad de monotonía implica una relación no decreciente entre el coste de producción y el precio de los inputs. Un descenso del precio de un input no puede conducir a una subida de costes, ya que, el coste baja incluso cuando el productor no reacciona ante dicho cambio (ver ecuación 3.1). Bajo el supuesto de minimización de costes, si hace algo será para bajar el coste todavía más. Por el contrario, una subida en el precio de un input no puede conducir a una bajada de costes. Si existe una combinación de inputs que reduce el coste, ésta habría sido adoptada antes de la subida del precio por un productor que minimiza costes. De forma similar, la segunda propiedad indica que el coste no puede reducirse cuando aumenta la cantidad producida de output. Un incremento del output no se puede lograr con menos inputs debido a la monotonicidad de la función de producción. Por el mismo motivo, una producción menor requiere la utilización de menos inputs, lo que implica una disminución del coste.[50]

La propiedad de homogeneidad lineal en el precio de los inputs refleja el hecho de que lo relevante para las empresas son los precios relativos y no los precios absolutos. Si los precios de todos los inputs cambian en la misma proporción, las decisiones óptimas no cambian (dado que los precios relativos no varían), por lo que el coste varía en la misma proporción en que cambian los precios de los inputs. Un ejemplo claro es el del cambio de unidades de cuenta. Si los precios de los inputs se expresan en euros en vez de en pesetas las decisiones óptimas no deberían cambiar. Sin embargo, el coste será aproximadamente 166 veces más pequeño con las nuevas unidades monetarias.

La idea que está detrás de la propiedad de concavidad en el precio de los inputs se puede ver intuitivamente usando el concepto de función de costes pasiva[51] en el Figura 3.1. Geométricamente, la concavidad se caracteriza porque la representación gráfica de la función de costes va por debajo de su tangente geométrica. Cuando sube el precio de un input de w_i^1 a w_i^2, una de las opciones que tiene el productor es no cambiar los inputs que usa. En este caso, el coste se incrementa linealmente con el precio del input que ha subido (gráficamente, se moverá de A hasta B por la tangente geométrica).[52] Sin embargo, generalmente existe la posibilidad de sustituir el input cuyo precio ha subido por otro para reducir el impacto sobre los costes, por lo que si el productor minimiza el coste y tiene posibilidades de sustitución el coste se incrementará en menor medida que sin sustitución. Por tanto, la función de costes irá por debajo de su tangente.

---

[50] En la práctica es frecuente, sin embargo, encontrase que la derivada de la función de costes respecto al output (precio de los inputs) es negativa en algunos puntos o para algunas empresas. Esto no implica necesariamente que la teoría no se cumpla. Más bien está reflejando la existencia de problemas de especificación. Así, por ejemplo, una reducción en la cantidad producida (lo cual libera inputs) acompañado con un aumento en la calidad (que exige emplear más inputs) puede manifestarse como una relación negativa entre los costes (que aumenta) y el output (cuya cantidad disminuye). El problema aquí se encuentra en que realmente se están produciendo dos "outputs": la cantidad y la calidad, pero este último no es observable.

[51] Véase, Varian (1992; p. 87).

[52] El incremento de coste pasivo es el resultado de multiplicar la cantidad usada del input x_i^1 por el incremento de precio de ese input.

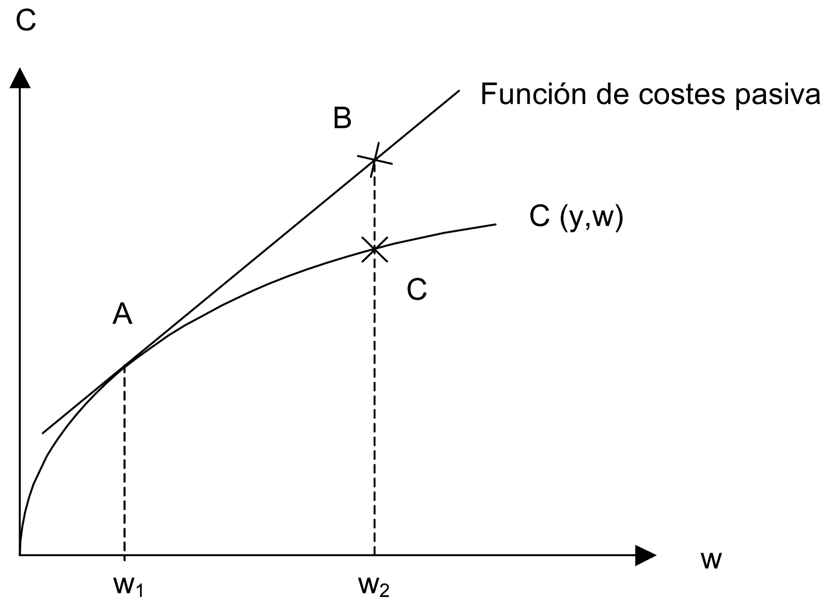

Figura 3.1. *Función de costes pasiva*

3.2.2. La demanda de factores

Como se vio en el apartado anterior, el proceso de minimización de costes determina las cantidades optimas de factores en un proceso productivo. Estas cantidades óptimas dependen de los precios de los inputs y la cantidad de output. La relación funcional entre la cantidad óptima de un input, su precio, el precio de otros inputs y la cantidad de output se conoce como función de demanda de factores.

La función de demanda de un input se puede obtener derivando la función de costes con respecto al precio de ese input:

$$\frac{\partial C(y,w)}{\partial w_j} = x_j(y,w) \qquad (3.\ 7)$$

Esta propiedad conocida como el Lema de Shephard[53] es, desde un punto de vista numérico, una propiedad casi trivial. Cuando el precio de un input aumenta en una unidad el coste aumenta en el número de unidades que se usan de ese input. Este resultado es cierto para todos los precios de los inputs y nivel de output con lo

[53] En algunos libros de texto esta propiedad se presenta en el contexto de la teoría de la dualidad. La razón es que Shephard (1953) usa una versión de este lema para analizar las propiedades de la función de costes que aseguran que tiene origen en una tecnología regular. Sin embargo, este resultado se puede usar separadamente de la teoría de la dualidad.

cual se establece una relación funcional entre la función de costes y las demandas de inputs.

Es necesario reconocer, sin embargo, que el lema de Shephard puede resultar poco intuitivo para una persona acostumbrada a pensar en términos económicos, ya que el razonamiento anterior se apoya en una actitud pasiva por parte del productor, mientras que el cambio en el precio de un input suele manifestarse en un cambio en las cantidades usadas de todos los inputs. La demostración matemática del lema arroja algo de luz en este problema. La función de costes definida en (3.2) es igual a la función de Lagrange de la ecuación (3.3) evaluada en el óptimo. Por tanto, la derivada de la función de costes con respecto al precio de un input puede analizarse derivando la función de Lagrange. Es decir,

$$\frac{\partial C(y,w)}{\partial w_j} = \frac{\partial L}{\partial w_j} = x_j + \sum_k \left(w_k - \lambda f_k(x)\right)\frac{\partial x_k}{\partial w_j} + \frac{\partial \lambda}{\partial w_j}(y - f(x)) \qquad (3.\ 8)$$

El hecho de que el lema de Shephard se cumpla en puntos óptimos hace que los dos últimos términos se anulen (ver ecuación 3.3.). Es decir, los cambios en las cantidades óptimas de inputs no afectan, marginalmente, al cambio en el coste. Este resultado es una aplicación del denominado *teorema de la envolvente* (Silberberg, 2000; p. 127).

El lema de Shephard permite relacionar las propiedades teóricas de la función de costes con las propiedades de las demandas de inputs. Por ejemplo, la homogeneidad lineal en el precio de los inputs de la función de costes implica que las funciones de demanda de inputs, que son derivadas de la función de costes, son homogéneas de grado cero.[54] Es decir, una subida proporcional de los precios de todos los inputs no afecta, como ya se ha comentado, a la cantidad demandada de inputs en un proceso productivo en que se minimicen costes. Por otra parte, la concavidad de la función de costes implica una serie de restricciones sobre el signo y valor de las derivadas segundas de la función de costes. En concreto, el hessiano tiene que ser definido negativo. Estas restricciones sobre las segundas derivadas de la función de costes pueden reinterpretarse como restricciones sobre la primera derivada

[54] Este es un caso particular de una de las propiedades de la funciones homogéneas: las derivadas de una función homogénea de grado n son homogéneas de grado n-1.

de las funciones de demanda de inputs. Por ejemplo, la concavidad de la función de costes implica que:

$$\frac{\partial^2 C(y,w)}{\partial w_j^2} < 0 \tag{3. 9}$$

El lema de Shephard permite transformar esta propiedad de la función de costes en una propiedad de la función de demanda de inputs.

$$\frac{\partial^2 C(y,w)}{\partial w_j^2} = \frac{\partial\left(\dfrac{\partial C(y,w)}{\partial w_j}\right)}{\partial w_j} = \frac{\partial x_j(y,w)}{\partial w_j} < 0 \tag{3. 10}$$

Es decir, se establece que la demanda de inputs es decreciente con respecto al propio precio.

En definitiva, las propiedades de monotonicidad, homogeneidad y concavidad de la función de costes se han de traducir, en la práctica empírica, en el cumplimiento de una serie de restricciones que afectan a las derivadas tanto de la función de costes como de las funciones de demanda de inputs. Si se supone que dichas restricciones se cumplen, se pueden imponer en el proceso de estimación, lo cual permite obtener estimaciones más eficientes de los parámetros del modelo. Por otra parte, como se verá en el apartado siguiente, estas propiedades son esenciales para aplicar el principio de dualidad. Como se comentó anteriormente, el lema de Shephard es una pieza clave para descubrir cuál es el papel de las propiedades de la función de costes para tratar el problema básico de la dualidad: la integrabilidad.

3.2.3. Dualidad entre la función de producción y la función de costes

En las secciones previas se han analizado las propiedades de la función de costes, fruto del comportamiento minimizador de las empresas sujetas a una determinada tecnología. Sin embargo, en muchas ocasiones el problema que se plantea es justo el contrario. Si análisis empírico parte de una función de costes (lo mismo si fuera de beneficios o ingresos), la pregunta relevante es si la función de costes usada tiene su origen en una tecnología con las propiedades discutidas en el capítulo 2. En otras palabras, ¿es posible conocer las propiedades de la tecnología subyacente a partir de los parámetros estimados de la función de costes? La respuesta es afirmativa bajo ciertas condiciones.

La teoría de la dualidad establece las propiedades que deben cumplir las funciones de coste y beneficio para asegurar su relación con una tecnología razonable. En el caso de la función de costes, la homogeneidad de grado uno en el precio de los inputs y la concavidad en precios de los inputs aseguran la existencia de una tecnología subyacente regular.[55] La dualidad tuvo sus orígenes en los años 30 con Hotelling y fue desarrollada en los 40 por Roy, Hicks y Samuelson, aunque el primer tratamiento sistemático de la dualidad en la producción vino de la mano de Shephard (1953) en su libro *Cost and Production Functions.*

Silberberg (2000; p. 234) propone una interesante demostración de este resultado, a medio camino entre la intuición y las matemáticas. El lema de Shephard proporciona una relación entre inputs usados y outputs obtenidos a través de la función de costes. En un proceso productivo con dos inputs se puede escribir como:

$$\begin{aligned} x_1(w_1, w_2, y) &= \frac{\partial C(w_1, w_2, y)}{\partial w_1} \\ x_2(w_1, w_2, y) &= \frac{\partial C(w_1, w_2, y)}{\partial w_2} \end{aligned} \qquad (3.\ 11)$$

La homogeneidad de grado uno en precios de inputs de la función de costes se traduce en la homogeneidad de grado cero en precios de inputs de las funciones de demanda de inputs. Esta propiedad permite escribir las funciones de demanda de inputs como:

$$\begin{aligned} x_1(w_1, w_2, y) &= x_1\left(\frac{w_1}{w_2}, 1, y\right) = h_1(w, y) \\ x_2(w_1, w_2, y) &= x_2\left(\frac{w_1}{w_2}, 1, y\right) = h_2(w, y) \end{aligned} \qquad (3.\ 12)$$

donde w es el precio relativo de los inputs. La ecuación (3.12) representa un sistema de dos ecuaciones en las que es posible, en principio, despejar el output (y) en función de los inputs (x_1, x_2). Es decir:

$$g(x_1, x_2, y) = 0 \qquad (3.\ 13)$$

[55] En el caso de la función de beneficios, la homogeneidad de grado uno y la convexidad en los precios de inputs y outputs aseguran la existencia de esta tecnología.

Esta operación es posible cuando el Jacobiano del sistema de ecuaciones sea distinto de cero. Los términos del Jacobiano son derivadas de las funciones de demanda de inputs con respecto a los inputs. Es decir, los valores de estos términos están relacionados con las derivadas segundas de la función de costes. Como se recordará, los signos de estas segundas derivadas vienen definidos por la concavidad de la función de costes. En general, se puede demostrar que la concavidad de la función de costes asegura la no nulidad del Jacobiano y, por tanto, que se puede establecer la relación matemática entre inputs y outputs partiendo de la función de costes.[56] En resumen, las propiedades de homogeneidad y concavidad en precios de la función de costes bastan para asegurar la existencia de una tecnología subyacente cuyas características se pueden analizar a partir de la estimación de la función de costes.

Un ejemplo sencillo del principio de dualidad puede plantearse con la función de costes Cobb-Douglas que en el caso de dos inputs y un output se puede escribir como:

$$C(w_1, w_2, y) = Bw_1^{\alpha} w_2^{1-\alpha} y^{\beta} \qquad (3.\ 14)$$

Por construcción, la función es homogénea de grado 1 en los precios de los inputs. Las demandas de inputs pueden ser obtenidas usando el lema de Shephard.

$$\begin{aligned} x_1(w_1, w_2, y) &= \alpha B w_1^{\alpha-1} w_2^{1-\alpha} y^{\beta} \\ x_2(w_1, w_2, y) &= (1-\alpha) B w_1^{\alpha} w_2^{-\alpha} y^{\beta} \end{aligned} \qquad (3.\ 15)$$

Las demanda de factores y su cociente pueden escribirse como funciones del precio relativo de los factores:

$$\begin{aligned} & x_2(w_1, w_2, y) = (1-\alpha) B \left(\frac{w_2}{w_1} \right)^{1-\alpha} y^{\beta} \\ & \frac{x_1}{x_2} = \frac{\alpha}{1-\alpha} \frac{w_2}{w_1} \Leftrightarrow \frac{w_2}{w_1} = \frac{1-\alpha}{\alpha} \frac{x_1}{x_2} \end{aligned} \qquad (3.\ 16)$$

Finalmente, se puede despejar el output como función de los inputs en la expresión (3.16). Como resultado, se obtiene que la tecnología subyacente en una función de costes Cobb-Douglas se representa mediante una función de producción:

---

[56] Un razonamiento parecido permite establecer que la homogeneidad de grado uno en precios y la convexidad de la función de beneficios permiten asegurar que representan un proceso con una tecnología regular.

$$y = f(x) = \left[\frac{1}{\alpha B} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} \right]^{\frac{1}{\beta}} x_1^{\frac{\alpha}{\beta}} x_2^{\frac{1-\alpha}{\beta}} \tag{3. 17}$$

Nótese que la función de producción (3.17) tiene la "forma" de una Cobb-Douglas. Es decir, si la función de costes es Cobb-Douglas, también lo es la función de producción subyacente. Por esta razón, se dice que la función Cobb-Douglas es *autodual*. Esta conveniente propiedad no la tienen otras funciones más flexibles como la translog o la generalizada de Leontief.

3.2.4. Aplicaciones de la función de costes

La función de costes ha sido ampliamente usada en el análisis empírico de la producción.[57] A continuación se comentan algunos de los trabajos más relevantes. Nerlove (1963) estimó una función de costes con datos de empresas generadoras de energía eléctrica en Estados Unidos, en lo que probablemente constituye la primera aplicación empírica de la dualidad. A partir de este estudio pionero, la función de costes fue usada para estudiar economías de escala, cambio tecnológico o sustitución entre factores.

Una de las principales aplicaciones empíricas de la función de costes ha sido el estudio de las economías de escala, explotando la dualidad entre la función de costes y la función de producción. Como se ha visto en el Capítulo 2, los rendimientos a escala hacen referencia a cambios en la producción provocados por cambios en la cantidad empleada de todos los inputs. En concreto, la elasticidad de escala se puede escribir como una suma de elasticidades parciales, que para la función de producción (3.17) se puede expresar como sigue:

$$e = \frac{\partial \ln y}{\partial \ln x_1} + \frac{\partial \ln y}{\partial \ln x_2} = \frac{\alpha}{\beta} + \frac{1-\alpha}{\beta} = \frac{1}{\beta} \tag{3. 18}$$

Esta expresión indica que la elasticidad de escala es constante e igual a la inversa del parámetro que acompaña al output (β) en la función de costes (3.14). De aquí se deduce que es posible analizar los rendimientos a escala, una característica de la tecnología a largo plazo, a partir de la estimación de una función de costes. En efecto, el resultado anterior se puede generalizar fácilmente para cualquier función de costes de la forma siguiente:

$$e = \left[\partial \ln C / \partial \ln y\right]^{-1} = \varepsilon_{Cy}^{-1} \quad (3.\ 19)$$

es decir, la elasticidad de escala es igual a la inversa de la *elasticidad tamaño* (Hanoch, 1975) que mide en qué porcentaje varían los costes cuando el output cambia en un uno por ciento. Christensen y Greene (1976) emplean en su conocido estudio sobre las economías de escala en el sector eléctrico una medida equivalente, que es igual a (1-ε_{Cy}), que toma valores positivos cuando hay economías de escala y negativos cuando hay deseconomías.

Las cuestiones sobre economías de escala en la función de costes están relacionadas con la forma de la función de costes medios. Esto se puede ver fácilmente a través de la relación existente entre los costes medios y marginales, por una parte, y la elasticidad de tamaño, por otra. Es decir:

$$\varepsilon_{Cy} = \frac{\partial \ln C}{\partial \ln y} = \frac{\partial C}{\partial y}\frac{y}{C} = \frac{CMa}{CMe} \quad (3.\ 20)$$

Por tanto, la elasticidad de tamaño es igual al cociente entre el coste marginal y el coste medio. Este resultado es muy intuitivo ya que cuando el coste marginal es mayor que el coste medio la función de costes medios es decreciente y viceversa. Por tanto, cuando ε_{Cy}<1 existirán economías de tamaño, como puede verse en la Figura 3.2.

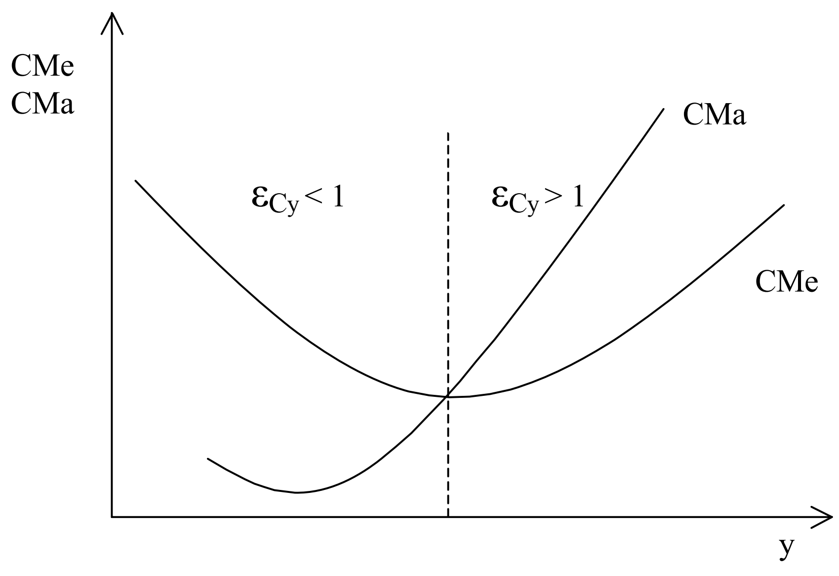

Figura 3.2. *Elasticidad de tamaño, costes medios y costes marginales*

[57] Un *survey* que cubre las primeras aplicaciones de funciones de costes es el de Walters (1963).

Otra de las aplicaciones de la dualidad ha sido para analizar cambios en la tecnología. Los avances tecnológicos (ordenadores más potentes, etc.) permiten obtener el mismo output con menos inputs y, por tanto, con menos costes. Una forma de modelizar el efecto del cambio tecnológico sobre los costes es permitiendo que el parámetro B de la función de costes (3.14) varíe a lo largo del tiempo, esto es:

$$C(w,y,t) = B(t)\cdot C(w,y) = B(t)\cdot\left[w_1^{\alpha} w_2^{1-\alpha} y^{\beta}\right] \qquad (3.\ 21)$$

donde t representa un instante de tiempo.[58] Tomando logaritmos en (3.21) y derivando con respecto al tiempo se obtiene:

$$\dot{C} = \frac{\partial \ln B(t)}{\partial t} + \left[\alpha \dot{w}_1 + (1-\alpha)\dot{w}_2 + \beta \dot{y}\right] \qquad (3.\ 22)$$

donde el punto encima de una variable indica una tasa de variación proporcional.

En síntesis, la expresión (3.22) pone de manifiesto que la variación del coste puede deberse a un desplazamiento de la función de costes (esto es, la contribución del cambio tecnológico al cambio en los costes) o a un movimiento a lo largo de la misma cuando varía el nivel de producción o los precios de los inputs.

A continuación se puede explotar la dualidad entre la función de costes y la función de producción con objeto de medir la contribución del cambio tecnológico al crecimiento de la producción. Para ello, la ecuación (3.17) se escribe de la siguiente forma:

$$y = B(t)^{\frac{-1}{\beta}}\cdot f(x) = B(t)^{\frac{-1}{\beta}}\cdot\left[\frac{1}{\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\right]^{\frac{1}{\beta}} x_1^{\frac{\alpha}{\beta}} x_2^{\frac{1-\alpha}{\beta}} \qquad (3.\ 23)$$

Tomando logaritmos en (3.23) y derivando con respecto al tiempo se obtiene:

$$\dot{y} = -\frac{1}{\beta}\cdot\frac{\partial \ln B(t)}{\partial t} + \left(\frac{\alpha}{\beta}\dot{x}_1 + \frac{1-\alpha}{\beta}\dot{x}_2\right) \qquad (3.\ 24)$$

[58] Solow (1957) utilizó esta variable para recoger el efecto del cambio tecnológico, entendido éste en sentido amplio, esto es, como “una expresión taquigráfica de *cualquier tipo de desplazamiento* de la función de producción [costes]. Así, los descensos de actividad, las aceleraciones, las mejoras de la educación de la fuerza de trabajo y todo tipo de cosas, se considerarán como *cambio tecnológico*” (Solow, 1957, p. 312).

donde el primer término de la derecha es la derivada parcial de la función de producción (3.23) con respecto al tiempo y mide la parte del crecimiento del output atribuible a un avance tecnológico. El segundo término, por su parte, mide la contribución de los inputs a las variaciones del output. Por tanto, la ecuación (3.24) indica que un avance tecnológico ahorrador de costes (es decir, $\partial \ln B(t)/\partial t<0$) se puede interpretar como un aumento en la producción dado el vector de inputs. La ecuación (3.24) indica, además, que ambas interpretaciones están inversamente relacionadas a través del parámetro β o la elasticidad de tamaño ε_{Cy}.[59]

Por último, el lema de Shephard permite expresar la elasticidad de sustitución de Allen como función de derivadas de la función de costes.[60] En concreto:

$$\sigma_{jk} = \frac{C(y,w)\dfrac{\partial^2 C(y,w)}{\partial w_j \partial w_k}}{\dfrac{\partial C(y,w)}{\partial w_j}\dfrac{\partial C(y,w)}{\partial w_k}} \qquad (3.\ 25)$$

expresión que, en general, es mucho más sencilla que la equivalente en el primal.

3.2.5. Análisis a corto plazo

El análisis realizado hasta el momento considera que el productor puede elegir las cantidades de inputs que minimicen el coste. Sin embargo, es razonable pensar en casos donde algunos de los inputs no puedan ser cambiados en el periodo analizado, es decir, se consideran fijos. Un ejemplo de factor fijo puede ser la capacidad de gestión del empresario. Así, por ejemplo, en el sector agrario es frecuente observar que los agricultores no aumentan su capacidad de gestión. La razón es que la principal fuente de conocimiento del agricultor es exógena: cursillos de formación,

---

[59] Ohta (1975) fue el primero en demostrar que las dos interpretaciones del cambio tecnológico, primal y dual, están estrechamente relacionadas a través de los rendimientos a escala. En concreto, en el caso de un único output, la relación viene dada por:

$$\varepsilon_{Cy}\cdot\frac{\partial \ln f(x,t)}{\partial t} = -\frac{\partial \ln C(y,w,t)}{\partial t}$$

de donde se deduce que, sólo cuando la tecnología presenta rendimientos constantes a escala, la definición primal y dual de cambio tecnológico coinciden. Más adelante, Caves, Christensen y Swanson (1981) extendieron esta demostración para el caso de empresas multiproducto.

[60] Bravo-Ureta *et al.* (1993) utilizan este enfoque para estimar elasticidades de sustitución. Este artículo constituye un buen ejemplo de análisis de la producción a través de la función de costes.

asesoramiento externo, etc. La experiencia ("learning by doing") queda relegada a un segundo lugar debido a la complejidad de los procesos agronómicos que se manejan. Sin embargo, la dificultad de definición y medida de este factor dificulta enormemente el análisis empírico.[61]

La existencia de inputs fijos añade algunas peculiaridades al análisis empírico dual. Por una parte, la función de coste a largo plazo (3.1) no existe. Si las empresas minimizan el coste de los inputs variables, entonces existe una función de coste total a corto plazo, que depende tanto de la cantidad como del precio de los inputs fijos:

$$\begin{aligned} C(y,w,z,r) = \quad & \min_x \sum_j w_j x_j + rz \\ \text{s.a.} \quad & y = f(x,z) \end{aligned} \tag{3. 26}$$

donde, por simplicidad, z denota la cantidad empleada de un input fijo y r representa su precio. A corto plazo, el coste puede descomponerse en el coste fijo (r·z) y el coste variable (CV) que se puede representar como: [62]

$$\begin{aligned} CV(y,w,z) = \quad & \min_x \sum_j w_j x_j \\ \text{s.a.} \quad & y = f(x,z) \end{aligned} \tag{3. 27}$$

Si el input fijo fuera utilizado de forma óptima (por casualidad o por que realmente se está en un equilibrio a largo plazo) no sería posible reducir los costes totales escogiendo una cantidad z distinta. Por tanto, en este caso, se tendría que cumplir que:

$$\frac{\partial C(y,w,z,r)}{\partial z} = 0 \tag{3. 28}$$

o, lo que es lo mismo,

$$-\frac{\partial CV(y,w,z)}{\partial z} = r \tag{3. 29}$$

De las condiciones de optimalidad (3.28) o (3.29) se puede obtener entonces la cantidad óptima del input z=z(y,w,r). Insertando esta cantidad en (3.26), la función de costes totales a corto plazo se convertiría en la función de costes a largo plazo (3.2).

---

[61] De hecho, el input "gestión" se suele medir de forma residual modelizando la parte estocástica del modelo. A este respecto, véase también el Capítulo 5.

[62] Aplicando el lema de Shephard a los inputs variables es posible "completar" esta función con las ecuaciones de demanda o las ecuaciones de participación de los inputs variables.

Aunque las funciones de coste total (3.26) y coste variable (3.27) se definen a corto plazo, esto no significa que no se puedan analizar las características de la tecnología subyacente a largo plazo. Es decir, la dualidad analizada en la sección anterior entre producción y costes se mantiene a corto plazo. Así, por ejemplo, las economías de escala pueden estudiarse a partir de la función de costes a corto plazo (3.26).[63] Derivando con respecto al output y reordenando se obtiene:

$$\varepsilon_{Cy} = \frac{d\ln C}{d\ln y} = \frac{\partial \ln C(y,w,z,r)}{\partial \ln y} + \frac{\partial \ln C(y,w,z,r)}{\partial \ln z} \cdot \frac{d\ln z}{d\ln y} \tag{3. 30}$$

lo cual es equivalente a

$$\varepsilon_{Cy} = \varepsilon_{Cy}^{CP} + \varepsilon_{Cz}^{CP} \cdot \frac{d\ln z}{d\ln y} \tag{3. 31}$$

donde ε_{Cy} es la elasticidad tamaño a largo plazo (la inversa de la elasticidad de escala), mientras que el primer término de la derecha se puede interpretar como la elasticidad tamaño a corto plazo. Estas expresiones ponen de manifiesto que, en equilibrio a largo plazo de todos los inputs, el nivel de producción influye sobre el coste total tanto directamente (primer término de la derecha) como indirectamente a través de su relación con el input z (segundo término). Una forma conveniente de expresar la relación entre la elasticidad tamaño a largo plazo y corto plazo es la siguiente:

$$\varepsilon_{Cy}^{CP} = \varepsilon_{Cy}\left(1 - \varepsilon_{Cz}^{CP} \cdot \frac{d\ln z / d\ln y}{\varepsilon_{Cy}}\right) = \varepsilon_{Cy}\left(1 - \varepsilon_{Cz}^{CP} \cdot \frac{d\ln z}{d\ln C}\right) \tag{3. 32}$$

Si la tecnología es homotética, un incremento en el output aumenta a largo plazo la utilización de todos los inputs en la misma proporción. Por lo tanto, el incremento proporcional del input z es igual al de los costes, es decir, $d\ln z/d\ln C=1$. Teniendo en cuenta que la elasticidad de escala es la inversa de la elasticidad tamaño a largo plazo, se puede escribir la elasticidad de escala e(x,z) como sigue:

$$e(x,z) = \varepsilon_{Cy}^{-1} = \frac{1 - \varepsilon_{Cz}^{CP}}{\varepsilon_{CY}^{CP}} \tag{3. 33}$$

---

[63] En Morrison (1993) se puede encontrar un tratamiento muy intuitivo de la dualidad entre la función de producción y las funciones de costes a corto plazo. En este libro, además, se explica con mucha claridad multitud de aspectos relacionados con las economías de escala, el cambio técnico, la productividad, problemas que aparecen en la práctica empírica, etc.

Veamos un aplicación de todo lo dicho para el caso de una tecnología Cobb-Douglas. Para ello, se transforma (3.14) en una función de costes a corto plazo:

$$C(w,y,z,r) = CV(w,y,z) + rz = \left[B \cdot w_1^{\alpha} w_2^{1-\alpha} y^{\beta}\right] \cdot z^{-\gamma} + rz \tag{3. 34}$$

Reemplazando B en (3.17) por $B \cdot z^{-\gamma}$ podemos escribir la función de producción dual de la función de costes (3.34) como:

$$y = f(x) \cdot z^{\frac{\gamma}{\beta}} = \left[\frac{1}{\alpha B}\left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha}\right]^{\frac{1}{\beta}} x_1^{\frac{\alpha}{\beta}} x_2^{\frac{1-\alpha}{\beta}} \cdot z^{\frac{\gamma}{\beta}} \tag{3. 35}$$

donde la elasticidad de escala (la suma de elasticidades parciales de los tres inputs x_1, x_2 y z) es igual a:

$$e = \frac{\alpha}{\beta} + \frac{1-\alpha}{\beta} + \frac{\gamma}{\beta} = \frac{1}{\beta}(1+\gamma) \tag{3. 36}$$

Las elasticidades de la función de costes a corto plazo (3.35) respecto al output y al input fijo toman la siguiente forma:

$$\varepsilon_{Cy}^{CP} = \frac{\partial \ln CV}{\partial \ln y} \cdot \frac{CV}{C} = \beta \cdot \frac{CV}{C} \tag{3. 37}$$

$$\varepsilon_{Cz}^{CP} = \left(\frac{\partial CV}{\partial z} + r\right) \cdot \frac{z}{C} = \left(\frac{\partial \ln CV}{\partial \ln z} - 1\right) \cdot \frac{CV}{C} + 1 = -(\gamma+1) \cdot \frac{CV}{C} + 1 \tag{3. 38}$$

Reemplazando las elasticidades que aparecen en (3.33) por las expresiones anteriores nos queda finalmente que la elasticidad de escala se puede calcular a partir de la estimación de una función de costes variables (o totales a corto plazo) como sigue:

$$e = \varepsilon_{Cy}^{-1} = \frac{1 + (\gamma+1) \cdot \frac{CV}{C} - 1}{\beta \cdot \frac{CV}{C}} = \frac{\gamma+1}{\beta} \tag{3. 39}$$

Tanto la función de coste total como de coste variable definen totalmente el comportamiento a corto plazo del productor y la tecnología subyacente. Sin embargo, la función de coste variable puede ser más fácil de tratar empíricamente. De hecho, los precios de los inputs fijos relevantes para la empresa no suelen ser observables o no se corresponden necesariamente con los precios de mercado. Por ejemplo, una industria química hace una fuerte inversión en una sofisticada planta. A partir de ese momento toma decisiones periódicas sobre producción y uso de inputs variables (trabajo y energía) dependiendo de los precios de mercado. A partir de la instalación de la planta (factor fijo) ocurren dos cosas. Por un lado, no está claro cuál es el precio del factor fijo ya que no existe un mercado para ese factor. Por otro lado, ese precio es irrelevante para la toma de decisiones (a corto plazo) sobre inputs variables.

3.3. La función de beneficios

La dualidad analizada en las secciones anteriores se puede extender prácticamente sin grandes cambios a empresas cuyo objetivo es la maximización de beneficios. En este caso, es posible analizar las características de la tecnología subyacente a partir de la estimación de una función de beneficios.

Las decisiones sobre outputs producidos e inputs usados pueden modelizarse como el resultado de maximizar los beneficios de la empresa sujeto a las restricciones de la tecnología y de los precios de los inputs y de los outputs.

$$\begin{aligned} \Pi(p,w) = \max_{x\,y} \quad & py - \sum_j w_j x_j \\ \text{s.a} \quad & y = f(x) \end{aligned} \qquad (3.\ 40)$$

donde $\Pi(.)$ es la función de beneficios y p es el precio del output. La ecuación de Lagrange que permite resolver el programa de optimización anterior es:

$$L = py - \sum w_j x_j - \lambda(\, y - f(x)) \qquad (3.\ 41)$$

Las condiciones de primer orden se obtienen derivando (3.42) con respecto a las cantidades de inputs y output y con respecto a λ e igualándolas a cero.

$$\begin{aligned} & -w_j + \lambda f_j(x)) = 0 \quad \forall j \\ & p - \lambda = 0 \\ & y - f(x) = 0 \end{aligned} \qquad (3.\ 42)$$

De nuevo, las condiciones de primer orden definen un sistema de ecuaciones cuya solución da lugar a las cantidades de inputs y output que maximizan el beneficio. Las condiciones de segundo orden se cumplen si la función de producción es cóncava, esto es, la función de producción presenta rendimientos decrecientes a escala.

Las cantidades óptimas de inputs y outputs para la maximización del beneficio cambian con los precios de los inputs y los outputs. Es decir, la maximización de beneficios implica una relación funcional entre el beneficio obtenido y los precios de los outputs y los inputs. Adicionalmente, la teoría económica proporciona una serie de propiedades que la relación funcional citada debe poseer para poder ser considerada una función de beneficio.

La función de beneficios debe: 1) ser monótona creciente en los precios de los outputs; 2) monótona decreciente en los precios de los inputs; 3) homogénea de grado 1 en el precio de inputs y outputs; y 4) convexa en precios de inputs y outputs.

La primera propiedad indica que el beneficio no puede reducirse al aumentar el precio de un output. De hecho, el beneficio aumenta sin cambiar la decisión óptima. Por supuesto, aumenta al menos en la misma magnitud cuando se adapta la decisión de inputs y outputs al nuevo precio ya que sólo se llevará a cabo dicha adaptación para mejorar el beneficio Por razones similares, el beneficio no puede crecer cuando aumenta el precio de un input. De otro modo, hubiera sido posible aumentar el beneficio antes de que subiese el input y esto indicaría que la decisión anterior no era óptima.

La propiedad de homogeneidad lineal en precios indica que la variable relevante son los precios relativos y no los precios absolutos. Es decir, las cantidades demandas de inputs y ofertadas de outputs no cambian si todos los precios de los inputs y los outputs cambian en la misma proporción, por lo que los beneficios aumentan exactamente en la proporción en la que ha aumentado los precios.

La propiedad de convexidad puede ser motivada intuitivamente usando, al igual que en los costes, el concepto de función de beneficios pasiva. Cuando sube el precio del output de una de las opciones que tiene el productor es no cambiar la cantidad de output que usa. En este caso, el beneficio se incrementa linealmente con el precio del output. Sin embargo, si se puede aumentar el beneficio en mayor cuantía aumentando la cantidad de output producida la función de beneficios irá por encima de su tangente.

La maximización de beneficios permite establecer una relación funcional entre output producido y precios de inputs y outputs (relación que se conoce como función de oferta de outputs) y una relación funcional entre cantidades óptimas de inputs y precios de inputs y outputs (denominadas funciones de demanda de inputs). A continuación se analizan sus principales características.

El lema de Hotelling establece que la derivada de la función de beneficios con respecto al precio de un output es igual a la función de oferta de ese output. Del mismo modo, la derivada de la función de beneficio con respecto al precio de un input es la función de demanda de ese input cambiada de signo. En términos matemáticos este resultado puede expresarse como:

$$\begin{aligned} \frac{\partial \Pi(w,p)}{\partial p_i} &= y_i(w,p) \\ \frac{\partial \Pi(w,p)}{\partial w_j} &= -x_j(w,p) \end{aligned} \qquad (3.\ 43)$$

Cuando aumenta el precio de un input en una unidad el beneficio se reduce en el número de unidades que se usan de ese input. Cuando aumenta el precio de un output el beneficio aumenta en el número de unidades que se producen de ese output. Sin embargo, esta propiedad se cumple para todo precio de input y output. Por lo tanto, implica una relación funcional entre función de beneficio, oferta de output y demanda de inputs.

Al igual que el lema de Shephard, el lema de Hotelling puede ser poco intuitivo. Sin embargo, la aplicación del teorema de la envolvente en la ecuación (3.42) permite ver con claridad el fenómeno.

El lema de Hotelling permite deducir las propiedades de la oferta de outputs y demanda de inputs a partir de las propiedades de la función de beneficios. En primer lugar, las funciones de oferta y demanda son las derivadas de una función homogénea de grado 1. Por tanto, son homogéneas de grado cero. Es decir, las cantidades demandas de inputs y ofertadas de outputs no cambian si todos los precios de los inputs y los outputs cambian en la misma proporción. Por otra parte, la convexidad con respecto al precio del output de la función de beneficio establece una restricción sobre las segundas derivadas de la función de beneficio con respecto a los precios de los outputs. En concreto, el hessiano tiene que ser positivo definido. Estas restricciones sobre las segundas derivadas de la función de beneficio pueden interpretarse cómo restricciones sobre las primeras derivadas de las funciones de oferta de output. Por ejemplo, la convexidad de la función de beneficio implica que:

$$\frac{\partial^2 \Pi(w,p)}{\partial p_i^2} > 0$$

El lema de Hotelling permite transformar esta propiedad de la función de beneficios en una propiedad de la función de oferta de output.

$$\frac{\partial^2 \Pi(w,p)}{\partial p_i^2} = \frac{\partial\left(\frac{\partial \Pi(w,p)}{\partial p_i}\right)}{\partial p_i} = \frac{\partial y_i}{\partial p_i} > 0 \qquad (3.\ 44)$$

Es decir, se establece que la oferta de output es creciente con respecto al propio precio.

Bibliografía

Bravo-Ureta, B.E., W. Ahmad y R. Wackernagel, (1993), "Sustitución entre Factores de Producción en Explotaciones Lecheras: Una Comparación de las Elasticidades de Tipo Allen y Morishima", *Investigación Agraria: Economía*, 8, 349-362.

Caves, D.W., L. Christensen y J. Swanson, (1981), "Productivity Growth, Scale Economies, and Capacity Utilisation in U.S. Railroads 1955-74", *American Economic Review*, 71, 994-1002.

Chambers, R., (1988), *Applied Production Analysis*, Cambridge University Press.

Christensen, L. y W. Greene, (1976), "Economies of Scale in U.S. Electric Power Generation", *J. of Political Economy*, 84(4), 655-676.

Griliches, Z. y J. Mairesse, (1997), "Production Functions: the Search for Identification", Working paper. NBER.

Hanoch, G., (1975), "The Elasticity of Scale and the Shape of Average Costs", *American Economic Review*, 65(3), 492-97.

Morrison, C.J. (1993): *A Microeconomic Approach to the Measurement of Economic Perfomance: Productivity Growth, Capacity Utilization and Related Perfomance Indicators*, Springer-Verlag, New York.

Nerlove, M., (1963), "Returns to Scale in Electricity Supply", en C. Christ *et al*. (eds.), *Measurement in Economics: Essays in Memory of Yehuda Grunfeld*, Stanford University Press.

Otha, M. (1975), "A Note on the Duality Between Production and Cost Functions:Rates of Return to Scale and Rates of Technical Progress", *Economic Studies Quarterly*, 25, 63-65.

Paris, Q. y M.R. Caputo, (1995), "The Rethoric of Duality", *J. of Agricultural and Resource Economics*, 20(1), 195-214.

Shepard, R.W., (1953), *Cost and Production Functions*, Princeton University Press.

Silberberg, E., (2000), *The Structure of Economics*,

Solow, R. (1957): "Technical Change and the Aggregate Production Function", *Review of Economics and Statistics*, 39, 312-320.

Varian, H., (1992), *Microeconomic Analysis*, Antoni Bosch, Barcelona.

Walters, A.A., (1963), "Production and Cost Functions: An Econometric Survey", *Econometrica*, 31 (1-2), 1-66.

CAPITULO 4

LA ESPECIFICACION DEL MODELO EMPIRICO

En este capítulo se analizan los problemas más relevantes relacionados con la especificación del modelo empírico. En concreto se empieza estudiando en detalle la importancia de la forma funcional, los problemas de las formas funcionales restrictivas y la noción de flexibilidad. Después se define el concepto de forma funcional flexible y se desarrolla con detalle la más usada en trabajos empíricos: la translog. También se describen otras formas funcionales flexibles, como la generalizada de Leontief y la cuadrática. A continuación, se discute el problema de la elección de forma funcional. Por último, se presentan unas formas funcionales especiales propuestas para analizar la producción de algunos sectores específicos.

4.1. La importancia de la forma funcional

La teoría económica proporciona descripciones de los fenómenos económicos como una relación funcional entre variables, tal como:

$$y = f(x) \tag{4. 1}$$

donde *y* es la variable explicada, *x* es la variable explicativa y f(x) representa la forma funcional, es decir, f(x) es una descripción de la relación entre *x* e *y*.

El análisis empírico de la producción necesita hacer explícita la forma de la función *f*. Este es un tema importante puesto que la forma funcional empleada restringe los resultados que se pueden obtener. Por ejemplo, en la función de producción Cobb-Douglas, las elasticidades output son constantes para cualquier vector de inputs usados y para cualquier nivel de producción. Asimismo, la elasticidad de sustitución de Allen entre cualquier par de inputs es igual a uno. Estos resultados provienen exclusivamente de la forma funcional especificada y no guardan necesariamente relación con la tecnología que generó los datos que se usan para el análisis empírico. En la Tabla 4.1 se puede ver cómo las formas funcionales tradicionales en el análisis económico (lineal, Cobb-Douglas, CES) restringen el valor de las principales características económicas de la función de producción.

Tabla 4.1. *Características económicas de varias funciones de producción*

Función	Ecuación	Productividad marginal de X_j	Elasticidad de Escala	Elasticidad de Sustitución
Lineal	$y = \alpha_1 x_1 + \alpha_2 x_2$	α_j	1	∞
Cobb-Douglas	$y = A x_1^{\alpha_1} x_2^{\alpha_2}$	$\alpha_j \dfrac{y}{x_j}$	$\alpha_1+\alpha_2$	1
CES	$y = \gamma(\delta x_1^{-\theta} + (1-\delta) x_2^{-\theta})$	$\theta\delta\gamma^{-\frac{\rho}{\theta}} \dfrac{y^{1+\frac{\rho}{\theta}}}{x_j^{1+\rho}}$	ρ	$(1+\theta)^{-1}$

La función Cobb-Douglas surge en el año 1928 y pasan más de tres décadas hasta que aparece la forma funcional CES en 1961. La Tabla 4.1 permite observar que el avance no es muy grande ya que los rendimientos a escala siguen siendo independientes de la cantidad de output y de los inputs. En realidad, la principal diferencia consiste en que la elasticidad de sustitución en la CES puede ser cualquier número real (no necesariamente 1).

Las limitaciones de estas formas funcionales hicieron que durante los años 60 y principios de los 70 se desarrollaran otras formas funcionales que permiten reducir algunas de las restricciones impuestas por la Cobb-Douglas y la CES. Una de ellas es la función de producción generalizada, sugerida por Zellner y Revankar (1969), que es una transformación de la Cobb-Douglas. Para el caso de dos inputs puede escribirse de la siguiente manera:

$$ye^{\theta y} = Ax_1^{\alpha_1}x_2^{\alpha_2} \tag{4. 2}$$

donde θ es un parámetro a estimar. Una particularidad de esta función es que los rendimientos a escala varían con el nivel de output. Tomando logaritmos y derivando con respecto a los inputs en logaritmos se tiene que:

$$\begin{gathered}
\ln y + \theta e^{\ln y} = \ln A + \alpha_1 \ln x_1 + \alpha_2 x_2 \\
\frac{\partial \ln y}{\partial \ln x_1} + \theta e^{\ln y} \frac{\partial \ln y}{\partial \ln x_1} = \alpha_1 \quad \Rightarrow \quad \frac{\partial \ln y}{\partial \ln x_1} = \frac{\alpha_1}{1+\theta y} \\
\frac{\partial \ln y}{\partial \ln x_2} + \theta e^{\ln y} \frac{\partial \ln y}{\partial \ln x_2} = \alpha_2 \quad \Rightarrow \quad \frac{\partial \ln y}{\partial \ln x_2} = \frac{\alpha_2}{1+\theta y}
\end{gathered} \tag{4. 3}$$

Por tanto, dado que los rendimientos a escala son la suma de las elasticidades-output para todos los inputs, se comprueba que son una función del output:

$$e = \frac{\partial \ln y}{\partial \ln x_1} + \frac{\partial \ln y}{\partial \ln x_2} = \frac{\alpha_1 + \alpha_2}{1+\theta y} \tag{4. 4}$$

La función de producción generalizada potencial introducida por de Janvry (1972) es también una generalización de la función de producción Cobb-Douglas.

$$y = A\prod_i x_i^{f_i(x)} \exp g(x) \tag{4. 5}$$

Cuando $f_i(x)=\alpha_i$ y $g(x)=0$ la función se convierte en una función de producción Cobb-Douglas.

Otro avance lo supuso el desarrollo de las funciones de producción radio-homotéticas por Färe (1975). Su nombre proviene de que estas funciones son una transformación homotética de una función de producción radio-homogénea. Aunque esta función no es lineal, una forma sencilla de la misma puede escribirse como (ver Färe, Jansson y Lovell, 1985):

$$\exp(y) = A\prod_j x_j^{(\alpha_j x_j)/\sum x_i} \tag{4. 6}$$

que puede linealizarse tomando logaritmos:

$$y = \alpha_0 + \sum_j \alpha_j x_j^* \ln x_j \tag{4. 7}$$

donde $X_j^*=xj/\Sigma_i x_j$ es la parte proporcional del input j sobre el resto de los inputs.

Una de las ventajas de esta forma funcional es que permite una mayor flexibilidad para la elasticidad de escala. Como ya se vio en el Capítulo 2, la elasticidad de escala en las funciones homogéneas es independiente del output y de la combinación de factores, mientras que en las funciones homotéticas depende del output, pero no de la combinación de inputs. Para las funciones radio-homogéneas la elasticidad de escala varía con la combinación de inputs, pero no con el output, mientras que en las funciones radio-homotéticas la elasticidad de escala depende tanto del output como de la combinación de factores. Efectivamente, la elasticidad de escala es $e = (\sum \alpha_j x_j^*)/y$.

Sin embargo, todos estos intentos de generalización de la función de producción Cobb-Douglas no fueron enteramente satisfactorios. Las funciones de producción mencionadas anteriormente son difíciles de estimar y plantean problemas de interpretación de los parámetros. Por este motivo, los investigadores continuaron buscando formas funcionales cada vez menos restrictivas. Este proceso ha llevado al desarrollo de las llamadas formas funcionales flexibles, cuyo concepto se expone en el apartado siguiente.

4.2. El concepto de flexibilidad

Como se ha visto en el apartado anterior, las formas funcionales pueden condicionar los resultados que se obtienen en el trabajo empírico, por lo que parece necesario encontrar formas funcionales que, en la medida de lo posible, superen esas limitaciones. Las restricciones que imponen las formas funcionales están relacionadas con el número de parámetros que tienen. Por ejemplo, la función de producción lineal tiene k parámetros en un proceso productivo con k inputs. El que las productividades marginales sean constantes en esta función de producción es una consecuencia directa del número de parámetros. Por ejemplo, si se añade una interacción entre las variables se tiene:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 \tag{4. 8}$$

Con esta modificación las productividades marginales no son constantes. El parámetro adicional α_{12} hace que la productividad marginal de un input dependa de la cantidad del otro factor. Es decir:

$$\begin{aligned} \frac{\partial y}{\partial x_1} &= \alpha_1 + \alpha_{12} x_2 \\ \frac{\partial y}{\partial x_2} &= \alpha_2 + \alpha_{12} x_1 \end{aligned} \tag{4. 9}$$

Sin embargo, como se puede ver en las segundas derivadas, las dos productividades son constantes con respecto al cambio del input analizado y cambian del mismo modo cuando cambia el otro input.

$$\begin{aligned} &\frac{\partial^2 y}{\partial x_1^2} = 0 \quad \frac{\partial^2 y}{\partial x_2^2} = 0 \\ &\frac{\partial^2 y}{\partial x_1 \partial x_2} = \frac{\partial^2 y}{\partial x_2 \partial x_1} = \alpha_{12} \end{aligned} \tag{4. 10}$$

Es decir, hay una restricción en las productividades marginales debido a que no se han añadido suficientes términos y, por tanto, suficientes parámetros. El problema desaparece añadiendo dos términos cuadráticos con sus respectivos parámetros:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \frac{1}{2}\alpha_{11} x_1^2 + \alpha_{12} x_1 x_2 + \frac{1}{2}\alpha_{22} x_2^2 \tag{4. 11}$$

En este caso, se tiene:

$$\begin{aligned}
&\text{a)} \quad \frac{\partial y}{\partial x_1} = \alpha_1 x_1 + \alpha_{12} x_2 \qquad \frac{\partial y}{\partial x_2} = \alpha_{12} x_1 + \alpha_2 x_2 \\
&\text{b)} \quad \frac{\partial^2 y}{\partial x_1^2} = \alpha_1 \qquad \frac{\partial^2 y}{\partial x_2^2} = \alpha_2 \\
&\text{c)} \quad \frac{\partial^2 y}{\partial x_1 \partial x_2} = \frac{\partial^2 y}{\partial x_1 \partial x_2} = \alpha_{12}
\end{aligned} \tag{4. 12}$$

Es decir, la productividad marginal de cada input depende de la cantidad usada de ambos (4.12a). Asimismo, la productividad marginal de cada input varía de forma distinta con la cantidad empleada del propio input (4.12b). Sin embargo, ambas productividades marginales cambian en la misma cuantía al cambiar la cantidad del otro input (4.12c).

Este ejemplo muestra cómo la flexibilidad depende del número de parámetros (términos) que tenga la función. La pregunta que surge es ¿cuántos términos hay que añadir para que la función sea flexible? El número de parámetros necesarios para lograr la flexibilidad depende del fenómeno que se modeliza. En este sentido, es frecuente modelizar el comportamiento de los productores como el resultado de un proceso de optimización (por ejemplo, minimización de costes o maximización de beneficio). En ese caso, las propiedades fundamentales del modelo están relacionadas con la estructura matemática del proceso de optimización, el cual, bajo ciertas condiciones de regularidad queda definido por el valor de la función, las primeras y las segundas derivadas (ver Chambers, 1988; p.161). Para un modelo con k variables explicativas, el número de términos a obtener será:

- valor de la función: 1
- primeras derivadas: k
- segundas derivadas: k^2

Este razonamiento puede aplicarse directamente al análisis empírico de la producción. En este caso, el objetivo es la representación de la información relevante para caracterizar exhaustivamente el comportamiento de los productores. En el caso de tecnologías de buen comportamiento, es decir, aquellas representables por una función de producción continua y dos veces diferenciable, un proceso productivo con k inputs puede ser representado por un vector que contenga los siguientes elementos:

- el nivel de producción
- k productividades marginales

- $k+(k^2-k)/2$ elasticidades de sustitución de Allen: hay $(k^2-k)/2$ elasticidades cruzadas distintas y k elasticidades de sustitución propias.

Por tanto, el mencionado vector contiene (k+1)(k+2)/2 elementos distintos que definen completamente el proceso de optimización.[64] De cara a la especificación del modelo, esto supone que las formas funcionales deberían tener (k+1)(k+2)/2 parámetros para poder modelizar todas las características de interés de un proceso productivo.[65] Esto excluye la modelización del cambio técnico. Fuss *et al*. (1978) indican que si se incluye cambio técnico en la función de producción es necesario añadir k+2 parámetros (dos para el cambio técnico neutral y k para las interacciones con los inputs), por lo que el número total de parámetros a estimar en este caso es de (k+2)(k+3)/2.

Por tanto, se puede hablar de formas funcionales flexibles como aquellas especificaciones que incluyen un número de parámetros igual al número de elementos necesario para caracterizar la tecnología que se pretende modelizar (Diewert, 1971), de forma que no restringen a priori las características de la tecnología. Es fácil comprobar que las formas funcionales analizadas con anterioridad contienen un número de parámetros inferior al número de características que quieren modelizar. Así, por ejemplo, la función de producción lineal tiene k parámetros en un proceso productivo con k inputs, la Cobb-Douglas tiene k+1, mientras que la CES tiene k+2. Por ello, es lógico que los efectos económicos de interés queden representados por una combinación de estos parámetros y que, por tanto, estén relacionados entre ellos. Es importante destacar que esta relación entre los efectos de interés sería una consecuencia del reducido número de parámetros en el modelo y no una característica esencial del fenómeno que se trata de modelizar.

La mayor parte de las formas funcionales flexibles usadas en la literatura están basadas en las aproximaciones en serie de Taylor. Las series de Taylor son aproximaciones polinómicas en un punto a una función arbitraria. El caso más sencillo es

[64] Fuss *et al*. (1978) hacen una descomposición distinta (pero equivalente) del número de efectos a modelizar. Señalas que los efectos están relacionados con los objetivos que tienen los estudios de producción: distribución de la renta (k-1 “shares”), escala (1 elasticidad de escala), sustitución entre factores (k(k-1)/2 elasticidades), elasticidades precio (k elasticidades) y el nivel del output (1).

el de la recta tangente a una curva. Esta recta es un polinomio de grado 1 que aproxima la curva en el punto de tangencia. De hecho, la aproximación es perfecta en el punto de tangencia, y buena en un entorno de ese punto. Es intuitivo también que la aproximación empeora a medida que nos alejamos del punto de tangencia.

La popularidad de estas aproximaciones se debe a tres motivos. En primer lugar, estas aproximaciones son lineales en los parámetros. En segundo lugar, se trata de una aproximación a una función arbitraria. Por último, una aproximación de Taylor de segundo orden a una función de producción con k inputs tiene (k+1)(k+2)/2 parámetros, es decir, es capaz de modelizar sin restricciones el número de efectos que se consideran relevantes en un proceso de optimización.

Como se desprende de la discusión anterior, las formas funcionales basan su flexibilidad en el elevado número de parámetros que contienen. Pero, al aumentar el número de parámetros disminuye la precisión con la que se puede estimar cada uno de ellos. Es decir, al querer estimar tantos efectos, algunos de ellos pueden no ser fácilmente estimables. En otras palabras, cuando se usan formas funcionales flexibles es de esperar que existan problemas de multicolinealidad.

El número de parámetros a estimar en una forma funcional flexible puede reducirse si se imponen restricciones. Por ejemplo, la función de costes es homogénea de grado 1 en los precios de los inputs. Por tanto, en una tecnología con k inputs, la función de costes tiene (k+1) argumentos (los k precios más el output), por lo que habría que modelizar (k+2)(k+3)/2 efectos. Sin embargo, si se impone la homogeneidad lineal, ese número se reduce a (k+1)(k+2)/2. De la misma manera, si en una función de producción con k inputs se impone homogeneidad de grado 1, el número de efectos distintos es k(k+1)/2 (Fuss *et al.*, 1978).[66]

4.3. Formas funcionales flexibles

---

[65] Puede extrañar que no se hayan incluido los rendimientos a escala como una característica de interés, tal y como se expuso en el Capítulo 2. El motivo es que aunque sea una característica de interés, toda la información necesaria para calcular los rendimientos a escala está en las productividades marginales.

Aunque en la literatura se han desarrollado un buen número de formas funcionales flexibles, en este apartado se describen las más empleadas en el análisis empírico: la translog, la generalizada de Leontief y la cuadrática normalizada.[67] Dado que la más usada de las tres es la translog, su descripción será más detallada que las de las otras funciones.

4.3.1. La función de producción translog

La función más frecuente en los análisis empíricos que usan funciones de producción flexibles es la función de producción logarítmica trascendental (Christensen, Jorgenson y Lau, 1973), más conocida por translog, cuya expresión es:[68]

$$\ln y = \beta_0 + \sum_{j=1}^{J} \beta_j \ln x_j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{K} \beta_{jk} \ln x_k \ln x_j \tag{4. 13}$$

Dado que $\ln x_k \cdot \ln x_j = \ln x_j \cdot \ln x_k$, necesariamente se cumple que $\beta_{kj} = \beta_{jk}$.[69] Por tanto, a veces, la función translog se escribe también de la siguiente forma:

$$\ln y = \beta_0 + \sum_{j}^{J} \beta_j \ln x_j + \frac{1}{2} \sum_{j}^{J} \beta_{jj} (\ln x_j)^2 + \sum_{j}^{J} \sum_{k}^{K} \beta_{jk} \ln x_j \ln x_k \tag{4. 14}$$

A continuación se estudia la función translog con cierto detalle. En concreto, se analizarán las características del proceso productivo, sus propiedades teóricas y la estructura de la tecnología.[70]

---

[66] Quizás es conveniente recordar que cada restricción permite estimar un parámetro menos. Este hecho es intuitivo en una restricción lineal dónde un parámetro se puede despejar en función de los demás.

[67] El lector interesado en otras formas funcionales flexibles puede consultar el *survey* de Thompson (1988).

[68] La translog puede interpretarse como un caso particular de la función de producción propuesta por de Janvry, donde:

$$f_j(x) = \beta_j + \frac{1}{2} \sum_{j}^{J} \beta_{jk} \ln x_j \quad \text{y} \quad g(x) = 0$$

[69] Por este motivo, en algunos trabajos se dice que la función translog se estima imponiendo simetría en los parámetros.

[70] Los lectores interesados en profundizar en las características de la translog encontrarán en la monografía de Boisvert (1982) una buena referencia. Otras referencias son el capítulo 6 de Heathfield y Wibe (1987) y el capítulo 12 de Chung (1994).

4.3.1.1. *Características del proceso productivo*

En una translog las características económicas del proceso productivo (elasticidades de producción, productividades marginales, elasticidad de escala y elasticidades de sustitución) no se pueden deducir directamente de los coeficientes de la función sino que hay que obtenerlas algebraicamente. En otras palabras la flexibilidad se consigue haciendo que las características de la tecnología sean función de los inputs usados.

Derivando (4.14) con respecto a $\ln x_j$ se obtiene la elasticidad de producción para el input j:

$$e_j = \beta_j + \beta_{jj} \ln x_j + \sum_{k \neq j} \beta_{jk} \ln x_k \tag{4. 15}$$

El producto marginal del input *j* es:

$$\frac{\partial y}{\partial x_j} = \left[\beta_j + \beta_{jj} \ln x_j + \sum_{j} \beta_{jk} \ln x_k \right] \frac{y}{x_j} \tag{4. 16}$$

La elasticidad de sustitución de Allen en la translog es una expresión complicada donde intervienen los productos marginales y las cantidades de inputs. Por tanto, es una medida específica de la empresa.

Las expresiones de las elasticidades output y de los productos marginales indican que ambas dependen, no sólo, de los parámetros de la función de producción (comunes para todos los individuos) sino de las cantidades de inputs usadas (específicos para cada individuo). Por tanto, la estimación de la función de producción translog permite obtener elasticidades output y productos marginales específicos para cada individuo en la muestra.

En estas expresiones se observa que los efectos económicos de interés (elasticidades output y productos marginales) son una función del valor de los parámetros y de los inputs usados por el productor. Por tanto, es complicado interpretar cada uno de los parámetros por separado. Esta dificultad de interpretación tiene un paralelismo con el poco interés que puede tener realizar contrastes estadísticos sobre los parámetros separadamente cuando se estima una función translog. Existe un caso particularmente interesante cuando los valores de los inputs son 1. En ese caso, las elasticidades output son iguales a los parámetros de primer orden de la función de producción. Este característica será explotada en el próximo capítulo cuando se discutan cuestiones prácticas de estimación.

La función translog puede interpretarse de dos formas alternativas. En un primer caso, se supone que la tecnología viene dada exactamente por la función translog en todo el dominio de datos y se dice que la función se interpreta en forma exacta. En el segundo caso, se interpreta como una aproximación de segundo orden a una función arbitraria; esto es, la expresión (4.12) está aproximando la verdadera función alrededor de un punto. La calidad de la aproximación será mejor cuanto más cerca se encuentre del punto de expansión. Es importante destacar que aunque los coeficientes estimados de la función translog en forma exacta y en forma aproximada son distintos, las características económicas que se estiman son las mismas.[71] Es decir, si se estima la translog en forma exacta y se evalúan las elasticidades de producción en la media geométrica de las variables originales, se obtienen los mismos valores que en la forma aproximada.

4.3.1.2. *Propiedades teóricas*

La función translog no satisface la monotonía y la cuasiconcavidad globalmente, es decir para cualquier valor de las variables explicativas. Berndt y Christensen (1973) demuestran que existe al menos un punto donde esas propiedades no se cumplen. Por tanto, estas propiedades han de ser investigadas en cada punto (forma exacta) o en el punto de aproximación (forma aproximada).

En su variante local, la condición de cuasiconcavidad requiere que la matriz hessiana de derivadas segundas tenga todos sus menores principales no-negativos. Los elementos de esta matriz son de la forma:

$$\frac{\partial^2 y}{\partial y_j \partial y_k} = H_{jk} \qquad (4.\ 17)$$

donde cada uno de los elementos, distinguiendo entre los que pertenecen a la diagonal principal y el resto, quedan del siguiente modo:

[71] En realidad, sólo son distintos los coeficientes de los términos de primer orden.

$$\begin{aligned} H_{jj} &= e_j(e_j - 1) + \beta_{jj} \\ H_{jk} &= e_j e_k + \beta_{jk} \end{aligned} \tag{4. 18}$$

siendo e_j la elasticidad de producción con respecto al input j.

4.3.1.3. *Estructura de la tecnología*

La propiedad de homogeneidad puede imponerse en la función translog mediante la siguiente restricción paramétrica:

$$\sum_k \beta_{jk} = 0 \qquad \forall j \tag{4. 19}$$

La anterior expresión indica que existen tantas restricciones como número de factores de producción en la función.

Para imponer rendimientos constantes de escala, hay que añadir a las restricciones de homogeneidad la siguiente condición:

$$\sum_j \beta_j = 1 \tag{4. 20}$$

Si existen (o se imponen) rendimientos constantes a escala, la translog puede escribirse, en el caso de dos inputs, como:

$$\ln\frac{y}{x_1} = \beta_0 + \beta_1 \ln\frac{x_2}{x_1} + \beta_2\left[\ln\frac{x_2}{x_1}\right]^2 \tag{4. 21}$$

La anterior ecuación es equivalente a la linealización de la función CES sugerida por Kmenta. Por tanto, la función translog con rendimientos constantes a escala es equivalente a la función CES.

La separabilidad débil entre el par de factores *j,k* con respecto al factor *h* requiere que la relación marginal de sustitución de *j* y *k* no se vea afectada por la cantidad de *h* (Berndt y Christensen, 1973). La relación marginal de sustitución puede escribirse como:

$$RMS_{jk} = \frac{\dfrac{\partial y}{\partial x_j}}{\dfrac{\partial y}{\partial x_k}} = \frac{\dfrac{y}{x_j} e_j}{\dfrac{y}{x_k} e_k} = \frac{x_k e_j}{x_j e_k} \tag{4. 22}$$

La condición de separabilidad débil implica que:

$$\frac{\partial RMS_{jk}}{\partial x_h} = \frac{x_k \dfrac{\partial e_j}{\partial x_h} - x_j \dfrac{\partial e_k}{\partial x_h}}{(x_j e_k)^2} = \frac{x_k \dfrac{\beta_{jh}}{x_h} - x_j \dfrac{\beta_{jk}}{x_h}}{(x_j e_k)^2} \qquad (4.\ 23)$$

Esta condición se cumple si:

$$\beta_{jh} = \beta_{kh} = 0 \qquad (4.\ 24)$$

4.3.2. La función de costes translog

La función de costes translog puede escribirse en el caso de un output como:

$$\ln C = \beta_0 + \beta_y \ln y + \frac{1}{2}\beta_{yy}(\ln y)^2 + \sum_{j=1}^{J}\beta_j \ln w_j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K}\beta_{jk}\ln w_k \ln w_j + \sum_{j=1}^{J}\beta_{yj}\ln y \ln w_j \qquad (4.\ 25)$$

El comportamiento optimizador de los agentes exige que la función de costes sea homogénea de grado uno en los precios de los inputs. Esta propiedad puede imponerse mediante las siguientes restricciones paramétricas:

$$\sum_{j=1}^{J}\beta_j = 1 \quad ; \quad \sum_{j=1}^{J}\beta_{jk}\ \forall k \quad ; \quad \sum_{j=1}^{J}\beta_{yj} = 0 \qquad (4.\ 26)$$

Por otra parte, la homoteticidad y la homogeneidad pueden imponerse mediante las siguientes restricciones:

$$\begin{aligned} &\text{a) Homoteticidad: } \beta_{yj} = 0 \\ &\text{b) Homogeneidad: } \beta_{yj} = 0\ ;\ \beta_{yy} = 0 \end{aligned} \qquad (4.\ 27)$$

La aplicación del lema de Shephard a la función de costes translog permite obtener las participaciones en los costes “cost shares” de cada factor:

$$\frac{\partial \ln C}{\partial \ln w_j} = \frac{\partial C}{\partial w_j}\frac{w_j}{C} = x_j \frac{w_j}{C} = S_j \qquad (4.\ 28)$$

A partir de la función de costes translog es muy sencillo calcular las elasticidades de sustitución de Allen:

$$\sigma_{jk} = \frac{\beta_{jk} + S_j S_k}{S_j S_k}$$
$$\sigma_{jj} = \frac{\beta_{jj} + S_j^2 - S_j}{S_j^2} \qquad (4.29)$$

Es evidente que si se anulan los términos de segundo orden de la función de costes translog (β_{jk}=0) la elasticidad de sustitución de Allen entre dos inputs es igual a 1, lo que refleja el hecho de que la translog se ha convertido en una función Cobb-Douglas.

La elasticidad tamaño, que es igual a la inversa de la elasticidad de escala, se obtiene derivando la función de costes con respecto al logaritmo neperiano del output:

$$\varepsilon_{Cy} = \frac{\partial \ln C}{\partial \ln y} = \beta_y + \beta_{yy} \ln y + \sum_j \beta_{yj} \ln w_j \qquad (4.30)$$

La función de costes translog ha sido muy usada en la literatura empírica para estudiar la existencia de economías de tamaño. En este sentido, se ha desarrollado un debate importante sobre si la función translog, al ser una función cuadrática, impone que la curva de costes medios tenga forma de U.[72] Esto invalidaría, por tanto, el uso de esa función para estudiar la influencia del tamaño en los costes medios. Sin embargo, esa crítica no es del todo cierta, como se explica a continuación. La condición necesaria para que la función translog exhiba deseconomías de tamaño es que β_{yy} sea mayor que cero. En primer lugar, es importante destacar que no hay ninguna regla que imponga que β_{yy} tenga que ser necesariamente positivo. Sin embargo, es posible generar datos a partir de una función que imponga la existencia de economías de tamaño (por ejemplo, la función recíproca) y, al estimar una función de costes translog encontrar que a partir de un cierto punto la translog detecta (incorrectamente) la existencia de deseconomías de tamaño. Este problema no es muy grave porque en estos casos en los que la translog se "engaña" con la curvatura de los datos, el mínimo de la función de costes medios siempre cae fuera del rango de los datos observados, por lo que no tiene verdadera relevancia empírica.

4.3.3. Otras formas funcionales flexibles

[72] Ver el artículo de Burton *et al*. (1993) y la respuesta de Hubbard (1993).

4.3.3.1. *La función generalizada de Leontief*

La función generalizada de Leontief (Diewert, 1971) es también una expansión de Taylor de segundo orden. La diferencia con la función translog es que para transformar de las variables independientes se toma raíz cuadrada en vez de logaritmos. Por tanto, la función puede escribirse como:

$$y = \alpha_0 + \sum_j \alpha_j x_j^{1/2} + \frac{1}{2}\sum_j \sum_k \alpha_{jk} x_j^{1/2} x_k^{1/2} \tag{4. 31}$$

Una ventaja de esta forma funcional sobre la translog es que, al no usar logaritmos, permite incluir observaciones nulas en los inputs. Morrison (1988) compara exhaustivamente la generalizada de Leontief con la translog.

En el caso de un proceso productivo con un solo output la función de costes generalizada de Leontief se puede escribir como:

$$C = y\left[\sum_k \sum_j \alpha_{jk} w_j^{1/2} w_k^{1/2}\right] \tag{4. 32}$$

La función de costes generalizada de Leontief ha sido muy usada en la literatura empírica. Algunos ejemplos son los trabajos de Morrison y Schwartz (1996) sobre la medición de la productividad del capital público.

4.3.3.2*. La función cuadrática*

La función de producción cuadrática (Lau, 1974) puede escribirse de la siguiente forma:

$$y = \gamma_0 + \sum_{i=1}^{n} \gamma_i x_i + \frac{1}{2}[\sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} x_i x_j] \tag{4. 33}$$

Esta función comparte con la generalizada de Leontief la ventaja sobre la translog de que los valores nulos de las variables no suponen ningún problema algebraico. Su principal desventaja es que no existe una restricción paramétrica que permita imponer la propiedad de homogeneidad.

La expresión de la elasticidad output respecto al factor *j* es:

$$e_j = (\gamma_j + \gamma_{jj} x_j + \sum_{k \neq j} \gamma_{jk} x_k) \frac{x_j}{y} \tag{4. 34}$$

El producto marginal respecto del input *j* es:

$$\frac{\partial y}{\partial x_j} = \gamma_j + \gamma_{jj} x_j + \sum_{k \neq j} \gamma_{jk} x_k \qquad (4.35)$$

La función de costes cuadrática para el caso de un output se puede escribir como:

$$C = \gamma_0 + \gamma_y y + \frac{1}{2}\gamma_{yy}\, y^2 + \sum_{j=1}^{J} \gamma_j\, w_j + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{jk}\, w_k w_j + \sum_{j=1}^{J} \gamma_{yj} w_j y \qquad (4.36)$$

En esta función no existe una restricción paramétrica que permita imponer la homogeneidad lineal en los precios de los inputs. Sin embargo, una variante de la cuadrática que es homogénea de grado 1 es la siguiente (Baffes y Vasavada, 1989):

$$C = y\left[\sum_{j=1}^{J-1} \gamma_j\, w^*_j + \frac{1}{2}\sum_{j=1}^{J-1}\sum_{k=1}^{K-1} \gamma_{jk}\, w^*_k w^*_j + \sum_{j=1}^{J-1} \gamma_{yj} w^*_j y\right] \qquad (4.37)$$

donde los w* representan precios relativos (normalizados por el precio del input k). En este caso, es evidente que si se multiplican todos los precios por una constante, el coste queda multiplicado por esa constante.

La popularidad de la función cuadrática en el análisis empírico se debe al desarrollo de la función cuadrática de beneficios por Lau (1976). Esta función ha sido usada en numerosas investigaciones empíricas.

4.4. Elección entre formas funcionales

No es difícil comprobar que si un mismo problema empírico se analiza empleando distintas formas funcionales, los resultados que se obtienen suelen ser distintos. Esto plantea la pregunta de cómo escoger una forma funcional concreta. Esta es una tarea difícil ya que el investigador ignora la verdadera estructura productiva.

Fuss *et al*. (1978) sugieren varios criterios para seleccionar entre distintas formas funcionales: (i) parsimonia en los parámetros: la forma funcional no debe contener más parámetros de los necesarios; (ii) facilidad de interpretación: los parámetros deben tener una interpretación directa; (iii) facilidad de estimación: las formas lineales en parámetros son más fáciles de estimar; (iv) robustez interpolativa:

dentro del rango de datos la función debe cumplir ciertas condiciones de regularidad; y (v) robustez extrapolativa: las hipótesis mantenidas deben verificarse fuera del rango de datos.

El criterio más difícil de cumplir por las tres formas funcionales flexibles estudiadas es el último. El problema se debe a que al ser aproximaciones en serie de Taylor de segundo orden (es decir, alrededor de un punto) a cualquier función arbitraria, estas funciones proporcionan una aproximación local, no global. Esto implica que alguna de las condiciones de regularidad (monotonía y curvatura) no se cumplen para algunas observaciones (normalmente, para las más alejadas del punto de aproximación).[73] Por este motivo, Gallant (1981) ha sugerido el uso de expansiones de Fourier para lograr aproximaciones globales a funciones desconocidas. Su complejidad de estimación hace que el número de aplicaciones empíricas sea muy escaso.

La aplicación de estos criterios de selección no permite afirmar que exista una forma funcional flexible superior a las demás. Aunque todas estas funciones permiten aproximar cualquier función arbitraria sin imponer restricciones sobre los parámetros a estimar, los resultados que se obtienen con cada una de ellas suelen ser bastante distintos. La voz de alarma en este sentido la dio el artículo de Wales (1977) en el que usó técnicas de Monte Carlo para comparar la translog (TL) y la generalizada de Leontief (GL).[74] Wales encontró que unas veces la TL era superior a la GL y viceversa, dependiendo del valor de las elasticidades de sustitución.

---

[73] Caves y Christensen (1980) estudiaron teóricamente el cumplimiento global de las propiedades teóricas de la translog y de la generalizada de Leontief. Sus resultados demuestran que las regiones regulares de ambas formas funcionales varían en función de la elasticidad de sustitución. De hecho, la GL satisface las condiciones de regularidad globalmente cuando no hay posibilidades de sustitución, mientras que la TL las satisface cuando la tecnología es Cobb-Douglas (elasticidad de sustitución igual a 1).

[74] En la técnica de Monte Carlo se parte de una forma funcional conocida (incluyendo los valores de los parámetros) y de unos datos (X) determinados. Los valores de la variable dependiente (y) se generan añadiendo realizaciones de una perturbación aleatoria que se toman de una distribución determinada. Para cada extracción de esta variable aleatoria se tiene una muestra distinta (pero con las mismas x's). Este procedimiento se repite hasta que se tiene un número elevado de muestras (miles) y entonces se estima el modelo con la forma funcional original en todas las muestras, se calculan las medias de los coeficientes estimados y se comparan con los parámetros conocidos de partida. De esta manera se puede saber la precisión con que una determinada forma funcional puede estimar la verdadera forma funcional que generó los datos.

Al estudio de Wales le siguieron otros cuyo objetivo era también comparar distintas formas funcionales flexibles usando técnicas de Monte Carlo. Así, por ejemplo, Guilkey, Lovell y Sickles (1983) compararon la TL, GL y la CD generalizada, encontrando que la TL era superior a las otras dos aunque la calidad de su aproximación a la tecnología verdadera se veía muy reducida cuando la elasticidad de sustitución entre inputs se alejaba de la unidad.

Por otro lado, los investigadores también han comparado distintas formas funcionales flexibles usando datos reales. Así, por ejemplo, Berndt y Khaled (1979) compararon la TL y la GL con datos del sector manufacturero de los Estados Unidos. Baffes y Vasavada (1989) usaron datos agregados de la agricultura USA para estimar tres funciones de costes (GL, TL y CN) y compararon la estimación de las elasticidades de sustitución de Allen y del cambio tecnológico no neutral. Asimismo, comprobaron algunas propiedades teóricas, como separabilidad, simetría, concavidad y monotonicidad. El resultado de la comparación fue que las tres formas funcionales proporcionaron resultados distintos, sin que ninguna función demuestre su superioridad frente a las demás.

4.5. Funciones de producción especiales

Una de las ventajas de tener una teoría de la producción muy general, es que permite la modelización de cualquier sector productivo, a pesar de las enormes diferencias que pueden existir entre ellos. Sin embargo, las fuertes especificidades de ciertas actividades productivas han motivado el desarrollo de funciones de producción específicas para ciertos sectores. La idea es que en aquellos sectores en los que haya características de la tecnología que comparten todas las empresas, es interesante incorporar esa información en la función de producción. En este apartado analizaremos dos de estos casos.

4.5.1. Factores limitantes

Existen procesos productivos en los que un factor parece tener una gran importancia. Por ejemplo, la leche líquida en una planta de embotellamiento de leche, la cantidad de madera en una serrería, las manzanas en un lagar de sidra, etc. En estos casos, donde la tecnología es relativamente sencilla, el output suele ser una proporción casi fija de estos inputs, por lo que su inclusión como variables explicativas en la función de producción suele hacer que el resto de las variables no resulten significativas.

Un modo de interpretar el papel de estos factores es considerarlos como factores de difícil sustitución. En un proceso productivo con capital y trabajo casi todo el capital puede ser sustituido con trabajo. Puede ser técnicamente difícil y económicamente inviable pero se puede hacer. Esto no ocurre con la leche en la planta de embotellamiento. Si no hay leche líquida, no hay producción. En otras palabras, la leche es un *factor limitante* de la producción de botellas de leche. De hecho, nunca se va a embotellar más leche de la que entre. En general, se embotellará menos leche de la que entre y las perdidas dependerán de otros factores productivos (trabajo, capital, energía,...).

Este problema no ha sido tratado en economía de la producción con la excepción de la modelización de las capturas en pesca. En este caso, existe un factor de producción, la cantidad (stock) de peces, que no sólo influye en las capturas sino que claramente las limita superiormente ya que no se puede pescar más peces de los que hay. Chambers y Strand (1998) proponen modelizar el fenómeno usando la siguiente función de producción:[75]

$$y = F(S,x) = S\left(1 - e^{-f(x)}\right) \quad , \quad f' > 0, \;\; f'' < 0 \tag{4. 38}$$

donde *S* representa el stock de peces y *x* es un vector de factores variables (número de tripulantes, longitud de las redes, cantidad de cebo, etc.). En esta función de producción se dan los siguientes casos extremos:

$$\begin{aligned} &F(S,0) = 0 \\ &\lim_{x \to \infty} F(S,x) = S \end{aligned} \tag{4. 39}$$

Es decir, sin usar factores variables no se obtiene ninguna captura pero un uso masivo de factor variable sólo logra acercar las capturas al valor del stock. A pesar de las buenas propiedades de esta función, la mayor parte de los estudios de producción en el sector pesquero han estimado la función de producción F(S,x), suponiendo que

[75] Esta función es conocida también como función de producción Beverton-Holt. La aportación de Chambers y Strand (1998) consiste en desarrollar la función de costes dual de esta función al tiempo que desarrollan un modelo que es consistente desde el punto de vista económico y biológico.

es Cobb-Douglas o translog. Estas funciones, ampliamente usadas en otros sectores productivos, no tienen en cuenta que el stock de peces actúa como un factor limitante y que, por tanto, la función de producción debe tener una asíntota superior al nivel del stock. Es decir, no se puede pescar más peces de los que hay, independientemente de la cantidad de inputs empleada.

La función de producción (4.38) se puede ver como una *función de pérdidas*, como sigue:

$$\ln\left(\frac{S-y}{S}\right) = -f(x) \qquad (4.\ 40)$$

El cociente dentro del paréntesis mide en términos proporcionales el volumen de peces que, siendo posible capturar dado el stock existente, no hemos capturado. Obviamente, esta pérdida depende inversamente del empleo de los inputs variables, de ahí el signo menos delante de f(x). Por lo tanto, para estimar procesos productivos en los que un input juega un papel limitativo, Chambers propone sustituir el output como variable dependiente por el logaritmo de una medida proporcional de las pérdidas.

4.5.2. Función de respuesta al fertilizante

A mediados del siglo XIX, von Liebig estudió la respuesta de los cultivos a las dosis de abono y expuso su "ley del mínimo", por la que el rendimiento de un cultivo es proporcional a la cantidad del nutriente más escaso. Es decir, si por ejemplo el nitrógeno es el factor limitante y se aumenta la cantidad de fósforo o potasio, el rendimiento no cambiará. Esto implica que las isocuantas tienen forma de ángulos rectos. Por tanto, las funciones de producción habituales, que implican productividades marginales positivas, no son adecuadas si la hipótesis de von Liebig es correcta. Paris (1992) encontró que las funciones de producción von Liebig eran superiores a otras en la modelización de la respuesta de la producción a la fertilización.

4.5.3. Función de producción con pesticidas

Un grupo importante de inputs en agricultura lo constituyen los pesticidas (herbicidas, insecticidas y funguicidas). El uso de estos productos químicos se ha generalizado de forma importante durante este siglo, contribuyendo en buena medida al aumento de productividad que ha tenido lugar en el sector agrario. Sin embargo, las características de los pesticidas son distintas de las de otros inputs agrícolas (fertilizantes, semillas,...) cuya contribución a la producción es directa. Efectivamente, los pesticidas no aumentan la producción sino que contribuyen a la producción evitando las pérdidas de output que tendrían lugar si no se controlasen las malas hierbas o los efectos de insectos u hongos dañinos.

La incorporación de estos inputs en la función de producción se hace a través de lo que se conoce como función de reducción de daño ("damage abatement function"). Por tanto, la función de producción puede escribirse como:

$$y = f[x, g(q)] \tag{4. 41}$$

donde X representa un vector de inputs "normales" en agricultura (tierra, trabajo, fertilizantes, riego,...) y g(q) es la función de reducción de daño que depende de un vector de pesticidas (q).

La especificación definitiva de la función de producción a partir de (4.41) requiere resolver varios problemas: a) Elección de forma funcional; b) Interacción de los inputs directos y los pesticidas; y c) Especificación de la estructura estocástica.

4.6. Problemas de medición de variables

En esta sección se comentan brevemente los problemas de medición de las variables (outputs, costes, beneficios, inputs, precios de inputs y precios de outputs) que entran como argumento en las funciones de producción, costes y beneficios presentadas en las secciones anteriores. Esta sección, por lo tanto, sirve de enlace entre el presente capítulo y el siguiente, en el que se abordan algunos problemas econométricos que surgen frecuentemente a la hora de estimar funciones de producción, costes y beneficios.

4.6.1. Medición de los outputs

En principio, la medición del output para una empresa que produce un único bien es algo sencilla. No obstante, incluso en este caso, pueden surgir algunas dudas. Es habitual que el producto de las distintas empresas no sea homogéneo, es decir, que existan diferencias de calidad entre empresas o a lo largo del tiempo. Este problema probablemente sea poco relevante en la producción de acero, pero en la producción de ordenadores, la calidad cambia a lo largo del tiempo. Pueden surgir, por tanto, dudas sobre la forma de tratar estas diferencias de calidad cuya medición suele ser complicada.

Una posibilidad es medir el output en términos monetarios aunque también en este caso surgen problemas. En concreto, la utilización del valor añadido en los estudios de productividad exige que la función de producción sea separable en los inputs intermedios o que los precios de los inputs varíen en la misma proporción que el precio del output. Ver, a este respecto, Bruno (1978).

Otro problema que se encuentra a la hora de especificar el output es que la mayoría de las empresas producen más de un output. Por ejemplo, en banca se suelen considerar tanto los créditos como los depósitos, o, en pesca, es normal capturar varias especies. Una solución a este problema, estudiada originalmente por Mundlak (1963), consiste en agregar todos los outputs en una única medida. La agregación lleva implícito un supuesto sobre la relación entre los outputs, por lo que es razonable preguntarse en qué medida puede afectar a los resultados la forma en la que se realiza la agregación. Si los distintos outputs están medidos en las mismas unidades el output agregado puede consistir simplemente en la suma de todos los outputs (por ejemplo, sumar distintos tipos de créditos o las capturas de distintas especies). Esta agregación trata a los distintos outputs de igual forma por lo que implícitamente está suponiendo que a la empresa le da lo mismo producir un output u otro. Por tanto, esta suma puede afectar seriamente a las estimaciones dado que no tiene en cuenta las diferencias existentes entre los distintos outputs. Por ejemplo, los créditos hipotecarios exigen utilizar menos recursos que los créditos a empresas, debido a las diferencias en plazos de amortización y en riesgo que existen entre ambos.

Por otra parte, cuando los outputs están medidos en distintas unidades la forma habitual de proceder es utilizar los ingresos (ventas), que es la estrategia seguida para medir el Producto Interior Bruto de un país. El uso de los ingresos requiere conocer los precios de todos los outputs, que no están disponibles en muchos casos. Esta estrategia es además criticable si las empresas actúan en mercados no competitivos, ya que si los precios no son estrictamente exógenos, existen errores en la cuantificación del "verdadero" output agregado. En este caso, una mayor producción puede conducir a un menor precio y, por tanto, a un sesgo a la baja en la medida del output.

Aparte de los problemas comentados, la agregación simple de outputs tanto en términos físicos como monetarios puede afectar a los resultados de un modo sustancial ya que no tiene apoyo teórico. Por este motivo, los investigadores han avanzado en el tratamiento teórico de los procesos de producción multiproducto. En este sentido, cabe señalar la expansión extraordinaria que ha tenido en los últimos años la función de distancia para modelizar tecnologías multiproducto.

4.6.2. Medición de los inputs

La medición del input capital presenta serias dificultades. Un primer problema surge porque el capital no es un concepto homogéneo, sino que es la suma de una serie elementos, principalmente máquinas e infraestructuras. La diversidad de bienes de capital que suele haber en una empresa obliga a realizar algún tipo de agregación y, dada la heterogeneidad de los mismos, la única solución es agregarlos por su valor y calcular el coste de uso anual de esos activos. En el caso de que el capital sea propiedad de la empresa, la forma habitual de aproximar el coste de uso es utilizar la amortización del activo según su coste de reposición. Sin embargo, si la tasa de uso del capital es distinta entre empresas, una tasa de amortización común para todas ellas no es una buena medida del uso del input.[76]

Otro problema radica en que en muchas ocasiones se suele medir el stock de capital y no el flujo de servicios del input, que sería la medida adecuada, lo que puede causar problemas si no se usa al cien por cien de su capacidad. Así, por ejemplo, un aumento en el nivel de producción causado por un aumento en el flujo de servicios del capital puede atribuirse erróneamente a un avance tecnológico dado que, con el mismo *stock* de capital, se produce más output. Dado que parece lógico imputar únicamente el coste de capital efectivamente usado, es necesario ajustar los datos por alguna medida de la capacidad utilizada, que, a nivel de empresa, es una información

[76] Existen otras formas de medir el capital. Por ejemplo, Griliches y Ringstad (1971) lo calculan a partir del valor dado en el seguro de incendios.

inexistente.[77] En algunos trabajos se calcula la cantidad efectiva de capital multiplicando la medida del input capital por la tasa de utilización de la capacidad que, para la economía española, calcula el INE. De esta manera se recoge la influencia del ciclo económico sobre la utilización del capital.

Otro problema es la variación en la calidad de los inputs entre distintas empresas. En la mayoría de los casos, no existen datos sobre esas diferencias de calidad. Por ejemplo, es frecuente tener el número de trabajadores pero no la cualificación de los mismos, o se tiene el número de hectáreas de terreno de las explotaciones agrarias pero no se sabe si el suelo es igual de productivo o la pendiente de las parcelas. Esas características de los inputs que no son medibles (o bien no se han medido) pasan a formar parte de lo que se conoce como "heterogeneidad inobservable", que termina incorporándose al término de error, por lo que si existe correlación entre alguna de estas características no medidas y las variables explicativas, los estimadores estarán sesgados (Griliches, 1957). En algunos trabajos, se propone evitar este problema estimando una función de producción que incluye como "input" el coste del mismo en lugar de la cantidad. Puesto que a mayor calidad, mayor precio, el coste recogerá no sólo diferencias en cantidad sino también en calidad.

Cuando se trabaja con inputs agregados, primero debe determinarse si la agregación está justificada desde el punto de vista teórico, lo cual viene condicionado por la existencia de algún tipo de separabilidad. En caso afirmativo, el segundo paso es seleccionar la forma correcta de proceder a la agregación, lo cual está estrechamente relacionado con la teoría de los números índices. Es bien conocido en la literatura de los números índices que la fórmula específica que debe utilizarse para agregar outputs, inputs o precios depende de las características de la tecnología que se pretende modelizar (Balk, 1998).

Una forma de ver intuitivamente las implicaciones de la tecnología para la agregación y en especial e concepto de separabilidad es a través del siguiente ejemplo. Supóngase que queremos trabajar con una función de producción y=f(K,L)

[77] Una notable excepción es la Encuesta Industrial, realizada por el INE y la Fundación Empresa Pública.

donde $K=g(K_1,K_2)$ es una medida agregada del capital. Si la tecnología subyacente puede escribirse a través de la función de producción $y=\alpha(K_1+K_2)+\beta L$, el capital se puede agregar sumando simplemente las cantidades de ambos tipos de capital, esto es, $K=K_1+K_2$. En este caso, la tecnología queda plenamente caracterizada con el trabajo (L) y la cantidad total de capital (K) ya que lo que ocurre "dentro" del capital agregado no depende del nivel de trabajo o del output producido. Si la tecnología no fuera separable en el capital, lo cual ocurriría si función de producción se pudiera escribir como $y=\alpha(K_1+L\cdot K_2)+\beta L$, no es posible utilizar una función agregadora del tipo $K=g(K_1,K_2)$ debido a que impondría unas ponderaciones "constantes" para ambos tipos de capital, mientras que lo "correcto" sería ponderar K_2 en función de la cantidad de trabajo.

Un problema común a la medición tanto de outputs como de inputs surge cuando éstos se miden por su valor ya que en contextos temporales es necesario "deducir" el efecto que los cambios en los precios han tenido sobre los cambios en el valor, las series han de deflactarse utilizando los índices de precios o deflactores relevantes.

4.6.3. Medición de los costes

Los costes variables no presentan grandes dificultades de cálculo desde el punto de vista contable, aunque puede surgir un problema si no se contabilizan adecuadamente las existencias a principio o final de período. Los costes fijos presentan mayores dificultades de valoración. Los principales problemas se plantean con los recursos propios, como la mano de obra familiar y la tierra o el capital en propiedad. Estos factores, aunque su uso no supone desembolso alguno para la empresa, tienen un coste, ya que desde un punto de vista económico los recursos deben ser valorados por su coste de oportunidad. Sin embargo, conocer ese coste de oportunidad es complicado pues no siempre existe un valor de mercado.

Este problema puede ser de escasa importancia si el objetivo de la investigación es el análisis a corto plazo de las decisiones de la empresa. De hecho, en este caso, la función de costes variables a corto plazo es suficiente para modelizar este comportamiento. Esta función tiene como variables explicativas las cantidades de factor fijo en vez de los precios de éstos. Las cantidades físicas de factores fijos son más fáciles de obtener en el análisis empírico aunque no están exentas de problemas, como se verá a continuación.

La función de costes a largo plazo es la herramienta básica de análisis del comportamiento de los productores a largo plazo. En este caso, todos los factores son variables y las variables explicativas son los precios de todos los factores productivos y el nivel de output. Evidentemente, en este caso el análisis empírico pasa por encontrar alguna medida del precio de los factores que son fijos a corto plazo. El problema es todavía más complicado ya que es razonable pensar que un incremento en un input fijo mejore los costes del periodo analizado y de periodos futuros. A su vez, el coste del input fijo no sólo es su coste de compra en el periodo de cambio sino que puede venir dado por interrupciones en la producción que pueden ocurrir a lo largo de varios periodos. Este coste puede variar dependiendo de la rapidez con que se haga el cambio[78]. En resumen, el análisis a largo plazo de los costes tiene serias dificultades empíricas tanto desde el punto de vista de la modelización como de la medición adecuada de los precios de los factores fijos.

4.6.4. Precios

En estudios de corte transversal es bastante habitual que los precios de los factores tengan relativamente poca variación entre las distintas empresas. Esto es así cuando las empresas de la muestra compran sus inputs en un mismo mercado. Sin embargo, en algunas ocasiones existe variación en los precios de los inputs que no se debe a que procedan de distintos mercados, sino a otros factores, como la calidad. Un ejemplo es el de la cualificación de la mano de obra. Normalmente, se tienen datos del total de salarios pagados por la empresa y del número de trabajadores, lo que permite calcular un salario medio. Sin embargo puede haber dos empresas con el mismo número de trabajadores pero con distinta proporción de trabajadores cualificados. En este caso la empresa con mayor número de trabajadores cualificados tendría un salario medio mayor pero probablemente tendría también trabajadores más productivos. Este tipo de situación, que se puede presentar con cualquier tipo de input, puede explicar que en algunas veces el efecto estimado del precio del input sobre el coste sea negativo. Para obtener este resultado bastaría con que el incremento de productividad de un input por la cualificación fuese superior al incremento de su precio. Como resultado, el coste disminuiría con un salario más elevado.

Otro problema adicional, es que las empresas adquieren materias primas y otros factores productivos en condiciones no competitivas, por lo que los precios no serían estrictamente exógenos como así exige la teoría. Así, por ejemplo, en los estudios sobre el sector bancario, se estiman funciones de costes en los que se incluye el precio de los depósitos como variable explicativa, aunque existe evidencia empírica en contra de la existencia de competencia perfecta en dicho mercado.

[78] Véase a este respecto la última sección del Capítulo 2 dedicado a los costes de ajuste.

Bibliografía

Baffes, J. y U. Vasavada, (1989), "On the Choice of Functional Forms in Agricultural Production Analysis", *Applied Economics*, 21,1055-1061.

Balk, B. (1998), *Industrial Price, Quantity, and Productivity Indices. The Micro-economic Theory and an Application*, Kluwer Academic Publishers.

Berndt, E. y L. Christensen, (1973), "The translog function and the substitution of equipment, structures and labor in US manufacturing", *J. of Econometrics*, 1, 81-114.

Berndt, E. y M.S. Khaled, (1979), “Parametric Productivity Measurement and Choice among Flexible Functional Forms”, *J. of Political Economy*, 87(6), 1220-1245.

Boisvert, R.N., (1982), *The translog production function: its properties, its several interpretations and estimation problems*, Dept. of Agr. Econ., Cornell University.

Bruno, M. (1978), "Intermediate Inputs and Value-Added", en M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*, Vol. 2, North-Holland, 3-16.

Burton, M., P.A. Ozanne y C. Collison, (1993), “Long-run Average Cost Curves in the England and Wales Dairy Industry – Comment”, *J. of Agricultural Economics*, 44, 502-506.

Caves, D.W. y L. Christensen, (1980), “Global Properties of Flexible Functional Forms”, *American Economic Review*, 70(3), 422-432.

Chambers, R., (1988), *Applied Production Analysis*, Cambridge University Press.

Chambers, R. e I.E. Strand, (1998), “A Dual, Bioeconomic Model of Short-Run Fisheries Production”, trabajo presentado en el IIFET IX, Tromso, Noruega.

Christensen, L., D. Jorgenson y L. Lau, (1973), "Trascendental Logarithmic Production Frontiers", *Review of Economics and Statistics*, 55, 28-45.

Chung, J. W. (1994), *Utility and Production Functions*, Blackwell Publishers, Cambridge Massachusetts.

de Janvry, A.C., (1972), “The Class of Generalized Power Production Functions”, *American J. of Agricultural Economics*, 54(2), 234-237.

Diewert, E., (1971), “An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function”, *J. of Political Economy*, 79, 481-507.

Färe, R., (1975), “A Note on Ray-Homogeneous and Ray-Homothetic Production Functions”, *Swedish J. of Economics*, 77, 366-372.

Färe, R. L. Jansson y K. Lovell, (1985), “Modelling Scale Economies with Ray-Homothetic Production Functions”, *Review of Economics and Statistics*, 55, 28-45.

Fuss, M., D. McFadden y Y. Mundlak, (1978), "A Survey of Functional Forms in the Economic Analysis of Production", en M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*, Vol. 1, North-Holland, 219-268.

Gallant, R., (1981), “On the Bias in Flexible Functional Forms and an Essentially Unbiased Form”, *J. of Econometrics*, 15, 211-245.

Griliches, Z., (1957), “Specification Bias in Estimates of Production Functions”, *J. of Farm Economics*, 39, 8-20.

Guilkey, D.K., C.A.K. Lovell y R. Sickles, (1983), “A Comparison of the Performance of Three Flexible Functional Forms”, *International Economic Review*, 24, 591-616.

Heathfield, D.F. y S. Wibe, *An Introduction to Cost and Production Functions*, MacMillan.

Hubbard, L., (1993), “Long-run Average Cost Curves in the England and Wales Dairy Industry – Reply”, *J. of Agricultural Economics*, 44, 507-510.

Lau, L., (1974), “Comments on Applications on Duality Theory”, en M.D. Intriligator y D.A. Kandrick (eds.), *Frontiers of Quantitative Economics*, Vol. II, North-Holland, 176-199.

Lau, L., (1976), “A Characterization of the Normalized Restricted Profit Function”, *J. of Economic Theory*, 12, 131-163.

Morrison, C., (1988), ”Quasi-Fixed Inputs in US and Japanese Manufacturing: A Generalizaed Leontief Restricted Cost Function Approach”, *Review of Economics and Statistics*, 70(2), 275-87.

Morrison, C., (1996), ”State Infraestructure and Productive Performance”, *American Economic Review*, 86(5), 1095-1111.

Mundlak, Y., (1963), “Specification and Estimation of Multiproduct Production Functions”, *J. of Farm Economics*, 45, 433-443.

Paris, Q., (1992), “The von Liebig Hypothesis”, *American J. of Agricultural Economics*, 74(4), 1019-1028.

Thompson, G.D. (1988), "Choice of Flexible Functional Forms: Review and Appraisal", *Western J. of Agricultural Economics*, 13(2), 169-183.

Wales, T.J. (1977), "On the Flexibility of Flexible Functional Forms", *J. of Econometrics*, 5, 183-193.

Zellner, A. y N.S. Revankar, (1969), “Generalized Production Functions”, *Review of Economic Studies*, 36, 241-250.

CAPITULO 5

ASPECTOS PRÁCTICOS DE LA ESTIMACIÓN ECONOMETRICA DE MODELOS DE PRODUCCIÓN

En este capítulo se abordan algunos problemas econométricos que surgen frecuentemente a la hora de estimar funciones de producción, costes y beneficios. Puesto que son muchos los problemas de tipo empírico a los que hay que hacer frente en la práctica diaria, hemos decido centrarnos en aquéllos que son comunes a todas las aplicaciones. Así, el capítulo comienza analizando los problemas de estimación de las formas funcionales más usadas. A continuación, se tratan el problema de la endogeneidad de las variables explicativas, la estimación en presencia de "heterogeneidad inobservable" y la modelización y estimación de funciones frontera. El capítulo termina con una reflexión sobre la contrastación estadística de las propiedades teóricas en modelos de producción.

5.1. Aspectos prácticos de la estimación de funciones de producción

En esta sección se describen las particularidades que presenta la estimación econométrica de las funciones de producción más empleadas en el análisis empírico: Cobb-Douglas y Translog.

5.2.1. Función Cobb-Douglas

Aunque la función de producción Cobb-Douglas (CD) no es lineal en los parámetros en su formulación original, la función puede "linealizarse" tomando logaritmos. En el caso de dos inputs la expresión toma la siguiente forma:

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + u \qquad (5.\ 1)$$

que es lineal en los parámetros y puede estimarse por mínimos cuadrados ordinarios. Al tomar logaritmos se reduce la escala de valores de las variables (es decir, se moderan las diferencias entre las empresas grandes y pequeñas), por lo que la estimación de la función CD en su forma logarítmica puede reducir los problemas de heteroscedasticidad.

Una formulación alternativa, que permite contrastar directamente la existencia de rendimientos constantes a escala, consiste en restar a ambos lados de la ecuación (5.1) el término $\ln x_1$. Adicionalmente, en la parte derecha de la ecuación se suma y resta $\beta_2 \ln x_1$. Agrupando términos, la función de producción CD puede escribirse finalmente como:

$$\ln \frac{y}{x_1} = \beta_0 + \gamma \ln x_1 + \beta_2 \ln \frac{x_2}{x_1} + u \qquad (5.\ 2)$$

donde $\gamma=\beta_1+\beta_2-1$, por lo que la hipótesis $\gamma=0$ permite contrastar la existencia de rendimientos constantes a escala de forma inmediata.

Un problema de la función Cobb-Douglas, compartido con la translog, es que en algunas ocasiones hay inputs que toman el valor cero, por lo que no se pueden tomar logaritmos. Esto puede suceder con algunos inputs muy específicos, como por ejemplo la fertilización. Hay algunas explotaciones agrarias de tipo familiar que no compran abonos químicos, por lo que el input fertilización toma el valor cero. Es necesario entender, no obstante, que este es un problema genérico que surge cuando se trata de hacer un análisis empírico con datos desagregados. De hecho, en un análisis agregado, el fertilizante puede ser parte de la partida de otros gastos y es prácticamente imposible que tome el valor de cero.

Cuando existe un número sustancial de observaciones nulas el problema es de gran dificultad empírica. Una solución (sencilla) consiste en suprimir las observaciones que presentan ceros, lo cual en principio no es aconsejable ya que se podría estar desaprovechando información relevante sobre el fenómeno analizado y sesgando los resultados si las empresas eliminadas tienen características especiales. Cabe preguntarse, en cualquier caso, si las empresas que persistentemente no emplean determinados inputs están realmente utilizando la misma tecnología, como así se supone implícitamente al estimar una función de producción común a todas las observaciones. Es decir, un cero puede representar todo un conjunto de diferentes situaciones. Desde el productor que está a punto de usar el input (utilizando una tecnología similar) al que se encuentra muy lejos de poder usarlo (dado que está utilizando una tecnología sustancialmente diferente).[79]

Una práctica habitual es sustituir los ceros por un número pequeño (por ejemplo, la unidad), de forma que se pueda tomar el logaritmo. Supongamos, por ejemplo, que el input x_2 de la función (5.1) tiene varios ceros. Battese (1997) sugiere sustituir x_2 por $x_2^*=\max(x_2,D_2)$ donde D_2 es una variable ficticia que toma el valor 1 cuando x_2 es cero o el valor 0 cuando x_2 es positivo. El modelo a estimar sería entonces:

$$\ln y = \beta_0 + \gamma_0 D_2 + \beta_1 \ln x_1 + \beta_2 \ln x_2^* + u \qquad (5.\ 3)$$

En muchas ocasiones la función Cobb-Douglas se especifica con una o más variables binarias. En este caso, la interpretación de los coeficientes de estas variables tiene una cierta complejidad. A continuación se describe un ejemplo en el que se especifica una función Cobb-Douglas con una variable ficticia (D) que indica la región geográfica en la que se sitúa la empresa.

$$y = A\prod x_i^{\alpha_i} e^{\gamma D} \qquad (5.\ 4)$$

La variable dummy (D) toma el valor 1 si la empresa está en la región 1 y cero si está en la región dos. Es decir:

[79] Este tema ha sido estudiado para el caso del consumo por Perali y Chavas (2001).

$$D = 1 \Rightarrow y_1 = A \prod x_i^{\alpha_{i.}} e^{\gamma}$$
$$D = 0 \Rightarrow y_2 = A \prod x_i^{\alpha_{i.}} \qquad (5.\ 5)$$

Restando una expresión de la otra, se obtiene:

$$y_1 - y_2 = A \prod x_i^{\alpha_{.}} (e^{\gamma} - 1) \quad \Rightarrow \quad \frac{y_1 - y_2}{y_2} = (e^{\gamma} - 1) \qquad (5.\ 6)$$

Por tanto, (e^{γ}-1) se interpreta como la tasa de cambio entre la producción de la región 1 y la 2 debida exclusivamente al hecho de la localización.[80] Alternativamente, para valores pequeños, existe una interpretación aproximada del parámetro γ. Operando en la expresión (5.6) se tiene que:

$$e^{\gamma} = 1 + \frac{y_1 - y_2}{y_2} \qquad (5.\ 7)$$

Tomando logaritmos neperianos en la expresión anterior para valores pequeños de la tasa de crecimiento se tiene que:

$$\gamma = \ln\left(1 + \frac{y_1 - y_2}{y_2}\right) \approx \frac{y_1 - y_2}{y_2} \qquad (5.\ 8)$$

5.3.2. Función Translog

Al igual que la función Cobb-Douglas, la función Translog (TL) es lineal en los parámetros y puede estimarse por MCO, pero al incluir términos de segundo orden no impone que las elasticidades (de escala, de sustitución, etc.) sean iguales para todas las empresas. Existen dos formas alternativas de interpretar la función translog. La primera es como una *forma exacta*, es decir, se supone que la tecnología subyacente tiene las características de una función con la estructura translog y la ecuación se estima por MCO sin ninguna transformación. Alternativamente, se puede considerar la translog como una aproximación de segundo orden a una función arbitraria, estimando lo que se conoce como una *forma aproximada*. Una aproximación de Taylor de segundo orden en un punto $\ln x^0$ puede escribirse como:

$$\ln y_i = \beta_0 + \sum_{j=1}^{J} \beta_j \left(\ln x_{ji} - \ln x_{ji}^0\right) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \left(\ln x_{ji} - \ln x_{ji}^0\right)\left(\ln x_{ki} - \ln x_{ki}^0\right) + u_i \qquad (5.\ 9)$$

o, lo que es lo mismo,

$$\ln y_i = \beta_0 + \sum_{j=1}^{J} \beta_j \ln\left(x_{ji}/x_{ji}^0\right) + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk}\left(x_{ji}/x_{ji}^0\right)\left(x_{ki}/x_{ki}^0\right) + u_i \qquad (5.\ 10)$$

donde el subíndice *i* representa una observación. Es decir, la aproximación de Taylor de segundo orden es una función de producción translog donde las variables se expresan como cocientes entre el punto observado y el punto de aproximación. Una práctica habitual es considerar la media geométrica de los datos como el punto de aproximación. Para ello se dividen todas las variables independientes por su media geométrica. Es decir, se construyen unas nuevas variables X* de la siguiente forma:

$$x_i^* = \frac{x_i}{x^M} = \frac{x_i}{\sqrt[N]{\prod_{i=1}^{N} x_i}} = \frac{x_i}{\prod_{i=1}^{N} x_i^{\frac{1}{N}}} \qquad (5.\ 11)$$

La ventaja de hacer esta transformación es que la media geométrica de las nuevas variables es igual a 1.

$$\prod_{i=1}^{N}\left(x_i^*\right)^{\frac{1}{N}} = \prod_{i=1}^{N} \frac{\left(x_i\right)^{\frac{1}{N}}}{\left(\prod_{i=1}^{N} x_i^{\frac{1}{N}}\right)^{\frac{1}{N}}} = \frac{\prod_{i=1}^{N} x_i^{\frac{1}{N}}}{\prod_{i=1}^{N} x_i^{\frac{1}{N}}} = 1 \qquad (5.\ 12)$$

Por tanto, cuando se emplea la translog como una aproximación en serie de Taylor, al evaluar las elasticidades usando como punto de expansión la media geométrica de los datos transformados, el logaritmo de las cantidades utilizadas de inputs es cero, por lo que las expresiones anteriores se simplifican enormemente. En este caso, los coeficientes de primer orden son las elasticidades de producción de cada input. Efectivamente, como se puede ver en la ecuación (5.9) al tomar neperianos en la media geométrica de las variables transformadas, la elasticidad del input j es β_j. Esto constituye una importante ventaja de la forma aproximada sobre la exacta ya que los estadísticos t asociadas a los coeficientes de primer orden en la forma aproximada permiten el contraste estadístico de si las elasticidades de producción evaluadas en la media geométrica son significativamente distintas de cero. Por el contrario, los coeficientes de la translog en forma exacta no tienen una interpretación directa, por lo

[80] Suits (1983) tiene una sencilla exposición sobre la interpretación de los coeficientes de variables ficticias.

que si queremos analizar las elasticidades (u otras características económicas relevantes de la función de producción) es necesario obtener primero las elasticidades y, además, calcular las desviaciones típicas de sus estimadores, cuya expresión es relativamente complicada.[81]

Es preciso destacar, no obstante, que la tecnología estimada es la misma tanto si se estima una función TL en forma exacta o aproximada. De hecho, la única diferencia entre la forma exacta y la aproximada es que en esta última se han reescalado las variables independientes, dividiendo cada variable por una constante (su media geométrica).[82] El efecto de este cambio de escala afecta a los coeficientes estimados de los términos de primer orden (los términos cuadráticos y los productos cruzados no cambian), pudiendo llegar a cambiar incluso de signo.[83]

Esto se puede ver fácilmente en la Figura 5.1, donde se representa, en términos logarítmicos, una función TL con un único input. Cuando estimamos en forma exacta los parámetros de primer orden miden la pendiente de la TL (esto es, la elasticidad) cuando $\ln X=0$, es decir, cuando las empresas emplean sólo una unidad de input, valor que no es nada relevante si la mayoría de las empresas utilizan claramente más de una unidad de dicho input. Cuando estimamos en forma aproximada, se está desplazando el eje vertical hacia la derecha hasta el valor medio de los datos, por lo que los parámetros de primer orden miden ahora la pendiente de la TL en la media geométrica de los datos, cuando $\ln X=\ln X^M$. Obsérvese, no obstante, que la curvatura (concavidad) de la función es la misma en la forma exacta y aproximada debido a que los parámetros de segundo orden (términos cuadráticos y productos cruzados) (que son los que determinan el grado de curvatura de una función) son idénticos.

---

[81] La expresión general de la varianza de las elasticidades se puede ver en Anderson y Thursby (1986).

[82] En el caso de un modelo lineal, el cambio de escala en una variable se traduce en un cambio en su coeficiente, mientras que en una Cobb-Douglas, sólo afecta al término independiente.

[83] Una sencilla nota que explica con claridad las consecuencias de los cambios en las unidades de medida en los modelos translog es el artículo de Hunt y Lynk (1993).

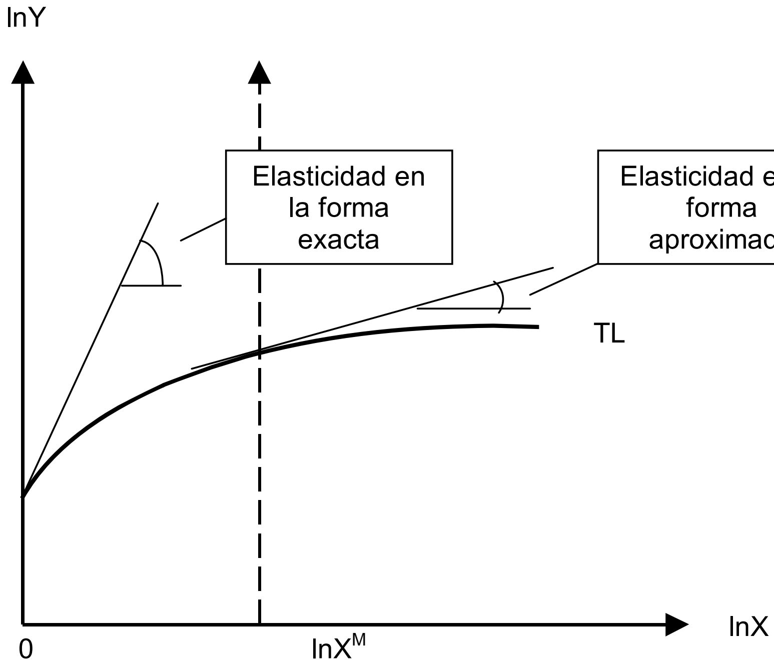

Figura 5.1. *Forma aproximada vs. forma exacta*

White (1980) analiza las dificultades (cuando no la imposibilidad) de estimar por MCO una aproximación de Taylor a una función arbitraria. Si aceptamos este argumento, la práctica habitual de estimar una función translog en el centro de la distribución no pasa de ser un conveniente cambio de unidades que permite una interpretación directa de los coeficientes de primer orden y estadísticos t de las elasticidades output evaluadas en el centro de la distribución.

Comentarios similares se pueden realizar cuando se estima una función de costes translog que, en el caso de múltiples outputs, puede escribirse como:

$$\begin{aligned}\ln C_i = \quad & \beta_0 + \sum_{l=1}^{M} \beta_l \ln y_{li} + \frac{1}{2}\sum_{l=1}^{M}\sum_{m=1}^{M} \beta_{lm} \ln y_{li} \ln y_{mi} + \sum_{j=1}^{J} \theta_j \ln w_{ji} + \\ & \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \theta_{jk} \ln w_{ji} \ln w_{ki} + \frac{1}{2}\sum_{l=1}^{M}\sum_{j=1}^{J} \gamma_{lj} \ln y_{li} \ln w_{ji} + u_i\end{aligned} \tag{5. 13}$$

La función de costes es homogénea de grado uno en los precios de los inputs. Esta propiedad se cumple si:

$$\begin{aligned}&\sum_{j=1}^{J}\theta_j = 1\\&\sum_{j=1}^{J}\theta_{jk} = \sum_{k=1}^{J}\theta_{jk} = 0\\&\sum_{j=1}^{J}\gamma_{lj} = 0 \quad \forall l\end{aligned} \qquad (5.\ 14)$$

Este sistema de restricciones se puede utilizar con dos finalidades: 1) para contrastar la propiedad de homogeneidad lineal en precios asociada a la minimización de costes; 2) para reducir el número de parámetros a estimar. En efecto, estimar la función de costes (5.13) imponiendo el cumplimiento de (5.14), es equivalente a estimar la siguiente ecuación:

$$\begin{aligned}\ln C_i = \; & \beta_0 + \sum_{l=1}^{M}\beta_l \ln y_{li} + \frac{1}{2}\sum_{l=1}^{M}\sum_{m=1}^{M}\beta_{lm}\ln y_{li}\ln y_{mi} + \sum_{j=1}^{J-1}\theta_j \ln\left(w_{ji}/w_{Ji}\right) + \\ & \frac{1}{2}\sum_{j=1}^{J-1}\sum_{k=1}^{J-1}\theta_{jk}\ln\left(w_{ji}/w_{Ji}\right)\ln\left(w_{ki}/w_{Ji}\right) + \frac{1}{2}\sum_{l=1}^{M}\sum_{j=1}^{J-1}\gamma_{lj}\ln y_{li}\ln\left(w_{ji}/w_{Ji}\right) + u_i\end{aligned} \qquad (5.\ 15)$$

donde ahora el coste total y los precios de los inputs aparecen normalizados por el precio de uno de los inputs. Obsérvese que el sumatorio en los precios se reduce en un término, lo cual significa que el número de parámetros a estimar disminuye cuando se impone la propiedad de homogeneidad lineal. La ecuación (5.15) sugiere, además, una forma sencilla de imponer el conjunto de restricciones (5.14), que en muchas aplicaciones puede llegar a ser algo muy laborioso. Es decir, basta con seleccionar un input como numerario, dividir el coste total y el precio del resto de inputs por el precio de dicho input, y "construir" la función de costes translog con una variable menos.

Una cuestión que suele surgir cuando se impone la homogeneidad en precios mediante la normalización del coste y los precios es si ésta debe hacerse antes o después de normalizar las variables por su media geométrica si se quiere estimar la función de costes en su forma flexible. En otras palabras, ¿es correcto imponer primero la homogeneidad y dividir después los precios *relativos* (w_j/w_J) por su media geométrica? La duda surge por el hecho de que tanto la propiedad de homogeneidad como la especificación flexible de la función de costes implican una normalización. El origen es, sin embargo, de muy distinta naturaleza. Y aquí está precisamente la respuesta a la pregunta anterior. En efecto, la homogeneidad lineal en precios es equivalente a la imposición de una serie de restricciones sobre los parámetros de una función de costes, tanto si ésta está especificada en forma exacta o flexible. De aquí se deduce, por tanto, que la homogeneidad en precios debe imponerse *después* de seleccionar una especificación u otra. Es decir, la normalización por la media geométrica debe hacerse, en su caso, *antes* de dividir los costes y los precios de los inputs por el precio del input numerario.

Es común que en un fenómeno económico se determinen simultáneamente un conjunto de variables. Por ejemplo, el coste y las demandas de inputs en el modelo de minimización de costes o el beneficio y las ofertas de outputs y demandas de inputs en el modelo de maximización del beneficio. Esta simultaneidad se traduce en la existencia de relaciones teóricas entre las distintas ecuaciones que describen estas variables. En concreto, la aplicación del lema de Shephard para obtener las demandas de inputs lleva a que los parámetros que aparecen en las demandas de inputs sean los mismos que en la función de costes. Puesto que las distintas ecuaciones comparten los mismos parámetros (o un subconjunto de ellos), con frecuencia se estiman conjuntamente aludiendo a argumentos econométricos de eficiencia de estimación. Este procedimiento se puede asimilar a tener más datos para estimar los mismos parámetros, lo que no puede dar lugar a “peores” estimaciones en el sentido de mayores varianzas de los estimadores. Las ganancias de eficiencia en la estimación son aún mayores si las perturbaciones aleatorias de las distintas ecuaciones estén correlacionadas. Este es precisamente el caso de un sistema de ecuaciones ya que, al proceder todas la variables dependientes del mismo fenómeno económico (proceso generador de datos), las perturbaciones aleatorias de cada una de las ecuaciones tienen el mismo origen y, por lo tanto, es de esperar que exista una correlación importante entre ellas.

A continuación, se describen algunos detalles prácticos de la estimación de sistemas de ecuaciones en el caso de una función de costes translog.[84] En este caso, dado que la variable dependiente está expresada en logaritmos, las demandas de factores tienen expresiones relativamente complicadas, por lo que es preferible aplicar el lema de Shephard en logaritmos. Es decir,

[84] Comentarios similares se podrían realizar para el caso de sistemas de ecuaciones formados por funciones de producción y las condiciones de primer orden para la maximización del beneficio, o por funciones de beneficios, de oferta de outputs y demanda de inputs.

$$\frac{\partial \ln C}{\partial \ln w_j} = \frac{\partial C}{\partial w_j}\frac{w_j}{C} = x_j \frac{w_j}{C} = S_j \tag{5. 16}$$

donde S_j representa la proporción de costes destinada a adquirir el input *j* o participación del input *j* en el coste total ("*cost share*"). Aplicando el lema de Shephard a la función de costes (5.13), se obtienen las siguientes ecuaciones de participación:

$$S_j = \theta_j + \sum_{k=1}^{J} \theta_{jk} \ln w_k + \sum_{l=1}^{M} \gamma_{jl} \ln y_l \tag{5. 17}$$

Puesto que las ecuaciones de participación se obtienen derivando la función de costes, es importante destacar que los coeficientes de la ecuación (5.17) son los mismos que los que aparecen en la función de costes (5.13), por lo que a la hora de estimar hay que imponer las correspondientes restricciones paramétricas de igualdad de los coeficientes.[85] A estas ecuaciones hay que imponerles además las restricciones derivadas de la homogeneidad de grado cero en precios, indicando que las participaciones en los costes no varían cuando el precio de todos los inputs varía en la misma proporción. Este segundo conjunto de restricciones se impone implícitamente si se aplica el lema de Shephard sobre la función de costes (5.15) y no sobre la función (5.13). Nótese en este caso que la función de costes depende (del logaritmo) de precios relativos ($w_k^*=w_k/w_J$) y no de los precios en niveles (w_k), al tiempo que los sumatorios excluyen el input seleccionado como numerario. Estas peculiaridades no modifican la discusión anterior si se tiene en cuenta que la función de costes (5.10) se puede escribir de forma muy simplificada como sigue:

$$\ln C = \ln C \left[\ln w^* , \ln y \right] + \ln w_n \tag{5. 18}$$

Aplicando la regla de la cadena, las ecuaciones de participación se pueden obtener como:

$$\frac{\partial \ln C}{\partial \ln w_j} = \frac{\partial \ln C}{\partial \ln w_j^*} \cdot \frac{\partial \ln w_j^*}{\partial \ln w_j} \quad , \quad j = 1,\ldots,J-1 \tag{5. 19}$$

---

85 La intuición es que la restricción expresa un parámetro como función de otro conjunto de parámetros. Esto permite sustituir en el proceso de estimación ese parámetro por el resto de parámetros en la estimación. En otras palabras, cada restricción permite estimar un parámetro menos. Por ejemplo, si se tiene que $a_1+a_2=1$, sólo es necesario estimar a_1 ya que cuando aparece a_2 en el modelo se puede escribir como 1- a_1.

$$\frac{\partial \ln C}{\partial \ln w_J} = \sum_{j=1}^{J-1} \frac{\partial \ln C}{\partial \ln w_j^*} \cdot \frac{\partial \ln w_j^*}{\partial \ln w_j} + 1 \qquad (5.\ 20)$$

De la primera expresión se deduce que las ecuaciones de participación de los primeros (n-1) inputs se pueden obtener directamente derivando la función (5.14) con respecto a los precios relativos (lnw*) ya que el segundo término es, por construcción, igual a la unidad. El segundo término en (5.19) es igual a –1, por lo que la participación del último input se puede obtener de forma residual como la unidad menos la participación global del resto de inputs.

Cuando se especifica un sistema de ecuaciones formado por una función de costes translog y las correspondientes ecuaciones de participación, se supone que las perturbaciones aleatorias de los distintos individuos en la muestra no están correlacionadas, aunque pueden haber correlación entre las perturbaciones aleatorias de las ecuaciones para un mismo individuo (correlación “contemporánea”). El sistema de ecuaciones contiene perturbaciones aditivas para todas las ecuaciones, que siguen una distribución normal conjunta, pudiendo escribirse como:

$$\ln(C / w_J) = \ln C \left[\ln w^*, \ln y\right] + u \qquad (5.\ 21)$$

$$S_j = S_j(\ln w^*, \ln y) + v_j \quad , \quad j = 1, \ldots, J-1 \qquad (5.\ 22)$$

$$S_J = 1 - \sum_{j=1}^{J-1} S_j(\ln w^*, \ln y) + v_J \qquad (5.\ 23)$$

Este sistema puede estimarse por mínimos cuadrados generalizados usando el método de ecuaciones aparentemente no relacionadas (“Seemingly Unrelated Regression”) de Zellner o por máxima verosimilitud.[86] Como pone de manifiesto la ecuación (5.18), los parámetros de una de las ecuaciones de participación se pueden obtener a partir de los parámetros del resto de ecuaciones de participación. Es decir, el sistema es singular. Para evitar este problema se deberá eliminar la ecuación de participación del input numerario (5.23). Es preciso señalar que la elección del input numerario, y la eliminación consiguiente de su ecuación de participación, no influye en

[86] El primer método consiste básicamente en estimar cada ecuación por MCO para obtener una estimación de la varianza de las distintas perturbaciones aleatorias así como de la correlación entre las perturbaciones de cada ecuación, esto es, de la matriz de varianzas y covarianzas. Si esta información se incorpora en la estimación por MCO y el método se repite hasta que los cambios en la matriz de varianzas y covarianzas son muy pequeños, los resultados son equivalentes a los de máxima verosimilitud.

los resultados si el sistema se estima por máxima verosimilitud (Barten, 1969). No ocurre lo mismo si se estima por mínimos cuadrados generalizados.

Como se puede ver, las ecuaciones de demanda de factores incluyen todos los parámetros de la translog menos los que afectan a las variables del output. Si no se está interesado en esos parámetros, el sistema anterior puede estimarse prescindiendo de la ecuación de costes (5.21), como sugieren Berndt y Wood (1975). En concreto, estos autores sugieren estimar sólo las ecuaciones de participación bajo el supuesto de RCE, ya que de esa manera toda la información económica relevante está en los “shares”.

5.3. Variables explicativas endógenas

El trabajo original de Cobb y Douglas fue criticado por Marschak y Andrews (1944), los cuales argumentaron que las cantidades de inputs no son variables exógenas, por lo que están correlacionadas con la perturbación aleatoria, dado que son decididas por el productor conjuntamente con el nivel de producción. Por tanto, no se puede estimar la función de producción por mínimos cuadrados ordinarios ya que los estimadores son sesgados e inconsistentes, produciéndose lo que se conoce como sesgo de simultaneidad.[87]

Un sencillo ejemplo puede proporcionar un poco de intuición sobre la naturaleza del problema. En condiciones de laboratorio (en un invernadero con temperatura y humedad controlada) añadir agua a una planta afecta positivamente a su crecimiento. Sin embargo, un agricultor observa la meteorología (perturbación aleatoria) y la usa para tomar sus decisiones. Si el analista no observa la meteorología puede darse un fenómeno bastante curioso. Si hay sequía el agricultor usa mucha agua de riego y, sin embargo, la producción va a ser baja. De algún modo, el agua sólo sirve para reducir las perdidas provocadas por la sequía. Si hay un periodo de lluvias, el agricultor usa poca agua de riego y, sin embargo, obtiene una buena producción. En este ejemplo, la relación entre agua de riego y producción agrícola

[87] Esta crítica fue continuada en siguientes trabajos por Hoch (1958), Mundlak (1963), Mundlak y Hoch (1965). El tema ha sido analizado más recientemente en los artículos de Mundlak (1996) y Griliches y Mairesse (1997).

puede ser negativa. Esto es debido, a que la variable explicativa se ve afectada por la misma perturbación aleatoria (la sequía) que la producción.

En estos casos, se debe modelizar tanto el nivel de producción como las cantidades de inputs a través de un modelo de ecuaciones simultáneas, por lo que Marschak y Andrews sugieren estimar un sistema. Para solventar el problema del sesgo de ecuaciones simultáneas, Klein (1953), utilizando las condiciones de primer orden para la maximización del beneficio, usó las proporciones de factores para estimar los coeficientes de la función de producción. Las condiciones de primer orden para la Cobb-Douglas pueden escribirse como:

$$\beta_j \frac{y}{x_j} = \frac{w_j}{p} \tag{5. 24}$$

Por tanto, se puede calcular una "estimación" de β_j como la proporción del valor del input *j* en el valor del producto total:[88]

$$\beta_j = \frac{w_j x_j}{py} \tag{5. 25}$$

Añadiendo perturbaciones aleatorias multiplicativas y tomando logaritmos neperianos, se obtiene:

$$\ln y = -\ln \beta_j + \ln \frac{w_j}{p} + \ln x_j + u_j \tag{5. 26}$$

que puede simplificarse de la siguiente forma:

$$\ln \frac{y}{x_j} = -\ln \beta_j + \ln \frac{w_j}{p_i} + u_j \tag{5. 27}$$

Existe por tanto, una ecuación para cada input. Estas ecuaciones contienen los mismos parámetros que la función de producción (los β_j) y pueden estimarse como un sistema, con o sin la ecuación de la función de producción. En este procedimiento, las variables endógenas (inputs y outputs) pasan a ser variables dependientes. Las variables explicativas son los precios de los factores que en competencia perfecta son

[88] Tyner y Tweeten (1965) sugirieron que debido a problemas de ajuste, las proporciones de factores observadas no tenían por qué coincidir con las óptimas. Por tanto, especifican un modelo de ajuste parcial para las "*factor shares*" y estiman econométricamente las proporciones en función de las proporciones del período anterior.

exógenos. De este modo, se corrige el problema de endogeneidad de las variables explicativas.

La clave del problema está en si la empresa conoce la perturbación aleatoria en el momento de tomar la decisión sobre las cantidades de inputs. Marschak y Andrews supusieron que sí, mientras que Mundlak y Hoch (1965) formulan el problema suponiendo que sólo una parte de la perturbación es conocida por el productor.

Zellner, Kmenta y Drèze (1966) sugieren que si la actividad productiva se desarrolla bajo condiciones de riesgo, es razonable suponer que el objetivo del productor va a ser maximizar los beneficios esperados. En ese caso la correlación deja de existir y los estimadores recobran sus propiedades originales. El argumento de estos autores se desarrolla en el contexto de la producción agraria y consiste en suponer que una buena parte de lo que contiene la perturbación aleatoria de la función de producción (alteraciones del clima, plagas,...) no puede ser anticipado por los agricultores, por lo que no puede estar correlacionado con los inputs. El argumento de Zellner, Kmenta y Drèze ha sido citado en innumerables trabajos empíricos para justificar la estimación de la función de producción sin las condiciones de primer orden. Sin embargo, aunque el argumento pueda tener su lógica en el sector agrario, no parece que pueda trasladarse a otros sectores, por lo que muchos autores han criticado esta “sobreutilización” del argumento.[89]

Este problema también se presenta en la estimación de funciones de costes cuando el output es endógeno. Hay que destacar que el output que aparece en la función de costes es predeterminado ya que la función de costes se obtiene suponiendo que las empresas minimizan el coste de producir un output planeado, que es no-estocástico.[90] Por este motivo, a veces se dice que la función de costes es una función “ex-ante”. Sin embargo, en la práctica, la variable independiente de la mayor parte de los estudios econométricos es el output realizado, y no el output planeado

[89] Griliches y Mairesse (1998) expresan esta visión con cierta crudeza: “A pesar de la eminencia de los autores y del lugar de publicación (Econometrica), casi nadie ha empleado este razonamiento para usar MCO sin sentirse culpable por ello”.

[90] En algunas industrias el output es exógeno. Por ejemplo, aquellos sectores en los que las empresas producen el output que se demanda en un momento determinado (electricidad, transporte, ...). Ver el pionero estudio de Nerlove (1963) sobre la producción de electricidad.

que indica la teoría. Como el output realizado es estocástico, puede suceder que el coste medio observado para dos empresas idénticas sea diferente debido a la aleatoriedad del proceso de producción. Este problema puede ser especialmente grave en las actividades agrícolas, donde la mayoría de los recursos se asignan al comienzo de la temporada, antes de que el nivel de output ex-post se conozca, y donde la producción tiende a ser más variable que en otros sectores.

Desde el punto de vista econométrico, esta situación se traduce en un problema de 'errores en las variables' que conduce a obtener estimaciones sesgadas de los parámetros de las funciones de coste. La solución a este problema fue propuesta por Walters (1960) y Martin (1983) y consiste en utilizar una variable instrumental para el output planeado, que a diferencia del observado dependería de los factores usados y carecería de componentes aleatorios.[91]

5.4. Estimación en presencia de heterogeneidad inobservable

Cuando se estima un modelo económico es frecuente encontrar situaciones en las que alguna variable explicativa importante es inobservable. Uno de los casos más frecuentes es el del input gestión en la estimación de funciones de producción.[92] También es posible que, aún incluyendo todos los inputs en la función de producción, existan diferencias entre individuos en la calidad de los inputs que no son observables.[93] En todos estos casos se dice que existe una heterogeneidad inobservable que no es posible incluir explícitamente en el modelo, bien porque no se pueden medir, porque su inclusión en el modelo lo complica en exceso o, simplemente, porque se desconoce su relación con

[91] Hubbard y Dawson (1987) utilizaron esta técnica con una muestra de explotaciones lecheras del Reino Unido para demostrar las diferencias entre las relaciones *ex-ante* y *ex-post*. Como variable instrumental utilizan el valor calculado del output tomado de una función de producción translog. En una segunda etapa, estiman una función de costes translog encontrando que existen grandes diferencias entre la curva de costes *ex-post* y la curva *ex-ante*. Las estimaciones correspondientes a la curva *ex-post* sobreestiman el coste medio mínimo y subestiman el nivel óptimo de producción.

[92] Otro input inobservable o de difícil medición es el capital humano de la mano de obra. Schultz (1960, 1961) fue el impulsor de la investigación en este importante concepto.

[93] Este caso es mucho más frecuente de lo que cabe pensar. El input tierra se suele medir por el número de hectáreas de las que dispone un agricultor, sin embargo, es de esperar que la calidad de la misma no sea igual. El input trabajo, se suele medir por número de empleados o por horas trabajadas, por lo que en ambos casos se supone que la productividad marginal de todos los individuos es igual. Por otra parte, los servicios del factor capital se suelen aproximar por la amortización de la maquinaria, sin considerar que las máquinas no son de la misma "generación" y, por tanto, no son igual de productivas.

la variable dependiente. En consecuencia, la información contenida en las variables omitidas o las diferencias no observables formarían parte del término de error del modelo. Esto se puede comprobar fácilmente usando el siguiente modelo:

$$y_i = \alpha + \beta x_i + \gamma z_i + \varepsilon_i \tag{5. 28}$$

donde z_i es una variable que recoge diferencias inobservables que afectan a la variable dependiente y_i, y ε_i es una perturbación aleatoria que cumple las propiedades habituales. Dado que z_i no es observable, lo que en la práctica se estima es:

$$y_i = \alpha + \beta x_i + u_i \tag{5. 29}$$

donde $u_i = \varepsilon_i + \gamma Z_i$. Si la variable Z_i no está correlacionada con x_i, se puede estimar el modelo (5.28) por MCO sin problemas. Sin embargo, si las variables omitidas están correlacionadas con las que han sido incluidas, entonces el término de error del modelo estará correlacionado con las variables incluidas, $E(x_i u_i) \neq 0$, haciendo que las estimaciones mínimo cuadráticas de los parámetros del modelo sean sesgadas e inconsistentes (Griliches, 1957).

Este problema no se puede evitar cuando se estima con datos de corte transversal. Sin embargo, si se dispone de datos de panel, es posible eliminar el efecto de la variable omitida (la heterogeneidad inobservable) bajo el supuesto de que la heterogeneidad es invariante en el tiempo, proporcionando así una solución al problema del sesgo que afecta a los estimadores.

Los datos de panel consisten en observaciones de un mismo individuo en distintos momentos del tiempo. Por tanto, en los modelos con datos de panel la variación entre observaciones se produce en dos dimensiones: la de los individuos y la temporal. De acuerdo con esto, el modelo (5.27) puede expresarse bajo el supuesto de que la heterogeneidad es constante a lo largo del tiempo como sigue:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \tag{5. 30}$$

donde el subíndice i=1,...,N indica individuos mientras que t=1,...,T indica tiempo. La heterogeneidad inobservable se recoge en un modelo con datos de panel en los $\alpha_i = \alpha + \gamma z_i$. Los α_i, denominados efectos individuales, son los coeficientes de todas aquellas variables omitidas del modelo que toman valores diferentes para cada individuo pero son invariantes en el tiempo. Hay N efectos individuales (tantos como individuos).

Existen dos estrategias básicas para estimar este modelo. La primera consiste en considerar los efectos individuales como parámetros a estimar en el modelo y da lugar al denominado estimador de efectos fijos. La segunda estrategia, consistente en considerar los efectos individuales como variables aleatorias que pasan a formar parte de la perturbación aleatoria del modelo, da lugar al denominado estimador de efectos aleatorios.

El estimador de efectos fijos se basa en el hecho de que, al ser los efectos individuales invariantes en el tiempo, cualquier diferencia que se realice en (5.27) haría desaparecer el efecto individual. Así, por ejemplo, tomando primeras diferencias se obtiene:

$$y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \qquad (5.\ 31)$$

La estimación de esta ecuación por MCO permite obtener estimaciones insesgadas y consistentes de los parámetros. Esto es debido a que, por un lado, el efecto individual que puede estar correlacionado con la variable explicativa ha sido eliminado al tomar diferencias y, por otra, la variable explicativa (x_{it}-x_{it-1}) no está relacionada con el término de error, dado el supuesto de partida E($\varepsilon_{it}\cdot x_{it}$)=0.

Una alternativa al uso de las primeras diferencias es restar a la ecuación de cada periodo la ecuación de las medias individuales: La ecuación de medias individuales puede escribirse como:

$$\bar{y}_i = \alpha_i + \beta\bar{x}_i + \bar{\varepsilon}_i$$

donde,

$$\bar{y}_i = \frac{1}{T}\sum_t y_{it}$$

$$\bar{x}_i = \frac{1}{T}\sum_t x_{it}$$

$$\bar{\varepsilon}_i = \frac{1}{T}\sum_t \varepsilon_{it}\ .$$

En este caso, el modelo resultante es:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

Por las mismas razones comentadas previamente, la estimación de esta ecuación por MCO permite obtener estimaciones insesgadas y consistentes de los parámetros. Se puede demostrar que el resultado de estimar este modelo por mínimos cuadrados ordinarios es equivalente a estimar el modelo inicial sin término independiente con una variable binaria identificando a cada individuo. En ese caso, las estimaciones de los parámetros de las variables binarias individuales son estimaciones insesgadas de los efectos individuales.

La ventaja de esta aproximación es que la estimación del parámetro β no está afectada por la correlación entre α_i (alternativamente, z_i) y las variables explicativas x_{it}. De hecho, en ambos casos, este término desaparece al tomar las diferencias. La desventaja es que con este método no se puede estimar el efecto de una variable invariante en el tiempo, ya que, esta variable desaparece al tomar las oportunas diferencias.

El modelo de coeficientes aleatorios considera los efectos individuales como parte del término de perturbación aleatoria del modelo. El efecto individual (α_i) se escribe como α_i=α+u_i. Como resultado, la ecuación (5.29) se escribe como:

$$y_{it} = \alpha + \beta x_{it} + v_{it}$$

donde v_{it}=u_i+ε_{it}.

Los parámetros en este modelo pueden ser estimados consistentemente por MCO si v_{it} no está correlacionada con las variables explicativas. Puesto que ε_{it} se ha supuesto inicialmente que no está correlacionado con las variables explicativas la condición para la estimación consistente es que el efecto individual (u_i) no esté correlacionado con las variables explicativas. La eficiencia de la estimación puede ser incrementada si se tiene en cuenta la peculiar estructura de la matriz de varianzas de v_{it} (no es una matriz escalar). Es decir, este modelo se estima generalmente por mínimos cuadrados generalizados.

La ventaja fundamental del estimador de coeficientes aleatorios es que permite la estimación del efecto de variables que no cambian en el tiempo. Por ejemplo, los años de estudios de un individuo adulto. La desventaja es que la estimación se hace bajo el supuesto de que los efectos individuales no están correlacionados con las variables explicativas. Este supuesto es un tanto restrictivo en muchas ocasiones. Por ejemplo, es difícil de mantener si los efectos individuales representan habilidad o capacidad de gestión de un individuo o empresa y las variables explicativas representan los factores de producción elegidos por ese individuo o empresa.

5.5. Funciones frontera

Las funciones de producción, de costes y de beneficios son funciones frontera. Este concepto se entiende mejor, si se recuerda primero la definición de cada una de estas funciones:

- Función de producción: proporciona el *máximo* output que se puede obtener para cada combinación de factores.
- Función de costes: proporciona el *mínimo* coste de producir cada nivel de output, dados unos precios de los factores.
- Función de beneficios: proporciona el *máximo* beneficio que se puede obtener dados los precios de inputs y outputs.

Como se muestra en la figura 5.2, las tres funciones proporcionan una cota superior (inferior) para la producción o beneficios (costes), esto es, “envuelven” los datos, no permitiendo que haya empresas por encima (debajo) de la función de producción y beneficios (costes). Por esta razón, se dice que son funciones “frontera”.

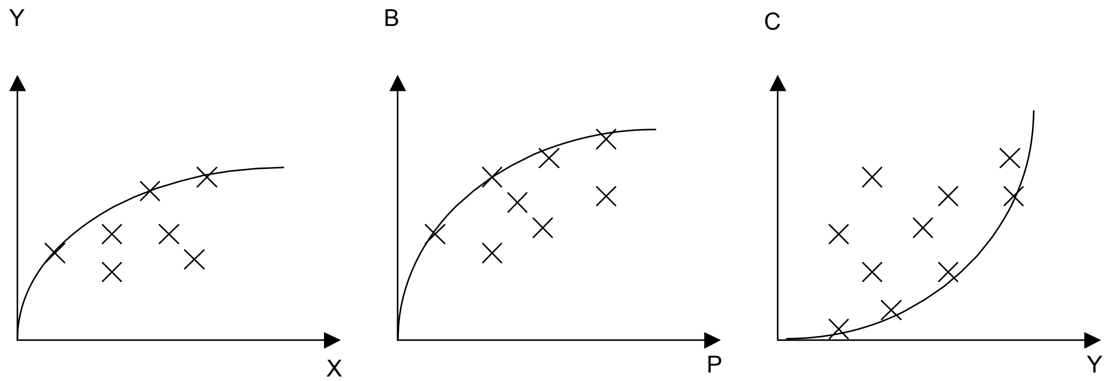

Figura 5.2. *Funciones frontera de producción, beneficios y costes*

De acuerdo con lo anterior, en un mundo determinístico sin factores exógenos que influyan en la producción, beneficios o costes, las desviaciones de las fronteras deberían ser únicamente de un signo: negativas cuando se estima una de producción o de beneficios, y positivas cuando es una función de costes. Sin embargo, cuando se estima una de estas funciones por mínimos cuadrados ordinarios se obtiene una función media, es decir, con residuos positivos y negativos. Esta práctica contradice la definición teórica de estas funciones, puesto que permite que haya empresas que estén por encima y por debajo de la frontera. La diferencia entre las funciones medias y frontera puede ser muy importante. En la Figura 5.3 se representa una nube de puntos, una función de costes *media* que "atraviesa" la nube de puntos por el centro y la función de costes *frontera* que "envuelve" los datos. Como se puede observar, las implicaciones de ambas para el cálculo de economías de tamaño son completamente diferentes. Mientras que la función de costes media parece sugerir la existencia de rendimientos constantes a escala, la función frontera indica que, efectivamente, existen economías de tamaño o rendimientos a escala decrecientes.

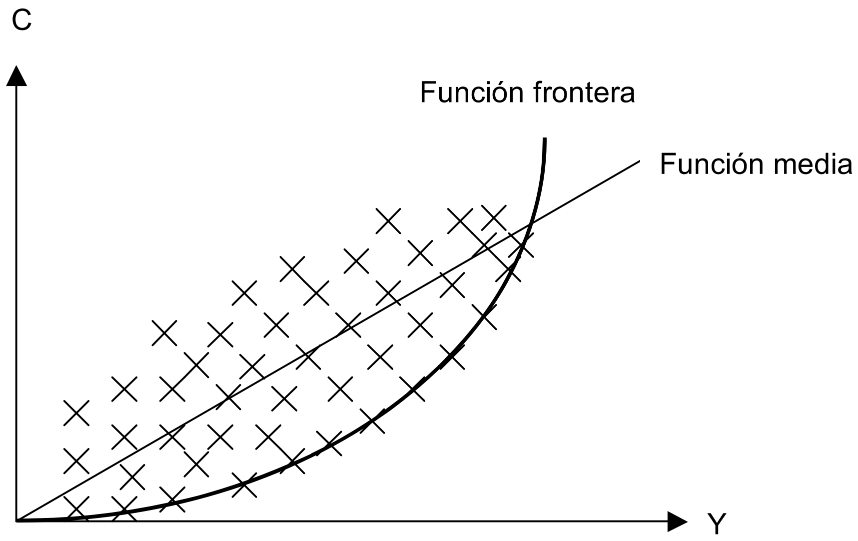

Figura 5.3. *Función media y función frontera*

La diferencia entre una frontera y una función media se encuentra, en la práctica, en que la primera concede "más importancia" a las desviaciones de un signo (positivo, en una función de costes), mientras que en una función media las desviaciones positivas y negativas tienen la "misma importancia".[94] Un caso extremo es la frontera de costes de la Figura 5.1 donde el peso de las desviaciones negativas es nulo. Esta se consigue imponiendo que la perturbación aleatoria siga una distribución de una cola. De esta manera, en el caso de las funciones de costes las

[94] De hecho, un caso extremo es la frontera de costes de la Figura 5.2 ya que el peso de las desviaciones negativas es incluso nulo.

empresas estarán en la frontera o por encima de ella, mientras que en el caso de las funciones de producción y beneficios, todas las empresas están en la frontera o por debajo de ella. Es decir, las funciones frontera se pueden especificar de la siguiente forma:

- Función de producción frontera: $y = f(x) - u, \qquad u \geq 0$
- Función de costes frontera: $C = C(y, w) + u, \qquad u \geq 0$
- Función de beneficios frontera: $B = B(p, w) - u, \qquad u \geq 0$

Como se puede ver la parte determinística de una función frontera es igual a las que se habían considerado hasta el momento y sólo difieren en la parte estocástica. Este es un buen ejemplo para entender que la especificación de un modelo requiere pensar no sólo sobre la parte determinística sino también sobre la estocástica.

Las anteriores ecuaciones pueden estimarse por máxima verosimilitud una vez que se ha hecho un supuesto sobre la distribución de u_i. Los más habituales son la semi-normal, normal truncada, exponencial y gamma. Un sencillo método para estimar una función frontera, sugerido por Greene (1980), consiste en estimar en primer lugar una función media por MCO y corregir el término independiente añadiéndole el máximo residuo positivo obtenido en la estimación si es una función de producción o de beneficios o el máximo residuo negativo en el caso de una función frontera de costes. De esta forma, todas las observaciones se encontrarán por debajo (encima) de la frontera de producción o de beneficios (costes), a excepción de la correspondiente al máximo residuo (negativo) que será considerada como la más eficiente.

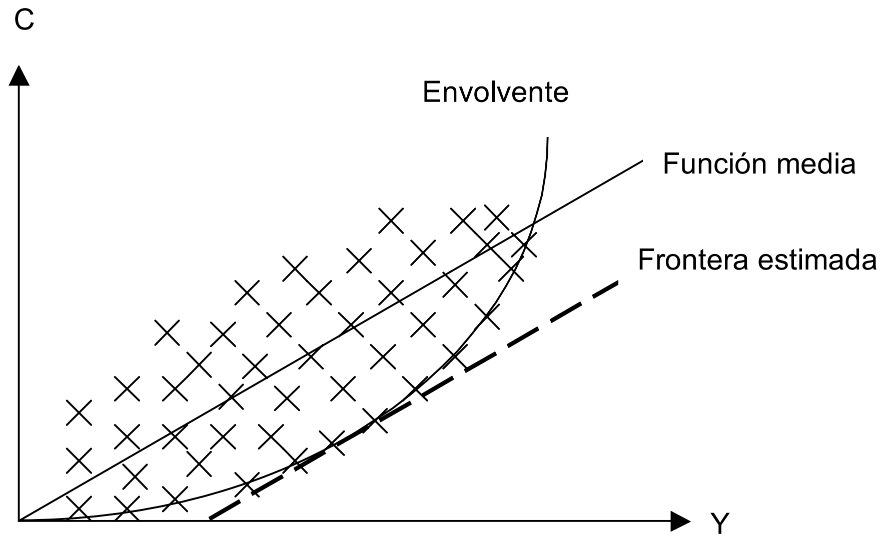

Figura 5.4. *Mínimos cuadrados corregidos*

El uso de funciones frontera ha tenido un gran auge en las dos últimas décadas debido principalmente a que en estos modelos, la perturbación aleatoria se interpreta como una medida de la ineficiencia de la empresa. Así, por ejemplo, la distancia de una empresa a la frontera de producción (costes o beneficios) estimada se considera como ineficiencia técnica (en costes o en beneficios).

Las fronteras analizadas hasta el momento reciben el nombre de fronteras determinísticas puesto que atribuyen toda la desviación de la frontera a la ineficiencia, ignorando el hecho fundamental de la naturaleza estocástica de la producción. Al suponer que la distancia a la frontera es totalmente atribuible a la ineficiencia de la empresa, no se tiene en cuenta que las empresas pueden verse afectadas por shocks exógenos (que no están bajo su control), los cuales no inciden de igual forma sobre todas las empresas. Aigner, Lovell y Schmidt (1977) introducen el concepto de frontera estocástica, cuya característica principal es que la perturbación aleatoria está dividida en dos componentes. En el caso de una frontera de producción estocástica, la ecuación es:

$$y = f(x) + v - u \qquad (5.\ 32)$$

donde *y* es la producción observada de una empresa, f(x) representa la producción máxima obtenible dado el vector de inputs *x*, *v* es un término de error simétrico que representa sucesos que no son controlables por la empresa (suerte, la climatología, etc.), mientras que el término de error *u* es no-negativo y se supone que se distribuye independientemente de v, siguiendo una distribución de una cola. Este término recoge la distancia de cada empresa a su frontera estocástica (que se define como la suma de la función de producción f(x) y la perturbación aleatoria *v*), representando, por tanto, una medida de su ineficiencia. Las implicaciones a nivel conceptual de que la frontera de producción sea estocástica son muy importantes para la interpretación de la ineficiencia. Como dicen Aigner, Lovell y Schmidt: "...el agricultor cuya cosecha es devastada por la sequía o una tormenta es desafortunado con nuestra medida, pero ineficiente con la medida habitual".

En el caso de las fronteras estocásticas, el problema para estimar por máxima verosimilitud es que, para formar la función de verosimilitud de la variable dependiente, hay que calcular previamente la función de densidad de ε=v-u. Aigner, Lovell y Schmidt (1977) la calculan para el caso en que *v* se distribuye según una normal con media cero y varianza σ_v^2 y *u* se distribuye según una seminormal, esto es, una normal N(0, σ_u^2) truncada positiva:

En el análisis empírico puede darse el caso de que σ_u^2 no sea significativamente distinta de cero. En otras palabras, que no se rechace la hipótesis nula de que $\sigma_u^2=0$. Esto puede interpretarse como evidencia de que todas las empresas son eficientes, siendo las desviaciones de la frontera (que coincide con la estimación por MCO) consecuencia de factores puramente aleatorios y ajenos a las empresas. Por tanto, Schmidt y Lin (1984) proponen como test de la existencia de frontera estocástica de eficiencia el contraste de asimetría de los residuos de la estimación por MCO.

5.6. Contraste de las propiedades teóricas

La teoría de la producción tiene fuertes implicaciones para el análisis empírico. Por ese motivo, en muchos trabajos se imponen y en otros se contrasta si las funciones estimadas cumplen las propiedades deseables. Esta fuerte conexión ha sido fuente de importantes debates en el seno de la profesión.

Así, por ejemplo, Appelbaum (1978) encontró en un estudio empírico que algunas propiedades (simetría, homogeneidad lineal) no se cumplían y que las conclusiones obtenidas del modelo primal y del dual eran distintas. Inspirados por este trabajo, Fox y Kivanda (1994) compararon artículos que estimaban funciones de coste o beneficio entre 1976 y 1991, comprobando si se contrastaban las hipótesis de homogeneidad, monotonicidad, curvatura y simetría. El resultado fue que el 45% de los artículos no contrastaban ninguna de esas propiedades y que en los que se contrastaban, se rechazaban en un porcentaje elevado. Shumway (1995) revisó esta misma base de datos ampliándola en dos años, confirmando también el alto número de veces que se rechazaban esas hipótesis, siendo la propiedad de monotonía la menos rechazada.

¿Cuál es la conclusión que se debe sacar de estos estudios? Dado el elevado número de estudios en los que no se cumplen las propiedades, uno puede pensar en principio que existe un importante divorcio entre la teoría de la producción y la realidad, por lo que se necesitaría modificar de alguna manera la teoría existente. Sin embargo, hay razones bastantes para justificar esa divergencia sin tener que recurrir a cambiar la teoría.

En primer lugar, Chambers (1989) indica que una desventaja del análisis paramétrico es que la hipótesis mantenida de la forma funcional no se contrasta por lo que el rechazo de una hipótesis nula puede indicar que la forma funcional no representa adecuadamente el proceso que se modeliza. Por ejemplo, en el caso de fertilización presentado en el primer capítulo una forma funcional lineal podría dar lugar al resultado de que el fertilizante no afecta a la producción de hierba. Asimismo, los datos pueden ser la causa de ese rechazo ya que cuando se contrasta una hipótesis se están contrastando también los datos. Es decir, los problemas de toma de datos pueden esconder las características fundamentales del proceso básico que se pretende analizar. Shumway (1995) indica que el rechazo de hipótesis como la curvatura, basado en el signo de los menores principales del hessiano no implica necesariamente un rechazo estadístico. Este contraste se basa en la inspección visual de los signos de unos determinantes. Las propiedades estadísticas de estos determinantes son, en principio, desconocidas y, por tanto, no es posible hacer afirmaciones como que se rechaza la hipótesis nula de que el signo de un determinante es significativamente distinto de cero. Love (1999) sostiene que los tests de homogeneidad están sesgados hacia el error de tipo I. En resumen, puede verse que el contraste de hipótesis es una tarea complicada que merece una mayor atención de la que normalmente se le dedica en los artículos.

Bibliografía

Aigner, D., C.A.K. Lovell y P. Schmidt, (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", *J. of Econometrics*, 6, 21-37.

Alvarez, A. (2001), "Concepto y medición de la eficiencia productiva", en A. Alvarez (coord.), *La medición de la eficiencia y la productividad*, Pirámide.

Anderson, R.G. y J.G. Thursby (1986), "Confidence Intervals for Elasticity Estimators in translog Models", *Review of Economics and Stat*istics, 68, 647-656.

Appelbaum, E., (1978), "Testing Neoclassical Production Theory", *J. of Econometrics*, 7, 87-102.

Arrow, K., H. Chenery, B. Minhas y R. Solow, (1961), "Capital-Labor Substitution and Economic Efficiency", *Review of Economics and Statistics*, 29, 155-73.

Barten, A. (1969), "Maximum Likelihood Estimation of a Complete System of Demand Equations", *European Economic Review*, 1, 7-73.

Battese, G.E., (1997), "A Note on the Estimation of Cobb-Douglas Production Functions When Some Explanatory Variables Have Zero Values", *J. of Agricultural Economics*, 48(2), 250-3.

Berndt, E. y L. Christensen, (1973). "The translog function and the substitution of equipment, structures and labor in US manufacturing", J. of Econometrics, 1, 81-114.

Berndt, E. y D. Wood (1975), "Technology, Prices, and the Derived Demand for Energy", *Review of Economics and Statistics*, 57, 376-384.

Chambers, R., (1989), "Recent Developments in Production Economics", *The Economic Record*, (Sept), 243-264.

Cobb, C.W. y P.H. Douglas, (1928) "A Theory of Production", *American Economic Review*, 18(1), 139-72.

Fox, G. y L. Kivanda, (1994), "Popper or Production?", *Canadian J. of Agricultural Economics*, 42, 1-13.

Greene, W. (1980), "Maximum Likelihood Estimation of Econometric Frontier Functions", *Journal of Econometrics*, 13, 27-56.

Griliches, Z., (1957), "Specification Bias in Estimates of Production Functions", *J. of Farm Economics*, 39, 8-20.

Griliches, Z. y J. Mairesse, (1998), "Production Function: The Search for Identification", en S. Strom (ed.) *Econometrics and Economic Theory in the 20th Centry: The Ragnar Frisch Centennial Symposium*, Cambridge University Press.

Hoch, I., (1958), "Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function", *Econometrica*, 26, 566-78.

Hubbard, L. y P. Dawson, (1987), "Ex-ante and ex-post long-run average cost functions", *Applied Economics*, 19, 1411-1419.

Hunt, L.C. y E.L. Lynk, (1993), "The Interpretation of Coefficients in Multipicative-logarithmic Functions", *Applied Economics*, 25, 735-738.

Jorgenson, D.W., (1986), "Econometric Methods for Modelling Producer Behavior", en *Handbook of Econometrics*, Vol. 3, Z. Griliches y M. Intriligator (eds.), Elsevier.

Just, R., D. Zilberman and E. Hochman, (1983), “Estimation of Multicrop Production Functions”, *American Journal of Agricultural Economics*, 65, 770-780.

Love, H.A., (1999), “Conflicts between Theory and Practice in Production Economics”, *American J. of Agricultural Economics*, 81(3), 696-702.

Lovell, C. A. K., Richardson, S., Travers, P., Wood, L.L., (1994), “Resources and Functionings : A New View of Inequality in Australia”, en W. Eichhorn (ed.), *Models and measurement of Welfare and inequality*, Springer-Verlag.

Klein, L., (1953), *A Textbook of Econometrics*, Row, Peterson and Co.

Kmenta, J., (1967), “On Estimation of the CES Production Function”, *International Economic Review*, 8, 180-189.

Marschak, J. y W.H. Andrews, (1944), ”Random Simultaneous Equations and the Theory of Production”, *Econometrica*, 12, 143-205.

Martin, W., (1983), "A Note on Cost Functions and the Regression Fallacy", *Rev. of Marketing and Agricultural Economics*, 51(3), 249-57.

Mundlak, Y., (1961), “Empirical Production Function Free of Management Bias”, *J. of Farm Economics,* 43, 44-56.

Mundlak, Y., (1996), “Production Function Estimation: Reviving the Primal”, *Econometrica*, 64(2), 431-438.

Mundlak, Y. e I. Hoch, (1965), “Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions”, *Econometrica*, 33, 814-828.

Nerlove, M., (1963), “Returns to Scale in Electricity Supply”, en C. Christ *et al*. (eds.), *Measurement in Economics: Essays in Memory of Yehuda Grunfeld*, Stanford University Press.

Nerlove, M., (1968), *Estimation and Identification of Cobb-Douglas Production Functions*, Chicago, Randy McNally.

Schmidt, P. y T.F. Lin, (1984), "Simple Test of Alternative Specification in Stochastic Frontiers Models", *J. of Econometrics*, 13, 57-66.

Schultz, T.W., (1960), “The Formation of Human Capital by Education”, *J. of Political Economy*, 68.

Schultz, T.W., (1961), “Investment in Human Capital”, *American Economic Review*, 51, 1-17.

Shumway, C.R., (1995), “Recent Duality Contributions in Production Economics, *J. of Agricultural and Resource Economics*, 20(1), 178-194.

Suits, D., (1983), “Dummy Variables: Mechanics vs. Interpretation”, *Review of Economics and Statistics*, 177-180.

Tintner, G., (1944), “A Note on the Derivation of Production Functions from Farm Records”, *Econometrica*, 12(1), 26-34.

Tyner, F.H. y L.G. Tweeten, (1965), “A Methodology for Estimating Production Parameters”, *American J. of Agricultural Economics*, 47, 1462-1467.

Walters, A.A., (1960), "Expectations and the Regression Fallacy in Estimating Cost Functions", *Rev. of Economics and Statistics*, 42(2), 210-15.

Zellner, A., J. Kmenta y J. Drèze, (1966), "Specification and Estimation of Cobb-Douglas Production Function Models", *Econometrica*, 34, 784-795.

White, H. (1980), “Using Least Squares to Approximate Unknown Regression Functions”, *International Economic Review*, 21(1), 149-170.