

1. Historia intelectual de la Eficiencia en Economía¹

En la primera mitad del siglo XX, la teoría microeconómica se aproximó al concepto de eficiencia desde la perspectiva de Pareto. El criterio de Pareto se cumple si ninguna persona puede quedar mejor sin que alguien empeore. El primer teorema del bienestar clásico sostiene que la eficiencia de Pareto se logra si y sólo si:

- Existen mercados para todos los productos posibles.
- Los mercados son perfectamente competitivos.
- Los costos de transacción son insignificantes.
- No hay externalidades.

El supuesto implícito era que las empresas siempre toman decisiones óptimas en el uso de insumos, y que cualquier ineficiencia en una economía se origina en la forma en que se asignan los recursos entre las empresas, más que dentro de las mismas. Las dos amenazas principales a la eficiencia en este paradigma eran los monopolios y las restricciones al comercio (internacional).²

En la segunda mitad del siglo XX, fue desafiado el supuesto de que las empresas siempre toman decisiones óptimas. Se aceptó que, además de la eficiencia "social" o "distributiva" original, la eficiencia interna de la empresa también era digna de ser analizada. Este había sido tradicionalmente un campo de la investigación operacional (IO), que se ocupaba del "análisis de actividades", donde el gerente pasa a ser el tema de interés; de ahí el término "eficiencia en la gestión."

Durante la década de los 1950s, varios estudiosos³ trataron de formalizar los dos tipos de eficiencia. Estos son referidos a veces como la escuela *neo-walrasiana*. Dentro de la escuela neo-walrasiana el artículo seminal sobre medición de la eficiencia es la definición de eficiencia productiva de Farrell. Farrell se inspiró en el trabajo de Koopmans, "Análisis de actividades",⁴ y su medida de eficiencia técnica es similar al "coeficiente de utilización de recursos" de Debreu. La novedad del enfoque de Farrell es que su medida de eficiencia permite explícitamente la inclusión de múltiples insumos y productos, mientras que los trabajos anteriores (por ejemplo, de números índice) se limitaban a menudo a insumos o productos individuales (por ejemplo, la productividad media del trabajo).

La definición de empresa eficiente de Farrell es "que logra producir el mayor producto posible a partir de un conjunto determinado de insumos." Farrell introduce la función de producción eficiente como un caso especial de la función de producción paretiana tradicional, definida como "el producto que una empresa perfectamente eficiente podría obtener de cualquier combinación dada de insumos."

¹ Sección incluida como Apéndice técnico de Southern California Evidence-based Practice Center—RAND Corporation, [Identifying, Categorizing, and Evaluating Health Care Efficiency Measures](#), Final Report, 2008.

² Leibenstein H. [Allocative efficiency vs "X-efficiency"](#), The American Economic Review 1966; 53(3): 392-415.

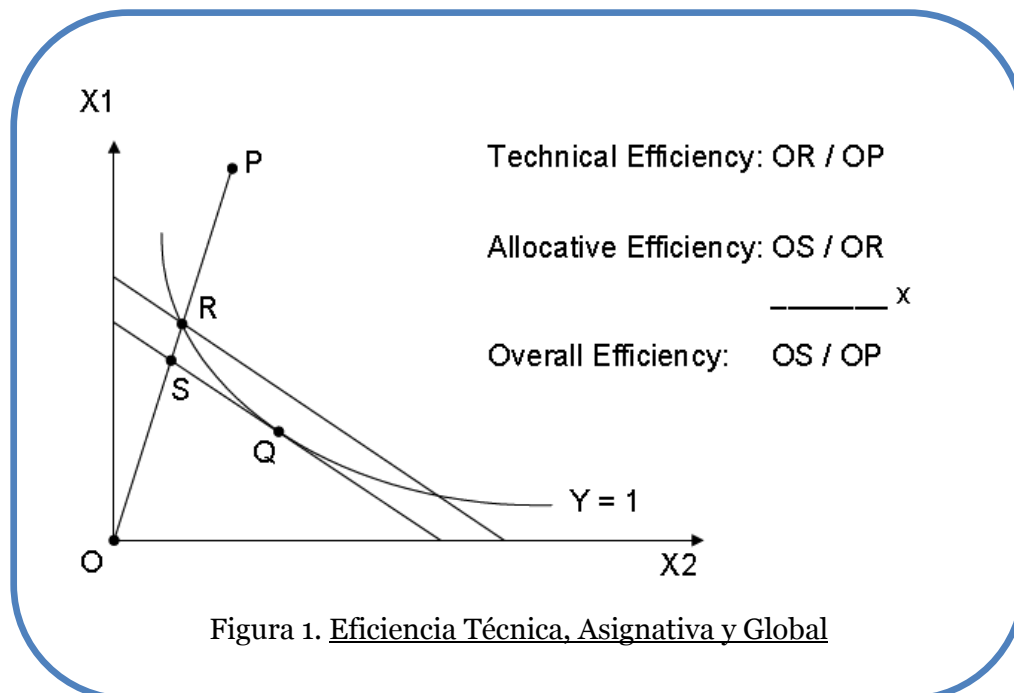
³ Debreu G. [The coefficient of resource utilization](#), Econometrica 1951; 19(3): 273-92; Farrell M.J. [The measurement of productive efficiency](#), Journal of the Royal Statistical Society Series A 1957; 120(3): 253-90.

⁴ Koopmans T. C. [Efficient allocation of resources](#), Econometrica 1951; 19(4): 455-65.

Farrell distingue entre eficiencia técnica, de precios, y global. La eficiencia técnica se define como el éxito de una empresa en obtener el máximo rendimiento de un determinado conjunto de insumos, es decir, producir en la "frontera técnica." La eficiencia de precios se define como el éxito de la empresa en elegir un conjunto óptimo de insumos, es decir, el conjunto que minimiza el costo si la empresa estuviera produciendo en la frontera técnica. La eficiencia global (comúnmente conocida como eficiencia productiva) es el producto de la eficiencia precio y la técnica. De haber ineficiencia técnica o precio, cada una implicará ineficiencia global (tal como la define Farrell).

Muchos economistas definen eficiencia técnica al igual que Farrell, pero definen eficiencia productiva como la minimización de costos, es decir, subsumiendo la eficiencia técnica. Bajo este enfoque la ineficiencia técnica implica ineficiencia productiva, lo que a su vez implica ineficiencia de Pareto.

La Figura 1 muestra el marco de análisis clásico de Farrell que hace que sea posible descomponer la eficiencia global en eficiencia técnica y de asignación (precio). Consideremos el caso de una producción simple (Y) que se obtiene mediante dos insumos (X_1 , X_2). Bajo el supuesto de que la función de producción $Y = f(X_1, X_2)$ es linealmente homogénea, la isocuenta unidad eficiente, $Y = 1$, muestra todas las combinaciones técnicamente eficientes. En la Figura 1, P representa una firma, país, individuo, etc., que también produce en $Y = 1$, pero utilizando mayores insumos y por lo tanto menos eficiente en sentido técnico. La magnitud de la eficiencia se puede expresar como la relación entre el uso de recursos óptimo y real (OR / OP). Tomando en cuenta la línea isocosto (que representa el precio relativo de los factores), podemos identificar la eficiencia asignativa. Cualquier punto de la línea $Y = 1$ tiene eficiencia técnica, pero sólo Q implica eficiencia técnica a un costo mínimo. La eficiencia asignativa (precio) se puede expresar como la relación entre el costo mínimo y el real (OS / OR), y la eficiencia global es el producto de la eficiencia técnica y de asignación.



Leibenstein (obra citada) hace una distinción similar, aunque menos formal que Farrell, y propone el término eficiencia-X, que es esencialmente lo mismo que la eficiencia técnica de Farrell. Aigner and Chu⁵ muestran que desde una perspectiva empírica la (in)eficiencia puede ser modelada ya sea mediante programación lineal o programación cuadrática, y que los supuestos originales de Farrell sobre rendimientos a escala de la función de producción de la industria no son necesarios.

A partir de los 1970s aparecen los primeros trabajos empíricos que estiman la eficiencia técnica en un marco de regresión o usando el Análisis de Envoltente de Datos (DEA).

La eficiencia, en particular la eficiencia técnica, está usualmente asociada a mediciones realizadas en un punto dado en el tiempo. Los cambios en el tiempo de la frontera técnica suelen estudiarse en el marco del análisis de la productividad, que en su forma moderna también tiene origen en los años 1950s.

2. Análisis de Envoltente de Datos (DEA)⁶

2.1 *Introducción*

Este libro se ocupa de evaluaciones de rendimiento y en especial se ocupa de la evaluación de las actividades de organizaciones tales como empresas comerciales, agencias gubernamentales, hospitales, instituciones educativas, etc. Estas evaluaciones toman diversas formas en los análisis habituales. Algunos ejemplos son el costo por unidad, la ganancia por unidad, la satisfacción por unidad, etc., que son medidas indicadas bajo la forma de una relación como la siguiente,

$$(1) \quad \text{Producto/Insumo.}$$

Esta es una medida de eficiencia comúnmente usada. La medida habitual de "productividad" también adopta una forma de cociente cuando se utiliza para evaluar el desempeño del empleado o trabajador. La "producción por hora trabajada" o la "producción por trabajador empleado" son ejemplos con las ventas, ganancias u otras medidas de producción que aparecen en el numerador. Estas medidas se refieren a veces como "medidas de productividad parcial." Esta terminología está destinada para distinguirlas de "medidas de productividad total de los factores," ya que estas últimas intentan obtener un valor de la relación producción-insumo que tenga en cuenta todos los productos e insumos. Pasar de las medidas de la productividad parcial a las de productividad total de los factores mediante la combinación de todos los productos y todos los insumos para obtener una única relación ayuda para evitar imputar ganancias a un factor (o un producto) que sean realmente atribuibles a algún otro insumo (o producto). Por ejemplo, un aumento en la producción como consecuencia de un aumento del capital o la mejora de la gestión podría ser atribuido por error a la mano de obra (cuando se utiliza una única proporción de producción a insumo) a pesar de que el rendimiento del trabajo se deterioró durante el período considerado. Sin embargo, un intento de pasar de una medida parcial a otra medida de productividad total de los factores se encuentra con dificultades tales como la elección de los insumos y productos que deben ser considerados y los ponderadores a ser utilizados con el fin de obtener una relación de una sola producción a un solo insumo que se reduce a una forma similar a la expresión (1).

⁵ Aigner DJ and Chu S.F. *On estimating the Industry Production Function*, The American Economic Review 1968; 58(4): 826-39.

⁶ William W. Cooper, Lawrence M. Seiford, and Kaoru Tone, [Data Envelopment Analysis – 2d Edition](#), 2007, Capítulo 1, *General Discussion*.

También se incurre en otros problemas y limitaciones con los intentos tradicionales para evaluar la productividad o eficiencia cuando hay múltiples productos e insumos a ser tomados en cuenta. Algunos de los problemas que deben abordarse serán descritos a medida que se avance con el Análisis de Envoltente de Datos (DEA), el tema de este libro. El enfoque relativamente nuevo encarnado en DEA no requiere que el usuario fije ponderaciones de cada insumo y producto, como en el enfoque usual de los números índice, y tampoco requiere fijar a priori las formas funcionales necesarias en los enfoques de regresión estadística de estos temas.

DEA utiliza técnicas como la programación matemática que puede manejar un gran número de variables y relaciones (restricciones) y esto relaja los requisitos que se encuentran a menudo cuando uno se limita a la elección de sólo unas pocas variables debido a las técnicas empleadas porque de lo contrario encontrará dificultades. Relajar las condiciones sobre el número de candidatos a ser utilizados en el cálculo de las medidas de evaluación deseadas hace que sea más fácil tratar problemas complejos y hacer frente a otras consideraciones que pueden plantearse en muchos contextos de políticas empresariales y sociales. Por otra parte, el extenso cuerpo de teoría y metodología de la programación matemática puede ser usado para guiar los análisis e interpretaciones. También puede ser utilizada en la ejecución de los cálculos porque gran parte de lo que se necesita ya se ha desarrollado y adaptado para su uso en muchas aplicaciones previas de la DEA. Mucho de esto se encuentra disponible en la literatura sobre investigación en la DEA y una gran parte de esto se ha incorporado en códigos informáticos disponibles en el mercado que se han desarrollado para su uso con la DEA.⁷

DEA ofrece una serie de oportunidades adicionales para su uso. Esto incluye oportunidades para la colaboración entre los analistas y responsables, que se extienden desde la colaboración en las opciones de los insumos y productos que se utilizarán e incluye elegir los tipos de preguntas "qué pasaría si" que se abordarán. Este tipo de colaboraciones se extienden al "benchmarking" de comportamientos de los competidores "qué pasaría si" y son la identificación de potenciales (nuevos) competidores que puedan surgir para su consideración en algunos de los escenarios que se puedan generar.

2.2 Un único insumo, un único producto

Para comenzar nuestro estudio de la DEA y sus usos, volvemos al caso de un único producto y un único insumo y aplicamos la fórmula (1.1) con el siguiente ejemplo sencillo. Supongamos que hay 8 sucursales que identificamos desde la A a la H en la cabecera de cada columna en la tabla 2.1.

Tabla 2.1 Un insumo, un producto

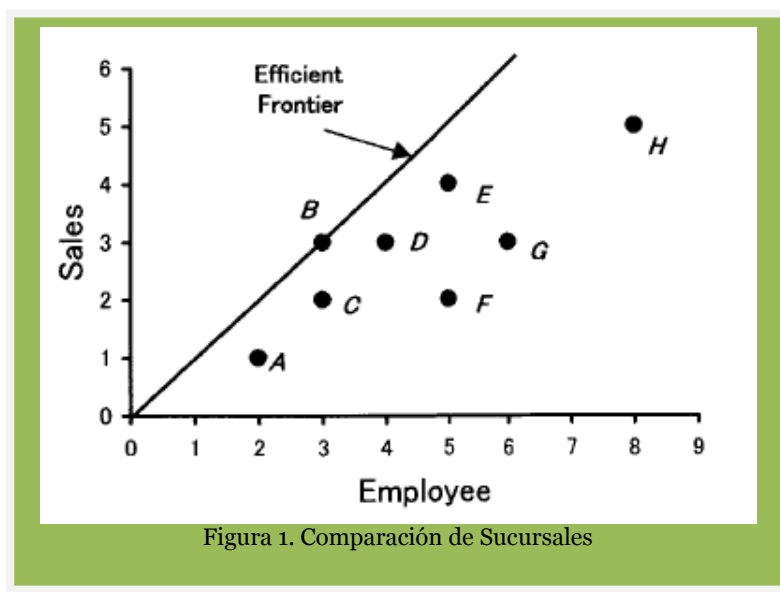
Sucursal	A	B	C	D	E	F	G	H
Empleados	2	3	3	4	5	5	6	8
Ventas	1	3	2	3	4	2	3	5
Ventas/Empl.	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

El número de empleados y de ventas (medidas en cientos de miles de dólares) son como se registra en cada columna. La línea inferior de la Tabla 1.1 muestra las ventas por empleado - una

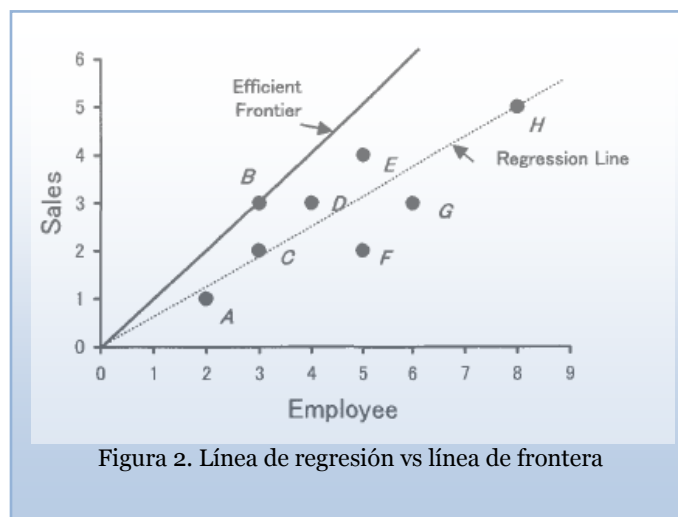
⁷ [Nota: El libro incluye un CD con software de apoyo e instrucciones al usuario.]

medida de la "productividad" a menudo usada en la gestión y análisis de las inversiones. Como se ha indicado, esto también puede ser tratado en un contexto más general de "eficiencia". Entonces, con esta medida, podemos identificar a B como la sucursal más eficiente y a F como la menos eficiente.

Vamos a representar estos datos como en la figura 1 poniendo al "número de empleados" y las "ventas" en los ejes horizontal y vertical, respectivamente. La pendiente de la línea que une cada punto con el origen corresponde a las ventas por empleado, y la pendiente más elevada es alcanzada por la línea desde el origen a B. Esta línea se llama la "frontera eficiente." Obsérvese que esta frontera toca al menos un punto y todos los puntos están, por lo tanto, sobre o por debajo de esta línea. El nombre de Análisis de Envoltente de Datos, tal como se utiliza en la DEA, proviene de esta propiedad, porque en el lenguaje matemático, se dice que una frontera tal "envuelve" a estos puntos.



Teniendo en cuenta estos datos, uno podría estar tentado a trazar una línea de regresión estadística ajustada a ellos. La línea de regresión en la Figura 2 muestra la recta de regresión que pasa por el origen que, en virtud del principio de mínimos cuadrados, se expresa por $y = 0.622x$. Esta línea, tal como se determina normalmente en estadística, pasa a través de la "Media" de estos puntos de datos y así podríamos definir los puntos superiores como *excelentes* y los puntos por debajo de ella como *inferiores* o *insatisfactorios*.



Se puede medir el grado de excelencia o inferioridad de estos datos puntuales por la magnitud del desvío de la línea ajustada de este modo. Por otro lado, la línea de frontera muestra el rendimiento de la mejor sucursal (B) y mide la eficiencia de otras tiendas por desviaciones respecto de ella. Por tanto, existe una diferencia fundamental entre los métodos estadísticos a través del análisis de regresión y DEA. El primero refleja un "promedio" o el comportamiento en torno a la "tendencia central" de las observaciones, mientras que el último trata con el mejor rendimiento y evalúa todas las actuaciones como desviaciones de la línea de frontera. Estos dos puntos de vista pueden dar lugar a grandes diferencias cuando se utilizan como métodos de evaluación. También pueden dar lugar a diferentes enfoques para la mejora. DEA identifica un punto como B para un futuro análisis o para servir como un "punto de referencia" a utilizar en la búsqueda de mejoras. El enfoque estadístico, por otra parte, hace un promedio de B y las otras observaciones, incluyendo F, como base para sugerir qué mejoras podrían ser buscadas.

Volviendo al ejemplo anterior, en realidad no es razonable creer que la línea de frontera se extienda hasta el infinito con la misma pendiente. Vamos a analizar este problema más adelante mediante el uso de diferentes modelos DEA. Sin embargo, suponemos que esta línea es efectiva en el rango de interés y llamamos a esto el supuesto de rendimientos constantes a escala.

En comparación con la mejor sucursal B, las demás son ineficientes. Podemos medir la eficiencia de otras sucursales respecto a B por

(2) $0 \leq (\text{Ventas por empleado de las otras sucursales})/(\text{Ventas por empleado de B}) \leq 1$
 y disponerlas en el orden siguiente con referencia a los resultados presentados en la Tabla 2:

$$1 = B > E > D > C > H > A = G > F = 0.4.$$

Por lo tanto, la peor sucursal, F, alcanza al 40% de la eficiencia de B.

Tabla 2.2 Eficiencia

Sucursal	A	B	C	D	E	F	G	H
Eficiencia	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

Ahora observamos el problema de cómo lograr que las sucursales ineficientes sean eficientes, es decir, la forma de moverlas hasta la frontera eficiente. Por ejemplo, la tienda A en la Figura 3 se puede mejorar de varias maneras. Una se logra mediante la reducción del insumo (número de empleados) a A₁ con coordenadas (1, 1) en la frontera eficiente. Otra es aumentando el producto (ventas en \$ 100.000 unidades) hasta A₂ (2,2). Cualquier punto del segmento de línea A₁A₂ ofrece una oportunidad para llevar a cabo las mejoras de una manera que supone que el insumo no se debe aumentar y el producto no debe disminuir al hacer eficiente la sucursal.

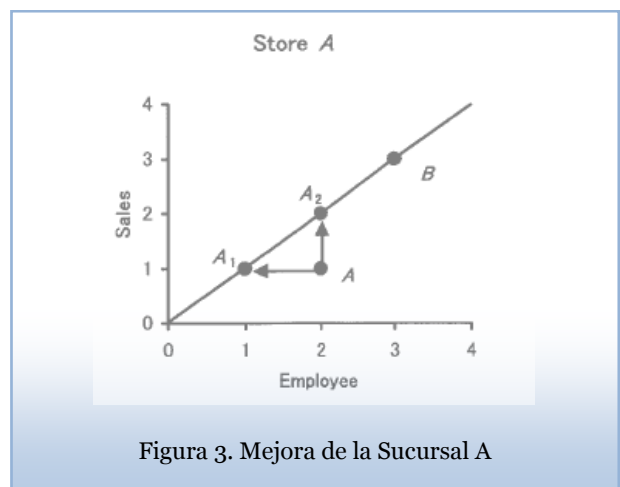


Figura 3. Mejora de la Sucursal A

Este sencillo ejemplo pasa de la relación en la tabla 2.1 a la "relación de relaciones" en la Tabla 2.2, lo que pone de manifiesto un punto importante. Los valores en (2.1) dependen de las unidades de medida utilizadas mientras que esto no es el caso de (2.2). Por ejemplo, si las ventas se expresaran en unidades de \$ 10,000, la relación de F cambiaría de $2/5 = 0,4$ a $20/5 = 4,0$. Sin embargo, el valor de (2.2) se mantendría sin cambios en $4/10 = 0,4$ y la puntuación de eficiencia relativa asociada a F no está afectada por esta elección de una unidad de medida diferente. Esta propiedad, a veces denominada "invariancia a las unidades" durante mucho tiempo ha sido reconocida como importante en ingeniería y ciencia. Ahí está, por ejemplo, el siguiente ejemplo del campo de la ingeniería de combustión donde las clasificaciones de eficiencia del horno se obtienen de la siguiente fórmula

$$(3) \quad 0 \leq E_r = y_r / y_R \leq 1, \text{ donde}$$

y_r = calor obtenido a partir de una unidad dada de combustible por el horno que se esté evaluando,

y_R = calor máximo que se puede obtener a partir de este mismo insumo de combustible.

El último, es el máximo de calor, que se puede calcular a partir de los principios termodinámicos por medio de análisis químico-físicos adecuados. El punto a destacar aquí, sin embargo, es que x , la cantidad de combustible utilizado, debe ser la misma para que, matemáticamente,

$$(4) \quad 0 \leq (y_r/x) / (y_R/x) = y_r / y_R \leq 1.$$

Por lo tanto, (3) se obtiene a partir de una relación de relaciones que es "invariante a las unidades".

Volviendo a las proporciones en la Tabla 2.2, podríamos observar que estos valores también están limitados por cero y la unidad. Sin embargo, las variaciones registradas en la Tabla 2.2 pueden ser resultado de una cantidad en exceso de producto o una deficiencia en el insumo. Por otra parte, esta situación es general en los negocios y en las aplicaciones socio-políticas (economía) que son de interés en este libro. Esta es una de las razones por las que podemos hacer poco uso de fórmulas como (4). Además, esta fórmula se limita al caso de un solo insumo y producto. Los intentos de extenderlo a múltiples insumos y múltiples productos se topan con los problemas que fueron identificados en nuestra discusión previa de medidas "parciales" y "productividad total de los factores".

2.3 El caso de dos insumos y un producto

Para pasar al caso de múltiples insumos y productos y su tratamiento, consideramos la Tabla 2.3 que enumera el rendimiento de 9 supermercados cada uno con dos insumos y un producto. El insumo x_1 es el número de empleados (unidad: 10), el insumo x_2 el área del piso (unidad: 1000m²) y el producto y las ventas (unidad: 100.000 dólares). Sin embargo, obsérvese que las ventas están normalizadas en 1 bajo el supuesto de rendimientos constantes a escala. Por lo tanto, los valores de los insumos se normalizaron para obtener 1 unidad de ventas. Graficamos los establecimientos, tomando al insumo x_1 / producto y y el insumo x_2 / producto y como ejes que podemos imaginar como "ejes normalizados" en la Figura 4.

Tabla 2.3 Caso de 2 insumos y 1 producto

Negocio	A	B	C	D	E	F	G	H	I
Empleados x_1	4	7	8	4	2	5	6	5.5	6
Área piso x_2	3	3	1	2	4	2	4	2.5	2.5
Ventas y	1	1	1	1	1	1	1	1	1

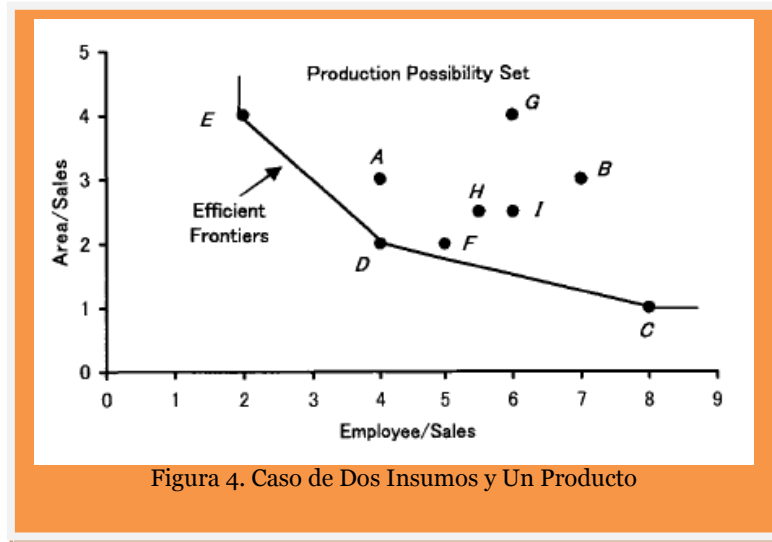


Figura 4. Caso de Dos Insumos y Un Producto

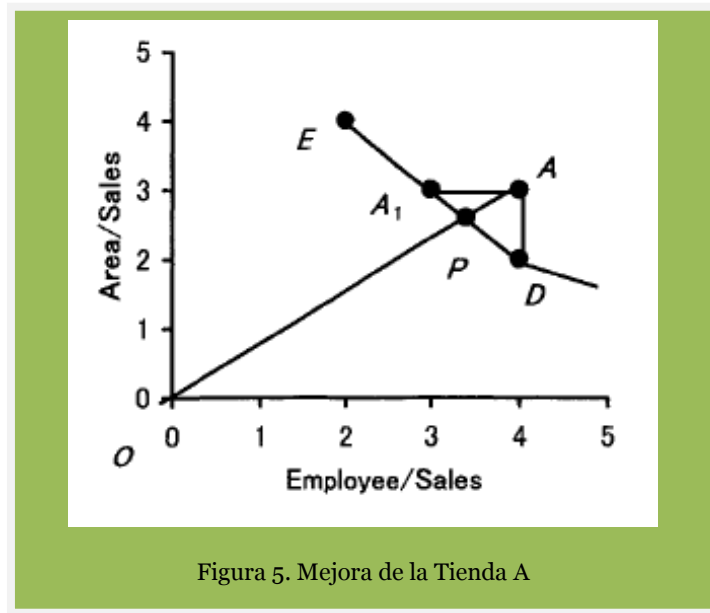
Desde el punto de vista de la eficiencia, es natural juzgar como más eficientes a los establecimientos que utilizan menos insumos para obtener un producto unitario. Por lo tanto, identificamos la línea de conexión C, D y E como la frontera eficiente. No hablamos de los *tradeoffs* entre estos tres establecimientos sino simplemente notamos aquí que ningún punto de esta línea de frontera puede mejorar uno de sus valores de un insumo sin empeorar el otro. Podemos *envolver* a todos los puntos de datos dentro de la región encerrada por la línea de frontera, la línea horizontal que pasa por C y la línea vertical a través de E. Llamamos a esta región el *conjunto de posibilidades de producción*. (Más precisamente, se debería llamar el supuesto de que el conjunto de posibilidades de producción es *lineal a trozos*, ya que no se garantiza que la (verdadera) frontera de esta región sea lineal por tramos, es decir, formada por segmentos lineales como el segmento de conexión ED y el segmento de conexión DC.) Esto significa que se supone que los puntos observados proporcionan evidencia (empírica) de que la producción es posible a las tasas especificadas por las coordenadas de cualquier punto en esta región.

La eficiencia de las tiendas que no están en la línea de frontera se puede medir haciendo referencia al punto de frontera de la siguiente manera. Por ejemplo, A es ineficiente. Para medir su ineficiencia dibujamos la línea OA, que va de cero a A, y cruza la línea de frontera en P (véase la Figura 5). Entonces, la eficiencia de A puede ser evaluada como

$$OP / OA = 0.8571.$$

Lo cual significa que la ineficiencia de A debe evaluarse por una combinación de D y E, debido a que el punto P está situado en la línea que conecta estos dos puntos. Los puntos D y E son llamados el *conjunto de referencia* de A. El conjunto de referencia para una tienda ineficiente puede variar de tienda en tienda. Por ejemplo, B tiene el conjunto de referencia compuesto por C y D en la figura 4. También podemos ver que muchas tiendas se juntan alrededor de D y por lo tanto se puede decir que D es una tienda eficiente que también es "representativa", mientras que C y E también son eficientes, pero también poseen características únicas en su asociación con segmentos de la frontera que están muy alejados de cualquier observación.

Ahora extendemos el análisis en la Figura 3 para identificar las mejoras refiriendo las situaciones ineficientes a la frontera eficiente en este caso de dos insumos (y un producto). Por ejemplo, A puede ser mejorado de manera efectiva por el movimiento hacia P con *Insumo x1 = 3.4* e *Insumo x2 = 2.6*, porque estas son las coordenadas de P, el punto de la frontera eficiente que antes hemos identificado con el segmento de línea OA en la figura 5. Sin embargo, también se puede utilizar cualquier punto del segmento DA₁ como un candidato para la mejora. D se alcanza mediante la reducción del *Insumo x2* (superficie), mientras que A₁ se consigue reduciendo el *Insumo x1* (empleados). Además, otra posibilidad de mejora se consigue mediante el aumento de la producción y el mantenimiento de la situación actual de insumos. Esto se discutirá más adelante.



2.4 El caso de dos productos y un insumo

La Tabla 2.4 muestra el número de clientes (unidad = 10) por vendedor y las ventas (unidad = 100.000 dólares) por cada vendedor de 7 sucursales. Para obtener una frontera conjunta en este caso, se divide por el número de empleados (= vendedores), que es considerado como el único insumo de interés. La frontera eficiente consiste entonces en las líneas que conectan B, E, F y G como se muestra en la Figura 6.

Tabla 2.4 Dos Productos y Un Insumo

Negocio	A	B	C	D	E	F	G
Empleados x	1	1	1	1	1	1	1
Clientes y_1	1	2	3	4	4	5	6
Ventas y_2	5	7	4	3	6	5	2

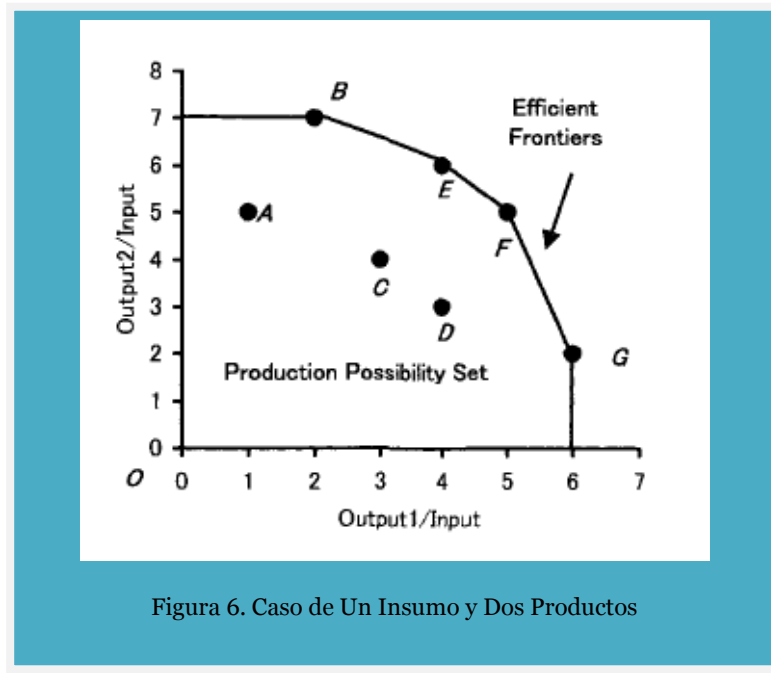


Figura 6. Caso de Un Insumo y Dos Productos

El conjunto de posibilidades de producción es la región limitada por los ejes y la línea de frontera. Las sucursales A, C y D son ineficientes y su eficiencia puede ser evaluada mediante referencia a las líneas de frontera. Por ejemplo, en la figura 7, la eficiencia de D se evalúa como

$$(5) \quad d(OD) / d(OP) = 0.75,$$

donde $d(OD)$ y $d(OP)$ significan respectivamente “distancia desde cero a D” y “distancia desde cero a P”.

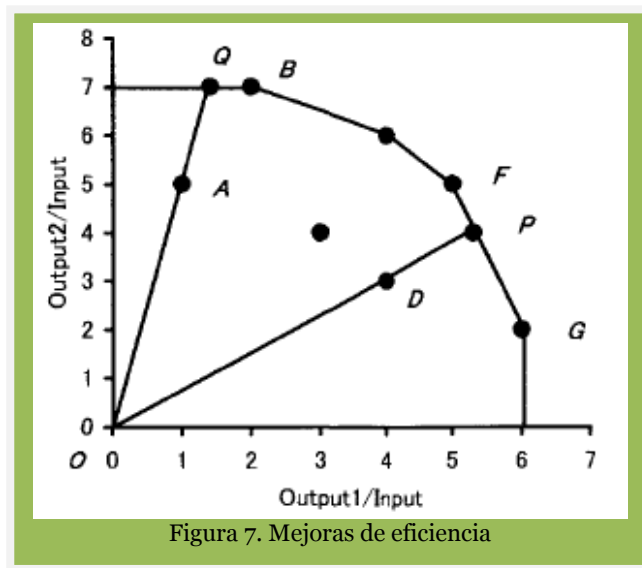


Figura 7. Mejoras de eficiencia

La relación anterior se conoce como "medida radial" y se puede interpretar como el cociente de dos medidas de distancia. La elección de las medidas de distancia no es única por lo que, debido a su familiaridad,⁸ seleccionamos las medidas euclidianas dadas por

$$d(OD) = \sqrt{3^2 + 4^2} = 5$$

$$d(OP) = \sqrt{[(16/3)^2 + 4^2]} = 20/3,$$

donde los términos bajo el signo radical son los cuadrados de las coordenadas de D y P, respectivamente, como se obtiene en la Tabla 2.4 para D y por la intersección de $y_2 = 3/4y_1$ e $y_2 = 20 - 3y_1$ para P. Como se dijo, luego de sustituir en (5) da

$$5 / (20/3) = 15/20 = 0.75.$$

Esta interpretación como una relación de distancias alinea los resultados con nuestra discusión anterior de tales proporciones. Debido a que la relación se forma en relación con la distancia euclidiana desde el origen sobre el conjunto de posibilidades de producción, siempre vamos a obtener una medida entre cero y la unidad.

También podemos interpretar los resultados para usos de gestión (u otros) de manera relativamente sencilla. La relación (5) siempre oscilará entre cero y la unidad. Como estamos interesados en producción, sin embargo, es más fácil de interpretar (5) en términos de su recíproca

$$d(OP)/d(OD) = (20/3) / 5 = 1.33.$$

Este resultado significa que, para ser eficiente, D tendría que aumentar sus insumos multiplicando por 4/3. Para confirmar que esto es así, basta con aplicar esta relación a las coordenadas de D y obtenemos

$$4/3 (4, 3) = (16/3, 4),$$

que así coincidiría con las coordenadas de P, el punto que está en la frontera eficiente usado para evaluar D.

Volviendo a (5) observamos que 0,75 hace referencia a la proporción de producto que P muestra que era posible de alcanzar. Es importante tener en cuenta que esto se refiere a la proporción de ineficiencia presente en los dos insumos de D. De este modo, el déficit de producto de D puede repararse mediante el aumento de los dos insumos sin cambiar sus proporciones - hasta que se alcanza P.

Como era de esperar, esta es sólo una de los distintos tipos de ineficiencia que serán de interés en este libro. Este tipo de ineficiencia que puede ser eliminada sin cambiar las proporciones se conoce como "ineficiencia técnica."

Otro tipo de ineficiencia ocurre cuando sólo algunos (pero no todos) los productos (o insumos) se identifican como exhibiendo un comportamiento ineficiente. Este tipo de ineficiencia se co-

⁸ Nuestras medidas de distancia y su uso se discuten en W.W. Cooper, L.M. Seiford, K. Tone and J. Zhu [*Some Models and Measures for Evaluating Performances with DEA: Past Accomplishments and Future Prospects*](#), Journal of Productivity Analysis, 2007.

noce como "ineficiencia de la mezcla", ya que su eliminación va a alterar las proporciones en las que se fabrican los productos (o se utilizan los insumos).⁹

Ilustramos el caso de "ineficiencia técnica" mediante el uso de D y P en la figura 7. Podemos utilizar Q y B para ilustrar la "ineficiencia de mezcla" o podemos usar A, B y Q para ilustrar tanto ineficiencia técnica como de mezcla. Luego, usando el último caso identificamos el componente de eficiencia técnica en el rendimiento de A por medio de la siguiente medida radial,

$$(6) \quad d(OA)/d(OQ) = 0.714.$$

Usando la recíproca de esta medida como sigue, y aplicándola a las coordenadas de A en (5):

$$1/(0.714) (1, 5) = (1.4, 7),$$

que son las coordenadas de Q.

Ahora podemos notar que los productos así ajustados, están en la relación $1.4 / 7 = 1/5$, que es la misma que la relación de A en la Tabla 2.4 - a saber, $y_1/y_2 = 1/5$. Esto aumenta tanto los productos de A sin empeorar su insumo y sin alterar las proporciones de producción. Esta mejora en la eficiencia técnica por el movimiento a Q no elimina todas las ineficiencias. A pesar de que Q está en la frontera no está en una parte eficiente de la frontera. Comparando Q con B se muestra un déficit en el producto 1 (número de clientes atendidos) por lo que puede alcanzarse un aumento adicional de este producto con un movimiento lateral de Q a B. Así, esta mejora también puede lograrse sin empeorar el otro producto o la cantidad del insumo. Pero la corrección del valor del producto, y_1 , sin alterar y_2 cambiará sus proporciones, por lo que podemos identificar dos fuentes de ineficiencias en la situación A: primero una ineficiencia técnica a través de la medida radial dada en (1.6), seguida de una ineficiencia de mezcla representada por el déficit de producto que permanece en y_1 después de que todas las ineficiencias técnicas se han eliminado.

Ahora introduciremos el término "ineficiencia puramente técnica", por lo que, en aras de la simplicidad, podemos utilizar el término "ineficiencia técnica" para referirnos a todas las fuentes de desperdicios - puramente técnicos y mezcla - eliminables sin empeorar cualquier otro insumo o producto. Esto también tiene la ventaja de conformarse a los usos que se aplican en la literatura. También simplificará las cosas cuando llegemos al análisis de precios, costos y otros tipos de ponderaciones que pueden ser asignadas a las diferentes fuentes de ineficiencia.

Comentario: El término "eficiencia técnica" está tomado de la literatura económica, donde se utiliza para distinguir los aspectos "tecnológicos" de la producción de otros aspectos, a los que se hace referencia generalmente como vinculados a la "eficiencia económica", que son de interés de los economistas.¹⁰ La última implica recurrir a información sobre precios, costos u otras consideraciones de valor que vamos a cubrir más adelante en este texto. Aquí, y en los próximos dos capítulos, nos centraremos en ineficiencias puramente técnicas y de mezcla que representan un "desperdicio" que se puede eliminar justificadamente sin requerir datos adicionales como precios y costos. Sólo se requiere asegurar que las mejoras resultantes sean valiosas, incluso aunque no les asignemos específicamente un valor.

⁹ A esta última se la conoce como ineficiencia de mezcla de los insumos y a la primera como ineficiencia de mezcla de los productos.

¹⁰ Ver por ejemplo H.R. Varian *Microeconomic Analysis* 2nd ed. (New York. W.W. Norton & Co., 1984.)

Tal como se utiliza aquí, el término ineficiencia de mezcla se toma de las literaturas de contabilidad donde también se le da otros nombres, tales como "variación física" o "variación en la eficiencia."¹¹ En este uso, la referencia es a los aspectos físicos de la producción que exceden una norma prescrita y por lo tanto representan usos excesivos de mano de obra, materias primas, etc.

2.5 Ponderaciones Fijas y Variables

Los ejemplos utilizados hasta aquí han sido muy limitados en el número de insumos y productos utilizados. Esto hizo posible el uso de representaciones gráficas sencillas para aclarar las cosas, pero, por supuesto, esto fue a expensas del realismo necesario para hacer frente a los múltiples insumos y múltiples productos que se encuentran comúnmente en la práctica. La clave es desarrollar enfoques que permitan hacer frente a tales aplicaciones sin tener que sobrecargar indebidamente a los usuarios con análisis o cálculos excesivos y sin requerir un gran número de supuestos (a menudo arbitrarios o cuestionables).

Consideremos, por ejemplo, la situación en la Tabla 1.5 que registra el comportamiento destinado a servir de base para evaluar la eficiencia relativa de 12 hospitales en términos de dos insumos, el número de médicos y número de enfermeras, y dos productos identificados como número de pacientes ambulatorios y pacientes hospitalizados (cada uno en unidades de 100 personas / mes).

Tabla 2.5 Caso de un Hospital

Hospital	A	B	C	D	E	F	G	H	I	J	K	L
Doctores	20	19	25	27	22	55	33	31	30	50	53	38
Enfermeras	151	131	160	168	158	255	235	206	244	268	206	284
P. ambulatorios	100	150	160	180	94	230	220	152	190	250	260	250
P. hospitalizados	90	50	55	72	66	90	88	80	100	100	147	120

Una forma de simplificar las cosas sería ponderar los diversos insumos y productos por pesos preseleccionados (fijos). La relación resultante daría entonces un índice para la evaluación de la eficiencia. Por ejemplo, la ponderación

$$v_1 \text{ (ponderación del doctor): } v_2 \text{ (ponderación de la enfermera) = 5:1}$$

$$u_1 \text{ (ponderación p. ambulatorios): } u_2 \text{ (ponderación p. hospitalizados) = 1: 3}$$

produciría los resultados que se muestran en la fila denominada "fijos" de la Tabla 2.6. (Nótese que estas relaciones se normalizan de manera que el máximo llega a ser la unidad, es decir, dividiendo por la relación de A.) Esto simplifica las cosas para su uso, sin duda alguna, pero plantea un sinnúmero de otras cuestiones tales como justificar la relación de 5 a 1 para el médico frente a

¹¹ Vide p.192 in W.W. Cooper and Y. Ijiri, eds., *Kohler's Dictionary For Accountants*, 6th Edition (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1981.)

la enfermera y / o la relación de 3 a 1 de las ponderaciones de los pacientes internos y externos. Por último, y aún más importante, están los problemas que pueden surgir con los resultados mostrados - ya que no está claro cuánto de los índices de eficiencia se debe a las ponderaciones y cuánta ineficiencia está asociada a las observaciones.

Tabla 2.6 Comparación de pesos Fijos y Variables

Hospital	A	B	C	D	E	F	G	H	I	J	K	L
Fijos	1	.90	.77	.89	.74	.64	.82	.74	.84	.72	.83	.87
CCR	1	1	.99	1	.76	.84	.90	.80	.96	.87	.96	.96

DEA, por el contrario, utiliza ponderaciones variables. En particular, los pesos se derivan directamente de los datos con la consecuencia de que se evitan numerosos supuestos a priori y cálculos implicados en la elección de pesos fijos. Por otra parte, las ponderaciones se eligen de forma de asignar un mejor conjunto de ponderaciones a cada hospital. El término "mejor" se utiliza aquí para significar que la proporción resultante de insumo-producto de cada hospital se maximiza con respecto a todos los demás hospitales cuando estos pesos son asignados a estos insumos y productos para todos los hospitales. La fila etiquetada CCR en la Tabla 2.6 muestra los resultados obtenidos a partir de DEA utilizando lo que se llama el "modelo CCR" ® en DEA. Como se puede observar, estos valores de eficiencia son siempre al menos tan grandes como los valores de la relación obtenidos a partir de las ponderaciones fijas anteriores. Por otra parte, este resultado de "mejor relación" es general, bajo las siguientes condiciones: (1) todos los datos y todos los pesos son positivos (o al menos no negativos), (2) la relación resultante debe estar entre cero y la unidad, y (3) estos mismos pesos para la entidad de destino (= hospital) se aplican a todas las entidades. En consecuencia, la entidad que se está evaluando no puede elegir un mejor conjunto de ponderaciones para su evaluación (en relación con las otras entidades). El significado de estos resultados está claro. En cada caso, la evaluación se realiza a partir de un punto de la frontera eficiente de manera que un valor como 0,88 para el hospital C significa que es 12% ineficiente. Es decir, en comparación con los miembros de un conjunto de referencia eficiente, es posible identificar una ineficiencia puramente técnica, de 12%, y posibles ineficiencias de mezcla, aun con el mejor conjunto de ponderaciones que cada uno de estos hospitales podría elegir para evaluar su propia ineficiencia.

Como veremos más adelante, las fuentes de ineficiencia, como la puramente técnica y la ineficiencia de mezcla, se identifican automáticamente para cada entidad y sus cantidades son estimadas por DEA. Por otra parte, el conjunto de referencia utilizado como referencia también es identificado. Por último, como también veremos, estos resultados se obtienen utilizando sólo supuestos a priori mínimos. Además de evitar la necesidad de una elección a priori de ponderaciones, DEA no requiere especificar la forma de la relación entre los insumos y productos de manera, tal vez, arbitraria y, aún más importante, no requiere que estas relaciones sean las mismas para cada hospital.

2.6 Sumario y conclusión

Hemos cubierto ahora una variedad de temas que serán perfeccionados y ampliados en este libro. Empleando medidas usuales de cocientes insumo-producto los hemos relacionado con temas tales como las mediciones de la productividad, así como con métodos de evaluación de la

eficiencia de uso comunes en economía, negocios e ingeniería. Posteriormente, vamos a introducir otros enfoques (no en forma de cociente), pero se lo hará de manera de mantener contacto con estas formas cociente.

Se examinaron extensiones a múltiples productos e insumos en términos de ponderaciones fijas a ser aplicadas de manera uniforme a los insumos y productos de todas las entidades a ser evaluadas, como en los índices económicos de la "productividad total de los factores." A continuación, este uso se contrastó con el uso de ponderaciones variables basadas en un mejor conjunto elegido para cada entidad a evaluar, como en DEA. También describimos interpretaciones y usos de este último como derivado de las fronteras de eficiencia con las que se han efectuado las evaluaciones. Después, esto se ha contrastado con las fuentes de ineficiencia mixtas, en general, poco claras, que están implícitas en el uso de enfoques de ponderación fija.

También fueron observadas ventajas adicionales de DEA en términos de (a) su capacidad de identificar fuentes y cantidades de ineficiencia en cada insumo y cada producto de cada entidad (hospital, tienda, horno, etc.) y (b) su capacidad de identificar miembros de referencia del conjunto eficiente utilizado para realizar estas evaluaciones e identificar estas fuentes (y cantidades) de ineficiencia.

Se supuso que todas las entidades evaluadas de ese modo usan los mismos insumos para producir los mismos productos. Además, se supone que todos los datos son positivos y las ponderaciones también se restringieron a valores positivos. Estas hipótesis se mantendrán en los siguientes capítulos y luego serán relajadas. También se supone que los insumos y productos son variables a discreción de los administradores o diseñadores. Esta hipótesis también se mantendrá y luego será atenuada, de forma que vamos a ser capaces de distinguir entre insumos y productos discretos y no discretos - para permitir diferencias en las circunstancias bajo las que operan las diferentes entidades. Entonces también introduciremos variables categóricas para permitir una mayor diferencia tales como categorías urbana o rural, etc., para obtener evaluaciones y puntos de vista más refinados.

La discusión en este capítulo se limitó a aspectos físicos de la eficiencia con distinciones entre ineficiencias "puramente técnicas" y "mezcla". A estas ineficiencias se las conoce como "desperdicio", ya que podrían ser removidas de todo insumo o producto sin empeorar ningún otro insumo ni producto. Otros tipos de ineficiencia cubiertos en este libro implicarán el movimiento a lo largo de las fronteras de eficiencia y por lo tanto también implicarán cambios o sustituciones. Tales movimientos pueden llevarse a cabo para lograr economías de rendimientos a escala o para mejorar las performances en términos de costos y beneficios. Todos estos movimientos a lo largo de las fronteras, sin embargo, implican la ausencia de ineficiencias técnicas (= puramente técnicas más mezcla). De ahí que este tema será el centro de atención en los capítulos inmediatamente siguientes.

3. Una introducción a los Métodos CNLS y StoNED de análisis de la eficiencia: Intuiciones económicas y aspectos computacionales¹²

3.1 *Introducción*

El Análisis de eficiencia es un campo interdisciplinario que abarca disciplinas como economía, investigación de operaciones, ciencias de la gestión e ingeniería. La teoría y los métodos de análisis de la eficiencia se utilizan en varios campos de aplicación, entre ellos agricultura, bancos, educación, medio ambiente, salud, energía, industria manufacturera, transporte y empresas de servicios públicos. El análisis de eficiencia se puede realizar a distintos niveles de agregación: las micro-aplicaciones van desde las personas individuales, equipos, plantas de producción, e instalaciones a las evaluaciones de eficiencia a nivel de empresa y de nivel de industria, mientras que las macro-aplicaciones van desde las evaluaciones comparativas de eficiencia de los sistemas de producción o industrias de diferentes países a la evaluación de la eficiencia de las economías nacionales. De hecho, la mejora de la eficiencia es uno de los componentes críticos del crecimiento de la productividad en el tiempo, que a su vez es el principal impulsor del bienestar económico. Como el desempeño de un país a nivel macro es simplemente un agregado del desempeño de las empresas individuales que operan dentro de ese país, los micro-fundamentos racionales del análisis de eficiencia son esenciales para la productividad a nivel macro.

Tradicionalmente, el campo de análisis de eficiencia estaba dividido entre dos paradigmas en competencia:

- a) El Análisis de Envolvente de Datos (Data Envelopment Analysis - DEA) (M.J. Farrell, [The Measurement of Productive Efficiency](#), 1957; A. Charnes, W.W. Cooper and E. Rhodes, [Measuring the efficiency of decision making units](#), 1978);
- b) El Análisis de Frontera Estocástica (Stochastic Frontier Analysis - SFA) (Dennis Aigner, C.A. Knox Lovell and Peter Schmidt, [Formulation and Estimation of Stochastic Frontier Production Function Models](#), 1977; Wim Meeusen and Julien van Den Broeck, [Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error](#), 1977; Peter Schmidt and C.A. Knox Lovell, [Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers](#), 1977).

DEA es un enfoque de programación axiomático y matemático para el análisis de eficiencia que no asume ningún tipo funcional particular de frontera o de distribución de la ineficiencia. La principal ventaja de DEA en comparación con las herramientas econométricas basadas en regresión es el tratamiento no paramétrico de la frontera, sobre la base de axiomas de teoría de la producción como libre disponibilidad (monotonía), convexidad (concavidad), y rendimientos constantes a escala (homogeneidad). Sin embargo, el principal inconveniente de DEA es que atribuye todas las desviaciones de la frontera a ineficiencia. Por el contrario, SFA utiliza técnicas de regresión paramétrica, que requieren especificaciones previas de las formas funcionales de la frontera y la distribución de la ineficiencia. La fortaleza de SFA es su modelo probabilístico de desvíos de la frontera, que se descomponen en un término de ineficiencia y otro de ruido que da cuenta de factores omitidos tales como la heterogeneidad inobservable de las empresas y de sus entornos operativos, los errores aleatorios de medición y procesamiento de datos, los errores de

¹² Andrew L. Johnson and Timo Kuosmanen, [An Introduction to CNLS and StoNED Methods for Efficiency Analysis: Economic Insights and Computational Aspects](#), Chapter 3, Subhash C. Ray, Subal C. Kumbhakar and Pami Dua, [Benchmarking for Performance Evaluation - A Production Frontier Approach](#), Springer, 2015. Los capítulos del libro pueden ser bajados en forma separada.

especificación, y otras fuentes de ruido. Hacemos hincapié en que los métodos DEA y SFA no deben ser vistos como competidores directos, sino más bien como complementarios: en el tradeoff entre DEA y SFA algo hay que sacrificar para conseguir algo. DEA no modela el ruido, pero es capaz de imponer propiedades axiomáticas y estimar la frontera de forma no paramétrica, mientras que SFA no puede imponer propiedades axiomáticas, pero tiene la ventaja de modelar la ineficiencia y el ruido.

Durante mucho tiempo, cerrar la brecha entre la DEA axiomática y la SFA estocástica fue uno de los problemas más acuciantes en materia de análisis de eficiencia. Los trabajos recientes sobre mínimos cuadrados no paramétricos convexos (CNLS) por Kuosmanen ([Representation theorem for convex nonparametric least squares](#), 2008), Kuosmanen and Johnson ([Data envelopment analysis as nonparametric least squares regression](#), 2010), y Kuosmanen and Kortelainen ([Stochastic non-smooth envelopment of data: Semiparametric frontier estimation subject to shape constraints](#), 2012) han llevado a una plena integración de DEA y SFA en un marco unificado de análisis de productividad, al cual nos referimos como Envolverte de datos estocástica no paramétrica (StoNED). El desarrollo de StoNED no es sólo una innovación técnica; se trata de un cambio de paradigma para el análisis de la eficiencia. Con StoNED, ya no tenemos que considerar si el ruido de modelado es más importante que la imposición de axiomas de la teoría de producción: StoNED nos permite hacer ambas cosas. El entorno de trabajo unificado de StoNED no sólo ofrece una visión más profunda de la intuición económica y los fundamentos de DEA y SFA, sino que también proporciona una plataforma más general y flexible para el análisis de eficiencia y temas vinculados, como la estimación de la frontera y el análisis de producción. Además, en las últimas décadas se han desarrollado varias extensiones a los métodos originales de DEA y SFA. El marco de StoNED unificado nos permite combinar las herramientas existentes de análisis de eficiencia en formas novedosas de todo el espectro de DEA-SFA, facilitando nuevas oportunidades para un mayor desarrollo metodológico.

El propósito de este capítulo es proporcionar una introducción a los estimadores CNLS y StoNED y revisar los fundamentos básicos de la economía de CNLS y StoNED a fin de introducir las formulaciones de programación matemática relacionadas y los códigos computacionales. Para una discusión más detallada sobre las propiedades teóricas y extensiones de CNLS y StoNED, remitimos al lector a Kuosmanen, T., Johnson, A.L., and Saastamoinen, A. ([Stochastic nonparametric approach to efficiency analysis: A unified framework](#), 2014).

Proporcionamos ejemplos detallados de códigos computacionales para dos lenguajes populares matemáticos de computación de alto nivel: el sistema general de modelización algebraica (GAMS: véase <http://www.gams.com>) y el laboratorio matriz (MATLAB: véase <http://www.mathworks.com/products/Matlab/>). Si bien otros lenguajes de computación o entornos tales como R, Python, o AIMMS pueden igualmente ser utilizados, computar el estimador CNLS incluso para una muestra relativamente pequeña de observaciones hace necesario el uso de un entorno de modelado matemático y de solucionadores de alto rendimiento de programación matemática para la programación cuadrática (QP) o programación no lineal (NLP), dependiendo de la formulación del modelo. Solucionadores integrados están disponibles para ambos GAMS y MATLAB, lo que hace de estos dos lenguajes informáticos matemáticos ambientes convenientes para el cálculo de los estimadores CNLS o StoNED. Aunque restringimos los ejemplos a formulaciones en este capítulo a estos dos lenguajes de computación, deseamos alentar a los profesionales con experiencia computacional a desarrollar sus propios códigos para otros lenguajes computacionales tales como R o Python.

