

H3
54
0474
1992

UNIVERSITAT POMPEU FABRA 

Indicadores de Eficiencia en el Sector Hospitalario

Guillem López
and
Adam Robert Wagstaff

Economics working paper 11. March 1992



Indicadores de Eficiencia en el Sector Hospitalario

Guillem López
Universitat Pompeu Fabra
y
Adam Robert Wagstaff
University of Sussex

Marzo 1991



1 Presentación

Las páginas que siguen constituyen la segunda y última parte del estudio "Elaboración de Indicadores para la Planificación de recursos sanitarios basados en criterios de eficiencia económica".

El presente trabajo consta de tres 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 utilizarán después en la construcción de los distintos indicadores. Finalmente, la tercera y última parte, procede a la estimación de indicadores para la planificación, a partir de la definición de funciones frontera de costes hospitalarios, respecto de las que se evalúa después la eficiencia en la utilización de los recursos sanitarios, para la muestra de centros INSALUD elegida a dichos efectos. Dicha aproximación se juzga, tras valorar las ventajas e inconvenientes de las alternativas existentes, como la más apropiada y consistente con el objetivo propuesto en la primera parte de esta investigación (véase G. Lopez y A. Wagstaff "Elaboración de Indicadores para la Planificación Sanitaria". Ministerio de Sanidad - AES (mimeo)). A los resultados obtenidos del análisis muestral efectuado, se añaden unas breves conclusiones que se derivan de su análisis.

1.2 Introducción

En la primera parte de esta investigación se habían analizado las alternativas existentes para la elaboración de indicadores para la planificación. Veíamos, a este respecto, que seguir un enfoque positivo y de carácter global parecía más coherente con los postulados de eficiencia económica requeridos (eficiencia en la asignación). Dado que éste no había sido hasta el momento el enfoque relevante en la planificación sanitaria, se imponía un mayor esfuerzo en analizar el grado de sustituibilidad entre factores productivos de acuerdo con criterios de economía en su asignación. Se trata con ello de buscar la combinación óptima de los recursos, a la vista de la polivalencia de las tareas efectuadas y el coste relativo de los factores productivos.

Dicho análisis reviste, como hemos dicho, carácter positivo y global, al mantener en favor de los centros gestores de gasto una esfera de autonomía en la gestión de los recursos, no forzando, con criterios normativos "ex ante", una determinada utilización de los recursos. En otras palabras, se trata con ello de que, una vez considerados los márgenes de sustitución técnica entre los distintos inputs sanitarios al servicio de unos mismos objetivos generales, posibilitar políticas descentralizadas de rentas y de empleo, dentro de los márgenes preestablecidos en las políticas envolventes de gasto.

Con el objetivo anterior, la primera parte del estudio analizaba la política de personal seguida en las dos últimas décadas por el Ministerio de Sanidad (Convocatoria de plazas, aumentos absolutos de plantillas, definición de especialidades, etc), así como las políticas salariales afectas. Dicha política se asociaba por otra parte, a indicadores de productividad- carga de trabajo asumida. La identificación de un efecto precio (ligada a

variaciones salariales) y un efecto cantidad (plantillas por niveles de trabajo asumido por tipo de personal equivalente), permitía mostrar la importancia, en términos de eficiencia económica, de un tipo u otro de planificación.

Nótese que, en cualquier caso, el objetivo de la investigación es derivar resultados que permitan aproximar la bondad con la que se utilizan los recursos sanitarios en el sector hospitalario en la actualidad, y los márgenes potenciales de mejora que se podrían producir en caso de incorporar los resultados derivados del análisis de la eficiencia económica a la planificación sanitaria.

En este sentido, se identifica aquí como una mejora en la utilización de los recursos, aquella reasignación de factores que ofrezca coeficientes de productividad relativa idénticos a sus costes relativos. Desde la óptica dinámica de un centro en particular, una asignación eficiente requeriría, además, que éste fuera desarrollando su actividad en la medida que, en la utilización de su capacidad, el "precio" iguale al coste marginal de producción a corto plazo, y expansionando ésta siempre que el coste marginal a largo plazo resultante sea inferior al "precio".

En la aproximación práctica a dicho propósito, en la primera parte de la investigación, se identificaron una serie de alternativas a la especificación de la función de producción/costes hospitalarios (López y Wagstaff, 1989 y Wagstaff, 1989). Los primeros resultados derivados de la estimación de las funciones Cobb-Douglas y Translogarítmica mostraban las virtualidades del enfoque propuesto, así como un conjunto de posibilidades de mejora en la formulación de los modelos y base estadística de aplicación (estimación de funciones frontera, con estimación de datos de panel, etc.).

Las páginas que siguen desarrollan dichas mejoras, analizando su relevancia a efectos de la planificación sanitaria.

2 Fundamentos teóricos del análisis

De acuerdo con la microeconomía convencional, los costes superan su nivel mínimo factible por una de las dos razones siguientes: O bien 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 que se está produciendo poco output, vista la combinación de factores empleados -lo que se conoce como ineficiencia técnica.

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 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 coeficiente de las productividades marginales entre los dos inputs¹.

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 de los dos inputs: $-p_2/p_1$, siendo p_i el precio de x_i . 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, de modo que la eficiencia en la asignación de los recursos requiere que

$$MP_2/MP_1 = P_2/P_1 \quad (1)$$

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; cuando esto suceda, (1) debe respetarse para cada par de inputs.

La denominada *eficiencia técnica* puede también mostrarse 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 un hospital en particular se determina, por tanto, exclusivamente 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 (es decir, muestren un nivel de eficiencia técnica superior) que otros. Las isocuantas y_A^0 e y_B^0 del gráfico 1 se asocian ambas a un mismo nivel de output.

Sin embargo, la isocuanta del hospital menos eficiente y_B^0 aparece situada más lejos del origen que la del hospital relativamente más eficiente y_A^0 . El hospital menos eficiente está, por lo tanto, utilizando más de ambos inputs para producir el mismo nivel de actividad que el hospital más eficiente. La ineficiencia técnica implica,

¹- 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.

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 escala², 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, $C1/C4$) y refleja el coste 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 asignación³.

Ambos tipos de ineficiencia pueden ser importantes en el sector sanitario. Por ejemplo, por lo que se refiere a la eficiencia asignativa, la combinación de categorías de profesionales sanitarios resulta bastante más amplia de lo que se acostumbra a postular en la planificación hospitalaria, en especial si se compara la utilización de recursos entre distintos sistemas sanitarios. La primera parte de ésta investigación ofreció evidencia a éste respecto. Sin embargo, que las substituciones potenciales de factores resulten efectivamente de interés dependerá de sus precios relativos. De ahí la importancia de las políticas salariales seguidas en el sector sanitario en los sistemas públicos de provisión. Cabe, por tanto, como objetivo de estudio en la planificación sanitaria, examinar si, a la vista de la productividad y nivel relativo de costes entre los distintos factores sanitarios (Cuadro 1), es posible diseñar pautas para una asignación más eficiente de los recursos sanitarios.

²- Beneficios constantes a escala implican que doblando todos los inputs se dobla el output.

³- 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.

CUADRO nº 1
COCIENTE DE COSTES ENTRE FACTORES PRODUCTIVOS
EN HOSPITALES 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

fuelle : Elaboración propia⁴.

Asimismo, en la medida en que se observan en el sector hospitalario niveles diversos de actividad (por ejemplo, en estancias medias, por tipos de pacientes y patologías) -véase la tabla 2, incluida más adelante-, pueden existir razones para considerar la posible presencia de ineficiencia técnica. En ambos supuestos se plantea, sin embargo, la conveniencia de su identificación y medida de modo sensible a las realidades sanitarias, a efectos de subministrar evidencia sobre la dirección de los cambios aconsejables en la planificación.

2.2 Aproximación práctica

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 predicciones muestrales (al estilo de las utilizadas por Feldstein, 1967) 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

⁴- 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.

las prácticas observadas.

Ambas aproximaciones se comentan a continuación.

2.2.1 Aproximaciones "no frontera"

Bajo esta óptica, la eficiencia global de un hospital puede evaluarse a través del examen de la función de producción, o de su dual de costes. Esta última no cabe identificarla con la ecuación de costes que, como se sabe, es una simple identidad contable, al definirse como un simple desglose de los componentes de coste (salarios, costes generales, de laboratorio, etc.). Una función de coste ha de ser consistente, a diferencia de la anterior, con los postulados de la teoría económica que la sustenta, expresando el coste en función de las variables que constriñen al hospital en sus intentos de minimizar costes. Dichas variables incluyen 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 del hospital es fijo-, la función de coste incluirá también dicho estoc, puesto que el carácter fijo de éste impide al hospital reducir sus costes a través de alterar su combinación factorial.

Pongamos por caso que la función de coste es de la forma

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

siendo c_i el coste hospitalario total o medio del centro i , la variable X_j refiere variables tales como el nivel de output y precios de sus inputs, y u_i el término de error. 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 en (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 National Health Service británico. 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 th 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 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 pueden por lo tanto interpretarse como el coste medio que hubiera observado un hospital con una combinación de casos particular si los costes específicos a su combinación de casos hubieran sido iguales a los de la media muestral. En efecto, un hospital con un residuo igual a cero registra en éste supuesto un coste medio que es exactamente la cantidad 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 hospitales 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 solo las variables "case mix" si no 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 variable⁵. Entre las diversas especificaciones estimadas por López Casanovas se incluyeron, así mismo, 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 Casanovas, 1984:134).

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

Por contra, la aproximación a funciones "no-frontera" 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 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 pueden reflejar también

⁵- Esta se corresponde con la ecuación estimada por Feldstein (1967) en su análisis de las economías de escala en el sector hospitalario.

perturbaciones estadísticas (cf Feldstein, 1967:50)⁶.

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 la noción de frontera: Son 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 si en una gran variedad de aspectos. (véase Forsund, Lovell and Schmidt, 1980; Schmidt, 1986).

Una de las diferencias se refiere a la interpretación del término "frontera". Algunos métodos, en este sentido, pretenden reflejar la frontera *absoluta*, como expresión del resultado en el caso de que la tecnología disponible se utilizara a plena ventaja. Otros métodos aspiran a identificar la frontera de la *práctica mejor* (Farrell, 1957), de acuerdo con 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. Esto 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), mientras 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 determinística (MFD) de Aigner y Chu (1968). La frontera de

⁶- El "ruido" o perturbación estadística recoge tanto errores en la medición como variables omitidas (cf Schmidt, 1986:340).

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 convierte ahora en

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

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 .

La 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).

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 factor docencia. El MFD se estimó por mínimos cuadrados corregidos. La ineficiencia media se estimó en 14740 ptas (precios de 1977) por caso ingresado, lo que equivalía a un 28.4 del coste medio para el año estudiado.

Pese a suponer sin duda una mejora respecto de la aproximación no frontera, el MFD resulta menos atractivo que la aproximación no paramétrica: se basa en una forma funcional específica y 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 ineficiencia⁷.

⁷- Una implicación de esta observación es que las dos aproximaciones producen idénticos "rankings" de eficiencia.

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

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

Aquí el término de error $v_i + u_i$ se compone de dos partes. La primera (v_i) es un componente doble (de dos lados), que captura los "shocks" aleatorios y el ruido estadístico. La segunda (u_i) refleja la ineficiencia (un solo lado), que se restringe a valores no negativos. 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 es común a todos los centros; la segunda (v_i), es un componente aleatorio que refleja el "ruido" estadístico y los "shocks" aleatorios, que varían entre hospitales. La ineficiencia del MFE se recoge en el término u_i y queda medida, por la tanto, en relación a la frontera estocástica

La aproximación más común, a efectos de diferenciar la ineficiencia de los "shocks" aleatorios y la perturbación estadística, se basa en examinar el sesgo, en su caso, de los residuos⁸. Si puede asumirse que el componente v_i del término de error se distribuye normalmente, el hecho de que la ineficiencia deba incrementar los costes significa que el conjunto del término de error haya de tener una distribución sesgada. En las aplicaciones se suele postular que la distribución del componente de la ineficiencia en el término del error es semi normal⁹. Bajo este supuesto resulta posible estimar el MFE a través de completar la información normalmente utilizada en la estimación del modelo regresional con la información 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). En cualquier caso, cualquiera que sea el método utilizado, se acaba considerando un residuo para cada hospital, una estimación de la media de u_i , pero 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 de estimación (cf Schmidt y Sickless, 1984) utiliza datos de panel. Se ofrece aquí tan sólo una breve presentación de dicho método, dado que su estudio pormenorizado se analizará en secciones subsiguientes. Cabe destacar, en cualquier caso,

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 Meussan y van den Breck (1977), es comunmente utilizado en el análisis aplicado.

En su lugar se supone que la ineficiencia permanece constante en el tiempo (aunque, naturalmente, puede variar entre hospitales). Ello significa que cambios en los costes hospitalarios a lo largo del tiempo no pueden deberse a cambios en la eficiencia. Una vía de solución consiste en eliminar el componente de ineficiencia no observable, a 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* ". Pero en ambos casos es posible computar la ineficiencia para cada hospital en particular de la muestra.

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 coste, era imputable a la variación en la eficiencia entre hospitales. Este resultado contrasta con el obtenido al utilizar el modelo DFM. La frontera de coste fue finalmente 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 significativa variación en la ineficiencia a lo largo de la muestra, ineficiencia que alcanzaba el 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 su ineficiencia relativa: el coeficiente de correlación de éstos era sólo 0.53¹⁰.

2.2.2.b La eficiencia técnica

La aproximación basada en la frontera de coste ofrece una estimación del grado de eficiencia de un hospital. 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 diseñadas para estimar la ineficiencia técnica. Estas pueden dividirse entre las aproximaciones (1) paramétricas y (2) no paramétricas.

Aproximaciones no paramétricas.

La aproximación no paramétrica, referida en la investigación operativa y en la ciencia de 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

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

(X_1 y X_2). Sea éste el caso en el que los siete hospitales 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. El hospital **B** utiliza más X_1 que el hospital **C**, pero menos X_2 . Del mismo modo, el hospital **D** utiliza más input X_2 que el hospital **C**, pero menos X_1 . Existe base suficiente, sin embargo, para considerar que el hospital **E** es ineficiente: Utiliza una mayor cantidad de ambos inputs que el hospital **C** y no produce un output mayor.

En otras palabras, para medir la eficiencia técnica de un hospital se requiere una estimación acerca de la ubicación de la isocuanta unitaria eficiente que sirva como referencia. Farrell partió, para ello 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 ambos. En términos del gráfico 2, ello significa que un hospital puede, por ejemplo, operar en **C** o en **D**, o en cualquier otro punto del segmento **CD**. Esos dos supuestos permiten separar las combinaciones de inputs en eficientes e ineficientes. Las eficientes serían aquellas que resultasen de seleccionar pares adyacentes de combinaciones, juntándolas con una línea. Si éste segmento tiene pendiente no-positiva y ninguna de las otras combinaciones en el mapa de isocuantas existe 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* todos los hospitales ineficientes (tales como **A**): De ahí el término AED.

Nótese que mientras que los vértices (tales como **B** o **C**) representan observaciones reales de hospitales, puntos entre dichos vértices suponen observaciones hipotéticas de hospitales hipotéticos, determinadas por las medias ponderadas de las combinaciones de inputs a partir de los centros 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 hospitales "dominados" por dos hospitales (por ejemplo **A**), se mide entonces a través de la comparación de la utilización de inputs que hacen con respecto a la de un hospital hipotético que utiliza los inputs en idénticas proporciones. En el caso de **A**, el hospital hipotético 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" por un sólo hospital (caso de **G**), ésta se mediría en relación al hospital situado al inicio de la porción plana de la isocuanta utilizado en la (**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 hospital se computa 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 del hospital A del gráfico 2 puede verificarse que el centro hipotético A* se sitúa 5/7 en el segmento BC (empezando desde B), que la eficiencia técnica de A es 6/7 y que a lo largo del segmento BC las productividades marginales de X₁ y X₂ son 1/6 y 1/3, respectivamente. El cociente $(-)(1/6)/(1/3) = (-)1/2$ determina la tasa marginal de sustitución técnica de X₂ por X₁ a lo largo del segmento BC. Esta sería la tasa de sustitución que hubiera registrado el hospital 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 el hospital 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*. En el gráfico 3 se sitúan cuatro observaciones de centros hospitalarios (A, B, C y D) que producen cada uno de ellos, dos outputs (Y₁ y Y₂) con un único input (X). Se supone que los cuatro centros utilizan la misma cantidad de X. Las combinaciones de output de los centros A B C y D claramente difieren. Ello no indica, sin embargo, que sean necesariamente ineficientes. El hospital B, al decidir producir más Y₁ que el hospital C, pero menos de Y₂ puede que esté reflejando simplemente el hecho de que la comunidad a la que sirve, prioriza Y₁ en relación a Y₂, a diferencia de la comunidad atendida por el hospital C. Lo mismo es cierto para los hospitales C y D, pero no para el hospital E: produce menos de ambos outputs que el hospital C y utiliza igual cantidad de input.

Con múltiples outputs, la medida de la ineficiencia procede de la misma forma que la vista anteriormente. 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 autoridades eficientes (por ejemplo C D), la técnica AED genera un conjunto de combinaciones de centros hipotéticos que "envuelven" a los hospitales 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 hospital que se compara¹¹. En este sentido, la eficiencia técnica del hospital A en el gráfico 3, por ejemplo, se mediría por el cociente OA/OA*.

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.

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 múltiples inputs existen a la vez que múltiples outputs, puede obtenerse también, para cada par de inputs, la tasa marginal de sustitución técnica a lo largo de la isocuanta eficiente. 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 hospital, manteniendo constante la utilización de todos los demás inputs. (cf, por ejemplo, Gravelle y Rees, 1981:175).

Los resultados de los cálculos efectuados ofrece también los valores de las tasas marginales de transformación (TMT) entre cada par de outputs a lo largo de la sección relevante de la FPP¹². La tasa marginal de transformación ofrece la pendiente de la FPP a lo largo del segmento en cuestión y es la TMT que hubiera registrado el hospital A si hubiera operado de una manera técnicamente eficiente. Indica, por lo tanto, las valoraciones relativas de los outputs implícitas en las decisiones del hospital. Por ejemplo, si la TMT estimada entre Y_1 y Y_2 es igual a 2, ello indicaría que el hospital 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 . Las autoridades sanitarias y los contribuyentes pueden entonces confrontar aquellas valoraciones para determinar si se adecúan a sus propias prioridades¹³. Vale la pena señalar en este contexto, que la técnica del AED postula que los hospitales no valoran sus inputs "per se". Sin embargo, puede suceder que algunos centros prefieran técnicas de producción trabajo-intensivas como medio de generar empleo en su comunidad.

Grosskopf y Valdamanis (1988) utilizan la aproximación no-paramétrica para comparar la eficiencia técnica de los hospitales públicos y los sin ánimo de lucro en los Estados Unidos. Utilizaron para dicho propósito cuatro outputs: 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 incluye 22 hospitales públicos y 60 hospitales sin ánimo de lucro de California. Los hospitales públicos producen, como

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.

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 hospitalarios sean eficientes desde el punto de vista técnico (cf Culyer, 1980: 297-300).

media, más de cada uno de los outputs. Sin embargo, emplean también más de cada uno de los inputs, con la excepción del factor capital. A pesar de ello, se estima para ellos en promedio, una eficiencia técnica mayor respecto de los hospitales sin ánimo de lucro en 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.

Aproximaciones paramétricas.

La aproximación paramétrica a la medida de la eficiencia técnica centra su atención en la función frontera de producción (FFP). Difiere, por tanto 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)

$$\ln Y_i = \beta_0 + \sum_j \beta_j \ln X_{ij} + u_i \quad u_i \leq 0 \quad (4)$$

Los hospitales pueden en este caso operar sobre o debajo de la frontera, pero no por encima. El alcance de la ineficiencia técnica viene indicado por el residuo de (4). 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

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

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.

Wagstaff (1987) utilizó esta aproximación para estimar la ineficiencia técnica para la muestra de 193 hospitales maternos utilizada en el estudio de Lavers y Whytes (1978). La especificación del componente estocástico se especificaba del mismo modo que en la ecuación (5), con el término de ineficiencia postulado como semi-normal. El modelo se estimó a través de máxima verosimilitud (cf Greene, 1982), lo que requiere computar, en primer lugar, el estimador de momentos. En este estadio 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 ineficiencia¹⁴. Schmidt y Lovell (1979), en su aproximación "frontera" a la medida de la ineficiencia asignativa, persiguen precisamente este extremo¹⁵.

Considérese el caso de una función de producción Cobb-Douglas. Supongamos por un momento que los hospitales no son ni técnica ni asignativamente ineficientes. Cada hospital estaría entonces minimizando sus costes, operando, por lo tanto, en su frontera de costes¹⁶. 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 del hospital: El objetivo del hospital consistiría en minimizar los costes de producir cada nivel de output, siendo la restricción la tecnología de producción (la función de producción). La función objetivo a minimizar es, por lo tanto, la ecuación de coste.

$$C = \sum_j p_j X_j \quad (6)$$

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

$$\ln Y = \beta_0 + \sum_j \beta_j \ln X_j + v \quad (7)$$

Siendo v los "shocks" aleatorios y la perturbación estadística. Minimizar la ecuación de

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 hospital superaban su mínimo factible si operaban de manera ineficiente, ya sea técnica o asignativamente. Sólo en el supuesto de ambos tipos de eficiencia era posible operar en la frontera de costes.

coste con la restricción de la ecuación (7) genera la función de coste de la forma

$$\ln C = K + (1/r) \ln Y + (1/r) \sum_j \beta_j \ln p_j - (1/r) v \quad (8)$$

siendo $r = \sum_j \beta_j$ la expresión de los beneficios constantes a escala postulados y K una función de los parámetros de la función de coste¹⁷.

La importancia de la ecuación (8) reside en el hecho de que suministra una "vara de medir" con la que se puede evaluar la ineficiencia de los centros. Supóngase que las elasticidades-output de cada uno de los inputs fuese conocida. (Los problemas asociados a su estimación se discuten más adelante). Es posible calcular, entonces, a partir de la ecuación (8), 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 sus costes exactamente.

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 (4) genera la función de coste de la forma siguiente

$$\ln C = K + (1/r) \ln Y + (1/r) \sum_j \beta_j \ln p_j - (1/r) [v-u] \quad (9)$$

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 en razón a la ineficiencia técnica.

La ecuación (9) modifica la frontera de costes básica al permitir 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

$$\beta_1 / \beta_j = p_1 X_1 / p_j X_j \quad j=2 \dots n \quad (10)$$

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

Si no se alcanza la eficiencia asignativa, la ecuación (10) no se cumple para ciertos pares de inputs. A efectos de considerar la posibilidad de ineficiencia asignativa, la ecuación (10) puede transformarse en

$$\ln p_1 X_1 - \ln p_j X_j + \ln(\beta_j/\beta_1) = \delta_j \quad j = 2 \dots n \quad (11)$$

de modo que δ_j representa la cuantía por la cual la condición de primer orden para la eficiencia asignativa deja de cumplirse. Schmidt y Lovell (op cit) muestran como, en presencia de ineficiencia asignativa, la ecuación análoga a (9) es

$$\ln C = K + (1/r) \ln Y + (1/r) \sum_j \beta_j \ln p_j - (1/r) [v-u] + (E - \ln r) \quad (12)$$

siendo

$$E = \sum_j (\beta_j/r) \delta_j + \ln[\beta_1 + \sum_j \beta_j e^{-\delta_j}] \quad (13)$$

El valor de $(E - \ln r)$ en la ecuación (13) 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, por 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 que di en la ecuación (11) puede ser tratada como una variable aleatoria con media cero- o si se considera deliberada -por lo que di tendría media no nula. El caso más interesante es cuando la ineficiencia asignativa se considera *deliberada y sistemática*. Si la ineficiencia asignativa se considera -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 12). 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 d_i a partir de la ecuación 11. 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 suele radicar, en el caso de estimar una función Cobb-Douglas, en que la multicolinealidad no permite estimar los parámetros de la función frontera con suficiente precisión.

2.3 Aplicación

En resumen, y a la vista de la revisión efectuada, tal como apuntábamos anteriormente, se ha optado en el ejercicio empírico que se reseña a continuación, por **estimar la función frontera de costes de los hospitales públicos del INSALUD** incluidos en la muestra. Dicha función se define como el "locus" de referencia para el estudio de la eficiencia con la que se utilizan los recursos sanitarios. Recuérdese que el objetivo primordial de esta investigación consiste en evaluar el grado de eficiencia de los centros hospitalarios como información substantiva previa a cualquier ejercicio de planificación que pretenda corregir, geográfica o funcionalmente, la distribución de los recursos hospitalarios. Se trata con ello de averiguar qué centros registran niveles particularmente bajos de ineficiencia, si ésta aparece de modo sistemático para determinados hospitales en el tiempo, con signos de reducción o ampliación en relación al resto de centros (existiendo o no, por tanto, efectos de causación acumulativa), a efectos de incorporar dicha evaluación en la planificación de los recursos sanitarios en la práctica.

A dicho propósito se dedican finalmente, la tercera parte de este estudio.

3 Método, Modelo, Estimación, Resultados y Análisis

Una primera ojeada a los datos sobre los que se pretendió apoyar la estimación pone en evidencia que existen diferencias substanciales en los indicadores económicos -en particular en el coste medio por caso-, entre las casi 50 Residencias Sanitarias de propiedad y gestión del INSALUD (todas ellas de más de cien camas y con oferta sanitaria no distorsionada por la proximidad de Ciudades Sanitarias) que constituye la muestra seleccionada para el análisis. Así, el coste medio por enfermo ingresado (1986) de Virgen del Puerto de Plasencia, por ejemplo, era de 101 mil ptas. Por contra, el coste medio por caso en Nuestra Sra. del Pino de Las Palmas de Gran Canaria era de 376 mil ptas. Sería absurdo concluir de esta evidencia parcial que el elevado coste por caso registrado en este último centro es el resultado de su ineficiencia. Dicha diferencia puede ser también debida, al menos en parte, al hecho de que su "case mix" ofrece una mayor complejidad y, por tanto, es más costoso. En efecto (véase Tabla 1), a diferencia de la mayoría de hospitales agudos en España, el hospital de Las Palmas parece tratar un número substancial de tipologías de enfermos poco usuales en los servicios de digestivo y respiratorio. Esto parece para el año en curso de una simple observación de la

información existente (Tabla 1). Asimismo, parte de la diferencia en el coste puede deberse también al hecho de que, ubicado en una isla alejada de la península, Nuestra Señora del Pino afronta unos mayores costes al ser un hospital universitario, a diferencia del de Virgen del Puerto. En otras palabras, influencias sistemáticas pueden explicar diferencias en los costes observados entre los distintos hospitales que componen la muestra. Estas deben tenerse en cuenta a la hora de inferir los niveles de ineficiencia.

Por encima de estas influencias sistemáticas en los costes, y tal como veíamos anteriormente, pueden existir también influencias aleatorias. Los hospitales pueden, por ejemplo, registrar un mal ejercicio. Cabe que operen, a su vez, factores que, sin reflejarse en la "case mix" o complejidad, estatus docente, etc. del hospital, influyan en el coste medio del centro. En definitiva, es deseable considerar esas influencias aleatorias además de las influencias sistemáticas que se puedan registrar para el análisis subsiguiente de la eficiencia hospitalaria.

Hemos visto en la sección precedente como la noción de función frontera de coste ofrece una manera de distinguir entre ineficiencia e influencias sistemáticas y aleatorias en los costes. En el ejercicio de estimación que a continuación se realiza, se emplea una versión del modelo frontera de costes que utiliza datos de panel o longitudinales, a efectos de desentrañar los anteriores determinantes de coste. Se considerarán las influencias sistemáticas, a través de especificar una función de costes hospitalarios que exprese el coste por caso como función del estoc de camas, complejidad de su "case mix", duración de la estancia media, etc. Las influencias aleatorias se considerarán igualmente al explotar la dimensión temporal y transversal de los datos de panel: básicamente, la idea implícita es la de que, mientras los "shocks" aleatorios pueden afectar adversamente a la actividad del hospital en un año cualquiera, las influencias aleatorias deben cancelarse entre si. En contraste con ello, la ineficiencia se presume que sí varía entre los hospitales, pero no a lo largo de los años. Sólo después de los anteriores ajustes, el residuo resultante, una vez considerados los efectos sistemáticos y aleatorios que influyen en el coste de los hospitales, se identifica, en la aproximación a la frontera de coste, como ineficiencia.

A continuación se ofrece, de un modo específico, la definición del modelo utilizado y los datos que se han hecho servir en la estimación.

3.1 Datos y definición del modelo

Para la estimación concreta de la eficiencia con la que operan los hospitales públicos españoles se emplearon datos relativos a 38 centros, listados en la tabla 1, para los años 1982-1986. Se tuvieron que excluir algunos hospitales de la muestra inicial al ser imposible contar con datos temporales completos para todas las variables de interés. La tabla 1 ofrece también diversos detalles acerca de dichos centros, para el último de los años estudiados.

La columna relativa a los costes totales "tot cst" cifra los gastos de hospitalización, en miles de ptas.. La columna referida como "av cst" contiene el coste por enfermo tratado,

y es igual al cociente entre los costes totales y el número de casos. Una simple inspección de "av cst" revela variaciones substanciales en el coste por caso entre los 38 hospitales, tal como señalábamos anteriormente. Nótese que buena parte del interés de esta investigación es identificar que parte de dichos costes pueden justificarse.

En la tabla 1 aparecen también datos sobre el tamaño del hospital (medido en términos del estoc de camas), el coeficiente de utilización, duración en días de la estancia media, así como información sobre su estatus docente (los hospitales universitarios aparecen codificados con un 1). Pese a que la muestra se seleccionó con criterios estrictos de homogeneidad "ex ante"¹⁸, pueden observarse importantes variaciones en las tres primeras variables: el tamaño de los centros va desde las 102 camas del hospital de Valdepeñas hasta las 658 camas del hospital de Valladolid; la tasa de utilización abarca desde el 52% de Virgen de la Luz de Cuenca, hasta el 93% de Valdepeñas, y la estancia media, desde los 7.2 días de Virgen del Monte Toro de Mahón hasta los 14.4 días de Nuestra Sra. del Pino de Las Palmas.

La tabla 1 presenta, finalmente, datos acerca de la "case mix" de los 38 hospitales. Resulta inmediatamente aparente de la Tabla que existen ciertos tipos de patologías -particularmente digestivo y respiratorio, electro, hematología y neurología- que se dirigen a unos hospitales en concreto, mientras que el resto de casos son tratados por todos los hospitales. De acuerdo con los postulados de la Teoría de la Información, que se analizan más adelante, ello indicaría que los primeros hospitales podrían clasificarse como relativamente más complejos que los segundos -al registrar estos últimos una mayor proporción de tipologías más comunes tales como pediatría, etc. Hospitales como Nuestra Sra del Pino de Las Palmas recibirían, además, una mayor valoración de su complejidad a la vista de su tipología en conjunto más compleja, y en relación a otros hospitales como Valdepeñas, para el que su actividad consiste, en mayor medida en el tratamiento de categorías de casos menos complejos. Ello viene reflejado en los valores de la columna final (TI): se trata de un índice basado en la teoría de la información, de finido como la complejidad de los casos Ω^n ¹⁹.

Tal como apuntábamos en la primera parte de ésta investigación, la estimación efectuada

¹⁸- En efecto, en su selección se evitó la posibilidad de que su oferta sanitaria estuviese distorsionada por la proximidad geográfica de hospitales maternos o por su inclusión en complejos hospitalarios más amplios, como en el caso de las Ciudades Sanitarias. Además se contó para la selección con la experiencia de staff del Ministerio, y en particular de Víctor Conde Rodelgo, a quien se desea agradecer su colaboración.

¹⁹- Esta aproximación fue inicialmente propuesta por Evans y Walker (1977). En TI, una categoría de casos se considera de baja complejidad si puede tratarse en cualquier hospital, y de alta complejidad si sólo en unos pocos. Cada categoría de casos recibe una puntuación sobre la base del alcance en el que la proporción real de dichos diagnósticos. La productividad o producto marginal de x_1 se define como el output adicional obtenido cuando se aumenta en una unidad la utilización de x_1 , manteniendo constante la del resto de inputs.

se basa en el modelo de *frontera de costes* de Schmidt y Sickless (1984), empujando *datos longitudinales o de panel*. Recuérdese que en esta aproximación no se requiere hipótesis alguna acerca del término que recoge el nivel de ineficiencia. En su lugar, se supone que la ineficiencia permanece constante en el tiempo (aunque, naturalmente, puede variar entre hospitales), a la vez que los "shocks" aleatorios se compensan en el tiempo. Ello significa que los "shocks" aleatorios pueden empujar a un hospital fuera de su frontera en cualquier ejercicio, pero dado que estos "shocks" se han de compensar en el tiempo, no pueden mantener constantemente a un hospital por encima de la frontera de coste. Al utilizar valores medios en el tiempo, la aproximación basada en datos de panel puede de este modo distinguir la ineficiencia de los "shocks" aleatorios.

Supóngase que cada uno de los H hospitales incluidos en la muestra es analizado a lo largo de T años. La función de coste puede escribirse:

$$c_{ht} = \beta_0 + \sum_j \beta_j X_{hj t} + v_{ht} + u_h \quad u_h \geq 0 \quad (14)$$

t refiere al período de tiempo considerado, siendo en este caso $i=h$ (hospital) en relación al análisis teórico anterior. Nótese que el término de ineficiencia u no tiene referencia temporal, lo cual refleja el postulado de que la ineficiencia varía entre los hospitales (h) pero es la misma en un hospital dado en cada uno de los años. La ecuación (14) contiene un error sesgado y tiene una media no nula. Sin embargo, al transformar la ecuación se obtiene otra similar a las utilizadas en los modelos de datos de panel. Esto es

$$c_{ht} = \beta_0^* + \sum_j \beta_j X_{hj t} + v_{ht} + u_h^* \quad (14')$$

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

Tal como vamos anteriormente en la aproximación a la ineficiencia basada en datos de panel, ésta 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 de las caracterizaciones, 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 externos ("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 ineficiencia²⁰.

La elección entre las dos caracterizaciones de la ineficiencia se limita 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 coste incluye variables que permanecen constantes a lo largo del tiempo, como suele ser el caso, por ejemplo, del estatus docente de un centro [véase Schmidt y Sickless, 1984:369]. En este supuesto, la caracterización de la ineficiencia como aleatoria debe ser adoptada en el modelo y estimarse a través de MCG de acuerdo 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 hospital 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 $\Sigma_{ht} = c_{ht} - \Sigma_j b_j X_{hjt}$ siendo b_j el estimador de β_j .

A la vista de la ecuación (14) resulta claro que estos residuos reflejan v_{ht} y u_h , así como β_0 . Sin embargo, al promediar los e_{ht} en el tiempo, v_{ht} desaparece y se obtiene el resultado

$$b_{oh} = (1/T) \sum_t X_{ht} = \beta_0 + u_h \quad (15)$$

de modo que los residuos promediados pueden utilizarse como una estimación de la suma de β_0 y u_h . Schmidt y Sickless sugieren entonces estimar β_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, tal como antes éste se definió²¹.

²⁰- Pitt y Lee (1981) y Schmidt y Sickless (op cit) consideran las ventajas e implicaciones de imponer hipótesis sobre las distribuciones de u y v .

²¹- A efectos de validar la variación en la ineficiencia a lo largo de la muestra, Wagstaff (1989) propone la utilización del test del multiplicador lagrangiano: dado que σ_u^* (la

3.2 Especificación y estimación del modelo

Se detalla a continuación la especificación efectuada de la función de costes hospitalarios. Tal como se señalaba en la sección 2.2., aunque el análisis de regresión múltiple puede utilizarse tanto para la estimación de la ecuación como de la función de costes, una investigación de la naturaleza aquí propuesta puede basarse tan sólo en la *función de costes*. Ello se debe a que, a través de una ecuación de costes resultaría posible única y exclusivamente evaluar la causa por la que el hospital A tiene un coste medio superior al hospital B (por ejemplo, debido a un empleo mayor de mano de obra por enfermo tratado). Sin embargo, en nuestro caso, el interés se centra en conocer si los mayores costes del hospital A pueden justificarse en el mayor output que produce y/o la elevada complejidad de su actividad. Para ello necesitamos una función de costes. La sugerencia de algunos autores (véase, por ejemplo, Straf, 1981) de que la investigación en esta área se beneficiaría de incluir variables tales como las que miden los factores productivos utilizados, deberían ser, por lo tanto, firmemente resistida. (cf Culyer y otros, 1982).

3.2.1 La medida del output hospitalario

Como hemos discutido ya, una de las variables (o conjunto de variables) que ha de incluirse en la parte sistemática de la frontera de coste es el output. Cabe determinar, por lo tanto cuál es el output de un hospital.

Qualys y valor añadido.

El objetivo al que se aspira con la actividad de un hospital es mejorar la calidad y/o duración de la vida de los pacientes por encima de la que resultaría en ausencia de tratamiento. Tiene sentido, por tanto, definir el output del hospital, a nivel conceptual al menos, como un valor añadido, resultado de los años de vida ganados ajustados por

varianza de u^*) y σ_u (la varianza de u) son iguales, la hipótesis nula $H_0: \sigma_u^2 = 0$, puede validarse utilizando el estadístico

$$LM = \frac{NT}{2(T-1)} \left(\frac{\sum_h (\sum_t e_{ht})^2}{\sum_h (\sum_t e_{ht}^2)} - 1 \right)^2$$

siendo e_{ht} los residuos a través de la estimación por mínimos cuadrados ordinarios para el modelo restringido

$$C_{ht} = \beta_0 + \sum_j \beta_j X_{hjt} + E_{ht}$$

Bajo la hipótesis nula

$$H_0: \sigma_u^2 = 0, LM \approx \chi^2(1)$$

(cf, por ejemplo, Breusch y Pagan, 1980; Judge y otros, 1985:526).

su calidad (QALYs o "quality-adjusted life years") [cf Culyer y otros (1971); Williams (1985). El output de un episodio de tratamiento en particular sería de este modo el número de QALYs ganados como resultado del tratamiento.

Ajuste por la combinación de patologías

Aunque la aproximación empírica en este campo es ya importante, éste se encuentra todavía en su infancia (véase, por ejemplo, Williams (op cit); Gudex, 1986). En efecto, los esfuerzos se concretan todavía en la estimación de la función de coste utilizando medidas de output intermedio ("throughput") como variable aproximativa al verdadero output del hospital. La medida más comunmente utilizada para ello es el número de casos tratados, así como el número total de días de estancia. Tal como apunta Breyer (1987), la preferencia por la primera sobre la segunda de las medidas se debe, probablemente a su más estrecha relación con el verdadero output del hospital -mejora en el estado de salud. En aquellos supuestos en que los enfermos tratados son relativamente homogéneos (como en el caso de los hospitales maternos, por ejemplo), utilizar el número de pacientes ingresados como medida del output es probablemente satisfactorio. Sin embargo, cuando las patologías tratadas por un hospital no son homogéneas (caso de la mayoría de centros de agudos y centros de práctica dental), el supuesto implícito de asignar ponderaciones o valoraciones iguales a cada caso resulta mucho menos atractivo.

La solución de Feldstein (1967) consistió en dividir los enfermos en grupos o categorías de "case mix", a la vista del departamento de admisión, y considerar posteriormente variaciones inter-hospitalarias en "case mix", ya sea a través de ponderar cada categoría por su coste medio esperado o a través de la introducción explícita del propio vector de las proporciones de "case mix" como regresor en la función de coste. La primera aproximación se basa en el supuesto bastante heroico de que el coste por caso puede considerarse una primera aproximación a su valoración marginal social; la segunda evita, en cierta medida, el problema de asignar ponderaciones, estando a resultas de las derivadas de la propia estimación.

En la presente investigación sobre el sector hospitalario español, se ha empleado la segunda de las aproximaciones sugeridas por Feldstein. Ya López-Casasnovas (1984) había utilizado, para su análisis inicial, seis categorías amplias de "case mix": medicina interna, cirugía, ginecología, pediatría, cuidados intensivos, y otros casos (éste último con carácter residual). Este tipo de aproximación no resulta, sin embargo, enteramente satisfactoria. En primer lugar, la introducción de un vector de variables "case mix" como regresor en la estimación de la función origina serios problemas de multicolinealidad²².

²²- Un análisis de los factores generadores de colinealidad de la estimación, basado en el indicador $CN = (\mu_k / \mu_1)^{1/2}$, siendo μ_k y μ_1 , respectivamente, el mayor y menor de los valores propios de la matriz $X'X$, señalaba al "case-mix" como la variable colineal principal. La restricción de rendimientos constantes a escala, en su caso, reducía el
(continued...)

En segundo lugar, no afronta el problema de las variaciones intra-categoría en la gravedad de los casos: no todos los pacientes ingresados en cirugía están igualmente enfermos. En tercer lugar, si se utiliza un pequeño número de categorías de "case mix", se corre el riesgo de ignorar algunas de las variaciones inter-hospitalarias en el "case-mix" (Barlow, 1968; Fuchs, 1969; Lave y Lave, 1970; Tatchell, 1983).

Teoría de la Información.

En la estimación aquí efectuada utilizamos la aproximación basada en la Teoría de la Información, tal como proponen Evans y Walker (1972), de un modo alternativo al ajuste tipo Feldstein realizado en estudios similares sobre idéntica base muestral. Como se ha comentado ya, la TI genera una medida del output hospitalario basado en la complejidad estimada de tratar diferentes categorías de patologías. En los trabajos empíricos, éstas se han definido en términos diversos, tales como los basados en los Grupos Relacionados por el Diagnóstico (DRG) o de acuerdo con la Clasificación Internacional de Enfermedades (ICD). Dada la falta de mejores datos, en nuestro trabajo hemos tenido que utilizar, en su lugar, categorías departamentales o por especialidades. Tal como apuntábamos antes, en la aproximación TI una tipología de casos se considera de baja complejidad si puede tratarse en cualquier hospital, y de elevada complejidad si sólo se puede tratar en unos pocos. A cada categoría de diagnósticos se le asigna un indicador de complejidad sobre la base de la información ganada al conocer que la proporción real de diagnósticos tratados en un hospital en particular difiere de la proporción que se hubiera registrado si aquellos se hubieran distribuido de modo idéntico entre todos los hospitales. El índice de complejidad de un hospital es, por lo tanto, la suma ponderada del indicador relativo de complejidad de las diversas categorías de diagnóstico, siendo sus ponderaciones las proporciones reales del "case mix" que registra el hospital.

Siguiendo a Farley (1989), supóngase que existe un registro de N pacientes, obtenidos de H hospitales y clasificados en G categorías. Sea N_{ih} el número de ingresos del tipo i en el hospital h , N_i el número total de casos del tipo i , S_{hi} la proporción de casos del hospital h de tipo i , y ϕ_i la proporción de casos esperados del tipo i para el hospital h en caso de una distribución uniforme de éstos. De acuerdo con ello, la complejidad de

²²(...continued)

indicador CN a niveles aceptables (al valor 22, cuando el máximo a partir del que la estimación resulta inestable es 30).

No es el anterior, además, el único inconveniente de la aproximación adoptada. Bases de datos de panel, con observaciones de los diferentes hospitales durante varios años, parecen mostrar, debido posiblemente al mayor consumo de recursos que se produce en el tiempo, un cierto deterioro de las posibilidades de producción. Esto es, la no incorporación del progreso técnico en la estimación genera la paradoja de un descenso en la productividad con la que se utilizan los recursos a lo largo del tiempo, a pesar de que es bien conocido que la medicina hospitalaria está sometida a un fuerte progreso tecnológico y que el envejecimiento de la población produce casos cada vez más complejos de tratar.

la categoría i , Ω_i se define como²³

$$\Omega_i = \sum_h p_{hi} \ln(S_{hi}/\Phi_{hi}) \quad (17)$$

Farley sugiere igualar ϕ_{hi} a N_h/N para todos los tipos de casos. En efecto, si la distribución de una categoría específica de casos entre los hospitales resulta ser idéntica a la distribución del conjunto de casos en general de los hospitales, dicha categoría podría considerarse de baja complejidad. Si, de otro modo, su distribución difiere grandemente de la distribución agregada (al estar concentrada en unos pocos hospitales), sería considerada de elevada complejidad. Resulta conveniente, en cualquier caso, estandarizar Ω_i . Siguiendo a Evans y Walker (op cit), sea

$$\Omega_i^* = \Omega_i / \sum_i \Omega_i (N_i/N) \quad (17')$$

de modo que la complejidad media de los casos entre los diversos hospitales H es igual a la unidad. La complejidad del hospital h , ϕ_h se define entonces como la media ponderada de los indicadores de complejidad, siendo los pesos iguales a las proporciones de "case mix" efectivamente registradas. Es decir:

$$\phi_h = \sum_i W_{hi} \Omega_i^*$$

$W_{hi} = N_{ih}/N_h$ recoge la proporción de ingresos del tipo i en el hospital h). Así por ejemplo, un hospital con una actividad compuesta en una gran proporción de categorías de casos definidos como complejos, recibiría una puntuación indicativa de elevada complejidad, mientras que otro con patologías poco complejas se consideraría de baja complejidad.

3.2.2 Otras variables de la función de costes

Existe en la literatura sobre el tema una considerable variación en cuanto a la especificación de la función de costes. La mayoría de los estudios, sin embargo, parecen centrarse en la función de costes medios, aunque también la función de costes totales es

²³- Cuando $S_{hi}=0$, $S_{hi} \ln(S_{hi}/\Phi_{hi})$ resulta no definida. Sin embargo, puede demostrarse que dicha expresión tiende a cero en la medida que S_{hi} tiende a cero.

a menudo analizada. Cuando ello se hace de modo conjunto, no dejan de observarse inconsistencias entre ambas especificaciones (véase, por ejemplo, Feldstein, 1967; McGuire y Williams, 1986).

Las inconsistencias entre las funciones de costes medios y totales dentro de un mismo estudio refleja la falta de consenso entre los economistas de la salud sobre la especificación apropiada de la función de costes de las instituciones sanitarias. A veces se registra una especie de actitud de "todo vale", incluyendo, por ejemplo, el número de camas, índice de rotación, coeficiente de utilización y estancia media como variables independientes, a pesar de estar todas ellas relacionadas a través de la restricción de capacidad o utilización. La mayoría de estudios utilizan asimismo variables ficticias o "dummy" para incluir en la estimación factores tales como el carácter docente de los hospitales, tipo de propiedad, índices salariales del personal empleado, nivel regional de renta, etc. Tal como Breyer (1987) apunta, si la ecuación que se estima es en realidad una función de costes, la justificación de la mayor parte de aquellas variables es cuestionable. Lo único que debiera de incluirse son variables que indiquen el output hospitalario, el precio de los inputs que se afronte y, si el horizonte temporal se limita al corto plazo (de modo que el estoc de capital resulta fijo), el número de camas.

En otros trabajos los propios autores (López Casanovas, 1984; Wagstaff, 1989b) habían adoptado la especificación de Feldstein (1967), en la que el coste medio se relacionaba al número de camas, su valor al cuadrado, el flujo de casos por cama y su valor al cuadrado, y un vector de "case mix" variables, definido en términos de proporciones de casos por especialidades médicas sobre el total. A la vista de los comentarios anteriores, se ha reconsiderado aquí la anterior formulación y se ha optado por seguir la aproximación propuesta por Barer (1982). Barer especifica con exquisito cuidado la función de coste de modo que resulte plenamente consistente con la perspectiva más teórica del análisis. Utiliza, además, TI y divide los costes totales en (i) costes fijos y (ii) costes variables, obteniendo una ecuación similar a la que será aquí utilizada

$$C_h = \alpha_1 B_h + \alpha_2 B_h^2 + \alpha_3 B_h^3 + \alpha_4 N_h \quad (18)$$

siendo C_h los costes totales de hospitalización del hospital h , B_h su estoc de camas y N_h el número de casos tratados por el hospital.

El primer componente de la función de costes, $\alpha_1 B_h + \alpha_2 B_h^2 + \alpha_3 B_h^3$, representa la parte correspondiente a los costes fijos dentro de los costes totales. Se puede esperar, a priori, que los costes fijos del hospital aumenten con el incremento en el número de camas, pero no necesariamente de un modo proporcional. La formulación (18) es, de hecho, más general que la de Barer, dado que a diferencia de la de éste, nosotros incluimos tanto valores al cuadrado del estoc de camas como valores al cubo. De este modo posibilitamos el supuesto plausible de que los costes van aumentando con el tamaño del hospital, pero a una tasa que decrece tras un punto determinado, para acelerar su crecimiento de nuevo más tarde.

El componente segundo de la función de coste $a_{4h}N_h$, representa la parte variable de los costes del hospital. Barer establece la hipótesis de que el coste marginal del hospital, a_{4h} , varía a lo largo de la muestra -de ahí el subíndice h. En particular, Barer propone que puede modelarse como una función lineal de varias variables, incluyendo la complejidad del "case mix", estancia media, gasto en actividades docentes, etc. Si definimos dicho vector de variables como $Z_{1h}, Z_{2h}, \dots, Z_{ih}$ resulta entonces

$$\alpha_{4h} = \tau_0 + \sum_i \tau_i Z_{ih}$$

Substituyendo esta expresión en (18), la función de coste total resulta

$$C_h = T_0 + \sum_i T_i Z_{ih} + \alpha_1 (B_h/N_h) + \alpha_2 (B_h^2/N_h) + \alpha_3 (B_h^3/N_h) + v_{hc} + u_h \quad u_h \geq 0; \quad (19)$$

siendo $C_h = C_h/N_h$ el coste por caso o enfermo ingresado.

3.3 Resultados de la estimación

Los coeficientes estimados de la función de costes para la muestra de hospitales INSALUD objeto de análisis tienen en todos los casos el signo esperado, siendo significativamente diferentes de cero (cf tabla 2). Ello implica, por ejemplo, que una estancia media mayor, una actividad de creciente complejidad y en el entorno de un hospital universitario se asocia con claridad con unos costes por caso más elevados. De la estimación resulta también que los costes fijos de un hospital son, en la práctica, una función creciente de su número de camas, aunque la relación entre los costes fijos totales y el estoc de camas no resulta lineal. (Véase el gráfico 4).

Nuestros resultados apuntan que la ineficiencia media en el sector hospitalario español, de acuerdo con la estimación efectuada es bastante alta: equivale a un 29% del coste medio por caso, existiendo una substancial variación en la ineficiencia entre los distintos centros. (véase el gráfico 5). El coste estimado de la ineficiencia para el Ministerio de Sanidad español de los 38 hospitales analizados, en ptas. de 1986 era, como media, de 44.530 ptas por cama, cifra próxima al 30% de dicho coste. Nuestros resultados confirman también el punto de vista de que, para evaluar la ineficiencia, no es posible mirar simplemente el coste por caso; hace falta mirar también los factores que ejercen una influencia sistemática en los costes (y, en particular, la complejidad de la actividad hospitalaria), así como las influencias aleatorias (simplemente resultado, por ejemplo, de un mal ejercicio irreplicable).

El gráfico 5 ordena los hospitales de acuerdo con su eficiencia. Este es el resultado especialmente relevante por sus implicaciones en términos de planificación de los recursos sanitarios y objeto primordial de análisis de ésta investigación. El código identificador

de cada uno de los hospitales puede encontrarse en la tabla 1. Asimismo, dicha tabla visualiza las variables descriptivas primarias de la actividad de los diversos hospitales. Un análisis detallado de los principales resultados de la estimación se ofrecen a continuación.

3.3.1 Análisis de los resultados

La tabla 2 ofrece los resultados básicos de la estimación y, en particular, los "intra" e "inter" estimadores obtenidos por Mínimos Cuadrados Generalizados del modelo Función Frontera de Costes. La transformación "intra" (o "within") expresa los datos en términos de desviaciones de las medias temporales. Por ello, tal como se indicó, la transformación elimina las variables que permanecen constantes en el tiempo; especialmente, el estatus docente de los centros y el índice de complejidad, regresionando tan sólo las variables B/N , B^2/N , B^3/N y LOS o estancia media. La transformación "inter" (o "between"), al contrario, expresa los datos simplemente en términos de medias temporales. Los estimadores MCG de la columna final son en efecto, medias ponderadas de los estimadores "inter" e "intra" -tal como se apuntaba en secciones precedentes- y se obtienen por MCO respecto de los datos expresados en términos de desviaciones de una fracción de sus medias temporales, siendo sus ponderaciones una función de las varianzas del error de las ecuaciones transformadas "inter" e "intra" (véase, por ejemplo, Judge y otros, 1982:490).

El componente de costes fijos.

Considérense primero los parámetros estimados del componente de costes fijos de la función de costes totales α_1 , α_2 y α_3 . Los tres parámetros resultan estadísticamente significativos en los intervalos habituales de confianza. Sus signos y magnitudes indican la relación entre los *costes fijos* y el *estoc de camas*, tal como ilustra el gráfico 4. Cuanto mayor es el número de camas de un hospital, más elevados resultan sus costes fijos. La relación no es, sin embargo, lineal. Un punto de inflexión en la pendiente creciente aparece en torno al hospital de 430 cama. El coste adicional asociado a una cama extra (es decir, el coste marginal asociado con una cama adicional) disminuye y más allá de este punto crece de nuevo. Estos resultados contrastan con los obtenidos por Barer (op cit), en los que el coste marginal de una cama adicional parecía incrementarse de modo continuo. En nuestra opinión, el resultado de Barer se deriva, al menos en parte, del hecho que su ecuación de costes totales no incluye un término que considere el valor cúbico de la variable en cuestión. En efecto, en dos de las especificaciones de Barer, el signo de los coeficientes implica que el coste marginal de una cama extra se convierte en negativo más allá de un cierto punto (esto es, el coeficiente del término elevado al cuadrado es negativo). Ello resulta difícilmente aceptable, y cabe sospechar que se debe a la restricción implícita en su estimación al asumir que $\alpha_3=0$.

En la literatura en economía de la salud, la relación parcial entre costes y estoc de camas (es decir, la relación entre costes y número de camas, considerando constante el número de casos) se interpreta, a menudo, como evidencia del alcance de las economías de escala en el sector hospitalario (por ejemplo, Feldstein, 1967). Esta interpretación resulta

extraña a la teoría microeconómica, en la que los beneficios a escala se definen en términos del output (número de casos), cuando todos los inputs (mano de obra e infraestructura, aquí camas) varían en igual proporción. Beneficios a escala son, por tanto, un fenómeno a largo plazo y requiere que tanto inputs como outputs varíen. Ello no es ciertamente lo que se está considerando en estas estimaciones comentadas. Al contrario, los resultados de la tabla 2 indican el efecto del tamaño del hospital en los costes fijos: en otras palabras, el factor trabajo y el output permanecen constantes en nuestra ecuación, al igual que en la mayoría de la literatura en esta área. No cabe, por tanto, interpretar dichos resultados en términos de la existencia o no de economías de escala.

El componente de costes variables.

El coeficiente afecto a la variable duración de la *estancia media* (LOS) es positivo y estadísticamente significativo. El valor del coeficiente estimado por MCG significa que un aumento en un día en la estancia media, permaneciendo el total de camas y el número de enfermos constante, genera un aumento en el coste por caso de tan sólo 7175 ptas. (es decir un 40% del total del coste por día de estancia). Dicho resultado indica también que no cabe esperar grandes ahorros de políticas que incidan en el margen, en reducciones adicionales en la estancia media.

En cualquier caso, el hecho de que estancias medias mayores se asocien a costes medios más elevados no resulta en ningún caso sorprendente: resultados similares aparecen en la mayoría de investigaciones. Lo que resulta menos obvio es que esta variable en concreto deba de incluirse en la función de coste. Existen dos argumentos para hacerlo. En primer término, los aspectos hoteleros de la asistencia hospitalaria pueden considerarse como un aspecto más del output hospitalario: En otras palabras, los días de estancia (nº de casos multiplicados por la estancia media) debieran contemplarse como "output" (véase Breyer, 1987). Dicho argumento tiene mayor sentido quizás en el estudio del comportamiento del sector privado de asistencia sanitaria. Nuestra opinión es, sin embargo, que en sistemas de asistencia pública es poco probable que una estancia media más elevada pueda, sin más, considerarse como algo positivo en si mismo. Ello sería el caso tan sólo cuando ésta se pudiera asociar con una relativamente más importante mejora en el estatus de los enfermos de la que de otro modo hubiera tenido lugar.

El punto anterior nos lleva al segundo de los argumentos, en favor todavía de incluir la variable LOS en la estimación: la duración de la estancia media puede, en cierta manera, considerarse una variable aproximativa de la complejidad de los casos atendidos por el hospital. Efectivamente, se ha argumentado en ocasiones que hospitales con LOS mayores tratan casos más severos dentro de cada categoría de "case mix" (por ejemplo, Barlow, 1968; Fuchs, 1969; Lave y Lave, 1970). El argumento es, por tanto, que aunque la Teoría de la Información permite distinguir entre categorías de casos más y menos complejos, no permite, sin embargo, identificar en nuestro caso, los hospitales que atienden los enfermos más complejos dentro de cada categoría "case mix". Ciertamente esta observación tiene una mayor fuerza cuando el número de categorías definidas es pequeño, como en nuestro análisis. No cuando la categorización es muy amplia, como en el caso en que se basa en los grupos de la Clasificación Internacional de Enfermedades.

El coeficiente positivo (y significativo) de la variable *complejidad* construida sobre la TI es consistente en nuestra estimación con la expectativa que "a priori" se tenía. Tal como se ha comentado en secciones precedentes, los hospitales con una carga de trabajo más compleja es probable que registren unos costes mayores. El resultado es asimismo coherente con el de Barer y otros autores que han utilizado la aproximación basada en TI. La magnitud del coeficiente implica que cada punto adicional en el índice IT de complejidad incrementa el coste medio en 7165 ptas. Si se compara, por ejemplo, el hospital Nuestra Sra. del Pino de Las Palmas y el Hospital de Valdepeñas, puede comprobarse como los 7.53 puntos adicionales del índice TI registrado por el primero de los hospitales supone, de acuerdo con nuestros resultados, un 21% del coste diferencial observado entre ellos.

La variable última en la ecuación estimada es el *estatus docente* del centro. Tal como se podía esperar, el coeficiente de esta variable en la estimación por MCG es positivo y significativamente diferente de cero. La magnitud de dicho estimador implica que el hecho de ser un hospital docente aumenta el coste medio en 28.275 ptas. (algo más de un 15% del coste total por caso tratado). Cabe señalar que, a este respecto, que nuestra especificación no diferencia, sin embargo, entre la diferente magnitud de los programas docentes en contra posición a la estimación de Barer.

La parte final de la tabla 2 indica, entre otros, el valor medio estimado del componente ineficiencia. En 1986, la ineficiencia media equivalía a un 29% del coste medio por caso. De modo equivalente, la ineficiencia en promedio suponía un aumento en el coste por caso tratado de un 41% por encima de su nivel mínimo factible. El coste estimado para las arcas públicas de este nivel de ineficiencia, en ptas. de 1986 y únicamente para la muestra considerada, equivalía a 621 millones; es decir, un 29% de los costes registrados.

En la tabla 2 aparecen también las magnitudes de las varianzas σ_u^2 y σ_v^2 : sus valores implican que un 54% de la variación del error compuesto $v_m + u_n$ viene explicada por la variación en el componente de ineficiencia u_n . La tabla indica así mismo, el valor del estadístico del multiplicador lagrangiano (LM), a efectos de validar la hipótesis $H_0: \sigma_u^2 = 0$; es decir, de que no existe variación entre hospitales en el nivel de ineficiencia. El valor del estadístico LM (63,83), supera el valor tabular del 1% de la distribución $\chi^2_{(1)}$ (6,635). El test sugiere, por lo tanto, que sí hay una variación significativa en el nivel de ineficiencia a lo largo de la muestra de hospitales.

Resulta ilustrativo en este punto comparar los resultados de la tabla 2, relativos a la ineficiencia, con los obtenidos por Wagstaff(1989) -el único otro trabajo hasta la fecha que utiliza la aproximación de Schmidt-Sickless en el análisis de la eficiencia hospitalaria. Como se ha comentado anteriormente, el estudio de Wagstaff difería de la investigación presente en varios aspectos clave: los datos se referían a los cinco años anteriores, precedentes al periodo quinquenal utilizado en esta investigación; el componente sistemático de la función de coste, utilizado en el estudio de Wagstaff, era el mismo que el utilizado por Feldstein (1967); finalmente, en lugar de utilizar la aproximación basada en TI para capturar las variaciones inter-hospitalarias en "case mix", Wagstaff adoptaba la aproximación de Feldstein, al incluir las proporciones de las

diferentes categorías de "case mix" directamente en la ecuación de la regresión efectuada. Aún con estas diferencias de partida, es interesante anotar que la ineficiencia media estimada en Wagstaff era bastante más elevada (un 42% del coste medio por caso) que la que se obtiene en esta investigación, aunque la varianza σ_u^2 expresada como proporción de la varianza total σ^2 era bastante inferior en la estimación de Wagstaff que la aquí resultante (33%).

3.3.2 La variación en la eficiencia entre hospitales

A través de la utilización de estimadores MCG de u_h para la ecuación (19) resulta posible analizar el alcance de las diferencias en la ineficiencia entre los distintos hospitales. El Gráfico 5 resume los resultados obtenidos para cada uno de los hospitales. En dicho gráfico, la altura de cada barra indica el coste por caso (en 1986) para cada hospital. Así, por ejemplo, el hospital de Ferrol (nº 22) registró en 1986 un coste por enfermo ingresado de 210.000 ptas (véase tabla 1). El área sombreada en cada barra representa la ineficiencia u_h . El gráfico permite, por lo tanto, una visión de conjunto del valor en ptas. de la ineficiencia (del tipo que sea) y la proporción que ésta constituye en el coste del hospital.

Lo primero que destaca del gráfico 5 es la importante variación en los niveles de ineficiencia: a la vez que el hospital 22 de Ferrol aparece con un alto nivel de eficiencia, el hospital 43 resulta el de menos con un coste estimado para la ineficiencia que registra cercano a las 110.000 ptas. por caso. La contundencia en la aplicación sistemática de dicho análisis puede exigir diversas cautelas pero indica suficientemente el grado en el que su eficiencia se aleja de su mínimo factible.

El segundo factor que emerge del gráfico en cuestión es que un alto coste por caso no es necesariamente sinónimo de ineficiencia. El hospital 22, por ejemplo, aparece con cero ineficiencia, aunque su coste medio es el octavo mayor de la muestra. De modo similar, el hospital 27 es el octavo más eficiente y, pese a ello, tiene el cuarto coste medio mayor por caso. Existen, claro está, hospitales con costes muy elevados que son también muy ineficientes: por ejemplo, parece que incluso en el caso de que ajustemos o tengamos en cuenta su relativo mayor tamaño, estatus universitario, elevada estancia media por enfermo y su actividad relativamente de alta complejidad, el hospital de Las Palmas (nº 41) aparece como un hospital muy poco eficiente en el período analizado.

4 Conclusión

En conclusión, el resultado de la estimación efectuada pone en evidencia que al evaluar la ineficiencia, uno no puede fijarse simplemente en el coste por caso; hace falta también contemplar tanto las influencias sistemáticas en los costes (la complejidad en la actividad hospitalaria) como las influencias aleatorias que afectaron al centro en el ejercicio en cuestión. En el modelo aquí utilizado, ambos tipos de influencias se tienen en consideración: los estimadores de ineficiencia del gráfico 5 no confunden sus efectos en el análisis de los costes hospitalarios, en relación con aquellos otros resultantes del nivel de ineficiencia observada. Dicha estimación supera, por lo tanto, las simples

comparaciones al uso de costes medios y utilización de recursos por actividad.

El modelo muestra con ello su potencial para la formulación de políticas de planificación y despliegue de recursos sanitarios de acuerdo con criterios de eficiencia económica, tema de atención, en definitiva, de ésta investigación.

TABLA 2

Table 2: Parameter estimates of hospital cost frontier

Variable	Specification		
	Within	Between	GLS
Constant		-94.644 (2.46)	-29.6423 (4.39)
R/N	13306.900 (9.09)	7546.640 (5.09)	8892.600 (9.33)
Beds sqrd/N	-36.022 (5.10)	-7.412 (1.36)	-16.774 (3.93)
Beds cubed/N	0.031 (3.46)	0.004 (0.69)	0.013 (2.52)
LOS	5.505 (2.22)	7.794 (1.27)	7.175 (3.01)
Case complexity		5.107 (1.75)	7.165 (2.33)
Teaching status		12.208 (1.07)	28.275 (2.58)
R ²	0.87	0.71	0.61
F	24.37	13.02	46.97
E(u)			52.99
σ_u^2			626.20
σ_v^2			536.95
σ^2			1163.15
LM			63.83

Note: Numbers in parentheses are t-values

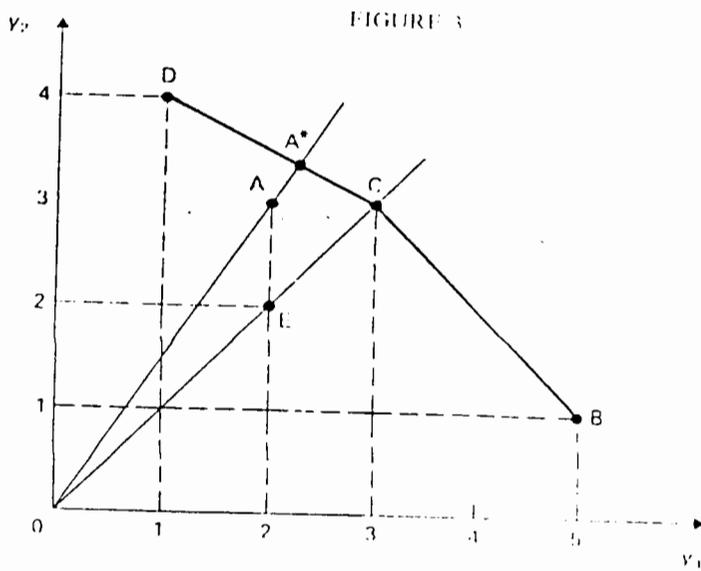
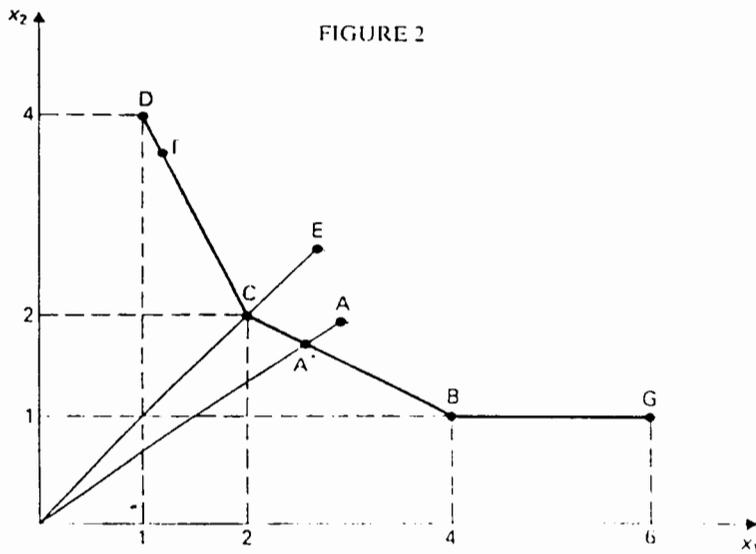
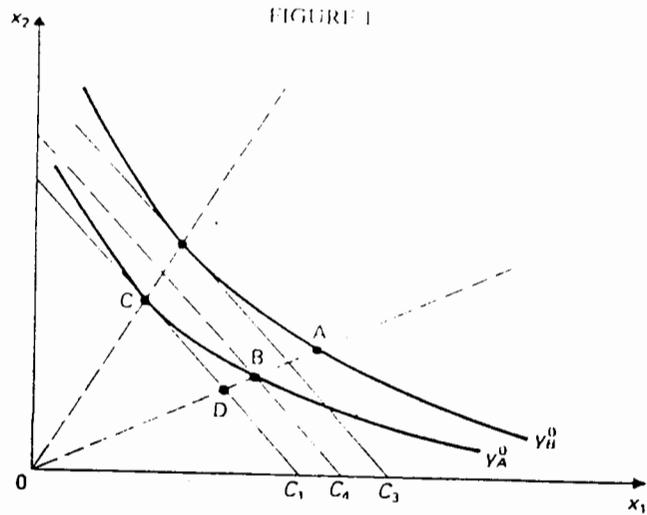


Fig 4: Relationship between fixed cost and stock of beds

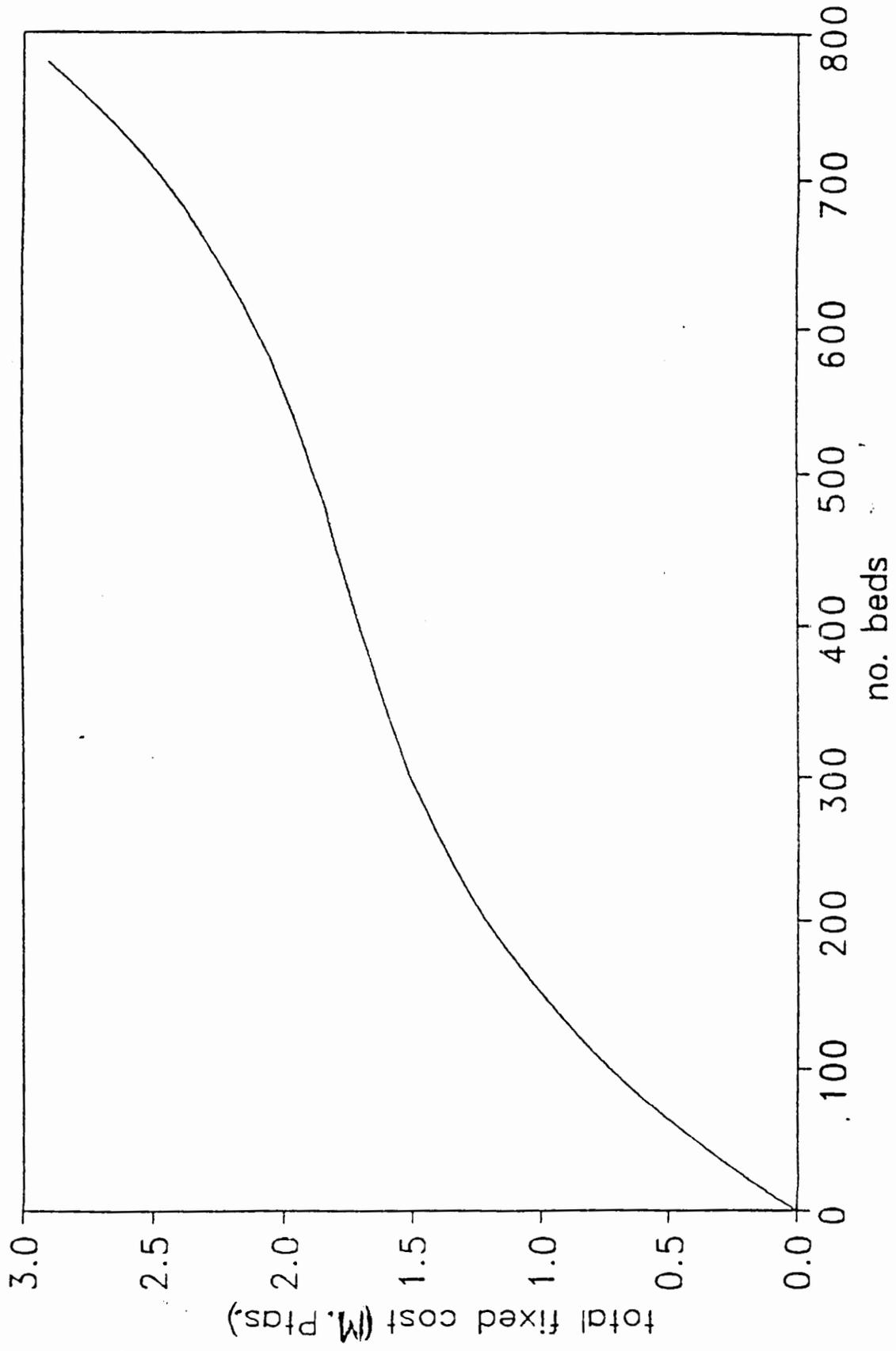
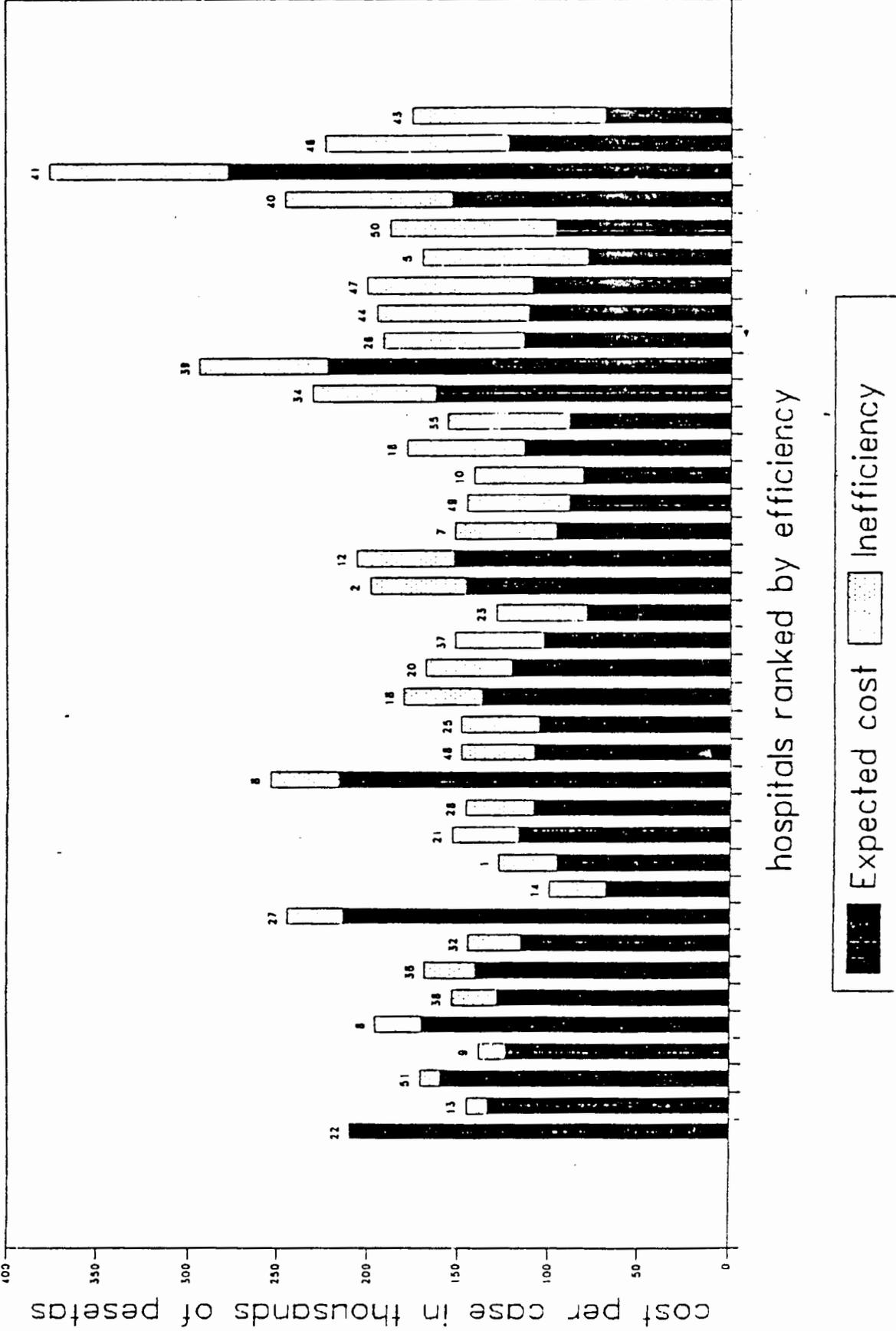


FIGURA 5

1986 costs broken down into expected cost and inefficiency components



Bibliografia

- AIGNER, D . J . and S . F . CHU (1968) . On estimating the industry production function. *American Economic Review* 58, 226-239.
- AIGNER, D . J ., C . A . K . LOVELL and P . SCHMIDT (1977) . Formulation and estimation of stochastic production function models. *Journal of Econometrics* ,6 21-37.
- BANKER, R . D ., R . F . CONRAD and 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.
- BYRNESS, P., R. FARE and S. GROSSKOPF (1984) . Measuring productive efficiency an application to Illinois strip mines. *Management Science* 30, 671-681.
- CHLARNES, A., COOPER, W.W. and RHODES, E. (1978) . Measuring the efficiency of decision-making units. *European Journal of Operational Research* 2, 429-444.
- CHLARNES, A., COOPER, W.W. and 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. (1980). *The Political Economy of Social Policy*. Martin Robertson, Oxford.
- CULYER, A.J., R.J. LAVERS and A. WILLIAMS (1971). Social indicators: health. *Social Trends* 2, 21-41.
- EAKEN, B . K . and T . J . KNIESNER (1982) . Estimating a non-minimum cost function for hospitals. *Southern Economic Journal*, 583-597.
- EVANS, R . G . and M . D . WALKER (1972) . Information theory and the analysis of hospital cost structure. *Canadian Journal of Economics* 5, 398-418.
- FARRELL, M . J . (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 for Health Service Efficiency. Health Services Research 3, 242-50 .
- FORSUND F . R ., C . A . K . LOVELL and P . SCHMIDT (1980) . A survey of frontier production functions and their relationship to efficiency measurement. Journal of Econometrics; 13, 5-25 .
- GRAY, A ., A . McGUIRE and P . STUART (1986) . Factor input in NHS hospitals. Discussion Paper 02/86 (Health Economics Research Unit, University of Aberdeen).
- GRAVELLE, H S.E. and R. REES (1981). Microeconomics (Longman, London).
- GREENE, W H (1980) Maximum likelihood estimation of econometric frontier functions. Journal of Econometrics 13, ~7-56
- GREENE, W H (1982) Maximum likelihood estimation of stochastic frontier production models. Journal of Econometrics 18, 285-289.
- GROSSKOPF, S . and V . VALDMANIS (1987) . Measuring hospital performance: a nonparametric approach. Journal of Health Economics 6, 89-108.
- GUDEX, C. (1986). QALYs and their use by the health service. Discussion Paper 19 (Centre for Health Economics, University of York).
- JOHNSTON, J. (1984). Econometric Methods. (McGraw-Hill, London).
- JONDROW, J., C.A.K. LOVELL, I. MATEROV and 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 . and 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 . and 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 . and A . WAGSTAFF (1988) . La combinación de los factores productivos en el hospital: una aproximación a la función de producción. *Investigaciones Económicas* 12, 305-327.
- McGUIRE, A . (1987) . The measurement of hospital efficiency. *Social Science and Medicine* 24, 719-724 .
- MEUSSAN, W and J VAN DEN BROECK (1977) . Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18 435—444.
- PITT, M M . and 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, 289-328.
- SCHMIDT, P . and 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. and R.C. SICKLES (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics* 2, 367-374
- STEINWACHS, D.M., J.P. WERNER, S. SHAPIRO, P. BATALDER, K. COLTON and F. WASSERMAN (1986). A comparison of the requirements for primary care physicians in HMOs with projections made by GMENAC. *New England Journal of Medicine* 314(4) 217-221.
- TATCHELL, M. (1983). Measuring hospital output: a review of the service-mix and case-mix approaches. *Social Science and Medicine* 17, 871-83.
- 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.
- WILLIAMS, A. (1985). The economics of coronary artery bypass grafting. *British Medical Journal* 291, 326-9.

RECENT WORKING PAPERS

1. Albert Marcet and Ramon Marimon, *Communication, Commitment and Growth* (June 1991).
2. Antoni Bosch, *Economies of Scale, Location, Age, and Sex Discrimination in Household Demand* (June 1991).
3. Albert Satorra, *Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures* (June 1991).
4. Javier Andrés and Jaume García, *Wage Determination in the Spanish Industry* (June 1991).
5. Albert Marcet, *Solving Non-Linear Stochastic Models by Parameterizing Expectations: An Application to Asset Pricing with Production* (July 1991).
6. Albert Marcet, *Simulation Analysis of Dynamic Stochastic Models: Applications to Theory and Estimation* (November 1991).
7. Xavier Calsamiglia and Alan Kirman, *A Unique Informationally Efficient and Decentralized Mechanism with Fair Outcomes* (November 1991).
8. Albert Satorra, *The Variance Matrix of Sample Second-order Moments in Multivariate Linear Relations* (January 1992).
9. Teresa Garcia-Milà and Therese J. McGuire, *Industrial Mix as a Factor in the Growth and Variability of States' Economies* (January 1992).
10. Walter Garcia-Fontes and Hugo Hopenhayn, *Entry Restrictions and the Determination of Quality* (February 1992).
11. Guillem López and Adam Robert Wagstaff, *Indicadores de Eficiencia en el Sector Hospitalario* (March 1992).

UNIVERSITAT POMPEU FABRA

Balmes, 132

Telephone (343) 484 97 00

Fax (343) 484 97 02

08008 Barcelona