

**“Evaluación de la eficiencia con factores exógenos
mediante un análisis semi-paramétrico”**

Cordero Ferrera, José Manuel
Universidad de Extremadura
(jmcordero@unex.es)

Pedraja Chaparro, Francisco
Universidad de Extremadura
(pedraja@unex.es)

Santín González, Daniel
Universidad Complutense de Madrid
(dsantin@ccee.ucm.es)

Resumen

En este trabajo se analiza el efecto que pueden tener los inputs no controlables en las unidades productivas sobre los índices de eficiencia estimados mediante DEA. Para controlar su efecto se propone una metodología semiparamétrica basada en el método de cuatro etapas de Fried *et al.* (1999) en la que se emplea el método *bootstrap* para mejorar la robustez de las estimaciones. Para ilustrar los beneficios de este procedimiento se plantea una aplicación empírica en el contexto educativo, donde la importancia de los inputs no controlables ha sido puesta de manifiesto en un buen número de estudios. Los resultados muestran las ventajas que aporta la metodología con respecto a otras alternativas propuestas.

1. INTRODUCCIÓN

Desde la publicación del trabajo realizado por Charnes, Cooper y Rhodes (1978) en el que se dio a conocer el Análisis Envolvente de Datos (DEA), esta técnica se ha aplicado en multitud de estudios y en sectores muy distintos para evaluar la eficiencia, siendo especialmente útil en contextos en los que resulta muy difícil modelizar la tecnología de producción mediante una forma funcional determinada.

Con la generalización de su uso, en la literatura han surgido multitud de trabajos que plantean extensiones del modelo original al objeto de conseguir una mayor precisión en las estimaciones del nivel de eficiencia que se derivan del uso de esta técnica. Entre ellas, una de las corrientes que ha despertado un mayor interés es el modo en el que pueden incorporarse a los análisis de eficiencia el efecto de los factores que están fuera del control de las propias unidades evaluadas.

Este tipo de variables son habituales en la práctica totalidad de los procesos productivos. De hecho, hay casos en los que su efecto sobre los resultados puede ser mayor que el producido por aquellos inputs controlados por las unidades. En consecuencia, si se pretende analizar el nivel de eficiencia con el que desempeñan su actividad un conjunto de unidades productivas debe tomarse en consideración el contexto en el que éstos operan. Esta es la única forma de que los resultados del estudio reflejen si el productor calificado como ineficiente lo es realmente o si aún haciendo todo lo que está en su mano, hay factores que no le permiten alcanzar los objetivos que otros sí logran.

El interés suscitado por el estudio de la influencia de estas variables ha dado lugar a la aparición de diversas propuestas metodológicas para su incorporación. Los primeros trabajos que abordaron esta cuestión plantearon modificaciones en la formulación estándar del DEA con el propósito de considerar de forma explícita el carácter no exógeno de algunas de las variables (Banker y Morey, 1986; Golany y Roll, 1993).

Otra opción muy extendida consiste en estimar los índices de eficiencia mediante DEA obviando la influencia de estas variables para, posteriormente, incluirlas en una segunda fase con el propósito de explicar potenciales ineficiencias de los productores (Ray, 1991; McCarty y Yaisawarng, 1993; Lovell *et al.*, 1994). Dentro de este enfoque, conocido habitualmente como análisis de segunda etapa, se han desarrollado recientemente extensiones basadas en el uso del método *bootstrap* que permiten superar una de las principales limitaciones de este enfoque (Simar y Wilson, 2007), representado por los problemas de sesgo que pueden producirse por la existencia de correlación entre los índices de eficiencia inicialmente estimados.

Este enfoque resulta apropiado cuando las variables consideradas son ambientales o de entorno, esto es, variables que no participan en el proceso productivo, pero que condicionan los resultados (como por ejemplo los factores climatológicos o geográficos). Sin embargo, no permite dar un tratamiento adecuado a otro tipo de factores, los denominados inputs no controlables, que sí participan en el proceso productivo pero están fijados exógenamente. Un ejemplo de este tipo de variables son las características del alumnado cuando las unidades evaluadas son un conjunto de centros escolares. En estos casos es imprescindible que los índices de eficiencia incluyan la información relativa a estos factores.

En este trabajo se plantea una nueva metodología que, tomando como referencia los modelos de segunda etapa, introduce algunas modificaciones en su estructura para conseguir que los índices de eficiencia reflejen realmente aquella parte del proceso productivo de la que se puede responsabilizar a cada productor, posibilitando así una comparación homogénea y equitativa entre las unidades.

El sector elegido para la realización de la aplicación empírica es el de la educación no universitaria, donde muchos estudios empíricos (véase Hanushek, 1986, 1997, 2003 para una recopilación) han puesto de manifiesto que los inputs no controlables, representados por indicadores relativos al

estatus socioeconómico y familiar de los estudiantes, tienen un efecto decisivo sobre los resultados.

La muestra utilizada está compuesta por 80 centros de educación secundaria de la Comunidad Autónoma de Extremadura, de los cuales se dispone de una fuente de información muy rica sobre las características de su alumnado. Esta base de datos se ha obtenido mediante una encuesta realizada directamente a los estudiantes, lo cual permite superar el problema habitual de escasez de información pública sobre los factores exógenos en educación.

La estructura del trabajo es la siguiente. En la sección segunda se expone la formulación básica de la metodología empleada para evaluar la eficiencia, el DEA, así como los diferentes enfoques para incorporar variables exógenas, prestando una especial atención a los modelos de varias etapas. A continuación se plantea una extensión que permite construir índices de eficiencia que pueden ser interpretados como verdaderos objetivos de producción para las unidades. En el tercer epígrafe se presentan los resultados obtenidos en la aplicación empírica en la que se evalúa el comportamiento de un conjunto de centros educativos. Por último, se presentan las principales conclusiones.

2. METODOLOGÍA

2.1. El Análisis Envolvente de Datos (DEA)

El DEA es una técnica de carácter no paramétrico que persigue la obtención de una envolvente que incluya a todas las unidades eficientes, junto con sus combinaciones lineales, quedando el resto de unidades por debajo de la misma. Dicha envolvente se identifica con la figura de la frontera eficiente, de manera que la distancia de las unidades ineficientes a la envolvente proporciona una medida de su nivel de ineficiencia. Con el DEA la construcción de la frontera no se establece *a priori* basándose en una función de producción teórica, sino que son los propios datos disponibles los que determinan la forma y la localización de la frontera productiva.

Desde el punto de vista de su formulación, el DEA plantea un problema de programación matemática para cada unidad observada, cuya resolución permite asignarle un índice de eficiencia. Un valor unitario implica que la producción observada y la potencial coinciden, es decir, el productor es eficiente. Si el índice es menor que uno, la unidad evaluada será ineficiente, ya que existen otras unidades en la muestra (las que forman el grupo de referencia en la comparación) que muestran un mejor comportamiento.

Al margen de ofrecer un índice que refleja el porcentaje de incremento de outputs (o reducción de inputs) necesario para que una unidad sea eficiente (eficiencia radial), el DEA también permite detectar posibles reducciones adicionales en los inputs o incrementos potenciales en los outputs mediante la incorporación al modelo dual de las denominadas variables de holgura o *slacks* (eficiencia no radial). Concretamente, para los inputs estas holguras representan la cantidad que se podría ahorrar cada productor en la utilización de los mismos en el caso de ser eficiente, mientras que, para los outputs, se identifican con cuánto podría incrementar la producción si alcanzara un comportamiento eficiente. De este modo, el modelo dual de maximización del output adopta la siguiente forma:

$$\begin{aligned}
 \text{Max} \quad & \theta + \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \\
 \text{s.a.} \quad & \sum_{j=1}^n \lambda_j x_{ij} + s_i^- = x_{i0} \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{rj} - s_r^+ = \theta y_{r0} \quad r = 1, 2, \dots, s \\
 & \lambda_j \geq 0; s_i^- \geq 0; s_r^+ \geq 0 \quad j = 1, 2, \dots, n
 \end{aligned}$$

donde i denota input y r output, θ_0 es el índice de eficiencia, λ_i son las ponderaciones y s_i^- y s_r^+ son las variables de holgura de inputs y outputs, respectivamente. En este caso, un productor es relativamente eficiente si y sólo si su índice de eficiencia es igual a la unidad y además todos los *slacks* son nulos. Con esta formulación del programa, además de asignar un índice de

eficiencia a cada unidad, se obtiene un valor que refleja la ineficiencia del productor (radial y no radial) en el uso de cada input o en la consecución de cada output, la cual puede deberse a diversos factores. Esta información es mucho más rica que la que ofrece el índice de eficiencia y puede ser de gran utilidad a la hora de identificar el origen de posibles ineficiencias por parte del productor.

2.2. Extensiones del DEA para incluir los factores exógenos

La literatura sobre DEA ha desarrollado diversas metodologías para tratar de incorporar a los análisis la influencia de los factores exógenos. Sin embargo, hasta el momento no existe un acuerdo entre los investigadores sobre la superioridad de alguna de ellas (Muñiz, 2002). En este epígrafe se describen las diferentes propuestas existentes en la literatura, planteando en algunos casos modificaciones que permitan mejorar la calidad de determinados modelos. Con el propósito de simplificar la exposición, las diferentes alternativas han sido agrupadas en tres categorías: modelos de una etapa, modelos de segunda etapa y modelos de valores ajustados.

a. Modelos de una etapa

El modelo más representativo de este enfoque es el propuesto por Banker y Morey (1986), cuya utilización está ampliamente extendida entre los investigadores ya que la mayoría de los programas informáticos desarrollados para el cálculo del DEA permiten incluir los inputs no controlables siguiendo esta opción (Barr, 2004). La formulación propuesta por estos autores es la siguiente¹:

¹ Este diseño sigue una orientación de minimización de inputs con rendimientos variables de escala.

$$\begin{aligned}
\min \quad & \theta - \varepsilon \left(\sum_{i=1}^M s_{i0}^+ + \sum_{r=1}^S s_{r0}^- \right) \\
\text{s.a.} \quad & \sum_{j=1}^n \lambda_j Y_j - s_0^- = Y_0 \\
& \sum_{j=1}^n \lambda_j X_j + s_0^+ = \theta X_0 \\
& \sum_{j=1}^n \lambda_j Z_j + s_0^f = Z_0 \\
& \sum_{j=1}^N \lambda_j = 1 \\
& \lambda_j \geq 0 \quad \forall j = 1, \dots, n \\
& s_0^+, s_0^-, s_0^f \geq 0
\end{aligned}$$

donde s_0^f representa el *slack* del input no controlable.

El programa resultante no persigue la reducción equiproporcional de todos los inputs, sino tan solo del subvector formado por los controlables. Esto supone que cada unidad es comparada únicamente con los productores que utilizan una cantidad igual o inferior del correspondiente input no controlable (Z_0)². Con esta modificación, los nuevos índices de eficiencia serán iguales o menores que los obtenidos si las variables exógenas fueran incluidas como inputs discrecionales.

La principal ventaja del tradicional enfoque unietápico es que permite introducir todas las variables relevantes en un único análisis DEA, simplificando en gran medida el cálculo de los índices de eficiencia. Sin embargo, esta opción presenta problemas metodológicos que deben ser tenidos en cuenta por los investigadores que pretendan emplear esta técnica para incorporar la información relativa a los inputs no controlables.

² El modelo permite incorporar variables exógenas con valores continuos, aunque resulta necesario conocer a priori el sentido que tiene su influencia sobre el proceso productivo, esto es, si un valor elevado de estas variables facilita el desempeño de la actividad o lo dificulta.

Un primer inconveniente es que la utilidad de este modelo está restringida a aquellos casos en los que se adopta una orientación de minimización de inputs, puesto que con una orientación de maximización de outputs todos los inputs quedan fijados y no existe posibilidad de dar un tratamiento diferenciado a los que están fuera del control de los productores. Esta limitación adquiere una enorme relevancia en el ámbito público y, en particular, en el sector educativo, donde las unidades administrativas suelen ser evaluadas en función de su capacidad para obtener los mejores resultados posibles con un presupuesto fijado previamente (enfoque de maximización del output).

Una segunda desventaja de este enfoque es que las unidades que son calificadas como eficientes son las mismas que las que resultarían en un DEA estándar en el que todas las variables se incluyeran como inputs controlados por el gestor, puesto que la frontera construida a través del DEA modificado es la misma que se obtendría con un DEA estándar. De este modo, los resultados finales dependen en mayor medida de que se incluyan más variables en la evaluación que de que éstas sean o no controlables.

Una tercera debilidad del modelo es que la inclusión de un mayor número de variables en el análisis incrementa automáticamente las posibilidades de que una unidad pase a ser calificada como eficiente. Este fenómeno se explica por la pérdida de grados de libertad que implica la inclusión de nuevas variables en el modelo, cuya consecuencia directa es la pérdida de poder de discriminación entre los productores cuanto mayor sea el número de variables consideradas.

Las limitaciones e inconvenientes señalados en las líneas anteriores nos lleva a valorar *a priori* la opción multietápica como un enfoque mucho más apropiado para abordar la incorporación de los inputs no controlables a los análisis de eficiencia.

b. Modelos de segunda etapa

Estos modelos aplican una regresión en la que se incluyen los índices de eficiencia iniciales (θ_i) como variable dependiente y las variables representativas de los inputs no controlables (Z_i) como regresores:

$$\theta_i = f(Z_i, \beta_i) + u_j$$

A partir de las estimaciones se pueden identificar cuáles son las variables influyentes (significativas) y su signo (positivo o negativo), además de poder corregir los índices estimados con el DEA estándar a partir de los parámetros estimados. La mayoría de los trabajos estiman la regresión a través de un modelo *Tobit* apoyándose en el hecho de que los valores de los índices de eficiencia están censurados y que un número importante de unidades se sitúa en la frontera (McCarty y Yaisawarng, 1993; Kirjavainen y Loikkanen, 1998).

Su principal virtud en relación al resto de modelos multietápicos es su sencillez, ya que utilizan como variable dependiente los propios índices de eficiencia obtenidos en la primera etapa, lo que facilita en gran medida la realización de los ajustes. Esta simplicidad tiene como limitación que la corrección realizada sobre los índices iniciales sólo tiene en cuenta el componente radial de la ineficiencia, obviándose las posibles ineficiencias derivadas de la existencia de holguras³.

No obstante, la principal crítica que han recibido estos modelos es que las estimaciones de los parámetros incorporan un sesgo porque los índices de eficiencia estimados en la primera etapa (variable dependiente en la regresión) se construyen a partir de la información de todas las unidades que componen la muestra, con lo que se incumple el requisito de independencia entre los errores (Xue y Parker, 1999). Este problema de correlación en los errores es

³ Fried *et al.* (1999) señalan que este hecho puede sesgar la estimación de los parámetros y conducir a conclusiones engañosas sobre el impacto de cada variable exógena sobre la eficiencia.

menor cuando el tamaño de la muestra es suficientemente grande, sin embargo, adquiere una mayor relevancia cuando las inferencias se realizan en muestras pequeñas, lo que convierte a estos modelos básicos de regresión en una alternativa poco recomendable desde el punto de vista metodológico. El único modo de evitar incurrir en estos problemas de sesgo es mediante la utilización de técnicas de *bootstrap* en la estimación de los parámetros de la regresión.

Aunque han sido varios los intentos de incorporar el *bootstrap* a la estimación de las regresiones de segunda etapa⁴, la metodología que ofrece una solución más satisfactoria es la propuesta por Simar y Wilson (2007). El modelo planteado por estos autores describe un proceso generador de datos (PGD) consistente con el procedimiento de estimación de regresiones en las que se incluyen variables explicativas de unos índices de eficiencia estimados previamente mediante técnicas no paramétricas. Tomando como referencia este PGD resulta posible la aplicación de técnicas de *bootstrap*, mediante las cuales se pueden obtener estimadores consistentes e insesgados de unos índices de eficiencia que incorporen el efecto de variables exógenas al proceso.

El modelo planteado comparte con los de segunda etapa la idea de que la eficiencia depende de las variables exógenas: $\theta = \psi(z_i, \beta) + \varepsilon_i$, donde ψ es una función continua, β un vector de parámetros, z_i es un vector que representa los valores de las variable exógena y ε_i es una variable aleatoria que sigue una distribución normal $N(0, \sigma_\varepsilon^2)$ con un truncamiento en $1 - \psi(z_i, \beta)$ ⁵. Sin embargo, en este caso se puntualiza que la variable incluida como dependiente en esta regresión no representa a los “verdaderos” índices de eficiencia (θ) porque éstos son desconocidos. Estos índices se reemplazan por unas estimaciones ($\hat{\theta}$) obtenidas a partir de las observaciones que componen

⁴ Entre estos trabajos cabe destacar los de Xue y Parker (1999) y y Hirschberg y Lloyd (2002), aunque la metodología que éstos plantean está basada en el remuestreo a partir de una distribución empírica, el cual resulta inconsistente en el contexto de la estimación de la eficiencia a través de métodos no paramétricos (Simar y Wilson, 1999a y 1999b).

la muestra disponible $S_n = (x_i, y_i)$ que, como ya se indicó al explicar los modelos de segunda etapa, están correlacionadas.

El propósito de este enfoque no es conseguir un estimador consistente para los índices de eficiencia, sino poder realizar inferencias válidas sobre ellos. Para tal fin, se plantea un algoritmo (*Algoritmo 1* en Simar y Wilson, 2007) basado en el empleo del *bootstrap* en el que las múltiples muestras necesarias se extraen del proceso generador de datos descrito. El procedimiento que se plantea es el siguiente:

1. Se calcula el índice de eficiencia ($\hat{\theta}_i$) para todas las unidades
2. Se estima la ecuación $\hat{\theta}_i = z_i\beta + \varepsilon_i$ por máxima verosimilitud, teniendo en cuenta que es una regresión truncada. A través de la estimación se obtienen los parámetros estimados de β y σ_ε ($\hat{\beta}$ y $\hat{\sigma}_\varepsilon$)
3. Se computan L estimaciones de β y σ_ε mediante bootstrap a través de las cuales se obtiene una matriz $A = \left\{ \left(\hat{\beta}^*, \hat{\sigma}_\varepsilon^* \right)_b \right\}_{b=1}^L$
 Para ello, partiendo del error ε_i de una distribución normal $[N(0, \hat{\sigma}_\varepsilon)]$ con un truncamiento a la izquierda en $(1 - z_i \hat{\beta})$, se calcula por máxima verosimilitud $\hat{\theta}_i^* = z_i \hat{\beta} + \varepsilon_i$ para cada $i=1, \dots, n$. De esta forma se obtienen los estimadores $\hat{\beta}^*, \hat{\sigma}_\varepsilon^*$.
4. A partir de los valores que componen la matriz A y las estimaciones originales ($\hat{\beta}$ y $\hat{\sigma}_\varepsilon$) se pueden construir intervalos de confianza para β y σ_ε e incluso se pueden calcular los valores predichos para los índices de eficiencia: $\hat{\theta}^* = \hat{\beta}^* z_i$

El mayor atractivo de esta propuesta metodológica es que permiten solucionar los problemas de sesgo asociados con los modelos básicos de

⁵ Esta truncamiento se justifica por el hecho de que los índices de eficiencia están censurados por el valor 1.

regresión⁶. Sin embargo, comparten con ellos algunas de sus principales limitaciones, como el hecho de que únicamente tienen en cuenta el componente radial de la ineficiencia. Además, el principal objetivo de esta propuesta es la obtención de unos estimadores de los parámetros de la regresión válidos, de forma que, a través de sus intervalos de confianza, pueda conocerse con mayor certeza cuáles son los factores que más influyen sobre la eficiencia de los productores. Por tanto, esta alternativa resulta mucho más apropiada en aquellos casos en los que las variables exógenas consideradas en la evaluación son variables de entorno.

Si se opta por la utilización de esta propuesta para la construcción de unos nuevos índices que incorporen el efecto de las variables exógenas, debe asumirse que éstos no reflejan unos verdaderos objetivos de producción para las unidades. Con este mecanismo, el orden de las unidades se modifica notablemente, pero el valor medio de los índices corregidos es prácticamente el mismo que el que se obtendría con un modelo que no incluyera el efecto de las variables exógenas⁷, siendo habitual que tras la realización del ajuste ninguna unidad sea calificada como eficiente⁸.

2.3. Modelo de varias etapas o de valores ajustados

Al margen de los modelos de segunda etapa, en la literatura existen otras propuestas multietápicas más complejas que permiten construir nuevos índices de eficiencia que incorporan el efecto de los inputs no discrecionales y que se basan en la utilización de los *slacks* totales (componente radial y no radial) obtenidos en la primera etapa.

El origen de este planteamiento se encuentra en el trabajo de Fried, Lovell y Vanden Eeckaut (1993), en el que se pone de manifiesto que la

⁶ Para muestras muy pequeñas respecto al número de variables Simar y Wilson (2007) proponen un *algoritmo 2* alternativo que presenta resultados más robustos.

⁷ El valor medio sería exactamente el mismo si la regresión se estima mediante MCO, pero con una regresión truncada la media de la eficiencia estimada es ligeramente superior.

⁸ Este fenómeno se produce en el trabajo realizado por Afonso y St. Aubyn (2005), en el que se aplica la metodología propuesta por Simar y Wilson (2007) en un análisis de los resultados educativos de un conjunto de países.

ineficiencia no es neutral entre las variables incluidas en la evaluación inicial, es decir, la actuación de las unidades difiere según la variable que se tome como referencia en la evaluación, con lo que la información proporcionada por los *slacks* totales de cada variable (componente radial y no radial) adquiere una enorme relevancia.

Este fenómeno no se tiene en cuenta en los análisis de segunda etapa, donde únicamente se considera el componente radial de la ineficiencia (reducción proporcional en todos los inputs o incremento proporcional en todos los outputs), asumiendo, erróneamente, que la influencia de las variables no discrecionales es la misma sobre todos los elementos del proceso (inputs y outputs), lo que implícitamente conduce a deficiencias en la propia medida de la ineficiencia obtenida con estos modelos.

Partiendo de esta idea, Fried, Schmidt y Yaisawarng (1999) proponen un modelo de cuatro etapas que analiza por separado el efecto que tienen las variables exógenas sobre las diferentes variables incluidas en el análisis inicial, utilizando para ello el valor de los *slacks* estimados en la primera etapa. Para ello se plantea en una segunda etapa un sistema de ecuaciones formado por n regresiones Tobit, una para cada variable, en las que las k variables dependientes son los *slacks* totales (componente radial y no radial) de cada output/input, dependiendo de la orientación, obtenidos en el DEA inicial (primera etapa), mientras que los regresores son los factores exógenos:

$$s_i^- = f(Z, \beta) + u \quad \text{o} \quad s_r^+ = f(Z, \beta) + u$$

Su propósito es identificar qué parte de estos *slacks* se explica por el efecto de los inputs no controlables y qué parte está asociada con la propia ineficiencia técnica de los productores. Esta descomposición se realiza utilizando los coeficientes estimados en la regresión, a partir de los cuales se pueden predecir los nuevos *slacks* de cada variable⁹, teniendo en cuenta la

⁹ Estos coeficientes nos ofrecen información acerca del efecto (positivo o negativo) de los inputs no controlables sobre los *slacks*. Este efecto puede ser diferente para cada *slack* e,

dotación de inputs no controlables de cada unidad, y se pueden ajustar (tercera etapa) los valores de las variables originales (inputs o outputs)¹⁰. Finalmente, la cuarta etapa consiste en la realización de un nuevo DEA en el que se incluyen los nuevos valores corregidos de las variables en función de la dotación de inputs no controlables de cada unidad, de manera que los nuevos índices estimados recogen exclusivamente el nivel de eficiencia con el que actúa cada productor.

Desde el punto de vista metodológico esta propuesta resulta muy atractiva, pues permite estimar índices que reflejan verdaderos objetivos de producción para las unidades. No obstante, su similitud con la estructura de los modelos de segunda etapa le hace compartir con ellos los problemas de sesgo que acompañan a la estimación de las regresiones, pues el cálculo de los *slacks* también se realiza a partir de la información de todas las unidades que componen la muestra, por lo que no se cumple el requisito de independencia en los errores de las regresiones que componen el sistema de ecuaciones.

Este problema no ha sido abordado hasta el momento en la literatura, pese a que, *a priori*, la solución del mismo no resulta excesivamente compleja ya que la estimación del sistema de ecuaciones de regresión original puede realizarse mediante la aplicación de técnicas de *bootstrap*. Esta posibilidad, inédita en los trabajos realizados en este campo, pasa por la aplicación de los mismos pasos que plantea el Algoritmo 1 planteado por Simar y Wilson (2007) para estimar los intervalos de confianza de los *slacks*. No obstante, para una correcta aplicación del mismo es necesario introducir unas pequeñas modificaciones en el proceso y que comentamos a continuación:

incluso, puede ocurrir que haya factores exógenos que influyan sobre unos *slacks*, pero no sobre otros.

¹⁰ Siguiendo el modelo original de Fried *et al.* (1999), los ajustes se realizarán restando al valor original de cada output la diferencia entre el mayor valor predicho y el valor predicho para cada unidad: $y_{j\ adj}^k = y_j^x - \left[\text{Max}^k \left\{ \text{ITS}_{j\ pred}^k \right\} - \text{ITS}_{j\ pred}^k \right]$.

- La sustitución del índice estimado en la etapa inicial $\hat{\theta}_i$ (variable dependiente en la regresión) por el valor estimado de los slacks \hat{ITS}_i^k para cada una de las variables.
- La modificación del punto de truncamiento de la distribución normal, que pasa a ser $-z_i\hat{\beta}_i$ en lugar de $(1-z_i\hat{\beta}_i)$ ¹¹.

La aplicación de este nuevo algoritmo en la estimación de cada una de las regresiones que componen el sistema de ecuaciones permite obtener unos intervalos de confianza para los parámetros de gran utilidad para identificar con claridad cuáles son las variables exógenas que más influyen (y en qué sentido) sobre los distintos outputs o inputs que forman parte del proceso. Pero su principal utilidad de cara a poder incorporar el efecto de los inputs no controlables en los índices de eficiencia, es que ofrece una estimación más precisa de los parámetros de las diferentes regresiones $\hat{\beta}_j$ (tomando el valor medio de las diferentes estimaciones) a partir de los cuales pueden realizarse los diferentes ajustes sobre los valores de las variables iniciales y, finalmente, calcular unos nuevos índices de eficiencia que reflejen unos objetivos de producción precisos para las unidades teniendo en cuenta su dotación de inputs no controlables.

Por tanto, con la modificación propuesta se evitan los problemas de sesgo relacionados con la correlación existente entre los valores de la variable dependiente, convirtiendo a este “*modelo de cuatro etapas con bootstrap*” en una alternativa muy atractiva para la construcción de índices de eficiencia que incorporen la información relativa a los inputs no controlables.

¹¹ Esta modificación en el punto de truncamiento está motivado por el hecho de que los slacks no están censurados por el valor 1, sino por el valor 0 (Oliveira y Santos, 2005).

3. APLICACIÓN EMPÍRICA EN EL SECTOR EDUCATIVO

3.1. Datos y variables

En esta aplicación se ha utilizado información relativa a 80 Institutos públicos de Enseñanza Secundaria de la Comunidad Autónoma de Extremadura. De estos centros se dispone de un gran volumen de información agregada relativa al rendimiento académico de sus alumnos, los factores productivos empleados y las características de los estudiantes.

El output escolar se mide a través de los resultados obtenidos por los alumnos en la Prueba de Acceso a la Universidad (PAU). A partir de esa información se han construido dos variables: “NOTAS” (nota media de los alumnos que aprobaron la PAU) y “APROBADOS” (porcentaje de alumnos aprobados en la PAU respecto a los matriculados a principio del curso). En cuanto a los inputs controlables, se han seleccionado dos de las variables más utilizadas en la literatura como representativas de los factores gestionados por los centros: “PROFESORES” (número total de profesores en el centro por cada 100 alumnos) y “GASTOS” (gasto total por alumno durante el ejercicio).

Los inputs no controlables están representados por las características del alumnado¹². Esta información se ha captado mediante encuestas realizadas a los estudiantes de todos los centros. A partir de ellas se obtuvieron veinte variables que ofrecen información sobre una gran variedad de circunstancias que pueden afectar al rendimiento académico de los alumnos. Tras un análisis exhaustivo, fueron seleccionadas once de ellas, todas correlacionadas significativamente con las variables representativas del output. En la Tabla 1 del Anexo se define cada una de estas variables y en la Tabla 2 se muestran los principales estadísticos descriptivos de las mismas.

¹² La elección de centro escolar público por parte de los padres depende en una elevada cuantía de la cercanía del domicilio del alumno a cada centro escolar. Además, existen comisiones de escolarización dependientes del gobierno local que en última instancia asignan los alumnos a los centros. Estos hechos refuerza todavía más el carácter no controlable que el centro tiene de las características de sus alumnos.

Estas variables aparecen asociadas a aquellos factores identificados por la literatura como los más influyentes en el rendimiento del alumno como son el entorno socioeconómico y familiar, sus aptitudes y su actitud hacia el estudio. En todo caso, dado el todavía elevado número de variables y el hecho de que ninguna en particular representa con precisión a los factores considerados por la literatura y mencionados con anterioridad, decidimos aplicar el *Análisis de Componentes Principales (ACP)*¹³. Esta técnica permite sintetizar toda la información disponible en un número reducido de variables con una pérdida mínima de información. Dichos valores serán los que, finalmente, se incorporarán a la evaluación de eficiencia como variables representativas de los denominados inputs no controlables.

El *ACP* nos permitió identificar tres componentes que recogen el 78 % de la información original (Tabla 3 del Anexo). Posteriormente, a partir del análisis de la matriz de componentes rotados (Tabla 4), se puede comprobar qué variables se asocian con cada uno de los componentes. El primer factor está formado por las cinco variables relativas a los ingresos familiares, los estudios y la profesión de los padres, es decir, por aquéllas que determinan el “*entorno socioeconómico del alumno*”. El segundo, por su parte, está asociado con las variables que incorporan información sobre la capacidad del alumnado y la imagen que tienen los padres de ellos o, lo que es lo mismo, las “*aptitudes del alumno*”. Finalmente, el tercer componente, el de menor peso relativo, incorpora la información que aportan las dos variables restantes, las horas empleadas en el estudio por los alumnos y sus aspiraciones académicas, lo que puede interpretarse como un indicador de la “*actitud del alumno*” hacia el estudio.

En definitiva, en la aplicación empírica se utilizan dos variables relativas al output escolar (NOTAS y APROBADOS), dos que ofrecen información acerca de los inputs controlados por los centros (PROFESORES y GASTOS) y tres representativas de los inputs no controlables (los tres componentes

¹³ Smith y Mayston (1987) fueron los primeros que recomendaron el uso de esta técnica para reducir el número de factores no discretos en la evaluación de la eficiencia de centros escolares.

obtenidos a partir de las once variables descritas anteriormente). La Tabla 5 del Anexo recoge los principales estadísticos descriptivos de estas variables.

3.2. Resultados sin variables exógenas

Una vez descritas las características de la muestra y las variables que van a ser empleadas en la aplicación, el siguiente paso consiste en la realización de un DEA inicial en el que se obvia el efecto de las variables exógenas, a través del cual se obtiene una primera evaluación del comportamiento de los centros. El modelo DEA empleado en esta primera etapa tiene una orientación de maximización de outputs y rendimientos variables de escala. La elección de esta orientación se basa en que los gestores de los centros tienen mayor capacidad para mejorar el rendimiento de los alumnos que para reducir la cantidad de inputs escolares empleados, por lo que parece más razonable que sean evaluados en función de las mejoras potenciales que puedan conseguir en el output escolar. La utilización de rendimientos variables de escala resulta obligada por el uso de ratios en las variables representativas de los inputs y los outputs (Hollingsworth y Smith, 2003).

Los resultados de la Tabla 6 del Anexo muestran ocho unidades eficientes (IES 12, 18, 24, 31, 44, 52, 56 y 71) que pueden, a su vez, dividirse en dos grupos. El primero (IES 12, 31, 56 y 71), formado por unidades que han obtenido unos mejores resultados en la PAU; el segundo (IES 18, 24, 44 y 52), representado por aquéllos con menores inputs. En cuanto a las unidades ineficientes, las que obtienen unos índices más bajos son las que presentan los peores resultados en la PAU (IES 10, 37, 53 ó 64) o las que, pese a contar con mayores recursos, alcanzan unos resultados mediocres (IES 15, 48 ó 62).

Los índices de eficiencia obtenidos en esta primera evaluación sólo podrían ser considerados como válidos si todos los centros tuviesen un alumnado de similares características, lo que no es cierto a la vista de la información contenida en el Cuadro 1. Es necesario, por tanto, incorporar estas variables al proceso de evaluación.

3.3. Resultados con variables exógenas

La opción más sencilla para proceder a la incorporación de los factores no controlables es a través de un modelo de segunda etapa en el que se incluyan los índices de eficiencia (componente radial de la ineficiencia) como variable dependiente en una regresión Tobit en la que las tres variables que representan a las variables exógenas fueran los regresores:

$$\theta = \beta_0 + \beta_1 CP1 + \beta_2 CP2 + \beta_3 CP3 + u_i$$

En el caso de optar por esta opción, la estimación de los parámetros de la regresión puede realizarse mediante una regresión Tobit estándar, sin embargo resulta mucho más apropiado emplear la metodología propuesta por Simar y Wilson (2007), mediante la cual se obtienen estimaciones insesgadas y consistentes de los parámetros. En la Tabla 7 del Anexo se presentan los parámetros estimados por ambos modelos (Tobit estándar y el algoritmo propuesto por estos autores).

Sin embargo, estos nuevos índices estimados a partir de los valores de los parámetros estimados ($\hat{\theta}^* = \hat{\beta}_1^* CP1 + \hat{\beta}_2^* CP2 + \hat{\beta}_3^* CP3$) presentan un grave problema de interpretación, ya que no reflejan verdaderos objetivos de producción para las unidades, sino qué parte de los índices de eficiencia iniciales se explica por el efecto de las variables exógenas. Por tanto, si el objetivo es encontrar una medida que se aproxime al nivel de eficiencia de las unidades evaluadas, esta aproximación no parece ser la mejor alternativa.

Pese a lo útil que resulta conocer cuáles son las unidades más perjudicadas y favorecidas por su entorno, lamentablemente este enfoque no permiten fijar objetivos reales para los productores. Así, en la Tabla 6 del Anexo se puede comprobar que, siguiendo este enfoque, a todas las unidades se les asignan unos índices de eficiencia por encima de la unidad (o por debajo del valor 100), siendo imposible construir una frontera por no existir unidades eficientes. Del mismo modo, tampoco existen conjuntos de referencia para las

unidades, lo que impide que éstas puedan utilizar como espejo la gestión llevada a cabo por otros productores como vía para conseguir mejoras en su propia actividad.

Ante estos problemas de interpretación práctica que plantean los índices obtenidos con los modelos de segunda etapa, cabe plantearse la aplicación del modelo de cuatro etapas. Esta opción permite diferenciar varios efectos de los inputs no controlables sobre cada uno de los outputs obtenidos en el proceso productivo y, además, contribuye a la construcción de unos índices que reflejan cuáles son los verdaderos objetivos de producción de las unidades considerando su dotación de factores exógenos. Para ello, se plantean dos regresiones Tobit en las que las variables dependientes están representadas por los *slacks* totales de cada output (componente radial y no radial de la ineficiencia) obtenidos en la etapa inicial (Tabla 6 del Anexo) y los regresores son los factores exógenos (tres componentes extraídos del ACP):

$$\begin{aligned} SlacksAPROBADOS &= \beta_0 + \beta_1 PC1 + \beta_2 PC2 + \beta_3 PC3 + \varepsilon_i \\ SlacksNOTAS &= \beta_0 + \beta_1 PC1 + \beta_2 PC2 + \beta_3 PC3 + \varepsilon_i \end{aligned}$$

Para evitar el sesgo de la estimación Tobit estándar, se ha procedido a estimar ambas regresiones empleando la metodología propuesta en el apartado anterior para los modelos de cuatro etapas, basada en la aplicación de técnicas de *bootstrap*.

En la Tabla 8 del Anexo se muestran los resultados obtenidos tanto con el Tobit estándar, como empleando el algoritmo basado en las técnicas de remuestreo, según el cual se toman como referencia los errores ε_i de una distribución normal $[N(0, \hat{\sigma}_\varepsilon)]$ con un truncamiento a la izquierda en $-z_i \hat{\beta}$. Posteriormente, se calcula por máxima verosimilitud $SlackPassed_i^* = z_i \hat{\beta} + \varepsilon_i$ y $SlackMarks_i^* = z_i \hat{\beta} + \varepsilon_i$ para cada $i=1, \dots, n$. De esta forma se obtienen los estimadores $\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*$ en cada regresión.

La primera conclusión que puede extraerse a partir de los resultados de ambas regresiones es que los coeficientes varían significativamente con la aplicación del *bootstrap*. La segunda conclusión, de mayor relevancia que la anterior, es que todas las variables son claramente significativas (p-values muy cercanos a 0). Por tanto, la corrección en función de la dotación de inputs no controlables está totalmente justificada, ya que estas variables tienen un efecto notorio sobre la ineficiencia total de los centros.

Los resultados de la regresión, además de confirmar la influencia que tienen estos factores sobre los índices estimados, nos indican que existe una relación negativa entre la ineficiencia (medida mediante los *slacks* de las variables) y los componentes representativos del entorno (CP1), la aptitud (CP2) y la actitud (CP3). Por tanto, cualquier mejora que se produzca en las aptitudes de los estudiantes o en su entorno socioeconómico implica una reducción del índice de eficiencia final, esto es, la unidad se aproxima a la frontera de posibilidades de producción.

Tras la estimación de la regresión mediante *bootstrap*, la tercera etapa consiste en la corrección de los valores iniciales de los outputs. Este ajuste se realiza restando al valor original de cada output la diferencia entre el mayor valor estimado y el valor estimado para cada unidad. Una vez ajustados los valores iniciales, el proceso finaliza con la realización de un nuevo DEA con los inputs discrecionales originales y los outputs corregidos. De este modo, los nuevos índices estimados reflejan únicamente el efecto de la ineficiencia de los productores una vez que se ha eliminado el efecto de los factores externos. En la Tabla 6 del Anexo se presentan los índices obtenidos con la estimación mediante un modelo Tobit normal (Modelo de cuatro etapas original) y el modelo que incorpora el *bootstrap* para garantizar la consistencia de las estimaciones (Modelo de cuatro etapas con *bootstrap*).

La observación de estos resultados nos permite apreciar que los índices de eficiencia estimados con este modelo representan verdaderos objetivos de producción para las unidades, teniendo en cuenta su dotación de inputs no controlables. Así, los índices que aparecen en las dos últimas columnas

muestran amplias diferencias respecto a los obtenidos en el DEA inicial, siendo posible identificar unidades que incrementan su índice respecto a los iniciales al tenerse en cuenta su dotación de factores exógenos (desfavorable) (IES 2, 28, 36 o 60) y otras que lo disminuyen porque en la primera evaluación no se tuvo en cuenta su mejor dotación de inputs externos (IES 20, 59 o 79). Además, los nuevos índices construidos permiten construir una nueva frontera de unidades eficientes, muchas de las cuales se mantienen respecto a la primera evaluación (IES 12, 18, 31 o 44), unas pasan a ser eficientes con la inclusión de estas variables (IES 34 o 42) y otras dejan de serlo (IES 24).

Finalmente, la comparación entre los resultados obtenidos con el modelo de cuatro etapas original y el que se plantea en este trabajo refleja unas diferencias relativamente pequeñas, aunque bastante significativas en algunos casos, como ocurre con los centros 22 y 56 en los que según el modelo empleado una unidad será eficiente o no. No obstante, el modelo que aquí se plantea ofrece unos resultados mucho más consistentes al eliminar el sesgo de utilizar una única muestra aleatoria, motivo por el cual consideramos que su aplicación resulta mucho más apropiada en aquellos casos en los que se pretenda incorporar los inputs no controlables al análisis de eficiencia mediante DEA.

4. CONCLUSIONES

En el contexto de la evaluación de la eficiencia es preciso distinguir entre la verdadera ineficiencia en la que incurren las unidades productivas evaluadas y el efecto que tienen los factores que están al margen de su control, para lo cual resulta de gran utilidad el uso de modelos de segunda etapa. En este trabajo se examinan estos modelos prestando un especial interés a las propuestas más recientes desarrolladas en este ámbito, en las que se utiliza el *bootstrap* para mejorar la precisión de las estimaciones de los índices de eficiencia obtenidos.

A través de este análisis se ha detectado que este enfoque no permite construir índices de eficiencia que puedan ser interpretados como objetivos de

producción para las unidades evaluadas en aquellos casos en los que las variables exógenas participan como inputs en el proceso de producción. Con el propósito de superar esta limitación se ha desarrollado una nueva propuesta con una estructura similar a la del modelo de cuatro etapas propuesto por Fried *et al.* (1999), en el que también se emplea el *bootstrap* para evitar posibles sesgos en el cálculo de los índices, mediante la cual se construyen índices de eficiencia que son interpretables.

Con el doble propósito de comprobar la validez de esta nueva propuesta y de comparar sus resultados con los que se obtienen al aplicar un modelo de segunda etapa se ha llevado a cabo una aplicación en el ámbito de la enseñanza secundaria. Para ello se ha utilizado una base de datos relativa a un conjunto de institutos de enseñanza secundaria sobre los que se dispone de una fuente de información muy rica sobre los factores no controlables que influyen sobre los rendimientos del proceso, representados por las características de su alumnado.

El análisis de los resultados permite apreciar que, a diferencia de lo que ocurre con los modelos de segunda etapa, los índices de eficiencia que se obtienen con la aplicación del modelo propuesto reflejan objetivos de producción para las unidades teniendo en cuenta el tipo de alumnado con el que cuentan.

Referencias bibliográficas

Afonso, A. y St. Aubyn, M. (2006): "Cross-country Efficiency of Secondary Education Provision: a Semi-parametric Analysis with Non-discretionary Inputs", *Economic Modelling*, vol. 23 (3), pp. 476-491.

Fried, H.O., Lovell, C.A.K. y Van Den Eeckaut, P. (1993): "Evaluating the Performance of US Credit Unions", *Journal of Banking and Finance*, vol. 17, pp. 251-265.

Fried, H., Schmidt, S. y Yaisawarng, S. (1999): "Incorporating the Operating Environment into a Nonparametric Measure of Technical Efficiency", *Journal of Productivity Analysis*, 12, pp. 249-267.

Hanushek, E. (1986): "The economics of Schooling", *Journal of Economic Literature*. Vol 24, nº3, 1141-1171.

Hanushek, E. (1997): "Assessing the effects of school resources on student performance: An update", *Educational Evaluation and Policy Analysis*. 19, 141-164.

Hanushek, E. A. (2003). "The failure of *input*-based schooling policies", *The Economic Journal*, 113, 64-98.

Hirschberg, J. y Lloyd, P. (2002): "Does the Technology of Foreign-Invested Enterprises Spill Over to Other Enterprises in China? An Application of Post-DEA Bootstrap Regression Analysis", in Lloyd, P.J. y ZANG, X.G. (eds.): *Modelling the Chinese Economy*, London: Edward Elgar Press.

Hollingsworth, B. y Smith, P. (2003): "Use of ratios in Data Envelopment Analysis", *Applied Economics Letters*, vol. 10, pp. 733-735.

Kirjavainen, T. y Loikkanen, H.A. (1998): "Efficiency Differences of Finnish Senior Secondary Schools: An Application of DEA and Tobit Analysis", *Economics of Education Review*, vol. 17 (4), pp. 377-394.

Mccarty, T. y Yaisawarng, S (1993): "Technical Efficiency in New Jersey School Districts", in Fried, H., Lovell, C.A.K. y Schmidt, S. (ed.): *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York.

Oliveira, M.A. y Santos, C. (2005): "Assessing School Efficiency in Portugal Using FDH and Bootstrapping", *Applied Economics*, 37, pp. 957-968.

Ray, S.C. (1991): "Resource Use Efficiency in Public schools: A Study of Connecticut Data", *Management Science*, vol. 37, nº 12, págs. 1.620-1.628.

Simar, L. y Wilson, P.W. (1999a): "Some Problems with the Ferrier/Hirschberg Bootstrap Idea", *Journal of Productivity Analysis*, 11, pp. 67-80.

Simar, L. y Wilson, P.W. (1999b): "Of Course We Can Bootstrap DEA Scores! But Does it Mean Anything? Logic Trumps Wishful Thinking", *Journal of Productivity Analysis*, 11, pp. 93-97.

Simar, L. y P. W. Wilson (2007): *Estimation and Inference in Two-Stage, Semiparametric Models of Production Processes*, Discussion Paper 0307, Institut de Statistique, Université Catholique de Louvain.

ANEXO

Tabla 1. Definición de las variables representativas de los INC

Nombre de la variable	Definición
APROTOD0	Alumnos que el curso pasado aprobaron todas las asignaturas
NOREPET	Alumnos que nunca han repetido curso
EXPDTE	Alumnos que aprobaron el pasado curso todas las asignaturas entre junio y septiembre con buenas notas
HORAS	Alumnos que estudian más de 10 horas semanales
ASPIRAC	Alumnos que desean cursar estudios universitarios
CONFIPAD	Alumnos cuyos padres tienen confianza en su éxito académico
INGRESOS	Alumnos cuyos padres tienen unos ingresos elevados
ESTPADRE	Alumnos cuyo padre tiene estudios superiores
ESTMADRE	Alumnos cuya madre tiene estudios superiores
PROFPADRE	Alumnos cuyo padre tiene una profesión cualificada
PROFMADRE	Alumnos cuya madre tiene una profesión cualificada.

Tabla 2. Estadísticos descriptivos de las variables representativas de los INC

	Mínimo	Máximo	Media	Desv. Tip.
APROTOD0	14.29	81.11	55.4881	12.07099
NOREPET	22.73	91.67	69.1862	12.45518
EXPDTE	8.91	55.33	30.8511	9.58687
HORAS	0.00	63.89	29.8171	11.26075
ASPIRAC	40.00	98.60	78.5088	11.53260
CONFIPAD	25.35	83.78	58.0298	8.70328
INGRESOS	2.00	49.41	18.8999	10.33527
ESTPADRE	0.00	47.01	12.9429	9.12190
ESTMADRE	0.00	38.03	11.3120	8.56715
PROFPADRE	6.25	74.81	36.2889	14.71205
PROFMADRE	0.00	44.83	16.3718	9.97005

Tabla 3: Análisis de Componentes Principales de los inputs no controlables

	Autovalores Iniciales			Suma de las saturaciones al cuadrado de la rotación		
Componente	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5.145	46.777	46.777	5.145	46.777	46.777
2	2.364	21.489	68.265	2.364	21.489	68.265
3	1.092	9.927	78.192	1.092	9.927	78.192
4	.665	6.048	84.241			
5	.459	4.169	88.409			
6	.371	3.373	91.782			
7	.271	2.465	94.247			
8	.208	1.887	96.135			
9	.193	1.757	97.892			
10	.152	1.381	99.273			
11	.079	.727	100.000			

Tabla 4: Matriz de componentes rotados

	COMPONENTE		
	1	2	3
INGRESOS	0,888	0,112	0.166
ESTPADRE	0,915	0,205	0.037
ESTMADRE	0,916	0,250	-0,042
PROFPADRE	0,899	0,020	0.190
PROFMADRE	0,901	0,114	0.049
APROTOD0	0.107	0.780	0.374
EDAD	0,072	0,690	0.482
EXPEDIENTE	0,085	0,883	0.126
CONFIPAD	0,274	0,773	-0.194
HORAS	0,057	0,088	0.909
ASPIRACIONES	0,255	0,416	0.505

Método de rotación: Normalización Varimax con Kaiser

Tabla 5: Estadísticos descriptivos de las variables utilizadas en el análisis

	MÍNIMO	MÁXIMO	MEDIA	DESV. TÍP.
NOTAS	50,28	72,52	59,59	4,0987
APROBADOS	12,50	86,42	40,33	14,0819
PROFESORES	7,68	13,21	9,64	1,1490
GASTOS	65,63	543,86	161,19	66,8621
CP 1 (Entorno)	0,00	4,60	1,3050	1,0000
CP 2 (Aptitud)	0,00	5,06	2,6922	0,9993
CP 3 (Actitud)	0,00	5,27	1,8900	1,0000

Tabla 6: Eficiencia radial (scores) y total (slacks) obtenidos en el DEA inicial

	SCORES		Slacks APROBADOS	Slacks NOTAS	Slacks PROFESORES	Slacks GASTOS
IES 1	83,44	1,166	26,18	12,01	3,48	73,08
IES 2	76,43	1,236	25,71	17,09	4,56	94,24
IES 3	87,20	1,128	6,35	9,28	1,32	32,09
IES 4	75,88	1,241	23,84	17,49	0,74	32,09
IES 5	84,23	1,158	8,32	11,37	0,79	37,16
IES 6	90,17	1,098	5,26	6,68	0,53	0,00
IES 7	85,12	1,149	8,48	10,50	0,63	0,00
IES 8	93,46	1,065	3,29	4,40	0,91	0,00
IES 9	85,85	1,142	8,34	9,71	0,33	0,00
IES 10	69,99	1,300	27,62	21,75	1,76	0,00
IES 11	92,38	1,076	3,05	5,04	0,86	0,00
IES 12	100,00	1,000	0,00	0,00	0,00	0,00
IES 13	77,92	1,221	16,50	15,92	1,40	0,00
IES 14	79,59	1,204	15,71	14,80	2,64	26,02
IES 15	75,44	1,246	17,79	17,81	2,58	59,98
IES 16	85,05	1,150	9,36	10,60	0,52	26,43
IES 17	92,30	1,077	5,12	5,42	0,54	28,33
IES 18	100,00	1,000	0,00	0,00	0,00	0,00
IES 19	79,86	1,201	11,59	14,41	2,33	34,21
IES 20	84,94	1,151	24,97	10,92	1,72	39,11
IES 21	82,68	1,173	8,68	12,26	1,28	0,00
IES 22	89,01	1,110	5,15	7,34	0,21	0,00
IES 23	79,41	1,206	24,27	14,93	1,49	7,15
IES 24	100,00	1,000	0,00	0,00	0,00	0,00
IES 25	93,49	1,065	8,13	4,72	1,75	0,48
IES 26	83,58	1,164	10,25	11,52	1,15	0,00
IES 27	84,65	1,154	10,02	10,45	0,06	0,00
IES 28	82,75	1,173	34,04	12,34	2,33	0,00
IES 29	83,11	1,169	8,30	11,98	2,18	0,00
IES 30	80,96	1,190	23,39	13,55	0,69	0,00
IES 31	100,00	1,000	0,00	0,00	0,00	0,00
IES 32	85,64	1,144	9,34	10,14	1,03	20,05
IES 33	83,98	1,160	26,82	11,62	1,22	7,09
IES 34	90,07	1,099	11,96	7,20	0,85	2,40
IES 35	83,88	1,161	24,02	11,65	1,17	0,00
IES 36	79,41	1,206	23,46	14,93	2,19	169,26
IES 37	69,33	1,307	16,61	22,24	2,91	83,18
IES 38	85,43	1,146	8,09	10,19	0,06	0,00
IES 39	81,83	1,182	12,81	12,72	0,21	96,06
IES 40	83,77	1,162	30,24	11,57	2,50	0,00
IES 41	80,44	1,196	8,58	13,23	0,62	0,00
IES 42	91,50	1,085	7,35	7,72	0,91	23,74
IES 43	89,23	1,108	7,37	7,81	2,10	60,04
IES 44	100,00	1,000	0,00	0,00	0,00	0,00
IES 45	91,33	1,087	4,51	6,25	1,09	26,81
IES 46	79,43	1,206	12,95	14,58	0,46	13,96
IES 47	88,67	1,113	5,67	7,64	1,88	0,00

IES 48	74,47	1,255	13,90	18,35	5,10	19,60
IES 49	77,36	1,226	19,04	16,42	2,20	92,92
IES 50	83,78	1,162	8,39	11,71	2,90	51,67
IES 51	82,68	1,173	9,31	12,46	2,14	14,51
IES 52	100,00	1,000	0,00	0,00	0,00	0,00
IES 53	74,58	1,254	12,48	18,43	1,85	72,69
IES 54	88,01	1,120	8,82	8,68	3,54	0,00
IES 55	86,53	1,135	7,31	9,69	0,82	15,51
IES 56	100,00	1,000	0,00	0,00	0,00	0,00
IES 57	84,58	1,154	11,54	11,01	1,73	0,00
IES 58	86,66	1,133	8,93	8,46	0,00	0,00
IES 59	92,49	1,075	3,20	4,96	0,00	0,00
IES 60	79,80	1,202	13,77	14,64	0,27	0,00
IES 61	81,66	1,183	14,86	13,30	2,46	23,50
IES 62	75,74	1,243	36,54	17,59	4,77	192,54
IES 63	83,01	1,170	13,19	12,32	1,78	26,32
IES 64	74,14	1,259	13,36	18,63	1,86	0,00
IES 65	79,81	1,202	10,44	14,64	3,92	1,44
IES 66	75,25	1,248	21,77	17,95	2,02	25,00
IES 67	84,77	1,152	8,84	10,46	0,93	0,00
IES 68	76,86	1,231	14,07	16,56	2,75	0,00
IES 69	80,54	1,195	9,95	14,06	2,74	49,89
IES 70	77,95	1,221	13,33	15,99	4,37	33,30
IES 71	100,00	1,000	0,00	0,00	0,00	0,00
IES 72	79,56	1,204	14,99	14,82	2,90	395,47
IES 73	77,10	1,229	15,04	16,61	2,95	44,05
IES 74	92,11	1,079	4,13	5,49	0,18	0,00
IES 75	87,45	1,126	6,42	9,07	2,73	66,32
IES 76	78,34	1,217	13,17	14,33	0,00	0,00
IES 77	83,67	1,163	17,13	11,84	2,67	119,60
IES 78	95,73	1,043	12,11	3,02	2,77	0,00
IES 79	93,24	1,068	4,10	4,31	1,68	0,00
IES 80	85,91	1,141	25,51	10,22	0,81	49,70

Tabla 7: Parámetros estimados en el modelo de regresión

VARIABLE DEPENDIENTE	Índices obtenidos en el DEA inicial	
	Tobit estándar	SW (2007)
C	1,2985** (0,0234)	1,2497** (0,0198)
CP1	-0,0378** (0,0064)	-0,0246** (0,0062)
CP2	-0,0235** (0,0064)	-0,0120** (0,0061)
CP3	-0,0187** (0,0064)	-0,0120** (0,0061)

Nota: Los errores estándar de los parámetros estimados se muestran entre paréntesis.
Los supraíndices ** indican que el parámetro estimado es significativo al 99 %.

Tabla 8: Parámetros estimados en el modelo de cuatro etapas

VARIABLE DEPENDIENTE	SLACKS APROBADOS (DEA INICIAL)		SLACKS NOTAS (DEA INICIAL)	
	Tobit estándar	Tobit con <i>Bootstrap</i>	Tobit estándar	Tobit con <i>Bootstrap</i>
C	32,608** (2,9177)	29,992** (1,3615)	22,472** (1,8694)	20,031** (1,4672)
CP1	-3,835 (0,8198)	-3,411** (0,3764)	-3,095** (0,5223)	-2,464** (0,4035)
CP2	-4,342 (0,8056)	-4,303** (0,3775)	-1,982** (0,5155)	-1,816** (0,4040)
CP3	-2,217 (0,7790)	-1,613** (0,3638)	-1,366** (0,5044)	-1,473** (0,4037)

Nota: Los errores estándar de los parámetros estimados se muestran entre paréntesis.
Los supraíndices ** indican que el parámetro estimado es significativo al 99 %.

Tabla 9: Índices de eficiencia estimados con los distintos modelos

	MOD. REGRESIÓN ORIGINAL	MOD. SIMAR Y WILSON (2007)	MOD. 4 ETAPAS ORIGINAL	MOD. 4 ETAPAS BOOTSTRAP
IES 1	79,65	83,07	92,42	92,95
IES 2	77,31	79,71	87,85	88,38
IES 3	88,58	85,29	87,05	88,68
IES 4	83,24	79,03	82,39	82,41
IES 5	85,43	82,71	88,65	89,50
IES 6	90,00	87,61	88,39	89,02
IES 7	90,87	83,04	81,65	83,20
IES 8	88,70	90,84	93,99	92,87
IES 9	86,59	84,52	92,18	90,45
IES 10	78,76	77,99	78,71	78,61
IES 11	93,90	90,26	84,94	87,80
IES 12	89,12	90,50	100,00	100,00
IES 13	82,06	79,82	84,26	83,97
IES 14	83,47	80,87	83,78	84,73
IES 15	79,95	79,24	84,76	84,49
IES 16	87,01	82,02	86,45	86,91
IES 17	86,97	87,37	95,67	96,07
IES 18	88,14	90,28	100,00	100,00
IES 19	79,47	80,02	90,56	89,95
IES 20	93,63	83,92	77,25	79,14
IES 21	80,81	82,68	92,27	92,10
IES 22	81,81	86,79	100,00	99,34
IES 23	83,63	80,48	84,37	84,24
IES 24	96,08	95,57	90,66	92,75
IES 25	84,33	90,27	100,00	100,00
IES 26	83,12	81,35	89,93	89,41
IES 27	84,38	81,30	89,77	89,95
IES 28	75,12	82,41	97,21	96,92
IES 29	84,00	82,78	86,34	86,16
IES 30	83,17	81,53	88,28	89,04
IES 31	90,43	91,95	100,00	100,00
IES 32	84,31	82,20	91,42	90,13
IES 33	79,01	83,05	96,16	96,10
IES 34	82,69	87,29	100,00	100,00
IES 35	83,79	82,95	89,68	90,88
IES 36	74,13	80,87	96,46	95,98
IES 37	81,20	78,27	75,63	75,34
IES 38	87,13	83,17	89,66	90,89
IES 39	84,10	78,61	87,83	86,85
IES 40	83,53	83,28	86,50	86,63
IES 41	81,92	81,88	89,03	87,55
IES 42	81,70	83,75	100,00	100,00
IES 43	84,56	86,88	94,26	94,29
IES 44	86,24	88,49	100,00	100,00
IES 45	87,30	88,19	93,65	94,76
IES 46	85,10	78,55	83,04	82,58
IES 47	86,90	86,92	88,78	88,87

IES 48	80,72	78,28	81,88	81,57
IES 49	89,16	80,05	73,20	74,11
IES 50	90,78	82,91	78,10	80,22
IES 51	86,26	81,86	84,52	85,15
IES 52	93,91	94,67	100,00	100,00
IES 53	79,80	78,91	84,43	83,96
IES 54	85,20	85,78	91,88	92,44
IES 55	86,60	84,13	89,89	90,60
IES 56	98,79	95,98	98,81	100,00
IES 57	83,00	83,69	89,98	90,58
IES 58	83,82	79,03	91,91	91,13
IES 59	97,55	89,67	82,70	85,02
IES 60	83,50	80,10	92,86	92,19
IES 61	89,64	81,97	77,18	79,06
IES 62	76,61	79,49	87,10	87,60
IES 63	85,04	82,61	86,02	86,38
IES 64	81,69	78,13	80,45	80,63
IES 65	82,74	80,58	85,62	85,94
IES 66	83,30	79,11	79,25	79,61
IES 67	85,61	83,18	87,07	88,00
IES 68	77,84	79,77	87,28	86,94
IES 69	84,29	80,99	84,70	86,34
IES 70	82,09	80,03	84,91	83,90
IES 71	85,94	89,13	100,00	100,00
IES 72	81,94	80,98	86,54	86,56
IES 73	80,27	79,85	85,61	85,95
IES 74	91,48	89,07	91,06	92,91
IES 75	83,79	85,31	94,59	94,35
IES 76	85,39	76,70	79,83	80,31
IES 77	82,01	83,24	90,66	91,51
IES 78	83,06	92,73	100,00	100,00
IES 79	93,96	88,10	82,42	84,72
IES 80	86,35	84,24	89,38	91,94