

## Indicadores de eficiencia para la gestión pública: Una revisión de métodos

**Guillem López Casasnovas - Adam Robert Wagstaff**

*Facultat de Ciències Econòmiques  
Universitat Pompeu Fabra  
c/ Balmes, 132  
08008 Barcelona*

**Indicadores de eficiencia para la gestión  
pública: Una revisión de métodos.**

**Efficiency Indicators for Public  
Management: A Review.**

### RESUMEN

### ABSTRACT

Las páginas que siguen constituyen una revisión de la literatura existente en el ámbito de la economía aplicada, relativa a la construcción de indicadores para la gestión de los recursos, basados en criterios de eficiencia económica. El propósito de dicha revisión es destacar las potencialidades que se derivan de la aproximación basada en la estimación de la función frontera de posibilidades de producción y de costes para la gestión de los recursos públicos, tanto en lo que atañe a las políticas de planificación como de presupuestación.

El presente artículo consta de dos partes. La primera avanza los contenidos básicos de los conceptos de eficiencia económica, técnica y asignativa. La segunda establece los mecanismos de aproximación empírica que se utilizan en la construcción de los distintos indicadores. Un breve apartado de conclusiones enfatiza las posibilidades de aplicación para la gestión de los recursos públicos.

This article offers a review of methods for constructing performance indicators in public management.

Particular attention is devoted to new indicators based on the estimation of the cost and production possibility frontiers. For this, surveys on public activity offer the empirical support required, once statutory differences in outputs are taken into account. Thereafter, evaluation criteria may result from these exercises which are relevant for planning and budgeting policies.

The article builds on the definition of concepts such as technical, allocative and economic efficiency. They are analysed in the first part of the paper. The second part offers some insights in the basics of the econometric methods required for putting the former concepts in operation. We finally discuss the advantages of the resulting evaluation mechanisms in public sector management.

## Indicadores de eficiencia para la gestión pública: Una revisión de métodos<sup>(\*)</sup>

### 1. INTRODUCCIÓN

Existe una clara tendencia en la planificación —y como consecuencia en la gestión— de los recursos públicos a considerar que las funciones de producción, y funciones de costes a éstas asociadas, presentan un grado de rigidez tan elevado, que su gestión ofrece márgenes nulos de sustituibilidad. Ello se ha reflejado en una planificación normativa de los factores productivos en términos de actividad, de modo aislado y sin análisis alguno de la combinación óptima de los recursos, a la vista de la polivalencia de las tareas efectuadas y el coste relativo de los factores productivos<sup>1</sup>.

Dichas políticas fracasan a la hora de considerar los márgenes de sustitución técnica entre los distintos factores al servicio de unos mismos objetivos, imponiendo una uniformidad que puede no resultar óptimo. Sólo por esta razón, la descentralización de la gestión pública, en contextos en los que las tecnologías de producción son imperfectamente conocidas, ya resultaría justificada<sup>2</sup>. Esta argumentación favorece, consecuentemente, la aplicación de políticas descentralizadas de rentas y de empleo, dentro de márgenes preestablecidos en las políticas envolventes de gasto.

(\*) Este artículo se ha realizado con ayuda financiera del Ministerio de Educación y Ciencia a través del Plan de Movilidad de Personal Investigador (1990-91). Asimismo, forma parte de la investigación financiada por la Dirección General de Planificación Sanitaria del Ministerio de Sanidad y Consumo (1990-91), que lleva el título genérico de *Elaboración de Indicadores para la Planificación Sanitaria basada en criterios de Eficiencia Económica*. Guillem López es prof. titular de la Universitat Pompeu Fabra (Cataluña) y Adam Wagstaff *lecturer* de la Univ. de Sussex (Reino Unido).

1. Este es el caso, en particular, de las políticas públicas de personal por lo que atañe a la evaluación de necesidades de plantillas en términos de outputs intermedios y actividad, de ratios entre tipos de personal, y a las políticas salariales afectas.

2. Véase, en este sentido, A.B. Atkinson y J.E. Stiglitz (1980) "*Lectures on Public Economics*". Mc Graw Hill, pp. 519-533.

La consecuente estimación de funciones de producción y/o coste de servicios públicos, posibilita evaluar la bondad con la que se utilizan los recursos en la práctica, y asimismo los márgenes potenciales de mejora que se podrían producir en su asignación, en caso de incorporar los resultados del análisis a la planificación y gestión de recursos. Esta es una de las vías que abren los nuevos desarrollos en la Hacienda pública en favor de competencia simulada para la consignación presupuestaria (véase, en este sentido, G. López Casasnovas, 1989).

Desde el análisis microeconómico se identifica como mejora en la utilización de los recursos, aquella reasignación de factores que tienda a reconducir los coeficientes de productividad observados hacia sus niveles relativos de coste. En su concepción dinámica relativa a un centro en particular, una asignación eficiente requeriría, además, que éste fuera desarrollando su actividad en la medida en que, con la utilización de su capacidad, el "precio" igualase al coste marginal de producción a corto plazo, y la expansionara siempre que el coste marginal a largo plazo resultante fuese inferior al "precio".

Conviene resaltar que para que sustituciones potenciales de factores resulten óptimas para la gestión, se han de considerar sus costes relativos. De ahí la importancia de determinar las políticas salariales en los sistemas públicos de provisión, de modo coherente con los objetivos de la planificación, y viceversa. Ello exige, en definitiva, examinar si, a la vista de la productividad y nivel relativo de costes entre los distintos factores, es posible diseñar pautas para una asignación más eficiente de los recursos.

El Cuadro 1 ofrece una ilustración de los costes relativos observados entre distintos factores sanitarios.

CUADRO Nº 1  
COCIENTE DE COSTES ENTRE FACTORES PRODUCTIVOS EN HOSPITA-  
LES PUBLICOS DEL INSALUD 1989

MEDICO ADJUNTO/ATS	1.76
MEDICO ADJUNTO/AUXILIAR CLINICO	2.80
MEDICO ADJUNTO/CELADOR	2.72
ATS/AUXILIAR CLINICO	1.59
ATS/CELADOR	1.55
AUXILIAR CLINICO/CELADOR	0.97

Fuente: Elaboración propia<sup>3</sup>.

3. Se trata del coste para el hospital y no de salarios recibidos. Corresponde a una muestra de hospitales públicos del Instituto Catalán de la Salud considerada al efecto.

Las páginas que siguen discuten los métodos utilizados para la elaboración de indicadores de eficiencia económica que, basados en la estimación de las funciones de producción y/o coste de determinados servicios, pueden incorporarse en la gestión pública.

## 2. FUNDAMENTOS TEÓRICOS DEL ANÁLISIS

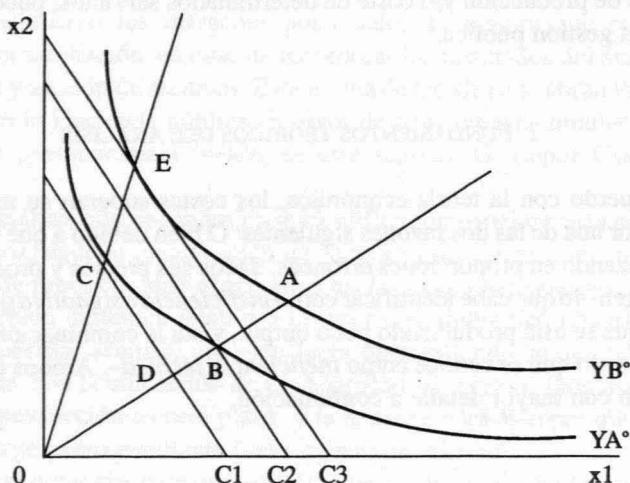
De acuerdo con la teoría económica, los costes superan su nivel mínimo factible por una de las dos razones siguientes: O bien debido a que los inputs se están utilizando en proporciones erróneas, dados sus precios y productividades en el margen—lo que cabe identificar como *ineficiencia asignativa* o de precios—, o bien a que se está produciendo poco output, vista la combinación de factores empleados—lo que se conoce como *ineficiencia técnica*—. Ambos conceptos se aproximan con mayor detalle a continuación.

### 2.1. Aproximación al concepto de eficiencia económica.

El gráfico 1 ilustra el caso de una función de producción con dos inputs,  $x_1$  y  $x_2$ , para un output  $y$ . La isocuanta  $y_A^0$  indica todas las combinaciones posibles de  $x_1$  y  $x_2$  que generan un mismo nivel de output,  $y_0$ . La pendiente de la isocuanta es negativa, reflejando el hecho de que una mayor utilización de  $x_1$  resulta en un aumento del output; a menos que  $x_2$  disminuya de modo compensatorio. La cantidad en la que  $x_2$  ha de disminuir por cada unidad de incremento de  $x_1$  es la pendiente de la isocuanta y se define como la tasa marginal de sustitución técnica entre  $x_1$  y  $x_2$ . Esta es a su vez igual al cociente de las productividades marginales de los dos inputs<sup>4</sup>.

4. La productividad o producto marginal de  $x_1$  se define como el output adicional obtenido del incremento en una unidad de  $x_1$ , manteniendo constante la utilización de todos los demás inputs.

GRAFICO 1



La línea isocoste, en el gráfico 1, identificada como  $C_1$ , indica las combinaciones de  $x_1$  y  $x_2$  que requieren un mismo nivel de gasto ( $C_1$ ). La pendiente de la línea de isocoste es igual al cociente del precio (coste) de los dos inputs:  $(-p_2/p_1)$ , siendo  $p_1$  el precio de  $x_1$ . La combinación de factores asignativamente eficiente para cada nivel de output será aquella que minimice el coste de producir el output en cuestión, o de modo equivalente, aquella combinación que maximice el nivel de output, o resultado obtenido, a partir de unos recursos financieros dados. En el gráfico 1, esto ocurre en el punto C. En este punto, la pendiente de la línea de isocoste y de la isocuanta coinciden. La eficiencia en la asignación de los recursos requiere, por lo tanto, que

$$(1) \quad MP_2/MP_1 = p_2/p_1$$

siendo  $MP_i$  la productividad marginal del input  $i$ . En efecto, si el coste unitario de  $x_1$  es doble del de  $x_2$ , entonces  $x_1$  debe ser, en el margen, dos veces más productivo que  $x_2$ . Esta regla puede generalizarse para cualquiera sea el número de inputs; para ello, (1) debe respetarse para cada par de inputs.

La denominada *eficiencia técnica* puede también ilustrarse utilizando el

diagrama de la isocuanta. Tal como se apuntaba anteriormente, cada isocuanta se asocia con un nivel de output: cuanto más lejos del origen esté la isocuanta, mayor será el nivel de output. La ubicación de la isocuanta para una función de producción en particular se determina exclusivamente, por tanto, de acuerdo con la tecnología. Es posible, sin embargo, considerar la posibilidad de que algunos centros sean más eficientes a la hora de transformar inputs en outputs que otros (es decir, muestren un nivel de eficiencia técnica superior). Las isocuantas  $y_A^\circ$  e  $y_B^\circ$  del gráfico 1 se asocian ambas a un mismo nivel de output. Sin embargo, la isocuanta del centro o unidad menos eficiente  $y_B^\circ$  aparece situada más lejos del origen que la del centro relativamente más eficiente  $y_A^\circ$ . La unidad menos eficiente está, por lo tanto, utilizando más de ambos inputs para producir el mismo nivel de actividad que la más eficiente. La ineficiencia técnica implica, consecuentemente, la utilización excesiva de todos los inputs.

Utilizando el gráfico 1 es posible medir también la ineficiencia (véase, Farrell, 1957). Supóngase que el hospital cuya eficiencia se está midiendo se sitúa en el punto A en lugar del punto C, de modo que no es eficiente ni desde un punto de vista asignativo ni técnico. Su nivel de eficiencia técnica se mide por el cociente  $OB/OA$  y se define como la proporción de la combinación de inputs utilizada que de hecho se requiere para producir el nivel de output observado. La ineficiencia técnica se aproxima como  $1-OB/OA$  y puede interpretarse como la proporción en la que el coste en el que se incurre al producir el nivel de output en cuestión podría disminuir, sin variación alguna en el cociente de inputs. Si existen beneficios constantes a escala<sup>5</sup>, la ineficiencia técnica puede interpretarse, aproximadamente, como la proporción en la que el output podría aumentar si el centro fuera eficiente al 100%.

El nivel de *eficiencia asignativa* del hospital se define como  $OD/OB$  (o de modo equivalente,  $C_1/C_4$ ) y refleja el coste en el que se hubiera incurrido si el hospital hubiera sido eficiente en relación con su coste actual. La ineficiencia asignativa se define como  $1-OD/OB$  y mide, el incremento proporcional en los costes que se origina debido a la ineficiencia asignativa.

La *eficiencia global* resultante se aproxima como el cociente  $OD/OA$  y corresponde al producto de la eficiencia técnica y la eficiencia asignativa. La ineficiencia global se define como  $1-OD/OA$  y mide la proporción en la que los costes superan su mínimo factible. En términos aproximados, se puede descomponer en la suma de la ineficiencia técnica y la ineficiencia en la asignación<sup>6</sup>.

5. Beneficios constantes a escala implican que doblando todos los inputs se dobla el output.

6. En sentido estricto, la ineficiencia global es la suma de  $(1-OB/OA)$  y  $(1-OD/(OB+AB))$ . Esta última expresión corresponde sólo de manera aproximada a la noción de ineficiencia asignativa.

Ambos tipos de ineficiencia son potencialmente importantes en el sector público, dadas las características que confluyen en su prestación y la naturaleza de los bienes proveídos (tales como seguridad, salud, educación, etc).

En la medida en que se observan en el sector público niveles diversos de actividad para la consecución de unos mismos objetivos, cabe interrogarse acerca de la posible presencia de ineficiencia técnica. En la medida en que difieran sus costes unitarios, tanto la ineficiencia técnica como la económica puede estar presente en la asignación de recursos. Ambos supuestos plantean, sin embargo, la necesidad de su identificación y medida, de modo sensible a las características de los outputs públicos; a efectos de suministrar información plausible sobre la dirección de los cambios aconsejables en la planificación y gestión de los recursos.

## 2.2. Una aproximación operativa.

Los intentos efectuados para aproximar en la práctica los anteriores conceptos han sido diversos. De entre ellos destacan dos métodos alternativos: los métodos basados en la aplicación de técnicas estadísticas de regresión múltiple para la estimación de medias muestrales y, más recientemente, las basadas en la definición de fronteras de coste o producción, a partir de métodos no necesariamente estadísticos, para la identificación del "locus" de eficiencia al que referir el análisis de las prácticas observadas.

Ambas aproximaciones se comentan a continuación.

### 2.2.1. Aproximaciones "no frontera".

Desde esta óptica, la eficiencia global de un hospital puede evaluarse a través del exámen de la función de producción, o de su dual de costes. Esta última no cabe confundirla con la ecuación de costes que, como se sabe, es una pura identidad contable, definida como un simple desglose de sus componentes (salarios, costes generales, de laboratorio, etc.). A diferencia de ésta, una función de costes ha de ser coherente con los postulados de la teoría económica que la sustentan, expresando el coste en función de las variables que determinan la optimización restringida de su actividad. Dichas variables incluyen, sin duda, el precio de los inputs -no el gasto en inputs- y el nivel de output producido. Si el horizonte relevante es el corto plazo -el estoc de capital de capital del centro es fijo-, la función de costes incluirá también dicho estoc, puesto que el carácter

fijo de éste impide al centro reducir sus costes a través de alterar su combinación factorial.

Si la función de costes es de la forma

$$(2) \quad c_i = \beta_0 + \sum_j \beta_j x_{ij} + u_i$$

siendo  $c_i$  el coste total del centro  $i$ . La variable  $x_j$  refiere factores tales como el nivel de output y precios de los inputs, y  $u_i$  el término de error. En el ámbito hospitalario público, Feldstein (1967) sugirió que la eficiencia de un hospital podía evaluarse comparando sus costes reales con los "esperados", siendo éstos últimos los resultantes de la predicción de la ecuación de costes estimada según (2). En otras palabras, el residuo  $u_i$  podía considerarse como una medición de la eficiencia: centros con residuos positivos se situarían por debajo del nivel de la eficiencia media (con sus costes reales en exceso de los *esperables*), mientras que aquellos con residuos negativos indicarían niveles superiores a la media.

Feldstein utilizó esta aproximación para investigar la eficiencia global de una muestra de 177 hospitales de agudos no docentes del Servicio británico de la Salud. En dicho estudio, la variable dependiente  $c_i$  se definió como el coste medio por ingreso (o coste por caso), y la variable  $x_j$  recogía la proporción de pacientes en cada una de las  $j$  categorías de *case mix* o abanico de patologías tratadas. Este último vector de variables pretendía capturar las diferencias en el output entre los distintos hospitales. El precio de los inputs no se incluía en la estimación, dado que eran los mismos, en principio, para todos los hospitales del NHS (con la excepción de los hospitales de Londres, que soportan unos mayores costes salariales).

En esta simple función de costes,  $\beta_0 + \beta_j$  pueden interpretarse como el coste marginal de tratar un paciente del tipo  $j$  (cf Feldstein, 1967:36). Los costes esperados corresponden, por lo tanto, al coste medio que hubiera observado un hospital, con una combinación de patologías determinada, si los costes específicos a dicha combinación hubieran sido iguales a los de la media muestral. En efecto, un hospital con un residuo igual a cero registraría, bajo éste supuesto, un coste medio exactamente igual al que cabría esperar de dicho hospital, sobre la base de los costes estimados para su combinación específica de casos. Los hospitales con residuo positivo (negativo), por contra, indicarían costes medios por encima de (inferiores) la cuantía esperable dado el coste medio específico para su *case mix*.

López Casanovas (1984) utilizó una aproximación similar para distinguir la



eficiencia de los hospitales para una muestra de centros del INSALUD. Como en el análisis de Feldstein, la variable dependiente considerada fue el coste medio por caso. Entre las variables explicativas se incluyeron, sin embargo, no sólo las variables *case mix*, sino también el estoc de camas utilizadas, su valor al cuadrado, el número de ingresos por cama y año y el valor al cuadrado de ésta misma variable. Entre las diversas especificaciones estimadas por López Casasnovas se incluyeron, asimismo, otras variables que tuvieran en cuenta el estatus docente, en su caso, del centro, el peso de la actividad de consulta externa, etc. Se trataba con ello de explorar las implicaciones en la gestión del gasto de los hospitales que se pudieran derivar como resultado de consignar los presupuestos en base a los costes esperados (*predichos* a partir de la estimación), en lugar de reembolsar los centros a partir de sus costes actuales. La racionalidad de dicha propuesta se basaba en el hecho de que para aquellos hospitales cuyos costes esperados superasen los reales, su mayor eficiencia relativa merecía un tratamiento diferenciado (cf. López Casasnovas, 1984:134)<sup>7</sup>.

En los dos casos anteriores, la aproximación seguida ofrece la ventaja de que, al fundamentarse en la teoría económica, resulta claro el tipo de eficiencia medido. Además, al incorporar variables que recogen la influencia sistemática de factores fuera del control específico de los hospitales, dicha aproximación hace más sensible la posterior identificación del residuo con el grado de eficiencia.

Por contra, la estimación de este tipo de funciones adolece de dos defectos. En primer lugar, no ofrece información sobre el nivel de eficiencia absoluto: Parece obvio que resulta importante conocer si los hospitales son muy ineficientes, o lo son tan sólo de modo marginal. En segundo lugar, esta aproximación asume que toda la variación que se registra en el residuo entre los distintos centros se debe a la variación de la eficiencia, cuando, en realidad, los residuos puede que también reflejen influencias aleatorias —fuera del control de los hospitales—, del mismo modo que perturbaciones estadísticas (es decir, tanto errores en la medición como variables omitidas cf. Feldstein, 1967:50).

7. Funciones similares de comportamiento de costes hospitalarios han sido estimadas posteriormente por J. Puig Junoy (Gaceta Sanitaria, prim. ep., nº 8, 1988) y L. Pohigas Santasusagagna (IV Jornadas de Economía de la Salud, Sevilla, 1986).

### 2.2.2. Aproximaciones "frontera".

En el transcurso de la última década se han propuesto nuevos métodos para la medición de la eficiencia económica, la mayoría de los cuales tiene como propósito principal superar los problemas inherentes a la aproximación *no-frontera* antes comentados. Todos tienen en común, sin embargo, la noción de frontera: esto es, se consideran centros eficientes aquellos que operan por encima de la frontera de producción o por debajo de la correspondiente frontera de costes. Más allá de compartir este principio, los métodos difieren entre sí en una gran variedad de aspectos (véase Forsund, Lovell and Schmidt, 1980; Schmidt, 1986).

Una de las principales diferencias se refiere a la interpretación del término "frontera". Algunos métodos pretenden reflejar la frontera *absoluta*, como expresión del resultado de la actividad alcanzable en el supuesto de que la tecnología disponible se utilizara a plena ventaja. Otros métodos aspiran a identificar la frontera de la *práctica mejor* (Farrell, 1957), a la vista de los resultados obtenidos por los propios centros que integran la muestra. Las estimaciones que intentan determinar la frontera absoluta pueden fácilmente encontrar un hospital en la muestra operando al 100% de eficiencia, mientras que ello no puede ser cierto para los métodos que determinan la mejor práctica. Aunque en realidad, tal como Forsund y otros (1980) apuntan, la distinción es poco probable que tenga una mayor significación práctica, dado que las dos fronteras han de converger en la medida que el tamaño de la muestra tienda a infinito.

Una diferencia más substancial se refiere al carácter estadístico o no del método en cuestión. Los métodos *estadísticos* establecen hipótesis acerca de las propiedades estocásticas de los datos, mientras que los métodos *no estadísticos* no. Otra diferencia concierne al carácter *paramétrico* o *no paramétrico* del método en cuestión: el primero de ellos impone una forma funcional particular (por ejemplo Cobb-Douglas o translog), mientras que el segundo no. Un rasgo atractivo de los métodos no estadísticos es que tienden a ser no paramétricos, a diferencia de las aproximaciones estadísticas que son casi siempre paramétricas.

#### 2.2.2.a. Eficiencia global.

La aproximación frontera a la medición de la eficiencia global se ha centrado mayormente en la *función frontera de coste*. El primer modelo frontera que

aparece en la literatura es el **modelo de frontera determinada (MFD)** de Aigner y Chu (1968). La frontera de coste MFD difiere de la función de coste (2) en la restricción de que el término de error ha de ser no negativo: en efecto, la ecuación (2) se especifica aquí como:

$$(3) \quad c_i = \beta_0 + \sum_j \beta_j X_{ij} + u_i \quad u_i \geq 0$$

Los centros hospitalarios pueden, por tanto, operar sobre o por encima de la frontera de coste, pero no por debajo de ella. El alcance de la ineficiencia viene indicada por el residuo de (3),  $u_i$ .

El MFD puede estimarse a través de una gran variedad de métodos, el más simple de los cuales es el de Mínimos Cuadrados Corregidos (MCC): éste método supone una traslación hacia abajo del estimador mínimo cuadrático ordinario de la constante, hasta que un residuo se convierte en cero y todos los restantes resultan positivos (cf. Forsund y otros, 1980; Schmidt, 1986).

De nuevo en el sector sanitario, Wagstaff (1989b) estimó el MFD para una muestra de 49 hospitales españoles públicos. La muestra incluía centros con y sin docencia, constituyendo una submuestra de la utilizada por López-Casasnovas y Wagstaff (1988). La especificación de la función de coste era similar a la de Feldstein (1967). Las variaciones en el output se recogían en un vector compuesto de seis categorías de variables *case mix*. Asimismo, la ecuación estimada incluía el número de camas, su valor al cuadrado, el índice de rotación y su valor al cuadrado, y una variable ficticia que recogía el impacto de la docencia. El MFD se estimó por mínimos cuadrados corregidos, resultando una ineficiencia media por caso ingresado, que se cifraba en un 28.4 del coste medio.

Pese a suponer sin duda una mejora respecto de la aproximación no frontera, el MFD resulta menos atractivo que la aproximación no paramétrica. Ello se debe, en principio, a que se basa en una forma funcional específica y que, en ningún caso, permite distinguir la ineficiencia de los *shocks* aleatorios y de la perturbación estadística. Tal como sucedía en el modelo de regresión *no frontera*, el MFD considera el residuo estimado en su conjunto como indicación de ineficiencia.

El **modelo de frontera estocástica (MFE)** ofrece un modo sofisticado de superar el problema antes referido, al especificar:

$$(4) \quad c_i = \beta_0 + \sum_j \beta_j X_{ij} + v_i + u_i \quad u_i \geq 0$$

Aquí el término de error  $v_i + u_i$  lo integran dos partes. La primera  $v_i$ , es un componente doble (de dos lados), que captura los *shocks* aleatorios y la perturbación estadística. La segunda,  $u_i$  refleja la ineficiencia, que se constriñe a valores no negativos (un solo lado). En el MFE, por tanto, la frontera se compone de dos partes: La primera,  $\sum_j \beta_j X_{ij}$ , constituye el núcleo no estocástico de la frontera, que resulta común a todos los centros; la segunda ( $v_i$ ), es un componente aleatorio que refleja el *ruido* estadístico y los *shocks* aleatorios, que varía entre hospitales. La ineficiencia en el MEF se captura en el término  $u_i$  y se mide, por lo tanto, en relación a la frontera estocástica  $\beta_0 + \sum_j \beta_j X_{ij} + v_i + u_i$ .

A efectos de diferenciar la ineficiencia de los *shocks* aleatorios y la perturbación estadística, la aproximación más común se basa en el examen del sesgo, en su caso, de los residuos<sup>8</sup>. Si puede asumirse que el componente  $v_i$  del término de error tiene una distribución normal, el hecho de que la ineficiencia deba incrementar los costes significa que el conjunto del término de error ha de tener una distribución sesgada.

En las aplicaciones se ha postulado, generalmente, que la distribución del componente de la ineficiencia en el término del error es semi normal<sup>9</sup>. Bajo este supuesto, resulta posible estimar el MFE a través de completar la información utilizada para la estimación del modelo regresional, con información acerca del alcance del sesgo en los residuos (véase Schmidt y Lovell, 1979). Una alternativa a lo anterior la ofrece el estimador de máxima verosimilitud (véase Greene, 1980, 1982). Sin embargo, sea cual sea el método utilizado, se acaba considerando un residuo para cada centro, una estimación de la media de  $u_i$ , pero no un estimador de  $u_i$ . Lo que puede estimarse, sin embargo, es  $E(u_i/v_i + u_i)$ ; esto es, el valor esperado de  $u_i$ , dado el valor del error compuesto (véase Jondrow y otros, 1982).

Un método alternativo consiste en la estimación (cf Schmidt y Sickless, 1984) basada en *datos de panel*. Esta aproximación no requiere hipótesis alguna acerca del término de error. Permite que la ineficiencia varíe o permanezca constante en el tiempo (aunque, naturalmente, en este último caso, pudiendo variar entre unidades). El supuesto de constancia en el tiempo de los niveles de ineficiencia significaría que cambios en los costes de la unidad a lo largo del tiempo no podrían deberse a cambios en la eficiencia. Un modo de tratar este problema consiste en eliminar el componente de ineficiencia no observable, a

8. Una alternativa a lo anterior es la utilización de datos de panel (véase Pitt y Lee 1981; Schmidt y Lee, 1984).

9. Este supuesto, que está en la base del modelo a menudo citado de Aigner, Lovell y Schmidt (1977) y de Meusan y van den Breck (1977), es comunmente utilizado en el análisis aplicado.

través de trabajar con datos definidos en términos de desviaciones sobre valores medios en el tiempo: ésta es la aproximación denominada de *efecto fijo*. Alternativamente, cabe tratar la ineficiencia como un *efecto aleatorio* o *random*. Pero en ambos casos es posible computar la ineficiencia para cada unidad en particular de la muestra.

Las propiedades de este tipo de modelos se ilustran a continuación.

Supóngase que cada uno de los  $H$  centros o unidades incluidos en la muestra se analiza a lo largo de  $T$  años. La función de costes puede especificarse como:

$$(4') \quad c_{ht} = \beta_0 + \sum_j \beta_j X_{hjt} + v_{ht} + u_h \quad u_h \geq 0$$

siendo  $c_h$  el coste medio por enfermo tratado en el centro  $h$ ,  $t$  se refiere al período de tiempo considerado,  $X_i$  el vector de variables explicativas incluidas en la estimación,  $\beta_i$  los parámetros asociados a dichas variables en la estimación,  $v$  y  $u$  el término de error, compuesto de dos partes ( $v$  captura los shocks aleatorios y la perturbación estadística, y  $u$  la ineficiencia).

Nótese que el término de ineficiencia  $u$  no tiene referencia temporal, lo cual refleja el postulado de que la ineficiencia varía entre los centros ( $h$ ), pero es la misma en un centro dado en cada uno de los años. La ecuación (4') contiene un error sesgado con media no nula. Sin embargo, al transformar la ecuación se obtiene otra similar a la utilizada en los modelos de datos de panel. Esto es

$$(4'') \quad c_{ht} = \beta_0^* + \sum_j \beta_j X_{hjt} + v_{ht} + u_h^*$$

siendo  $\beta_0^* = \beta_0 + \mu$  y  $u_h^* = u_h - \mu$ . Tanto  $v_{ht}$  como  $u_h^*$  tienen ahora media cero.

Consecuentemente, la aproximación a la ineficiencia basada en datos de panel puede caracterizarse de dos maneras. En la primera, la ineficiencia se trata como un efecto fijo, de modo que se considera completamente sistemática. Los coeficientes del modelo se estiman en este caso, utilizando una variable ficticia o *dummy* o estimador interno (*within estimator*); como en cualquier otro modelo de datos de panel de efecto fijo. Ello requiere transformar la ecuación anterior en otra expresada en términos de desviaciones de los valores medios a lo largo del tiempo (véase, por ejemplo, Judge y otros, 1985).

En la segunda caracterización, la ineficiencia se considera un efecto aleatorio; es decir, se tiene en cuenta que el nivel de ineficiencia puede venir

parcialmente determinado por el azar. Los coeficientes se estiman aquí a través de mínimos cuadrados generalizados (MCG), de modo similar a como lo hace la literatura de datos de panel: Los estimadores MCG son medias ponderadas de los estimadores externos (*between* o "entre" estimadores) y los internos (*within* o "intra" estimadores), siendo los primeros resultado de la estimación del modelo con datos expresados en términos de valores medios en el tiempo. En ningún caso se requiere supuesto alguno acerca de la forma precisa de la distribución del componente aleatorio del error ni de la forma de la distribución del término de ineficiencia.

La elección entre las dos caracterizaciones de la ineficiencia se realiza, en la práctica, a la vista de las circunstancias observadas. En concreto, no es posible tratar la ineficiencia como un efecto fijo si la función de costes incluye variables que permanecen constantes a lo largo del tiempo, como suele ser el caso, por ejemplo, de las variables relativas a los aspectos institucionales de un centro [véase Schmidt y Sickless, 1984:369]. En este supuesto, la caracterización de la ineficiencia como aleatoria debe adoptarse en el modelo y estimarse a través de Mínimos Cuadrados Generalizados (MCG), en línea con lo propuesto por Hausman y Taylor (1978). La razón de ello estriba en que la transformación "intra" elimina no sólo los efectos específicos al hospital, sino también las variables específicas al centro que no varían en el tiempo. Esto no supone un problema en el modelo de efectos aleatorios, dado que los estimadores de los coeficientes de las variables constantes en el tiempo pueden obtenerse a partir de los estimadores externos (o *between estimates*) [cf Hausman y Taylor, op cit].

El output de ambos métodos de estimación (MCG e *intra*) incluye estimaciones de los coeficientes de la función de coste. A través de ellos pueden recuperarse los estimadores del nivel de ineficiencia de cada hospital, tal como apuntan Schmidt y Sickless (op cit). En efecto, dichos autores sugieren calcular un conjunto de residuos  $\varepsilon_{ht} = c_{ht} - \sum_j b_j X_{hjt}$ , siendo  $b_j$  el estimador de  $\beta_j$ .

A la vista de la ecuación (4") resulta claro que estos residuos han de reflejar  $v_{ht}$  y  $u_h$ , así como  $\beta_0$ . Sin embargo, al promediar los  $\varepsilon_{ht}$  en el tiempo,  $v_{ht}$  desaparece y se obtiene el resultado

$$b_{oh} = (1/T) \sum_t \varepsilon_{ht} = \beta_0 + u_h$$

De este modo, los residuos promediados pueden utilizarse como una estimación de la suma de  $\beta_0$  y  $u_h$ . Schmidt y Sickless sugieren entonces estimar  $\beta_0$  a través de los valores mínimos de  $b_{oh}$  y  $u_h$  de modo que  $\hat{u}_h = b_{oh} - \min(b_{oh})$ .

Ello supone contabilizar el hospital más eficiente de la muestra como 100% eficiente, y medir el grado de ineficiencia de los demás hospitales en relación con el hospital más eficiente.

Wagstaff (1989b) estimó el MFE para la misma muestra de 49 hospitales públicos españoles utilizada para la estimación anterior del modelo MFD. La ineficiencia fue considerada inicialmente como una variable aleatoria con una distribución semi-normal, siendo el modelo estimado utilizando el estimador de momentos. El resultado derivado indicó que una proporción estadísticamente insignificante (un 10%) de la variación en el residuo de la función de costes, era imputable a la variación en la eficiencia entre hospitales. Este resultado contrasta abiertamente con el obtenido al utilizar el modelo DFM. La frontera de coste fue igualmente estimada utilizando datos de panel, combinando la serie transversal de 49 hospitales con datos temporales para el período 1977-81. Los resultados fueron aquí menos concluyentes, aunque sugerían que, en términos generales, un tercio de la variación en el residuo de la función de coste podía atribuirse a la variación interhospitalaria en la eficiencia. Aparecía, asimismo, una variación significativa en la ineficiencia a lo largo de la muestra, que se cuantificaba en un 42% del coste medio. Cabe destacar que el MFD y la versión de datos de panel del MFE generaban rankings diferentes por lo que se refiere a la ineficiencia relativa: el coeficiente de correlación era sólo 0.53<sup>10</sup>.

#### 2.2.2.b. La eficiencia técnica.

La aproximación basada en la frontera de coste ofrece una estimación del grado de eficiencia de un centro o unidad productiva. No indica, sin embargo, qué parte de la ineficiencia estimada viene generada por la ineficiencia técnica y qué parte por la ineficiencia asignativa. Vamos por ello a centrarnos, a continuación, en aquellas aproximaciones específicamente diseñadas para estimar la ineficiencia técnica. Estas pueden dividirse entre las aproximaciones (i) paramétricas y (2) no paramétricas a la función de producción.

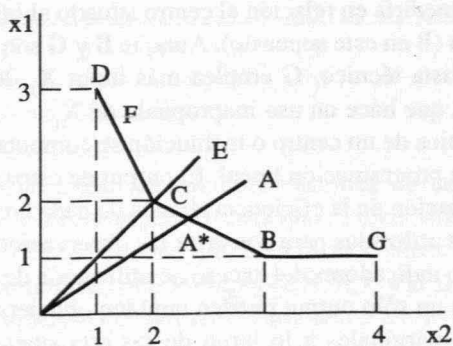
10. Dicho valor se basa no exactamente en la ineficiencia, si no en el porcentaje del coste actual en términos de los costes esperados.

### Aproximaciones no paramétricas.

La aproximación no paramétrica, referida en la investigación operativa y en general en la gestión como Análisis Envolvente de Datos (AED), tiene sus orígenes en el trabajo de Farrell (1957). Esta técnica puede ilustrarse con la ayuda del gráfico 2. Supongamos que existen siete centros (A... G), produciendo un único output ( $y$ ) a partir de dos inputs ( $X_1$  y  $X_2$ ). Sea éste el caso en el que los siete centros producen una misma *unidad de output*. En ausencia de información acerca de la ubicación de la isocuanta unitaria, no se posee criterio alguno para determinar si B C D son ineficientes desde el punto de vista técnico. La unidad B utiliza más  $X_1$  que la unidad C, pero menos  $X_2$ . Del mismo modo, la unidad D utiliza más input  $X_2$  que la unidad C, pero menos  $X_1$ . Existe base suficiente, sin embargo, para considerar que la unidad E es ineficiente: Utiliza una mayor cantidad de ambos inputs que el hospital C y no produce una mayor cantidad de output.

En otras palabras, para medir la eficiencia técnica de un centro desde esta aproximación, se requiere una estimación acerca de la ubicación de la isocuanta unitaria eficiente, que sirva como referencia para su evaluación. Farrell partió, para ello, de la constatación de que la isocuanta eficiente nunca puede tener pendiente positiva, siendo siempre convexa respecto del origen. La convexidad significa que si dos combinaciones de inputs pueden producir una unidad de output, también entonces pueden producirla una media ponderada de ambas.

GRAFICO 2





En términos del gráfico 2, ello significa que un centro puede, por ejemplo, operar en C o en D, o en cualquier otro punto del segmento CD, siendo posible separar, de este modo, las combinaciones de inputs en eficientes e ineficientes. Las eficientes serían aquellas que resultasen de seleccionar pares adyacentes de combinaciones, unidas por una línea. Si éste segmento tiene pendiente no positiva y no existe ninguna otra combinación en el mapa de isocuantas entre ésta y el origen, la combinación elegida se considera eficiente. Si éste no es el caso, la combinación se evalúa como ineficiente. Las combinaciones B y C, por ejemplo, se considerarían eficientes: la línea BC tiene pendiente negativa y no existe una combinación entre ésta y el origen. Nótese que, bajo este criterio, G resulta también eficiente: la línea BG tiene una pendiente no-positiva y no existe otra combinación entre ésta y el origen.

El segmento lineal que une a todas las combinaciones eficientes de inputs define el trazo de la isocuanta eficiente. Esta isocuanta *envuelve* todas las observaciones de centros ineficientes (tales como A): De ahí el término AED.

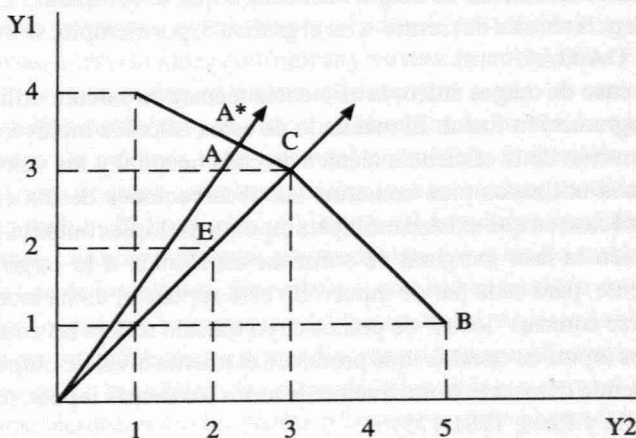
Nótese que mientras que los vértices (tales como B o C) representan observaciones reales de centros, puntos entre dichos vértices suponen observaciones hipotéticas de unidades hipotéticas, determinadas por las medias ponderadas de las combinaciones de inputs a partir de las observaciones existentes. El punto F, por ejemplo, se obtiene tomando 3/4 de la combinación de inputs contenida en D y 1/4 de la de C. La eficiencia técnica de los centros o unidades “dominados” por otras dos unidades (por ejemplo A), se mide entonces a través de la comparación de la utilización de inputs que hacen con respecto a la de una unidad hipotética que utiliza los inputs en idénticas proporciones. En el caso de A, la unidad hipotética relevante es  $A^*$  —una media ponderada de sus pares similares B y C— y su ineficiencia técnica se mide como  $OA^*/OA$ . En el caso de la eficiencia técnica de los centros “dominados” tan sólo por un único centro (caso de G), ésta se mediría en relación al centro situado al inicio de la porción plana de la isocuanta (B en este supuesto). Aunque B y G son ambos eficientes desde el punto de vista técnico, G emplea más input  $X_1$  del necesario y se considera, por tanto, que hace un uso inapropiado de  $X_1$ .

La eficiencia técnica de un centro o institución se computa, de este modo, a través de técnicas de programación lineal. El output de estos cálculos incluye, además de una estimación de la eficiencia técnica de cada hospital, los valores de las ponderaciones utilizadas para construir las observaciones de los centros hipotéticos, así como indicadores del exceso de utilización de cualquiera de los inputs. En el caso de un sólo output pueden también obtenerse estimadores de las productividades marginales a lo largo de las secciones relevantes de la isocuanta eficiente. En este sentido, para el supuesto de la unidad A del gráfico

2, podría verificarse: (i) que el centro hipotético  $A^*$  se sitúa  $5/7$  a lo largo del segmento  $BC$  (empezando desde  $B$ ), (ii) que la eficiencia técnica de  $A$  es  $6/7$  y (iii) que a lo largo del segmento  $BC$  las productividades marginales de  $X_1$  y  $X_2$  son  $1/6$  y  $1/3$ , respectivamente. El valor negativo del cociente  $(1/6)/(1/3) = (-) 1/2$  determina la tasa marginal de sustitución técnica de  $X_2$  por  $X_1$  a lo largo del segmento  $BC$ . Esta sería la tasa de sustitución que hubiera registrado la unidad  $A$  si hubiese actuado de una manera técnicamente eficiente. Si comparamos su valor con el cociente entre el precio de los inputs (cf ecuación (1)), puede determinarse también si la unidad  $A$  es eficiente desde el punto de vista asignativo (cf Charnes, Cooper y Rhodes, 1981).

La aproximación no paramétrica puede también ampliarse a la consideración de *múltiples outputs*.

GRAFICO 3



En el gráfico 3 se sitúan cuatro observaciones de centros o unidades productivas ( $A$ ,  $B$ ,  $C$  y  $D$ ) que producen cada uno de ellos, dos outputs ( $Y_1$  y  $Y_2$ ) con un único input ( $X$ ). Se supone que los cuatro centros utilizan la misma cantidad de  $X$ . Las combinaciones de output de  $A$ ,  $B$ ,  $C$  y  $D$  claramente difieren. Ello no indica, sin embargo, que sean necesariamente ineficientes. El centro  $B$ , al decidir producir más  $Y_1$  que el centro  $C$ , pero menos de  $Y_2$ , puede que esté reflejando, simplemente, el hecho de que la comunidad a la que sirve prioriza

$Y_1$  en relación a  $Y_2$ , a diferencia de la comunidad atendida por el centro C. Lo mismo es cierto para los centros C y D, pero nótese que no para el E: produce menos de ambos outputs que el centro C y utiliza igual cantidad de input.

Con múltiples outputs, la medición de la ineficiencia procede de la misma forma antes vista. Se separan las combinaciones de output ineficientes de aquellas eficientes a través de unir los pares de combinaciones con una línea. Si dicho segmento tiene pendiente no-positiva y ninguna de las demás combinaciones se sitúa en la zona noreste de éste, las combinaciones elegidas se declaran eficientes. Las combinaciones CD, por ejemplo, se considerarían eficientes: la línea CD tiene una pendiente negativa y no existe ninguna combinación al noreste de ésta. El punto A, por otro lado, es ineficiente. Si consideramos las medias ponderadas de las unidades eficientes (por ejemplo CD), la técnica AED genera un conjunto de combinaciones de unidades hipotéticas que “envuelven” a los centros ineficientes (por ejemplo A). Esta envolvente, que constituye de hecho la **frontera de posibilidades de producción (FPP)**, sirve entonces como *vara de medir* la combinación de outputs del centro que se compara<sup>11</sup>. En este sentido, la eficiencia técnica del centro A en el gráfico 3, por ejemplo, se mediría por el cociente  $OA/OA^*$ .

Como en el caso de output único, la eficiencia técnica se calcula utilizando métodos de programación lineal. El resultado de estos cálculos incluye, como antes, una estimación de la eficiencia técnica de cada hospital y los valores de las ponderaciones utilizadas para construir las observaciones de los centros hipotéticos. En el caso en que existen múltiples inputs y múltiples outputs, puede obtenerse también la tasa marginal de sustitución técnica a lo largo de la isocuanta eficiente, para cada par de inputs. En este supuesto, dicha isocuanta debe interpretarse como el “locus” de puntos cuyo trazado refleja las combinaciones de los dos inputs en cuestión que producen el mismo nivel de outputs del centro, manteniendo constante la utilización de todos los demás inputs. (cf, por ejemplo, Gravelle y Rees, 1981:175)

Los resultados de los cálculos efectuados ofrecen también los valores de las tasas marginales de transformación (TMT) entre cada par de outputs a lo largo de la sección relevante de la FPP<sup>12</sup>. La tasa marginal de transformación muestra la pendiente de la FPP a lo largo del segmento en cuestión, correspondiéndose

11. La frontera de posibilidades de producción indica la máxima cantidad que puede producirse de uno de los outputs, dada una cantidad determinada del otro output.

12. La tasa marginal de transformación indica la cantidad del output 2 a la que se debe renunciar para obtener una unidad adicional del output 1. Refleja, por tanto, el coste de oportunidad del output 1 en términos del output 2.

a la TMT que hubiera registrado el centro A si hubiera operado de una manera técnicamente eficiente. Indica, por lo tanto, las valoraciones relativas de los outputs implícitas en las decisiones del centro. Por ejemplo, si la TMT estimada entre  $Y_1$  y  $Y_2$  es igual a 2, ello indicaría que el centro está valorando implícitamente las unidades adicionales de  $Y_1$  como equivalentes a dos unidades de  $Y_2$ ; en otras palabras, se encuentra dispuesto a sacrificar 2 unidades de  $Y_2$  para obtener una unidad adicional de  $Y_1$ . Los gestores públicos y los contribuyentes pueden entonces confrontar aquellas valoraciones para determinar si se adecúan a sus propias prioridades<sup>13</sup>. Vale la pena señalar, en este contexto, que la técnica del AED postula que los centros no valoran sus inputs *per se*. Dicho extremo puede resultar controvertido en caso de existir evidencia de que determinadas instituciones públicas prefieran técnicas de producción trabajo-intensivas como medio de general empleo en su comunidad.

En el campo sanitario, Grosskopf y Valdamanis (1988) siguieron la aproximación no-paramétrica para comparar la eficiencia técnica de los hospitales públicos y los hospitales sin ánimo de lucro en los Estados Unidos. Utilizaron, para dicho propósito, cuatro outputs distintos: asistencia de agudos, cuidados intensivos, intervenciones quirúrgicas y asistencia ambulatoria y urgencias. Los dos primeros se miden en estancias, el tercero según las sesiones (sesiones ambulatorias incluidas) y el último en términos de visitas. Los inputs incluyen el número de médicos, otro personal no médico (expresado en término de equivalentes a tiempo completo) y los activos netos como medida del capital. La muestra incluía 22 hospitales públicos y 60 hospitales sin ánimo de lucro de California. Los hospitales públicos producían, como media, una cantidad mayor de cada uno de los outputs. Sin embargo, empleaban también mayor cantidad de todos los inputs, con la excepción del factor capital. A pesar de ello, se estimaba que los centros públicos, en promedio, mostraban una eficiencia técnica mayor respecto de los hospitales sin ánimo de lucro de la muestra: el nivel medio de eficiencia técnica de los hospitales públicos se estimaba en 0.942, mientras el de los hospitales sin ánimo de lucro era del 0.909.

Otras aplicaciones en el campo de la gestión pública se refieren al sector educativo (véase, en este sentido, Fare, Grosskopf y Weber, 1989). En España destaca el estudio de I. Vilardell (1989), relativo a la gestión de determinados servicios públicos locales (recolección y tratamiento de residuos urbanos).

13. Diversos autores (por ejemplo Lavers, 1972) han utilizado la programación lineal "inversa" para investigar este problema. La ventaja de utilizar la aproximación no-paramétrica, a diferencia de la programación lineal inversa, reside en que no postula que los centros sean eficientes desde el punto de vista técnico (cf Culyer, 1980:297-300).

### *Aproximaciones paramétricas.*

La aproximación paramétrica a la medida de la eficiencia técnica se basa en la estimación de la función frontera de producción (FFP). De modo similar a lo que veíamos en el caso de la función frontera de coste, difiere de la función de producción convencional en que restringe el error a valores no-positivos. La versión más simple de la FFP es el modelo de Frontera Determinística (MFD).

$$(5) \quad \ln Y_i = \beta_0 + \sum_j \beta_j \ln X_{ij} + u_i; \quad u_i \leq 0$$

Los centros o unidades pueden en este caso operar en o debajo de la frontera, pero no por encima. El alcance de la ineficiencia técnica viene indicado por el residuo de (5). El modelo se estima de modo similar a la frontera de coste MFD. Al igual que éste último, la frontera de producción MFD clasifica el residuo en su totalidad como ineficiencia. El modelo de función de producción frontera estocástica MFE contiene una mayor sofisticación

$$(6) \quad \ln Y_i = \beta_0 + \sum_j \beta_j \ln X_{ij} + v_i + u_i; \quad u_i \leq 0$$

Como en el caso de la frontera de coste MFE, la ineficiencia se captura en el error  $u_i$  y se mide, por lo tanto, en relación a la frontera estocástica ( $\beta_0 + \sum_j \beta_j \ln X_{ij} + v_i$ ). La estimación es básicamente similar al modelo frontera de coste MFE.

En el campo sanitario, Wagstaff (1987) utilizó esta aproximación para estimar la ineficiencia técnica para la muestra de 193 hospitales maternos utilizada en el estudio de Lavers y Whyne (1978). La especificación del componente estocástico se especificaba del mismo modo que en la ecuación (6), con el término de ineficiencia postulado como semi-normal. El modelo se estimó a través de técnicas de máxima verosimilitud (cf Greene, 1982), lo que requiere computar, en primer lugar, el estimador de momentos. El resultado derivado indicaba que aunque los residuos aparecían sesgados en la dirección equivocada para una frontera de producción (positiva, en lugar de negativa), el sesgo no era significativo. Ello significa que el estimador de Mínimos Cuadrados Ordinarios es de máxima verosimilitud y que no existe una variación transversal a lo largo de la muestra en la ineficiencia técnica.

### 2.2.2.c. Eficiencia asignativa.

En la mayor parte del trabajo empírico relativo al análisis de la ineficiencia asignativa, no se recoge intento alguno de cuantificar el coste de dicho tipo de ineficiencia<sup>14</sup>. Schmidt y Lovell (1979), en su aproximación "frontera" a la medida de la ineficiencia asignativa, persiguen precisamente este extremo<sup>15</sup>.

Considérese el caso de una función de producción Cobb-Douglas. Supongamos por un momento que los centros o unidades no son ni técnica ni asignativamente ineficientes. Cada centro estaría entonces minimizando sus costes, operando por lo tanto, en su frontera de costes<sup>16</sup>. En el caso de una minimización exacta de costes, la frontera de costes puede construirse a partir de la función de producción. La frontera de coste emerge como una solución al problema de la minimización restringida de la institución: El objetivo del centro consistiría en minimizar los costes de producir cada nivel de output, viniendo dada la restricción por la tecnología de producción (la función de producción). La función objeto a minimizar es, por lo tanto, la ecuación de coste.

$$(7) \quad C = \sum_i p_i X_i$$

C denota los costes totales,  $p_i$  el precio del input  $i$  y  $X_i$  refiere la cantidad de dicho input. Si la función de producción es Cobb-Douglas, la restricción en el problema de la minimización de costes viene dada por

$$(8) \quad \ln Y = \beta_0 + \sum_i \beta_i \ln X_i + v$$

14. Feldstein (1967) y López-Casasnovas y Wagstaff (1988) suponen excepciones a la afirmación anterior. Pero incluso sus estimaciones se refieren tan sólo al hospital medio de la muestra.

15. Eakin y Kniesner (1987) han sugerido, recientemente, una aproximación alternativa, que emplean para analizar la ineficiencia asignativa en una muestra de 331 hospitales de los Estados Unidos (véase el mencionado artículo para un estudio más detallado al respecto).

16. Recuérdese que los costes de un centro superaban su mínimo factible cuando operaban de manera ineficiente, ya sea técnica o asignativamente. Sólo en el supuesto de que se cumplan los requisitos de ambos tipos de eficiencia resultaba posible operar en la frontera de costes.

siendo  $v$  los *shocks* aleatorios y la perturbación estadística. Minimizar la ecuación de coste con la restricción expresada en la ecuación (8) genera la función de costes de la forma:

$$(9) \quad \ln C = K + (1/r)\ln Y + (1/r)\sum_i \beta_i \ln p_i - (1/r)v$$

siendo  $r = \sum_i \beta_i$  la expresión de los beneficios constantes a escala postulados y  $K$  una función de los parámetros de la función de coste<sup>17</sup>.

La importancia de la ecuación (9) reside en el hecho de que subministra una "vara de medir" con la que se puede evaluar la ineficiencia de los centros.

Supóngase, por ejemplo, que las elasticidades-output de cada uno de los inputs fuese conocida. Es posible calcular, entonces, a partir de la ecuación (9), los costes que corresponderían a un hospital que fuera 100% eficiente. En otras palabras, dadas las elasticidades output de cada uno de los inputs, es posible trazar la frontera de costes que se observaría si los hospitales minimizaran exactamente sus costes.

Supóngase ahora que los hospitales son eficientes desde el punto de vista asignativo, pero ineficientes en el sentido técnico. Supongamos además que su proceso de producción está sometido tanto a *shocks* aleatorios como a perturbaciones estadísticas. La minimización de costes en su caso, sujeta a la ecuación (5) genera la función de costes siguiente:

$$(10) \quad \ln c = K + (1/r)\ln Y + (1/r)\sum_i \beta_i \ln p_i - (1/r)[v-u]$$

El término  $(1/r)u$  mide el coste de la ineficiencia técnica: indica el porcentaje en el que los costes reales exceden la frontera de costes debido a la ineficiencia técnica.

La ecuación (10) modifica la frontera de costes básica al permitir la existencia de ineficiencia técnica. El paso siguiente es dar entrada a la ineficiencia asignativa. En el caso de la función de producción Cobb-Douglas, ello requiere que la igualdad en el cociente del gasto entre dos factores y el cociente de sus elasticidades-output respectivas se aplique a cualquier par de inputs, pudiendo reescribirse a este propósito, como:

17.  $K$  es efectivamente igual a  $\ln r - (1/r)\beta_0 - (1/r)\ln[\prod_i \beta_i]$  (véase Schmidt y Lovell, op cit: 347).

$$(11) \quad \beta_1/\beta_i = p_1 X_1 / p_i X_i \quad i = 2 \dots n$$

Si no se alcanza la eficiencia asignativa, la ecuación (11) no se cumple para ciertos pares de inputs. A efectos de considerar la posibilidad de ineficiencia asignativa, la ecuación (11) puede transformarse en

$$(12) \quad \ln p_1 X_1 - \ln p_i X_i + \ln(\beta_i/\beta_1) = \delta_i \quad i = 2 \dots n$$

de modo que  $\delta_i$  representa el coste que supone el que la condición de primer orden  $i$  para la eficiencia asignativa deje de cumplirse. Schmidt y Lovell (op cit) muestran como, en presencia de ineficiencia asignativa, la ecuación análoga a (9) es:

$$(13) \quad \ln C = K + (1/r)\ln Y + (1/r)\sum_i \beta_i \ln p_i - *(1/r)[v-u] + (E-\ln r)$$

siendo

$$(14) \quad E = \sum_j (\beta_j/r)\delta_j + \ln[\beta_1 + \sum_j \beta_j e^{-\delta_j}]$$

El valor de  $(E-\ln r)$  en la ecuación (14) indica el coste de la ineficiencia asignativa; es decir, el porcentaje en el que los costes reales superan la frontera de costes debido, precisamente, a la ineficiencia asignativa.

Nada se ha dicho, hasta el momento, acerca de la estimación de la ineficiencia en esta aproximación. Los problemas implícitos para este propósito se discuten ampliamente en Schmidt y Lovell (op cit). En gran parte, su tratamiento depende de (i) si se considera la eficiencia asignativa, y (ii) si éste es el caso, si se considera resultado de *errores* en la minimización de costes –de modo que  $y_i$  en la ecuación (12) puede ser tratada como una variable aleatoria con media cero–, o si se considera *deliberada* –por lo que  $y_i$  tendría media no nula. El caso más interesante aparece cuando la ineficiencia asignativa se considera deliberada y sistemática. Si éste es el caso –independientemente de que se postule su carácter sistemático o no– la aproximación seguida por Schmidt y Lovell consiste en



estimar la frontera de producción y las condiciones de primer orden, simultáneamente, para luego calcular los correspondientes parámetros de la frontera estocástica de costes (ecuación 13). La frontera de costes es efectivamente, por lo tanto, una frontera de coste sintética, dado que se “construye”, en vez de estimarse directamente.

Una aproximación alternativa más sencilla, sugerida por Schmidt (1986), consiste en estimar la frontera de producción estocástica (ecuación 5) y calcular, consecuentemente, los valores de  $\delta_i$  a partir de la ecuación 12. Los parámetros implícitos de la frontera estocástica de costes pueden calcularse entonces del mismo modo que en el método de Schmidt y Lovell ya comentado. El problema de seguir este tipo de aproximación radica, en caso de estimar una función Cobb-Douglas, en que la multicolinealidad no permite estimar los parámetros de la función frontera con suficiente precisión.

### 3. CONCLUSIÓN

Destaca quizás de la revisión efectuada de la literatura sobre la medición de indicadores de eficiencia, el grado de sofisticación econométrica que su estimación alcanza.

Su estudio pone en claro que, si bien resultan atractivas las estimaciones que ajustan en mayor medida por los costes incurridos, cabe enfatizar la necesaria consistencia exigible entre dichas estimaciones y la teoría microeconómica que las sustenta; ello parece ignorarse, en muchas ocasiones, en las aplicaciones prácticas.

Pese a la mayor complejidad que reviste su estimación, los métodos basados en la especificación de funciones frontera resultan claramente superiores, a efectos de evaluar las cuestiones de eficiencia en la gestión de los recursos, respecto de las estimaciones alternativas, basadas, comunmente, en comportamientos medios muestrales.

En caso de analizar funciones de coste –aproximación que suele resultar más factible en el estudio de la actividad de centros o unidades públicas–, las estimaciones que se basan en modelos estocásticos sirven mejor los objetivos que el análisis de eficiencia persigue. En particular, las formulaciones que utilizan la estimación de datos de panel, al ser capaces de identificar la parte de residuos que corresponde a influencias aleatorias y sistemáticas en la estimación de la ineficiencia, ofrecen un campo importante de investigación aplicada.

Esto parece especialmente cierto en el caso del análisis de la eficiencia en la gestión de los servicios públicos. En este contexto, la posibilidad de contar con

información recogida de modo sistemático en relación a centros o unidades que sirven a unos mismos objetivos y a lo largo del tiempo ofrece un ámbito de investigación apenas explorado en nuestro país. En los contextos aislados del mercado, dicha aproximación puede ser, por encima de todo, uno de los pocos controles factibles para una mejor asignación de los recursos presupuestarios.

En cualquier caso, la capacidad de cerrar el *gap* entre lo que suponen avances econométricos en las anteriores aproximaciones y la aplicabilidad de los métodos que han de acompañar su aplicación en la gestión operativa para la asignación eficiente de los recursos, constituye uno de los grandes retos de la economía aplicada en la actualidad. El avance en paralelo de ambos desarrollos ha de constituir, probablemente, en este sentido, su mejor validación.

#### BIBLIOGRAFÍA

- AIGNER, D.J. y S.F. CHU (1968). "On estimating the industry production function", *American Economic Review* 58, 226-239.
- AIGNER, D.J.; C.A.K. LOVELL y P. SCHMIDT (1977). "Formulation and estimation of stochastic production function models", *Journal of Econometrics* 6, 21-37.
- BANKER, R.D.; R.F. CONRAND y R.P. STRAUSS (1986). "A comparative application of Data Development Analysis and translog methods: an illustrative study of hospital production", *Management Science* 32, 30-44.
- BARLOW, R. (1968). "Review of Economic Analysis for Health Service Efficiency", *Economic Journal* 78, 921-923.
- BYRNES, P., R. FARE y S. GROSSKOPF (1984). "Measuring productive efficiency an application to Illinois strip mines", *Management Science* 30, 671-681.
- CHARNES, A.; COOPER, W.W. y RHODES, E. (1978). "Measuring the efficiency of decision-making units", *European Journal of Operational Research* 2, 429-444.
- CHARNES, A.; COOPER, W.W. y RHODES, E. (1981). "Evaluating program and managerial efficiency: an application of Data Envelopment Analysis to program follow through", *Management Science* 27, 668-697.
- CULYER, A.J.; R.J. LAVERS y A. WILLIAMS (1971). "Social indicators: health", *Social Trends* 2, 21-41.
- EAKEN, B.K. y T.J. KNIESNER (1982). "Estimating a non-minimum cost function for hospitals", *Southern Economic Journal*, 583-597.
- EVANS, R.G. y M.D. WALKER (1972). "Information theory and the analysis of hospital cost structure", *Canadian Journal of Economics* 5, 398-418.
- FARE, R.; S. GROSSKOPF y W.L. WEBER (1989). "Measuring School District Performance", *Public Finance Quarterly*, vol. 17, nº 4 Oct. (pp. 409-428).
- FARRELL, J.M. (1957). "The measurement of productive efficiency", *Journal of the Royal Statistical Society Series A* 120:253-266.

- FELDSTEIN, M.S. (1967). *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service* (North-Holland, Amsterdam).
- FUCHS, V.R. (1969). "Review of Economic Analysis of Health Service Efficiency", *Health Services Research* 3, 242-50.
- FORSUND, F.R.; C.A.K. LOVELL y P. SCHMIDT (1980). "A survey of frontier production functions and their relationship to efficiency measurement", *Journal of Econometrics*, 13-5-25.
- GRAVELLE, H. S.E. y R. REES (1981). *Microeconomics*, Longman, London.
- GREENE, W.H. (1980). "Maximum likelihood estimation of econometric frontier functions", *Journal of Econometrics* 13, 27-56.
- GREENE, W.H. (1982). "Maximum likelihood estimation of stochastic frontier production models", *Journal of Econometrics* 18, 285-289.
- GROSSKOPF, S. y V. VALDMANIS (1987). "Measuring hospital performance: a non-parametric approach", *Journal of Health Economics* 6, 89-108.
- JOHNSTON, J. (1984). *Econometric Methods*, (McGraw-Hill, London).
- JONDROW, J., C.A.K. LOVELL; I. MATEROV y P. SCHMIDT (1982). "On the estimation of technical inefficiency in the stochastic frontier production function model", *Journal of Econometrics* 19, 233-238.
- LAVE, J.R. y L.P. LAVE (1970). "Economic analysis for health service efficiency: a review article", *Applied Economics*, 1, 293-305
- LAVERS, R.J. (1972). "The implicit valuation of forms of hospital treatment", in: M.M. Hauser, ed., *The Economics of Medical Care*, Allen and Unwin, London.
- LAVERS, R.J. y D.K. WHYNES (1978). "A production function analysis of English maternity hospitals", *Socioeconomic Planning Sciences*, 12, 85-93.
- LÓPEZ CASASNOVAS, G. (1984). "The design of a budget-based contract as a tool for incentive motivation to improve efficiency in the allocation of resources in the health care sector, with especial reference to the public hospital sector in Spain", *unpublished DPhil dissertation*. Department of Economics, University of York.
- LÓPEZ CASASNOVAS, G. (1989). "Simulación de mercados y competencia en el sector público". *Papeles de Economía Española* nº 41. FIES.
- LÓPEZ CASASNOVAS, G. y A. WAGSTAFF (1988). "La combinación de los factores productivos en el hospital: una aproximación a la función de producción". *Investigaciones Económicas* 12, 305-327.
- LÓPEZ CASASNOVAS, G. y A. WAGSTAFF (1991). "Indicadores de eficiencia basados en la estimación de una función frontera estocástica de costes hospitalarios. Dirección General de Planificación Sanitaria". *Ministerio de Sanidad (mimeo)*.
- McGUIRE, A. (1987). "The measurement of hospital efficiency", *Social Science and Medicine* 24, 719-724.
- MEUSSAN, W. y J. VAN DEN BROECK (1977). "Efficiency estimation from Cobb-Douglas production functions with composed error", *International Economic Review* 18, 435-444.
- PITT, M.M. y L.F. LEE (1981). "The measurement and sources of technical inefficiency in the Indonesian weaving industry", *Journal of Development Economics* 9, 43-64.

- SCHMIDT, P. (1986). "Frontier production functions", *Econometric Reviews* 4, 89-328.
- SCHMIDT, P. y C.A.N. LOVELL (1979). "Estimating technical and allocative efficiency relative to stochastic production and cost frontiers", *Journal of Econometrics* 9, 343-366.
- SCHMIDT, P. y R.C. SICKLES (1984). "Production frontiers and panel data", *Journal of Business and Economic Statistics* 2, 367-374.
- TATCHELL, M. (1983). "Measuring hospital output: a review of the service-mix and case-mix approaches", *Social Science and Medicine* 17, 871-83.
- VILARDELL, I. (1989). "El análisis Envolvente de Datos y su aplicación a la gestión de los servicios públicos locales", *Tesis Doctoral leída en la Universidad Autónoma de Barcelona*.
- WAGSTAFF, A. (1987). "Measuring technical efficiency in the National Health Service: a stochastic frontier analysis", *Discussion Paper* 30 (Centre for Health Economics, University of York).
- WAGSTAFF, A. (1989a). "The measurement of hospital efficiency: a comment", *Social Science and Medicine* 28, 13-15.
- WAGSTAFF, A. (1989b). "Estimating efficiency in the hospital sector: a comparison of three statistical cost frontier models" *Applied Economics* 21, 659-672.