

EFICIENCIA EN LA INSPECCIÓN DE HACIENDA *

XOSÉ MANUEL GONZÁLEZ

DANIEL MILES

Universidade de Vigo

El objetivo de este trabajo es realizar un análisis de eficiencia de las Unidades Regionales de Inspección mediante el análisis envolvente de datos (DEA) y estudiar la precisión estadística de la estimación de eficiencia mediante técnicas *bootstrap*. La necesidad de racionalizar el uso de los recursos públicos ha hecho que el criterio de eficiencia económica haya recobrado protagonismo en los últimos años en este ámbito. Prueba de ello es el importante número de trabajos aplicados tratando de cuantificar eficiencia utilizando, fundamentalmente, el método *DEA*. Sin embargo, este análisis únicamente permite la estimación puntual de eficiencia técnica, sin que se pueda analizar la precisión de tal estimación. La novedad de este trabajo reside precisamente en la aplicación de dos estimadores *bootstrap* desarrollados recientemente en la literatura, para construir intervalos de confianza de la eficiencia técnica de cada unidad. De este análisis se observa que, si bien la estimación puntual de eficiencia para cada unidad es distinta, las diferencias entre unidades no son estadísticamente significativas para algunas de ellas.

Palabras clave: eficiencia, DEA, *bootstrap*.

Clasificación JEL: C14, D24, H59.

En los últimos años se han realizado un importante número de trabajos aplicados interesados en analizar la eficiencia técnica de unidades productivas [ver Cuenca (1994), Pedraja y Salinas (1996), Fuentesalz, Marcuello y Urbina (1996), Bosch, Pedraja y Suárez (1998), Mancebón y Bandrés (1999), entre otros]. La principal motivación de estos trabajos se encuentra en la necesidad de saber si se está haciendo un uso eficiente de los recursos disponibles en distintos centros productivos, fundamentalmente en aquellos vinculados a servicios públicos. Para ello, dentro de los métodos existentes, se ha utilizado fundamentalmente el análisis envolvente de datos.

Uno de los problemas de este método estriba en que no permite obtener medidas de precisión de la estimación de la eficiencia. Es decir, este procedimiento

(*) Deseamos agradecer a Francisco Pedraja sus amables sugerencias, así como las observaciones efectuadas por un evaluador anónimo, que han permitido mejorar sustancialmente este trabajo. No obstante, de persistir algún error sería responsabilidad de los autores.

únicamente reporta el estimador puntual de la eficiencia, no permitiendo analizar la sensibilidad de tales estimaciones, por lo que no es posible saber si dos niveles de eficiencia son significativamente distintos desde el punto de vista estadístico. Claramente, obtener la información estadística de esos estimadores puede resultar importante para complementar y/o relativizar las conclusiones económicas [ver McCloskey y Ziliak (1996)].

Recientemente, diversos trabajos han analizado las propiedades estadísticas de los estimadores *DEA*. Por ejemplo Korostelev, Simar y Tsybakov (1995) establecen la consistencia del estimador y derivan su tasa de convergencia. Además, diversos autores han propuesto aproximar la distribución del estimador mediante la distribución del estimador *bootstrap* [ver Gstach (1995), Lothgren y Tambour (1997), Simar y Wilson (1998), entre otros]. Dado que, en general, es difícil derivar analíticamente la distribución del estadístico *bootstrap*, estos trabajos proponen aproximarla mediante experimentos de Monte Carlo.

El objetivo de este trabajo es estimar la eficiencia técnica en un ámbito concreto como son las Unidades Regionales de Inspección (*URI's*) y complementar la información de la estimación puntual con la generada por la aproximación *bootstrap*. Este ejercicio nos sirve sobre todo para adentrarnos en una de las cuestiones que consideramos más relevantes en este tipo de análisis como es conocer la precisión estadística de la estimación de eficiencia. ¿La diferencia de la estimación puntual de la eficiencia entre dos unidades se debe a que los verdaderos parámetros de eficiencia son distintos entre estas dos unidades, o se debe simplemente a ruido muestral? Este aspecto no ha sido tratado en los trabajos empíricos de estimación de eficiencia citados anteriormente y parece de suma relevancia a la hora de juzgar las diferencias de eficiencia entre unidades productivas [ver González y Miles (1999)].

El trabajo se compone de 2 secciones. En la primera comentamos brevemente la técnica del análisis envolvente de datos, explicamos sus principios básicos de cálculo, ponemos de manifiesto su debilidad respecto a la fiabilidad de las estimaciones que suministra y presentamos la contribución de las técnicas de *bootstrapping* para solucionar este problema. En la segunda sección aplicamos el contenido que detallamos en la primera parte al ámbito específico de las *URI's*. La actuación de estas unidades tiene interés tanto desde la vertiente del gasto público, en lo que se refiere a racionalización del uso de recursos, como desde la vertiente del ingreso, ya que gran parte de su quehacer se dirige a minimizar el nivel de fraude. Por eso, en la segunda parte nos detenemos en comentar las peculiaridades del sector, seleccionamos las variables que nos aproximen al funcionamiento de su actividad y realizamos la medición de eficiencia, estudiando la precisión de las estimaciones y verificamos la utilidad de la técnica *bootstrap*¹. El trabajo finaliza con un último apartado donde se sintetizan las conclusiones obtenidas.

(1) Todos los programas y procedimientos utilizados en este trabajo han sido programados por los autores y están disponibles para quienes lo soliciten.

1. LA CUANTIFICACIÓN DE LA EFICIENCIA PRODUCTIVA

El estudio de la eficiencia de las entidades públicas se sitúa dentro del campo de la teoría de la producción y parte de la consideración de que los centros encargados de suministrar servicios públicos se pueden asimilar a organizaciones productivas que emplean diversos factores para obtener varios productos mediante una determinada tecnología de producción².

El análisis de eficiencia técnica de estos centros o unidades productivas se basa en conocer el conjunto de producción y su correspondiente función frontera. Tanto uno como otra se infieren de los datos correspondientes a las observaciones muestrales, con lo que se delimita una frontera correspondiente a la mejor práctica, en la que las observaciones situadas sobre dicha frontera presentan un nivel de eficiencia del cien por cien. Además, el análisis debe ajustarse a las características de incertidumbre y desconocimiento que rodea a la tecnología de producción pública, por lo que se recomienda realizarlo mediante aproximaciones no paramétricas, ya que estos modelos no requieren una especificación previa de la función de producción, sino que es suficiente con definir ciertas propiedades formales que debe satisfacer la tecnología de producción³.

El marco teórico de la estimación no paramétrica de eficiencia parte de considerar que la actividad productiva de un centro puede analizarse por medio del conjunto de producción alcanzable, a partir de pares *input-output* (x, y) ,

$$\Psi = \{(x, y) \in \mathbb{R}_+^{m+s} \mid x \text{ produce } y\}$$

Para un nivel de *output* dado, y , se define el conjunto de *inputs* que permiten alcanzar este nivel de producción como la correspondencia,

$$L(y) = \{x \in \mathbb{R}_+^m \mid (x, y) \in \Psi\}$$

Sobre estos conjuntos se realizan los supuestos habituales de ausencia de producción gratuita, convexidad y disponibilidad débil de *inputs* y *outputs* [Shepard, 1970)]. La frontera de producción eficiente se define como aquel subconjunto de $L(y)$ tal que con menores *inputs* no se puede alcanzar el mismo *output*.

$$Fr[L(y)] = \{x \mid x \in L(y), \theta x \notin L(y), \forall \theta \in (0, 1)\}$$

(2) Trabajos pioneros en el campo de la teoría de la producción que abordaron el problema de la eficiencia económica y su medida fueron los de Koopmans (1951), Debreu (1951) y Farrell (1957).

(3) La metodología alternativa a la aproximación no paramétrica consiste en plantear modelos paramétricos en los que se especifica *a priori* una forma funcional (ej. Cobb-Douglas, translog, etc) y se estiman sus parámetros de manera que las observaciones queden sobre o por debajo de las funciones de producción o costes. Con estos modelos se corre el riesgo de no representar bien la relación que están suponiendo y de calcular la ineficiencia con respecto a una formulación muy restrictiva del conjunto de producción.

De esta forma, si se dispone de una muestra aleatoria, $X = \{(x_k, y_k) \mid k = 1, \dots, n\}$ de n unidades productivas homogéneas y, además, esta muestra se extrae de una población con función de distribución desconocida, F , utilizando algún método M , se pueden estimar $\hat{\psi}$, $\hat{L}(y)$ y $\hat{Fr}[L(y)]$, y a partir de aquí determinar⁴:

$$\hat{\theta}_{kn} = \min \left\{ \theta \mid \theta x_k \in \hat{L}(y_k) \right\}$$

Con la resolución de estos programas tenemos una información particularizada de las unidades o centros analizados que puede ser utilizada por los gestores públicos en una fase inicial de diseño de incentivos. Así, mediante este análisis es posible disponer de índices individualizados de eficiencia de las unidades productivas, conocer los grupos de referencia eficientes con los que se debe comparar cada centro ineficiente y establecer los objetivos de consumo y producción para aquellos centros evaluados como ineficientes⁵.

No obstante, es necesario mantener alguna prudencia con la información obtenida y asegurarse una cierta homogeneidad en el conjunto de centros evaluados. Hay que tener en cuenta que esta metodología es muy sensible a los errores en los datos y a los *shocks* aleatorios o influencias de factores fuera del control de los centros. Además, la excesiva flexibilidad de la técnica, en el sentido de no existir ninguna restricción sobre las ponderaciones, puede condicionar los resultados y permitir que la eficiencia se pueda evaluar sin considerar todos los *inputs* y *outputs*, incorporando ponderaciones nulas en factores que pueden ser relevantes. Así, si un centro, por tener objetivos individuales o circunstancias particulares, es superior al resto en una *ratio output/input* (centro extremo), será evaluado como eficiente ya que podrá basar su análisis exclusivamente en dicha *ratio*, asignando ponderaciones nulas a todos los demás factores.

Estas cuestiones plantean cierta debilidad en la precisión estadística de la estimación *DEA* y aconsejan alguna cautela a la hora de establecer comparaciones entre centros productivos, cuando se toma únicamente como referencia las estimaciones puntuales. Precisamente para solucionar este problema y analizar la sensibilidad del estimador de la eficiencia ante variaciones muestrales, recientemente se han propuesto diversas aproximaciones *bootstrap*⁶. Estos métodos *bootstrap* permiten obtener intervalos de confianza para el verdadero valor del parámetro de eficiencia y esta nueva información ayuda a matizar las comparaciones de eficiencia entre unidades productivas

A modo introductorio, y siguiendo a Efron y Tibshirani, para realizar inferencia estadística sobre un cierto parámetro θ se obtiene una muestra $X = (x_1, \dots, x_n)$ de una población con distribución F desconocida y con dicha muestra se calcula

(4) El método de estimación fue formalizado por Charnes, Cooper y Rhodes (1978, 1979).

(5) Una valoración pormenorizada de las ventajas e inconvenientes del análisis envolvente de datos puede encontrarse en Pedraja y Salinas (1994) y un artículo en el que se muestra la riqueza de resultados del *DEA* es el de Mancebón (1998).

(6) Ver Efron y Tibshirani (1993) como lectura introductoria al *bootstrap*, o Davison y Hinkley (1997).

una estimación $\hat{\theta}$ a partir de un estimador $\theta = s(x)$. La distribución de este estimador dependerá de F , desconocida. Ahora bien, F_n , la función de distribución empírica es conocida, pues la calculamos a partir de la muestra X . De esta distribución empírica podemos obtener una nueva muestra X^* , una muestra *bootstrap*, y recalcular el valor del estadístico para obtener θ^* . Si este último procedimiento se repite B veces, tendríamos B valores del estadístico $\theta = s(x)$ evaluados en las muestras *bootstrap*. Dado que tenemos B estimaciones, podemos por ejemplo utilizarlas para calcular la precisión del estadístico $\theta = s(x)$, o el sesgo de este estadístico, e.g. estamos aproximando la distribución del estimador. Por una parte, lo que hemos comentado en este párrafo es la aproximación Monte Carlo a la distribución *bootstrap* de un estimador. Por otra, el procedimiento *bootstrap* no es adecuado en todos los contextos [ver Efron y Tibshirani (1993, pág. 87) o Davison y Hinkley (1997, pág. 37)].

En concreto, la propuesta de estos trabajos es aproximar la distribución del estadístico de eficiencia que depende de una función de distribución desconocida de la población. Es decir, como no es posible evaluar la función de distribución del estimador, $G_n(z, F) = \Pr(\hat{\theta}_{kn} \leq z | F)$, dado que F no es conocida, una forma de resolver esta cuestión es sustituir F por un estimador consistente como, por ejemplo, F_n la función de distribución empírica. Esto da lugar al *bootstrap*, que consiste en aproximar $G_n(z, F)$ a través de $G_n(z, F_n)$ [Horowitz (1995)].

Así, si F_n es un estimador consistente de F y si $X^* = \{(x_i^*, y_i^*) | i = 1, \dots, n\}$ es una muestra generada a partir de F_n , con esta última muestra y utilizando un método se pueden estimar $\hat{\psi}^*$, $\hat{L}(y)^*$ y $\hat{F}r[L(y)]^*$, y obtener posteriormente una estimación de la eficiencia de cada centro basada en X^* ,

$$\hat{\theta}_{kn}^* = \min \left\{ \theta \mid \theta x_k \in \hat{L}(y_k)^* \right\}$$

Al ser F_n conocido, las distribuciones anteriores son conocidas, pero su obtención analítica resulta bastante dificultosa. La forma práctica de solventar esta complicación consiste en aproximar esta distribución a través de simulaciones de Monte Carlo. Es decir, utilizando F_n , se generan muestras, X_b^* , para $b = 1, \dots, B$ y se estiman, $\hat{\theta}_{knb}^*$, para cada muestra y unidad productiva. Por tanto, al final de estas B replications se tendrá, para cada unidad productiva, (x_k, y_k) , una sucesión de valores $\left\{ \hat{\theta}_{knb}^* \right\}_{b=1}^B$.

La distribución empírica de $\left\{ \hat{\theta}_{knb}^* \right\}_{b=1}^B$ es la aproximación de Monte Carlo de la distribución del estimador *bootstrap* condicional en X . La media de esa distribución *bootstrap* $\bar{\theta}_{kn}^*$ se podría utilizar como una aproximación del estimador *DEA*, aunque esta aproximación del estimador no estaría libre de sesgo.

El procedimiento *bootstrap* permite también obtener un estimador insesgado, pues es posible obtener el sesgo de $\hat{\theta}_{kn}$, dado por $d_k = E(\hat{\theta}_{kn}) - \theta_k$, a partir de $\hat{d}_k^* = E(\hat{\theta}_{kn}^*) - \theta_k$, donde $E(\hat{\theta}_{kn}^*)$ puede estimarse por su análogo *bootstrap* $\bar{\theta}_{kn}^* = \sum_{b=1}^B \hat{\theta}_{knb}^* / B$. De esta forma, al introducir el sesgo definido como la diferencia entre la media *bootstrap* y el valor inicial del estimador, se puede calcular un estimador corregido, $\tilde{\theta}_{kn} = \hat{\theta}_{kn} - \hat{d}_k^*$, donde $\hat{d}_k^* = \bar{\theta}_{kn}^* - \hat{\theta}_{kn}$.

No obstante, es necesario señalar que esta corrección en la práctica también puede resultar relativamente problemática ya que, debido a la variabilidad del sesgo, se pueden obtener valores económicamente incorrectos, aunque estadísticamente válidos [Efron y Tibshirani (1993, pág. 138)].

Si queremos obtener una indicación de la precisión del estimador DEA original, debemos tener información de la desviación típica del estimador y de los intervalos de confianza que se puedan establecer. La desviación típica del estimador θ_{kn} puede aproximarse utilizando la desviación típica del estimador *bootstrap*,

$$\hat{\sigma} = \left(\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{knb}^* - \bar{\theta}_{kn}^*)^2 \right)^{1/2}$$

Por último, los intervalos de confianza de θ_{kn} se pueden aproximar utilizando la distribución empírica de $\hat{\theta}_{knb}^*$, una vez corregido este valor del sesgo. Es decir, con la obtención de $\tilde{\theta}_{knb}^* = \hat{\theta}_{knb}^* - 2\hat{d}_k^*$, para $b = 1, \dots, B$, se puede aproximar un intervalo de confianza mediante $(\tilde{\theta}_{knb}^{*(\alpha)}, \tilde{\theta}_{knb}^{*(1-\alpha)})$, donde $\tilde{\theta}_{knb}^{*(\alpha)}$ indica el 100 α percentil de la función de densidad $\hat{\theta}_{knb}^*$ para $b = 1, \dots, B$. Los percentiles serán los valores que acumulan, por ejemplo, el 5% y el 95% de la frecuencia, una vez que se ordenan los valores de eficiencia *bootstrap* [ver Simar y Wilson (1998, pág. 52)].

Para comprobar la operatividad de estos procedimientos, en la siguiente sección aplicaremos los métodos *bootstrap* desarrollados por Simar y Wilson (1998) y Löthgren y Tambour (1997) a una muestra de unidades productivas como son las Unidades Regionales de Inspección (*URI's*), de forma que obtendremos intervalos de confianza para el verdadero valor del parámetro de eficiencia⁷.

2. CARACTERIZACIÓN DEL ÁMBITO DE ESTUDIO: LAS UNIDADES REGIONALES DE INSPECCIÓN (*URI's*).

La Agencia Estatal de Administración Tributaria constituye un ámbito singular dentro de la administración pública española por configurarse como un Ente de Derecho Público, con personalidad jurídica propia y con una amplia autonomía en su política presupuestaria y de recursos humanos⁸.

(7) La razón por la que se utilizan los dos métodos es para comprobar si en su aplicación se aprecian diferencias sustanciales. Hay que tener en cuenta que el soporte de la medida de eficiencia está acotado entre cero y uno. Esto introduce un problema en el bootstrap dado que la función de distribución del remuestreo debe ser suavizada [véase Efron y Tibshirani (1993)]. Así, mientras Simar y Wilson (1998) recurren al suavizado de la función, Löthgren y Tambour (1997) aplican un remuestreo basado en la distribución empírica.

(8) La Agencia fue creada en virtud del artículo 103 de la Ley 31/1990 de los Presupuestos Generales del Estado e inició su andadura en 1992. Tiene encomendada la aplicación del sistema tributario estatal y del aduanero. Su objetivo es velar por el mandato constitucional de que todos los ciudadanos participen de acuerdo con sus posibilidades económicas en el sostenimiento de los servicios públicos, actuando básicamente en dos frentes: por un lado, prestar asistencia, ayuda e información a los contribuyentes y, por otro, aplicar diversos mecanismos de control que trasladan a los ciudadanos la conciencia del riesgo de defraudar y sus consecuencias.

La Agencia es la encargada de la gestión del sistema tributario estatal y aduanero. Esta gestión es fundamental para hacer efectiva la consecución de la recaudación prevista en los Presupuestos Generales del Estado de cada año. Su proceso de actuación incluye un conjunto amplio de actividades entre las que destaca la labor de control de cumplimiento tributario. En esta actuación desempeña un papel relevante las Dependencias de Inspección ya que su principal objetivo es detectar fraudes y, con ello, garantizar el cumplimiento de las obligaciones fiscales.

Las Dependencias de Inspección presentan una estructura organizativa descentralizada territorialmente en niveles complementarios. Así, por una parte operan las Unidades Regionales de Inspección (*URI's*), cuyo ámbito geográfico de actuación generalmente coincide con el ámbito de las Comunidades Autónomas, y su objetivo es la comprobación de aquellos contribuyentes considerados como “gran empresa” y aquellos otros que sin tener esa dimensión se considera oportuno que se encuentren bajo su competencia, en atención a la importancia o complejidad de sus operaciones. Por otra parte intervienen las Unidades “Provinciales” de Inspección, que operan en el ámbito geográfico de las Delegaciones y Administraciones de la Agencia (generalmente coincide con el ámbito provincial), y cuya actuación se circunscribe a personas físicas y entidades jurídicas domiciliadas fiscalmente en su territorio, que no son competencias de las Unidades Regionales ni de la Oficina Nacional de Inspección⁹.

En este trabajo nos centramos en evaluar, para cada una de las Delegaciones de la Agencia, el nivel de eficiencia en la actuación de las Unidades Regionales de Inspección. El funcionamiento de estas unidades organizativas se apoya en la utilización de medios técnicos (informáticos) pero sobre todo, el elemento clave es la cualificación, dedicación y profesionalidad de su personal. Así, los servicios inspectores propiamente dichos están organizados en equipos y unidades de inspección, compuestos por inspectores, subinspectores, agentes tributarios y personal auxiliar. Cada uno de estos equipos, siguiendo las directrices establecidas anualmente en el Plan Nacional de Inspección, desarrolla su actividad que se concreta en la comprobación de la conducta de contribuyentes. Como resultado de esta comprobación, se instruyen actas y expedientes con un determinado volumen de deuda en aquellos casos en los que se constaten conductas defraudadoras.

(9) La actividad inspectora se completa con la actuación de La Oficina Nacional de Inspección, que tiene competencia en todo el territorio nacional y permite una gestión integrada de grandes contribuyentes, controlando más adecuadamente sus obligaciones tributarias y posibilitando la existencia de un único interlocutor en la Agencia a través de dicha Unidad, que cuenta con sedes en Madrid, Barcelona, Sevilla, Valencia, Zaragoza, La Coruña, Cantabria y Oviedo.

Las Unidades Especiales encargadas de estudiar expedientes remitidos por otras unidades operativas en las que se aprecian indicios de delito fiscal realizan comprobaciones en materia de fiscalidad internacional y actúan en la colaboración y apoyo a los órganos judiciales.

La Unidad Central de Información se encarga de captar datos para apoyar actuaciones de comprobación.

2.1. *El proceso productivo de las URI's. Objetivos y selección de variables.*

Toda evaluación de eficiencia de una actividad requiere un ejercicio previo de clarificación de los objetivos que persigue un centro productivo ya que, según cuales sean estos objetivos, habrá que seleccionar los *inputs* y *outputs* que mejor representen la actividad que queremos analizar. En nuestro caso de estudio, las actividades de control del cumplimiento tributario realizadas por las *URI's* conllevarán la comprobación e investigación de posibles incumplimientos y actuaciones encaminadas a la efectiva recaudación de deudas liquidadas y no ingresadas. En correspondencia con estas tareas, para estas Unidades se plantea como principal objetivo la detección de fraudes tributarios¹⁰.

Para valorar este objetivo es necesario disponer de indicadores de carácter cuantitativo o cualitativo que nos reflejen lo mejor posible la cantidad y calidad de los resultados obtenidos por las distintas unidades. En la labor de inspección, la Agencia ha centrado el contenido del indicador en el importe de las liquidaciones practicadas que pueden determinar tanto una deuda a ingresar como una reducción de las devoluciones o una disminución en la compensación de bases. Por el contrario, el número de contribuyentes a inspeccionar se considera como una acción a desarrollar por las Unidades, pero no como un objetivo de las mismas ya que incluyen actuaciones claramente heterogéneas, y la correcta formulación del mismo exigiría, o bien la utilización de distintos indicadores, o bien la realización de complejas ponderaciones.

La valoración del objetivo computando únicamente el importe de las liquidaciones puede resultar problemático en la implantación de un sistema de incentivos, ya que se trata de una variable difícilmente controlable por las Unidades de Inspección. Hay que tener en cuenta que generalmente estas Unidades no participan en el proceso de selección, sino que en cada Delegación hay una Unidad encargada de seleccionar los contribuyentes a inspeccionar siguiendo un proceso aleatorio y tratando de responder a pautas, como por ejemplo intensificar la comprobación en un sector concreto, establecidas en el Plan Nacional.

Ahora bien, las liquidaciones se formalizan en distinto tipo de actas. Cada uno de estos documentos recoge el proceso y el resultado de cada inspección, de tal forma que por cada impuesto y, en determinados casos, por cada ejercicio investigado, se levanta un acta¹¹. Por tanto, parece razonable plantear que cuantas más actas realice una Unidad, mayor ha sido su intensidad en la búsqueda de fraudes tributarios. De este modo, el objetivo de detección de fraude que se establece para las *URI's* puede ser cuantificado de forma aproximada por dos variables: el número de actas instruidas en cada centro, que nos aproxima a la cantidad de tra-

(10) En virtud de lo establecido en el Reglamento General de Inspección (R.D. 939/1986 de 25 de Abril), el Ministerio de Economía y Hacienda establece para cada ejercicio un Plan Nacional de Inspección. En ese Plan se establece un objetivo genérico para la actuación inspectora. (Para el ejercicio 1999, se puede consultar el Plan de Objetivos de la Agencia en <http://www.aeat.es>).

(11) Hasta 1999, la inspección de impuestos directos suponía un acta por ejercicio e impuesto, mientras que los impuestos indirectos un acta podía englobar varios ejercicios. A partir de este año, esta última posibilidad se extiende a todo tipo de impuestos.

bajo realizado, y el volumen de deuda incoada, que nos informa de la calidad de esa labor inspectora.

Además, debemos tener presente que los resultados que las distintas Unidades pueden obtener en términos de deuda incoada están condicionados por variables del entorno como puede ser el nivel de riqueza de la Comunidad Autónoma en la que opera. De hecho, si analizamos cómo se comportan las variables deuda incoada y Valor Añadido Bruto de la comunidad podemos observar una alta correlación entre estas dos variables¹². Para aislar este efecto diferenciador, que nada tiene que ver con la productividad del personal dedicado a la inspección, relativizamos para cada *URI* la variable deuda respecto al valor añadido bruto de su respectiva comunidad y posteriormente determinamos la deuda relativa por acta.

En lo referente a los *inputs* sólo consideramos al factor trabajo mediante la *ratio* n.º Inspectores /Total Personal¹³. Dejamos de lado *inputs* que reflejen el capital y las compras corrientes, aunque si tenemos en cuenta que estas unidades organizativas son intensivas en el factor trabajo, estas ausencias no deberían condicionar demasiado los resultados.

En el cuadro 1 recogemos los datos globales de *inputs* y *outputs* para cada una de las 15 Delegaciones Especiales de la Agencia a partir de la información que para el año 95 proporciona la “Memoria de Actividades” del Departamento de Inspección Financiera y Tributaria¹⁴:

(12) En una regresión lineal simple tomando como variable dependiente la deuda y como regresor el Valor Añadido Bruto, se obtiene un R^2 de 0,7 y una *t-ratio* de 7,3.

(13) Evidentemente, el personal compuesto de inspectores, subinspectores y agentes tributarios tiene diferente grado de cualificación y, en consecuencia, puede abordar diferente tipo de tareas. Para plantear una aproximación en la que se tenga en cuenta los distintos niveles de especialización de las plantillas es por lo que construimos, para cada Dependencia, la relación entre inspectores y el número total de personal.

(14) Excluimos las Delegaciones Especiales de Navarra y País Vasco ya que, debido a sus regímenes tributarios especiales, sus datos no son extrapolables al resto de Delegaciones, al recogerse en estas Memorias únicamente información sobre tributos concertados que permanecen en la órbita de la inspección del Estado.

Cuadro 1: *INPUTS Y OUTPUTS* POR DEPENDENCIAS REGIONALES DE INSPECCIÓN

	Inspectores total personal ¹	Acta por contribuyente	Deuda/VAB por acta ²
Andalucía	0,41	1,99	1,49
Aragón	0,29	3,39	1,71
Asturias	0,33	4,13	4,31
Baleares	0,36	2,09	0,86
Canarias	0,44	1,84	3,37
Cantabria	0,33	2,46	9,44
C. La Mancha	0,38	5,21	1,84
C. León	0,36	1,66	5,01
Cataluña	0,47	3,47	0,70
Extremadura	0,25	2,61	3,26
Galicia	0,33	3,77	1,13
Madrid	0,38	2,75	1,31
Murcia	0,38	3,57	4,68
La Rioja	0,29	1,81	22,48
Valencia	0,35	3,66	1,33

Fuente: Memoria de Actividades del Departamento de Inspección Financiera y Tributaria.

(1) Los agentes tributarios no aparecen diferenciados según su pertenencia a UPI o URI, aunque a nivel global se sabe que de los 892 agentes, 805 (90%) pertenecen a UPI y 87 (10%) a URI, por lo que mantenemos esa proporción global para cada Dependencia.

(2) En millones de pesetas.

2.2. *Medición de la eficiencia productiva en las URI's*

Para cuantificar la eficiencia productiva en nuestro ámbito de estudio realizamos una aplicación en términos de minimización de *inputs* y asumiendo rendimientos variables. Este planteamiento parece el más adecuado porque, en primer lugar, si atendemos al diferente ámbito en el que operan las *URI's* y a su posible repercusión en el funcionamiento interno, es razonable evaluar la eficiencia de cada unidad tomando como referencia únicamente aquellas unidades que operen

en una escala similar.¹⁵ Además, en la medida en que las unidades no tienen garantizado un control absoluto de las dos dimensiones del *output*, parece conveniente evaluar la eficiencia en relación a combinaciones adecuadas o inadecuadas en las plantillas de las *URI*'s. De esta manera, considerando que la actividad inspectora se puede sintetizar en un *input* y dos *outputs*, nuestra evaluación de eficiencia se formula mediante el siguiente problema:

$$\begin{aligned}
 & \min \quad \theta \\
 & \text{sujeto a} \\
 & \sum_{j=1}^{15} \lambda_j x_{1j} - \theta x_{1k} \leq 0 \\
 & \sum_{j=1}^{15} \lambda_j y_{1j} - y_{1k} \geq 0 \\
 & \sum_{j=1}^{15} \lambda_j y_{2j} - y_{2k} \geq 0 \\
 & \sum_{j=1}^{15} \lambda_j = 1
 \end{aligned}$$

donde x_{1j} representa la proporción de inspectores sobre el total de personal para cada uno de los centros, y_{1j} recoge la relación entre actas instruidas y contribuyentes inspeccionados, y_{2j} la deuda relativa por acta y λ indica las ponderaciones a cada centro.

Los resultados obtenidos para las Unidades Regionales de Inspección aparecen detallados en la primera columna del cuadro 2. Esos datos pueden resumirse afirmando que de las 15 delegaciones analizadas, sólo la tercera parte de ellas (Aragón, Asturias, Castilla La Mancha, Extremadura y La Rioja) resultan eficientes, es decir, emplea sus *inputs* de personal de la mejor manera posible y obtiene los máximos resultados factibles en comparación con el resto de *URI*'s. Por otro lado, la eficiencia media del conjunto de *URI*'s se sitúa en 0,81. Es decir, por término medio, el conjunto de Unidades Regionales de Inspección podría desarrollar las mismas actividades mejorando la asignación de personal que actualmente emplea.

(15) En determinadas zonas, que comprenden un volumen importante de contribuyentes con operaciones fiscales complejas, se produce una cierta especialización y una mayor cualificación en las plantillas. Esta idea se refuerza si observamos la distribución de los inspectores por Unidades Regionales y comprobamos como 4 *URI*'s: Andalucía, Cataluña, Madrid y Valencia acaparan el 77% de los inspectores. Esta especialización y cualificación debería traducirse en los resultados conseguidos. Para contrastar esta hipótesis, siguiendo a Pedraja y Salinas (1996), examinamos la relación existente entre los índices de eficiencia y el tamaño de las unidades, utilizando como *proxies*, por un lado, el número de inspectores y por otra parte, el tamaño "ponderado" de la plantilla, elaborado a partir de los valores que utiliza la Agencia para crear unidades equivalentes de empleo (3,36 para inspectores, 1,62 para subinspectores y 1 para agentes). Ambas variables resultaron ser significativas en la explicación de los índices de eficiencia.

Variable	Coficiente estandarizado	t- Student
Inspectores	2,703 E-03	2,238
Plantilla "ponderada"	1,466 E-02	2,160

En esa columna también se puede apreciar como 4 Delegaciones (Cantabria, Galicia, Murcia y Valencia) presentan un índice de eficiencia comprendido entre el 75 y el 100%. El resto presenta un índice de eficiencia inferior al 75%. Entre estas últimas, las que peores resultados ofrecen son Canarias y Andalucía.

La discusión planteada en esta sección es la que habitualmente se realiza en los análisis de eficiencia basados en el método envolvente de datos: se compara y se discute la eficiencia a partir de la diferencia puntual de tales estimaciones. Ahora bien, dado que el método *DEA* utiliza una muestra para realizar el análisis de eficiencia, las diferencias en las estimaciones puntuales pueden deberse exclusivamente a ruido muestral más que a diferencias en los verdaderos niveles de eficiencia de las unidades. Como consecuencia de esto, puede resultar importante complementar las conclusiones económicas que surgen de la discusión de la estimación puntual con información sobre la precisión de tales estimaciones. Es decir, introducir información estadística que permita discernir sobre si las diferencias de la estimación puntual de eficiencia de dos centros es estadísticamente significativa (*i.e.* no se debe a ruido muestral). En la sección siguiente se aplican dos métodos *bootstrap* que permiten aproximar la precisión de la estimación *DEA*.

2.3. Análisis estadístico de sensibilidad en la medición de eficiencia

Los resultados de la aplicación del *bootstrap* a los datos de las Unidades Regionales de Inspección se presentan en las columnas 2 a 6 del cuadro 2. El esquema de la tabla es parecido al que presenta Simar y Wilson (1998). En la segunda columna aparece la media de los *bootstrap* por el procedimiento de Löthgren-Tambour y por el procedimiento Simar-Wilson. A continuación se detalla el sesgo y la desviación típica y por último se presenta el intervalo de confianza al 95% de la eficiencia técnica.

El aspecto más relevante de la tabla está dado por los intervalos de confianza corregidos, la última columna. El verdadero valor de la eficiencia técnica se encuentra dentro del rango definido por el intervalo de confianza y por ello, si dos intervalos no se intersectan, sería posible afirmar que las unidades tienen distintos niveles de eficiencia técnica. Esto es así, por ejemplo, con Andalucía, cuyo intervalo de confianza no se intersecta con el de Aragón, Asturias, Castilla La Mancha, Extremadura, Galicia, La Rioja o Valencia. Es decir, podemos afirmar, por una parte, con base en estos resultados que la eficiencia técnica de Andalucía es inferior a la de las restantes regiones. Por otra parte, las *ratios* de eficiencia técnica de estas últimas regiones no resultan diferentes entre sí.

En segundo lugar, al aplicar los procedimientos *bootstrap* a las *URI*'s con rendimientos variables podemos observar cómo la media de estos *bootstrap* es una variable bastante homogénea cuya práctica totalidad de valores se sitúa entre 0,8 y 0,9 con ambos procedimientos. La interpretación que se puede hacer de este resultado es que si se distribuyen aleatoriamente los *inputs* que están siendo utilizados entre las distintas unidades productivas y se repite este proceso múltiples veces, los resultados que se obtienen en términos de eficiencia son bastantes similares para los centros productivos. Por tanto, si utilizásemos la media *bootstrap* $\bar{\theta}_{kn}^*$ como una aproximación del estimador *DEA* $\hat{\theta}_k$, comprobamos cómo en la estimación inicial se infravalora la eficiencia en los centros más ineficientes y se so-

brevemente la eficiencia de los centros más eficientes. No obstante, la media *bootstrap* no está libre de sesgo.

Si tenemos en cuenta los sesgos y derivamos la eficiencia corregida, definida como diferencia entre la media *bootstrap* y el valor del estimador *DEA*, observamos que se produce una mayor dispersión de los resultados, al disminuir el estimador de eficiencia de los centros más ineficientes y al aumentar el estimador de los centros más eficientes. Esto está claramente explicado por el comportamiento de la media *bootstrap* que se sitúa por encima del estimador de eficiencia de los centros más ineficientes y por debajo para los centros eficientes.

La interpretación de los resultados de este estimador corregido debe de hacerse con cuidado. Por un lado, los valores inferiores a cero y mayores que uno tienen validez estadística, aunque carecen de interpretación económica al estar el estimador de eficiencia acotado entre cero y uno. Por otro, para interpretar el valor del estimador corregido es necesario utilizar la información del intervalo de confianza. Así, por ejemplo, para Baleares el estimador corregido toma el valor 0,58 en LT y 0,56 en SW. Este centro puede considerarse ineficiente y el valor de la ineficiencia está comprendido dentro del intervalo de confianza que se puede establecer para su nivel de eficiencia. En esta misma situación estarían Castilla León, Cantabria o Murcia, por lo que entre estos centros no podemos decir que existen diferencias significativas, ya que sus intervalos de confianza se solapan.

Por el contrario, para Galicia el estimador corregido toma el valor 1. Este centro puede considerarse totalmente eficiente ya que el valor de la eficiencia máxima está comprendido dentro del intervalo de confianza (0,7, 1,1) que se define para él. Pero lo mismo puede decirse para Valencia, puesto que, aunque sus estimaciones puntuales de eficiencia son diferentes, los resultados corregidos resultan ser estadísticamente semejantes.

Por tanto, después de este análisis estadístico, la comparación entre centros se hace más difusa, ya que los estimadores de eficiencia resultan bastante imprecisos, al presentar en la mayoría de los casos una desviación típica entre 0,2 y 0,3. Al final, sólo habría 3 grupos de centros sin que dentro de cada uno de ellos se puedan plantear diferencias de eficiencia que sean estadísticamente significativas¹⁶.

En definitiva, los resultados que se derivan de las aplicaciones *bootstrap* permiten matizar los comparaciones de eficiencia entre centros productivos. Así, una vez obtenida la estimación puntual de eficiencia por el método *DEA*, cuando se proceda a comparar varios centros habrá que tener en cuenta los intervalos de confianza y ver si se intersectan o no en un conjunto amplio de valores, para que pueda hablarse de diferencias significativas entre las medidas.

(16) También nos hemos planteado si esta situación se mantiene cuando ampliamos el tamaño muestral. Por tal motivo, hemos replicado este análisis a dos estudios previos, Pedraja y Salinas (1995) y Cuenca (1994) en los que se utilizó un mayor número de centros. Los resultados obtenidos sobre estas aplicaciones permiten constatar, en primer lugar, que las medidas de eficiencia son sensibles con respecto a la variación muestral. En segundo lugar, para buena parte de las unidades productivas los intervalos de confianza se intersectan, por lo que no podría decirse que estas unidades son significativamente distintas en términos de eficiencia productiva [González y Miles (1999)].

Cuadro 2: RESULTADOS DE LOS PROCEDIMIENTOS *BOOTSTRAP* [LÖTHGREN Y TAMBOUR (LT) Y SIMAR Y WILSON (SW)]

UNIDADES REGIONALES DE INSPECCIÓN CON RENDIMIENTOS VARIABLES

	$\hat{\theta}_k$	Media <i>Bootstrap</i>		Sesgo		Eficiencia corregida		Desv. típica		Intervalo confianza corregidos			
		$\tilde{\theta}_{kn}^*$		\hat{d}_k^*		$\tilde{\theta}_{kn}$		$\hat{\sigma}$		$(\tilde{\theta}_{knb}^{*(\alpha)}, \tilde{\theta}_{knb}^{*(1-\alpha)})$			
		LT	SW	LT	SW	LT	SW	LT	SW	LT	SW	LT	SW
Andalucía	0,60	0,82	0,85	0,22	0,24	0,39	0,36	0,15	0,14	0,14	0,56	0,08	0,51
Aragón	1,00	0,82	0,86	-0,18	-0,14	1,18	1,14	0,15	0,14	0,94	1,36	0,85	1,28
Asturias	1,00	0,87	0,89	-0,13	-0,11	1,13	1,11	0,14	0,13	0,84	1,26	0,83	1,22
Baleares	0,70	0,82	0,84	0,12	0,14	0,58	0,56	0,16	0,13	0,32	0,75	0,30	0,72
Canarias	0,57	0,83	0,86	0,26	0,29	0,31	0,28	0,15	0,13	0,06	0,48	-0,03	0,42
Cantabria	0,80	0,88	0,90	0,07	0,10	0,73	0,70	0,14	0,12	0,45	0,85	0,40	0,81
C. La Mancha	1,00	1,00	1,00	0,00	0,00	1,00	1,00	0,00	0,00	1,00	1,00	1,00	1,00
C. León	0,71	0,84	0,86	0,13	0,15	0,58	0,56	0,15	0,14	0,33	0,74	0,24	0,69
Cataluña	0,62	0,83	0,85	0,21	0,23	0,40	0,39	0,16	0,14	0,15	0,57	0,09	0,54
Extremadura	1,00	0,86	0,89	-0,14	-0,11	1,14	1,11	0,14	0,13	0,89	1,28	0,82	1,22
Galicia	0,92	0,84	0,85	-0,08	-0,07	1,00	0,99	0,15	0,14	0,74	1,16	0,70	1,14
Madrid	0,67	0,84	0,86	0,17	0,19	0,50	0,48	0,15	0,14	0,26	0,66	0,17	0,61
Murcia	0,80	0,84	0,87	0,05	0,07	0,75	0,72	0,15	0,14	0,48	0,90	0,41	0,85
La Rioja	1,00	1,00	1,00	0,00	0,00	1,00	1,00	0,00	0,00	1,00	1,00	1,00	1,00
Valencia	0,86	0,84	0,86	-0,02	-0,00	0,88	0,86	0,15	0,13	0,62	1,04	0,57	1,00

3. CONCLUSIONES

En este trabajo se ha realizado un estudio de eficiencia de las Unidades Regionales de Inspección, pertenecientes a la Agencia Estatal de Administración Tributaria. Para evaluar la actividad de estas unidades caracterizamos su proceso productivo mediante dos *outputs* (actas por contribuyente y deuda relativa por acta) y un *input* (n.º de inspectores respecto al total de personal). A partir de aquí, se ha utilizado en primer lugar una aproximación DEA tradicional para obtener una estimación de la eficiencia técnica de cada una de las unidades y establecer comparaciones. La eficiencia media del conjunto de URI's se sitúa en 0,81 y de las 15 delegaciones analizadas sólo la tercera parte de ellas resultan eficientes. Las restantes presentan ineficiencias en mayor o menor medida siendo Canarias y Andalucía las que ofrecen peores resultados.

Sin embargo, el análisis anterior únicamente permitía la estimación puntual de eficiencia técnica, sin que se pudiese analizar la precisión de tal estimación y, por tanto, en la comparación entre las unidades no era posible saber si la diferencia de eficiencia se debía exclusivamente a ruido muestral, lo que genera un problema de fiabilidad en la estimación. Por eso, en segundo lugar, y siendo esta la principal novedad del trabajo, se ha analizado la sensibilidad de tales estimaciones a partir de estimadores *bootstrap* desarrollados recientemente en la literatura para aproximar la distribución del estimador DEA. De esta forma, a partir de los métodos *bootstrap* desarrollados por Simar y Wilson (1998) y Löthgren y Tambour (1997) obtenemos intervalos de confianza para el verdadero valor del parámetro de eficiencia y matizamos las comparaciones, ya que de este último análisis se concluye que, si bien las estimaciones puntuales son distintas entre diferentes unidades, no puede decirse que sean estadísticamente distintas; o lo que es equivalente, la eficiencia técnica teórica entre las distintas unidades es similar.



REFERENCIAS BIBLIOGRÁFICAS

- Bosch, N., F. Pedraja y J. Suárez (1998): "La medición de eficiencia en la prestación de servicios locales. El caso del servicio de recogida de basuras". Documentos de trabajo. Fundación BBV.
- Charnes, A., W. Cooper y E. Rhodes (1978): "Measuring the efficiency of decision making units". *European Journal of Operational Research*, n.º 2, págs. 429-444.
- Charnes, A., W. Cooper y E. Rhodes (1979): "Short communication: measuring the efficiency of decision making units". *European Journal of Operational Research*, n.º 3, pág. 339.
- Coelli, T., D. Prasada y G. Battese (1998): *An introduction to efficiency and productivity analysis*. Kluwer Academic Publisher. Netherlands.
- Cuenca, A. (1994): "Eficiencia técnica en los servicios de protección contra incendios", *Revista de Economía Aplicada*, n.º 5, págs. 87-109
- Davison, A. y D. Hinkley (1997): *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics.
- Debreu, G. (1951): "The coefficient of resource utilization". *Review of Economics Studies*, vol. 9, págs. 300-312.
- Efron, B. y R.J. Tibshiriani (1993): *An introduction to the bootstrap*. Chapman and Hall. London.

- Farrell, M.J. (1957): "The measurement of productive efficiency". *Journal of the Royal Statistical Society*, serie A, vol. 120, págs. 253-281.
- Färe, R., S. Grosskopf y C. Lovell (1994): *Production Frontiers*. Cambridge University Press. Cambridge.
- Fuentesalz, L., C. Marcuello y O. Urbina (1996). "Eficiencia productiva en la prestación en la prestación de servicios de salud. Una aplicación a los centros de atención primaria". *Hacienda Pública Española*, n.º 138-3. págs. 29-36.
- Gstach (1995): "Comparing structural efficiency of unbalanced subsamples: a resampling adaptation of Data Envelopment analysis". *Empirical Economics*, vol. 20, págs. 531-542.
- Gonzalez, X.M. y D. Miles (1999): "Análisis Envolvente de Datos: Un estudio de sensibilidad". Documento de Trabajo. Departamento de Economía Aplicada. Universidade de Vigo.
- Horowitz, J.L. (1995): "Bootstrap methods in econometrics: Theory and numerical performance". Mimeo.
- Kalirajan, K.P. y R.T. Shand (1999): "Frontier production functions and technical efficiency measures". *Journal of Economics Surveys*, vol. 13, págs. 149-171.
- Koopmans, T.C. (1951): "Analysis of production as an efficient combination of activities", en *Activity Analysis of production and Allocation*, Koopmans, T.C. John Wiley and Sons, New York.
- Korostelev, Simar y Tsybakov (1995): "On estimation of monotone and convex boundaries". *Publications de l'Institut de Statistique de l'Universite de Paris*, n.º 39, págs. 3-18.
- Löthgren M. y M. Tambour (1997): "Scale efficiency and scale elasticity in DEA models. A Bootstrapping approach". Paper presented at the *Fourth European Workshop on Efficiency and Productivity Measurement*, Louvain, Belgium.
- Löthgren M. y M. Tambour (1997): "Bootstrapping the Data Envelopment Analysis Malmquist Productivity Index". *Applied Economics*, forthcoming.
- Löthgren M. (1998): "How to bootstrap DEA estimators: A Monte Carlo comparison". Mimeo.
- Mancebón, M.J. (1998): "La riqueza de los resultados suministrados por un modelo envolvente de datos. Una aplicación al sector de la educación secundaria". *Hacienda Pública Española*, n.º 145. págs. 165-186
- Mancebón, M.J. y E. Bandrés (1999): "Efficiency evaluation in secondary schools. The key role of the model specification and of ex post analysis of results". *Education Economics*. vol. 7, n.º 2.
- McCloskey D. y Ziliak (1996): "The standard error of regresions". *Journal of Economic Literature*, vol. XXXIV (March), págs. 97-114.
- Pedraja, F. y J. Salinas (1994): "El Análisis Envolvente de Datos (DEA) y su aplicación al sector público: una nota introductoria". *Hacienda Pública Española*, n.º 128, págs. 117-131.
- Pedraja, F. y J. Salinas (1996): "An assessment of the efficiency of Spanish Courts using DEA". *Applied Economics*, n.º 28, págs. 1.391-1.403.
- Shepard, R.W. (1970): *Theory of cost and production functions*. Princeton University Press. Princeton.
- Simar L. y P.W. Wilson (1998): "Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models". *Management Science*, vol. 44, n.º 1, págs. 49-61.
- Silverman, B.W. (1986): *Density estimation for statistics and data analysis*. Chapman and Hall. London.

Fecha de recepción del original: junio, 1999

Versión final: junio, 2000

ABSTRACT

In this paper we are concerned with studying the technical efficiency of Tax Monitoring Agencies using DEA estimation procedure. Recently, the discussion of technical efficiency has recovered its interest basically because of the need to rationalize the expenditures of the public sector. However, DEA procedure only allows the point estimation of the technical efficiency. That is, the distribution of the technical efficiency estimator is unknown and therefore it is not possible to compare how different are two units in terms of its efficiency. The novelty of this paper is the application of two bootstrap procedures to approximate the distribution of the technical efficiency estimator. We find that, despite of the fact that the point estimates of the technical efficiency are different, they are not significantly different.

Key words: efficiency, *DEA*, *bootstrap*.

JEL classification: C14, D24, H59.