

1. EL CONCEPTO DE EFICIENCIA: DEFINICIÓN Y CUANTIFICACIÓN

Antes de comenzar con el análisis de eficiencia referido en este trabajo parece conveniente precisar los conceptos básicos que van a ser utilizados en el análisis posterior. De este modo se estima oportuno comenzar centrando la atención en la noción de eficiencia así como en los diferentes técnicas existentes para cuantificarla.

La eficiencia es un concepto que posee diversas interpretaciones (**Dunlop (1985:2)**) que obliga a delimitar la acepción del mismo que se pretenda utilizar en una investigación.

Inicialmente, antes de entrar en sus diferentes acepciones, parece adecuado diferenciarlo de la noción de eficacia. Así, se entenderá como eficacia a la capacidad de establecer y lograr metas preestablecidas mientras que eficiencia (en términos genéricos, únicamente con el fin de distinguirla de la eficacia y sin perjuicio de su posterior e inmediato análisis conceptual pormenorizado) aludirá a la capacidad de obtener objetivos por medio de una relación deseable entre inputs y outputs o, en otros términos, de existencia de máxima productividad de los inputs empleados y/o de mínimo coste de obtención del producto (**Bardhan (1995:72)** y **Albi (1992:300)**).

1.1 La noción de eficiencia

Quizá la idea más extendida de eficiencia sea el concepto de óptimo de Pareto según el cual una asignación de recursos A es preferida a otra B si y sólo si con la segunda al menos algún individuo mejora y nadie empeora, es decir, un óptimo paretiano es una asignación de recursos que no puede modificarse para mejorar la situación de alguien sin empeorar la de otro/s (**Gravelle y Rees (1981:498 y 501)**). La garantía de la existencia de este tipo de equilibrio conlleva el cumplimiento de tres condiciones que están relacionadas con el término eficiencia: eficiencia productiva, de intercambio y global. La primera se cumple cuando existen iguales relaciones marginales técnicas de sustitución entre los

recursos empleados para generar los outputs. La segunda, cuando la relación marginal de sustitución entre los bienes son las mismas para todos los consumidores y, la tercera, necesita de la igualdad entre las relaciones marginales de sustitución entre pares de bienes y su relación marginal de transformación para la totalidad de los individuos (**Fuentes (1987:130-6)**).

Lindbeck (1971) consideró la diferenciación de tres extensiones adicionales de la idea de eficiencia: asignativa, técnica y coordinativa e informativa (teniendo las dos primeras, además, dimensiones estáticas y dinámicas).

Así, la primera, en su versión estática, coincidiría con la optimalidad paretiana, mientras que desde el punto de vista dinámico fuerza a que los inputs se agrupen en función de los gustos de los individuos ocasionando que la curva de transformación se expanda.

La segunda, la eficiencia técnica, surge de la interpretación de la función de producción como el conjunto de los puntos frontera del conjunto de producción, quedando particionado así el espacio de asignaciones en eficientes (las ubicadas justo sobre la función de producción), las ineficientes (las situadas debajo de la misma) y las imposibles (las localizadas más allá). En este sentido, se trata de un concepto puramente técnico puesto que contempla únicamente la relación entre las cantidades de insumos y productos y no sus valores. Éste es un elemento que la diferencia de la eficiencia asignativa o precio, la cual supone lograr el coste mínimo de producción de una cantidad determinada de output al cambiar las relaciones proporcionales de los inputs utilizados en función de sus precios y productividades marginales.

En definitiva, bajo el concepto de eficiencia técnica, la proporción de factores de una asignación eficiente puede variar si se modifica la técnica de producción pero no si cambian los precios y/o las productividades marginales. Además, la eficiencia técnica, en su versión estática, tendría, a su vez, una doble acepción. La primera de ellas, la macroeconómica, implicaría la reasignación de los recursos productivos para alcanzar un punto en la curva de transformación de una economía. La segunda, la microeconómica, haría referencia a la ubicación de cada unidad productiva en el conjunto de producción.

Por otro lado, en su versión dinámica, necesita del empleo urgente de nuevos métodos de producción así como del máximo posible incremento y dispersión de los nuevos outputs. Por último, acerca de la eficiencia técnica, cabe decir que su expresión puede realizarse en función de un punto de vista doble: al input o al output. Bajo el primero, reflejaría la cantidad mínima de inputs necesaria para producir un nivel determinado de output y, bajo la segunda, la cantidad máxima de producto obtenible de una cantidad determinada de insumos.

Un tipo particular de eficiencia técnica es la eficiencia X. Una asignación, bajo este concepto, también se consideraría ineficiente por emplear más factores de los necesarios o producir menos output del posible, pero los motivos no se basarían en el tipo de técnica empleada en la producción sino en el comportamiento de los individuos que forman parte de la unidad productiva (reducción de capacidad de esfuerzo, existencia de escasa presión competitiva, primacía de su nivel de utilidad sobre la obligación de reducción de costes al usar más recursos de los necesarios, etc.) propiciado por la carencia de alicientes profesionales en el ámbito del sector que se analice (**Salinas (1995:8-9), Albi (1992:301) y Leibenstein (1966:413)**).

En tercer lugar, la eficiencia coordinativa e informativa se alcanza mediante la minimización de los costes de la información necesaria para la toma de decisiones.

Del conjunto de las revisadas, en este trabajo se empleará la eficiencia técnica estática en su versión microeconómica².

1.2 La cuantificación de la eficiencia

Dado el papel crucial que desempeña el concepto de función de producción en el concepto de eficiencia (sobre todo en el de la técnica que es la que atañe) es lógico que en un principio hayan existido intentos que trataran de definir la eficiencia a partir de su

² La justificación de esta elección se efectuará más adelante tras comentar las características del sector sometido a análisis.

conocimiento previo³ puesto que, al fin y al cabo, es la expresión matemática de la relación técnica existente entre factores y productos. Así **Debreu (1951:273-5)** ofrece ya una definición de medida de eficiencia basándose en un ratio de distancias. Dicho ratio cuantificaría la proporción en que la situación obtenida en una economía se aleja de la óptima, considerando como tal aquella en la que fuera imposible aumentar la satisfacción de algún individuo sin, al menos, disminuir la de otro⁴. Este modo de concebir la cuantificación de la eficiencia, si bien no dependía de las unidades de medida, presentaba la dificultad de necesitar de la existencia de un sistema intrínseco de precios⁵ que homogeneizara las magnitudes de bienes comparadas en el proceso del cálculo del parámetro de eficiencia mediante el cómputo del ratio de distancias (**Debreu (1951:274)**).

Koopmans (1951:60 y 70-80) fue más genérico al demarcar un principio de eficiencia más amplio, evitando así la limitación de la idea de eficiencia de **Debreu**. Partiendo de la consideración de un marco de posibilidades técnicas muy similar al modelo Input-output de **Leontief**,⁶ define a un punto eficiente como aquella combinación de producto neto que, siendo factible, posee la propiedad de que cualquier incremento en una de sus coordenadas puede ser lograda sólo a costa de disminuir al menos una de las restantes⁷. En cualquier caso, en todo el capítulo escrito por **Koopmans** no existe ninguna referencia al modo de medir esa eficiencia. Tan sólo menciona su propia concepción de la misma. Con este panorama era necesario encontrar una noción que fuera genérica y mensurable. Éste fue precisamente uno de los propósitos de **Farrell (1957:253)** de quien **Koopmans (1951)** y **Debreu (1951)** pueden ser considerados como antecesores⁸ y cuya transcendencia queda patente por el hecho de que estudios sobre

³ Función de producción: Expresión matemática que relaciona la máxima cantidad de productos que pueden obtenerse a partir de un nivel determinado de factores.

⁴ Obsérvese el contenido paretiano de la filosofía definitoria del ratio de eficiencia de **Debreu**, tal y como reconoce en la página 278. El mismo contenido sería mantenido más tarde al definir el concepto de óptimo de una economía (**Debreu (1973:116)**).

⁵ Lo cual afectaría a la posibilidad de cómputo de índices de eficiencia en sectores en los que, como el público, habitualmente no operan con precios.

⁶ La diferencia fundamental es que **Koopmans** considera la posibilidad de la existencia de coeficientes técnicos negativos - en este caso el bien es usado y no producido por la rama pertinente- y el vector de producción final es definido en términos netos.

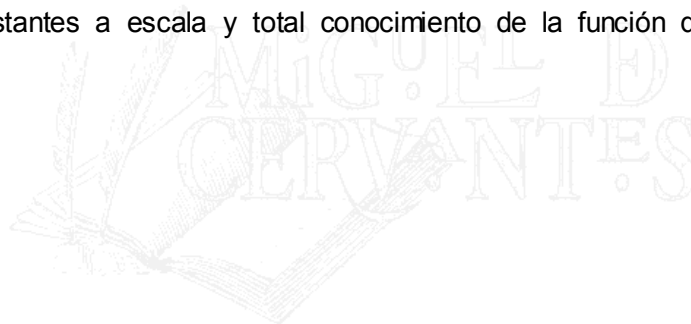
⁷ Obsérvese que también **Koopmans** mantiene la connotación paretiana en su definición de eficiencia.

⁸ En el caso de **Debreu (1951)**, el mismo **Farrell (1957:254)** reconoce la similitud de su medida de eficiencia técnica y la empleada por el primer autor.

medición de eficiencia posteriores al suyo han tomado como base su formulación teórica (**Mancebón (1996b)** y **Salinas (1995:10)**).

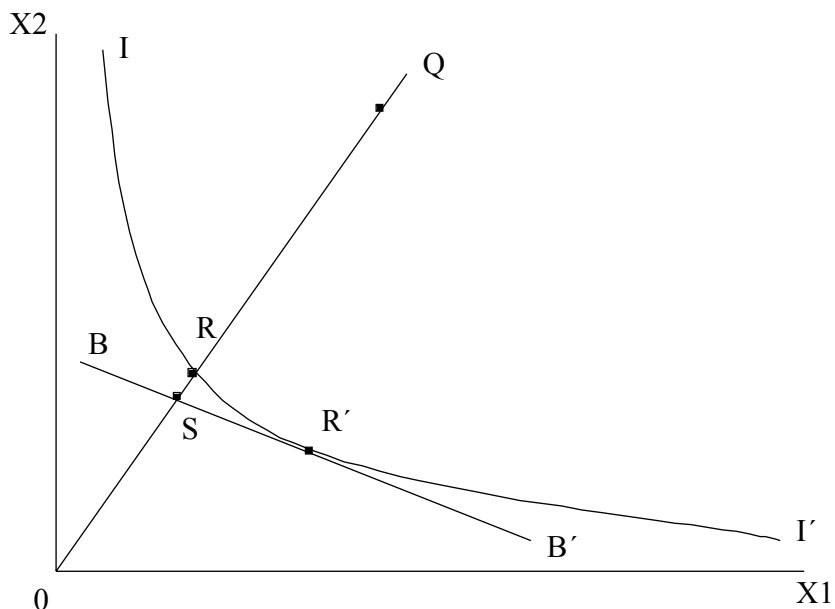
Farrell (1957:254-5) delimitó dos conceptos de eficiencia: eficiencia técnica y eficiencia precio. La primera la definió como la lograda al producir lo máximo posible a partir de unos inputs dados. La segunda entendió que la obtenía aquella unidad productiva que utilizara una combinación de inputs que, con el mínimo coste, alcanzara un output determinado a unos precios preestablecidos⁹.

En función de la primera y siguiendo su argumentación para el caso simple **Farrell (1957:254)** supuso una empresa que empleara dos factores para generar un output bajo rendimientos constantes a escala y total conocimiento de la función de producción (ver gráfico 1.1).



⁹ Las características propias del sector de la enseñanza secundaria pública (las cuales se analizan exhaustivamente en un capítulo posterior) sugieren que la definición de eficiencia más apropiada es la técnica ya que no involucra a los precios (ausentes en el proceso productivo de la enseñanza pública) y, en una de sus versiones, trata de maximizar el producto bajo la restricción de la existencia de unos inputs que están determinados (tal y como se expondrá que sucede con los institutos públicos).

Gráfico 1.1



Fuente: Farrell (1957:254).

En el gráfico anterior la curva I' es la isocuanta unitaria, de modo que representaría las combinaciones mínimas de inputs X_1 y X_2 necesarias para generar una unidad de producto. Es decir, cualquier combinación de inputs de esta isocuanta será eficiente para producir una unidad de output.

De este modo, R sería una asignación eficiente mientras que Q no pues emplea más insumos para lograr el mismo producto. En este sentido, la eficiencia técnica de Q vendría dada por OR/OQ .

Sin embargo, en el anterior razonamiento, no se han considerado en ningún momento los precios de los factores. Farrell (1957:254-55) los introdujo en su trabajo al considerar la eficiencia precio. Desde este punto de vista, la recta BB' reflejaría la relación existente entre los precios de los recursos mediante su pendiente. En este sentido, R' y no R sería la asignación eficiente puesto que tal vez ambas fuesen eficientes técnicamente pero sólo R' puede ser adquirida a los precios preestablecidos con el

mínimo coste posible (puesto que el punto de equilibrio del productor se deriva de la tangencia de la recta de restricción presupuestaria con la isocuanta).

En este sentido, la medición de la eficiencia precio o asignativa de la asignación R vendría dada por OS/OR. Es decir, si se deseara cambiar las proporciones de inputs hasta el mismo tipo que la reflejada por R' y mantener la eficiencia técnica constante, los costes deberían ser disminuidos en una proporción OS/OR.

Finalmente, **Farrell (1957:255)** definió la eficiencia global como el tipo de eficiencia que presentaría una asignación en caso de ser eficiente desde el punto de vista técnico y asignativo, estableciendo que sería igual al producto de ambas medidas de eficiencia: $(OR/OQ) \cdot (OS/OR) = (OS/OQ)$.

Además, **Farrell (1957:256)** también hizo explícito el modo de medición de la eficiencia¹⁰ y su interpretación geométrica para el caso en que la función de producción no fuera conocida. En este último caso, obtuvo una expresión analítica de medida de la eficiencia relativa¹¹ de diferentes unidades productivas bajo las hipótesis de convexidad de las isocuantas, rendimientos constantes a escala y pendiente no positiva de la isocuanta¹². La primera, la convexidad de las isocuantas, implicaba que si dos puntos eran obtenibles en la práctica, entonces la combinación convexa de ambos también. La segunda, los rendimientos constantes a escala, suponían la existencia de una relación invariable entre la modificación de los inputs y la obtención del output, lo cual meramente suponía que los procesos representados por las combinaciones de inputs y outputs de dos puntos cualesquiera no interfirieran entre sí¹³. Finalmente, la no positividad de la

¹⁰ Hay autores como **Mancebón (1996b:31)** que sostienen que **Farrell (1957:258-9)** no consideró adecuadamente la medición de la eficiencia bajo rendimientos no constantes a escala. No obstante, aunque pudiera ser así, sí merece la pena señalar que explicitó cómo hacerlo bajo rendimientos decrecientes a escala y propuso soluciones parciales a los problemas que presentaban los rendimientos crecientes a escala.

¹¹ Entiéndase por eficiencia relativa la obtenida por una unidad productiva en referencia a la conseguida por otra/s.

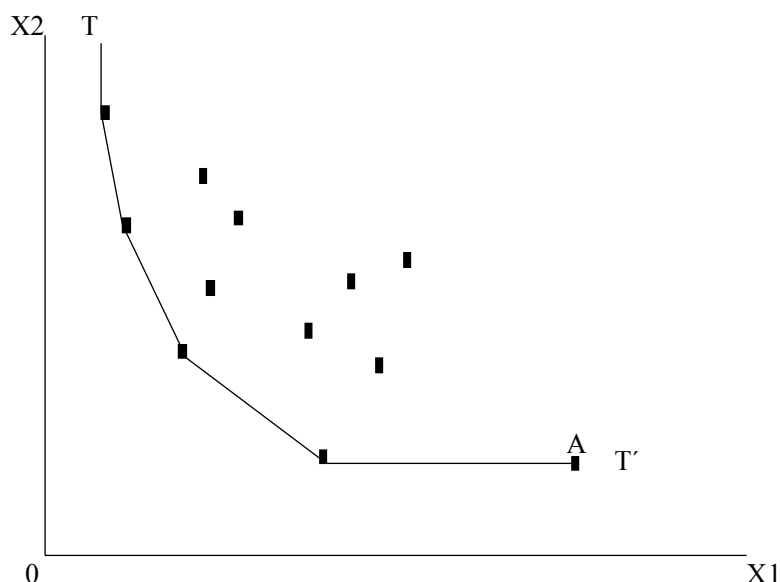
¹² El autor, implícitamente, también consideró el supuesto de libre disponibilidad de inputs y outputs al considerar que cualquier unidad de producción que consumiera más para generar lo mismo o produjese menos utilizando los mismos insumos pertenecería al conjunto de producción.

¹³ En definitiva, se trataba de eliminar la posibilidad de la existencia de efectos de escala en la eficiencia para que ésta fuera únicamente técnica.

pendiente de la isocuanta fue necesaria para evitar que cualquier incremento en ambos factores conllevara una reducción del output.

Así, estos supuestos, en tanto que restringían la forma de las curvas de nivel de la función de producción (isocuantas), introducían limitaciones en la pretensión inicial de lograr medidas de eficiencia sin conocer la relación funcional que ligara los inputs con los outputs. En cualquier caso, **Farrell (1957:256)** especificó que en un conjunto de observaciones tal y como el expuesto en el gráfico 1.2, la isocuanta escogida para ser considerada eficiente sería la TT' .

Gráfico 1.2



Fuente: Farrell (1957:256).

En el gráfico anterior se representan diferentes combinaciones de factores utilizadas para generar una unidad de producto. En definitiva, la isocuanta eficiente (TT') estaría constituida por el conjunto de puntos más cercanos al origen y las combinaciones convexas entre ellos (combinaciones hipotéticas o ficticias) puesto que cualquier vector de

recursos fuera de la misma emplearía más de al menos uno de los inputs para obtener la misma cantidad unitaria de output¹⁴.

Hechas estas consideraciones sobre su trabajo, a continuación **Farrell (1957:260)** estimó oportuno explicar algunos matices de su modo de obtener índices de eficiencia. Recalcó que era una medida realizada tomando como referencia un conjunto de unidades productivas, es decir, relativa. Además, era sensible a la variación del número de empresas incluidas en la comparación. También abundó en la hipotética homogeneidad de los factores de producción señalando que su cumplimiento no es imprescindible si la heterogeneidad se distribuye entre las empresas. Por otro lado, consideró que las diferencias en la calidad media de un factor podrían ser problemáticas ya que, en ese caso, el índice reflejaría tanto la eficiencia por la calidad de factores como por la adecuada gestión. No obstante, si las diferencias cualitativas fueran mensurables el problema también podría evitarse mediante la homogeneización de la calidad.

En cualquier caso, lo significativo es que el autor proporcionó una definición de eficiencia aún hoy empleada, un modo de medición de la misma y un método de aproximación empírica a la **frontera de eficiencia**¹⁵ cuando la función de producción es desconocida y lo único posible es utilizar las observaciones de insumos empleados y productos generados.

De todo lo anterior, se observa que para medir la eficiencia de un conjunto de unidades productivas es necesario conocer la función de producción o el conjunto de producción y la frontera de eficiencia. Para ello existen diversos métodos que pueden clasificarse en función de dos factores: su carácter paramétrico y/o determinístico. Los métodos paramétricos parten de la presunción de que la función de producción posee una determinada forma (los no paramétricos no presuponen ninguna forma de la función mencionada) y los determinísticos asumen que la distancia de la unidad analizada a la frontera es fruto de la ineficiencia (mientras que los estocásticos parten de la hipótesis de

¹⁴ Nótese que **Farrell (1957:255-6)** tan sólo mencionó la no positividad de la pendiente de la isocuanta. Sin embargo, parece lógico observar que la condición ideal hubiera sido la negatividad estricta para así evitar posibles tramos de pendiente nula que estarían ocupados por combinaciones de producción que emplearan más de uno de los dos inputs para generar el mismo output (en el gráfico 1.2, el punto A).

¹⁵ Conjunto de unidades eficientes.

que, al menos parte de esa distancia, es debida a perturbaciones aleatorias) (Hollingsworth et al. (1999:6 y 29) y Salinas (1995:15-6)).

Esquemáticamente:

Cuadro 1.1

Métodos analíticos	Paramétricos	No paramétricos
Determinísticos	Programación matemática paramétrica y Análisis de frontera determinístico	Análisis envolvente de datos
Estocásticos	Análisis estocástico de frontera	Análisis envolvente de datos estocástico.

Fuente: **Hollingsworth et al. (1999:29)**.

De todas estas posibilidades **Farrell (1957)**, en su propuesta de isocuanta del gráfico 1.2, estableció características determinísticas y no paramétricas al proceso. En este sentido, encajaría con el Análisis Envolvente de Datos (DEA).

El hecho de que el objeto del análisis del presente trabajo se centre en el Sector Público (Educación secundaria pública) hace evidente que existan una serie de características propias del mismo que condicionen el método de cuantificación de la eficiencia que se seleccione.

En general, puede afirmarse que, como características básicas, resaltan¹⁶: la ausencia de mercado que dificulta la valoración del producto público, el carácter monopolístico de su oferta, la existencia de múltiples criterios en la naturaleza de los objetivos del sector (eficiencia, equidad, estabilidad macroeconómica...), la carencia de competitividad y de mecanismos de expulsión de las unidades productoras ineficientes, la frecuente utilización de múltiples inputs y la generación de numerosos outputs en un

¹⁶ Para ampliar detalles sobre las mismas se puede, por ejemplo, consultar a **Wolf (1979, 1987 y 1988)**.

mismo centro y la inexistencia de mecanismos de incentivo hacia una producción eficiente.

Todo ello hace que, necesariamente, se deban introducir ciertas peculiaridades en el análisis de eficiencia de este sector. De este modo, no será posible incorporar información sobre precios, se tendrá que elegir un área de actividad en la que sea difícil justificar cualquier otro tipo de comportamiento que no busque la eficiencia, se deberá ajustar la definición de eficiencia a un contexto caracterizado por la ausencia de mercado y habrá que introducir la dificultad que surja de la imposibilidad de conocer la función de producción.

Por todo lo anterior, parece razonable seleccionar como método el Análisis Envolvente de Datos (DEA) ya que, como se expondrá a continuación, no necesita del conocimiento previo de la función de producción, no requiere del uso de información relacionada con precios y se adapta perfectamente a situaciones en las que las unidades productivas emplean múltiples inputs y generan varios outputs.

1.3. EL Análisis Envolvente de Datos (DEA)

Una vez vistas las diferentes posibilidades de modelos de análisis y justificados los motivos que conducen a la elección de DEA, se pasa a exponer los orígenes y fundamentos del método .

1.3.1 Orígenes

Si bien no es extraño encontrar la afirmación de que el modelo de análisis DEA fue desarrollado por primera vez en el año 1978 por **Charnes, Cooper y Rhodes, Seiford (1996:99)**, **Charnes et alter (1997:3)** afirman que el origen de esta técnica es debido a **Rhodes (1978)**, el cual la aplicó (en su tesis doctoral dirigida por **W.W. Cooper**) al análisis de eficiencia del programa de educación **Follow-Through** de las escuelas públicas de los Estados Unidos.

Fundamentalmente, este método sigue los conceptos básicos de **Farrell (1957)**¹⁷. No obstante, junto con este autor, hubo varios otros que proporcionaron los fundamentos necesarios para que DEA pudiera surgir y fuese utilizado (**Seiford (1996:99)**): **Charnes y Cooper (1962)**, **Aigner y Chu (1968)** y **Afriat (1972)**¹⁸.

Aigner y Chu (1968) trataron de continuar el trabajo seminal de **Farrell (1957)**. Tras distinguir entre diversos conceptos de función de producción que hacían complicado

¹⁷ En cualquier caso, no parece que **Rhodes (1978)** extendiera el concepto y medición de eficiencia de **Farrell (1957)** al caso de múltiples inputs y outputs tal y como afirman **Charnes et al. (1997:4)** puesto que, como se ha indicado en el apartado 1.1, **Farrell (1957)** ya abordó esta cuestión.

¹⁸ En particular, **Charnes y Cooper (1962)** proporcionaron un método de transformación lineal para la conversión de programas genéricos de optimización caracterizados por una función objetivo en forma fraccional con restricciones lineales con conjunto de soluciones acotado y no vacío. O sea del tipo:

$$\begin{aligned} \text{Max} \quad & \frac{C' \cdot X + \alpha}{D' \cdot X + \beta} && \text{(M.I)} \\ \text{S.a} \quad & A \cdot X \leq b \\ & X \geq 0 \end{aligned}$$

donde C' y D' son vectores traspuestos de coeficientes, A es una matriz de constantes, b un vector de constantes, α y β son escalares y constantes, y X es un vector de variables.

La transformación de variables fue:

$$Y \equiv T \cdot X \quad \text{(E.I)}$$

Donde $t \geq 0$ es escogido para que

$$D' \cdot Y + \beta \cdot t = \gamma \quad \text{(E.II)}$$

Con $\gamma \neq 0$.

La forma del programa transformado mediante $Y=T \cdot X$ fue:

$$\begin{aligned} \text{Max} \quad & C' \cdot Y + \alpha \cdot T \\ \text{S.a} \quad & A \cdot Y - b \cdot T \leq 0 \\ & D' \cdot Y + \beta \cdot T = \gamma \\ & Y, T \geq 0 \end{aligned} \quad \text{(M.II)}$$

donde γ es un número no especificado distinto de cero (**Charnes y Cooper (1962:181-3)**).

En definitiva, tal y como más adelante se podrá observar en un epígrafe posterior, expusieron lo que sería la transformación básica que después sería convenientemente rectificadas para trocar el programa fraccional original de optimización DEA en uno lineal y resoluble (Obsérvese que en el programa de partida, el fraccional, sólo se especificó un vector de variables (X) cuando en DEA se poseen dos (uno para los inputs - X - y otro para los outputs - Y -).

su entendimiento¹⁹ abordan la labor de completar el trabajo de **Farrell** en aquellos aspectos en los que este autor no logró ser lo suficientemente genérico (por ejemplo, en la estimación de la frontera eficiente bajo la ley de las proporciones variables) utilizando métodos de programación matemática (**Aigner y Chu (1968: 830-1)**). Sin embargo, cabe mencionar que aunque estos autores perseguían una generalización del método de **Farrell (1957)** introdujeron también un elemento restrictivo al considerar concreciones predeterminadas de las funciones de producción ya que una de las características que hacían amplia a la idea de **Farrell** era la no necesidad de considerar previamente una forma específica de función de producción. Además, un problema adicional con su procedimiento es que los test habituales de significatividad de las estimaciones se basaban en supuestos muy restrictivos acerca de las perturbaciones (**Dunlop (1985:13-14)**).

Un enfoque diferente es el que realizó **Afriat (1972)** al desarrollar un método de análisis de la producción que, en la filosofía de **Farrell (1957)**²⁰, evitaba la consideración de especificaciones concretas de la función de producción y, al mismo tiempo, convertía al método de **Farrell** en un caso particular (**Afriat (1972:568-9)**). De todos modos, este autor sí que basa su análisis en consideraciones específicas acerca de determinadas propiedades (no decrecimiento, concavidad...) que deben tener las funciones de producción para que puedan cumplirse las conclusiones de los teoremas enunciados.

En definitiva todos los precedentes comentados generarían un método que compara entre sí unidades de decisión (DMU) homogéneas respecto a inputs y outputs, dando así una medida de la eficiencia relativa. La eficiencia técnica relativa de cada DMU

¹⁹ Los autores mencionan diferentes definiciones de función de producción: la de una empresa, la de la industria, la agregada de la industria y la media. El primero hace referencia a la máxima cantidad de producto obtenible por una empresa a partir de una combinación dada de factores durante el periodo de tiempo requerido para producir el output. El segundo – que los autores asimilaron al concepto de función de producción eficiente de **Farrell** - se refería al límite máximo de output que una empresa puede esperar obtener a partir de una cierta combinación de factores con el nivel técnico existente durante el periodo de producción, siendo este output máximo aplicable a todas las empresas de una misma industria. La función de producción agregada de la industria expresaría la relación entre el output agregado y los inputs agregados de esa industria y, finalmente, la función de producción media sería catalogada por los autores como un concepto ambiguo en el que no estaría determinado el objeto de aplicación del término "media" -¿sería aplicado sobre inputs, sobre outputs, sobre la técnica...?- De hecho, acaban apoyando el uso de la función de producción frontera como método para determinar la máxima capacidad productiva de una industria o para medir el output potencial de una economía (**Aigner y Chu (1968:826-30)**).

²⁰ No obstante, aunque diferente, el método de **Afriat (1972)** también supone una continuidad a la aproximación indirecta a la función de producción de **Nerlove (1961)** el cual proponía la construcción de una familia de funciones de costes para, posteriormente, estimar los valores concretos de sus parámetros.

es calculada computando el ratio definido por el cociente entre la suma ponderada de los outputs y la suma ponderada de los inputs, siendo los pesos calculados en función de criterios paretianos y considerando que la eficiencia de ninguna entidad puede superar la unidad²¹ (**Charnes et alter (1997:6)**).

1.3.2. Fundamentos

Una vez expuestos sus orígenes, se trataría de continuar con sus fundamentos. Para ello, primero se procederá a exponer una serie de conceptos relacionados con la filosofía del modelo y después, en el siguiente epígrafe, se comentarán los detalles del mismo. Finalmente, se comentarán las ventajas y desventajas de esta técnica.

DEA es una técnica de medición de la eficiencia basada en la obtención de una frontera de eficiencia a partir del conjunto de observaciones que se considere sin la estimación de ninguna función de producción, es decir, sin necesidad de conocer ninguna forma de relación funcional entre inputs y outputs²². Es en definitiva una alternativa para extraer información de observaciones frente a los métodos paramétricos²³ cuyo objetivo es la obtención de un hiperplano que se ajuste lo mejor posible al conjunto de observaciones. DEA, por el contrario, trata de optimizar la medida de eficiencia de cada unidad analizada para crear así una frontera eficiente²⁴ basada en el criterio de **Pareto (Charnes et alter (1997:4))**. De este modo, primero se construye la frontera de producción empírica y después se evalúa la eficiencia de cada unidad observada que no pertenezca a la frontera de eficiencia. Así, además de no ser un método paramétrico (por no presuponer la existencia de una función que relacione inputs con outputs) tampoco es

²¹ En la versión al input pues, como se explicará posteriormente, en la versión al output el ratio de eficiencia podrá ser igual o mayor que uno.

²² No obstante sí es necesario, tal y como se explica posteriormente, realizar algunos supuestos sobre esa relación funcional: convexidad y continuidad. Este hecho contrasta con la afirmación de **Charnes et alter (1997:5)** acerca de que DEA no necesita ninguna hipótesis acerca de la forma funcional que relaciona inputs con outputs.

²³ Son ya numerosos los trabajos realizados que combinan DEA con métodos paramétricos con el fin de lograr resultados sinérgicos (**Cooper y Gallegos (1991)**, **Lovell et al. (1997: 329-52)** o **Bardhan (1995)** por ejemplo). Además, tal y como se expone en un capítulo posterior de esta tesis, los resultados de DEA pueden ser utilizados para estratificar en dos o más clases la eficiencia de las DMUs y obtener regresiones para cada estrato produciendo así mejores resultados que con la aplicación directa de las regresiones (**Sengupta (1987a:2280-1)**).

²⁴ Hay que destacar que dicha frontera es perfectamente alcanzable puesto que está constituida por unidades de decisión reales, de modo que sería una frontera eficiente y factible.

estadístico puesto que no asume que la eficiencia no captada siga algún tipo de distribución probabilística (al estilo de los tests de consistencia de inputs y outputs observados con la frontera de producción implementados por **(Hannoch y Rothschild (1972)²⁵ y (Sengupta (1987:2280))**).

De cara al proceso de evaluación, se considera que una unidad productiva es eficiente y, por tanto, que pertenece a la frontera de producción, cuando produce más de algún output sin generar menos del resto y sin consumir más inputs, o bien, cuando utilizando menos de algún input, y no más del resto, genere los mismos productos **(Charnes Cooper y Rhodes (1981:669))**.

Lo anterior explicaría el tipo de elementos que componen la frontera eficiente, pero deja sin aclarar cómo evaluar las DMUs que no formen parte de ella. La idea es comparar cada unidad no eficiente con aquella que lo sea y, a la vez, tenga una técnica de producción similar; es decir, que utilice inputs similares para producir outputs parecidos.

No necesariamente debe ocurrir que esa entidad, eficiente y homogénea técnicamente con la evaluada, deba tener su reflejo en la realidad. Puede ocurrir (de hecho es lo usual) que la unidad con la que se compare la entidad en evaluación no sea real sino una combinación lineal de otras existentes. Esta peculiaridad es perfectamente coherente con el planteamiento de **Farrell (1957:254-8)** explicado antes y asume dos requisitos: la posibilidad de utilización de insumos de modo continuo y la convexidad de la frontera de eficiencia²⁶. El conjunto de unidades reales eficientes combinadas para generar otra unidad eficiente, pero ficticia, se denomina grupo de referencia y su identificación permite planificar las mejoras de las DMUs ineficientes sobre la base de niveles efectivamente alcanzados.

²⁵ Estrictamente los tests mencionados tenían como finalidad comprobar la validez de determinadas hipótesis acerca de la función de producción, tales como cuasi-concavidad, monotonidad y homotecidad, a partir de las observaciones sobre inputs y datos evitando cualquier tipo de parametrización de la función de producción **(Hannoch y Rothschild (1972:256))**.

²⁶ La continuidad garantiza que los inputs sean perfectamente divisibles y la convexidad que la combinación lineal de dos o más DMU pertenecientes al conjunto factible también pertenezca al mismo.

En cualquier caso, para medir la eficiencia de una unidad hay dos opciones. La primera, comprobar la cantidad de inputs utilizada para obtener el mismo output (orientación al input) y, la segunda, lograr el máximo output manteniendo los inputs (orientación al output). Escoger una vía u otra debe depender de las características concretas del problema a analizar²⁷.

1.3.3. El modelo básico.

Inicialmente dicho modelo fue propuesto por **Rhodes (1978)**²⁸ y posteriormente publicado por **Charnes et alter (1978)**²⁹. La medida de eficiencia que adoptaron relacionó la suma ponderada de inputs con la de outputs de cada unidad de decisión (DMU) y utilizó modelos de optimización lineal para calcular las ponderaciones. En cualquier caso, el modelo original no era lineal; sino que era fraccional (**Charnes et alter (1978:430)**):

$$\begin{aligned}
 & \text{Max}_{u,v} \quad h_o = \frac{\sum_{r=1}^s U_r \cdot Y_{ro}}{\sum_{i=1}^m V_i \cdot X_{io}} \\
 & \text{S.A.} : \\
 & \frac{\sum_{r=1}^s U_r \cdot Y_{rj}}{\sum_{i=1}^m V_i \cdot X_{ij}} \leq 1 \quad \forall j : 1 \dots n \\
 & U_r, V_i \geq 0 \quad \forall r : 1 \dots s \quad \forall i : 1 \dots m
 \end{aligned} \tag{M.1}$$

donde:

h_o : función objetivo. Medida de la eficiencia.

Y_{rj} : output i -ésimo de la DMU j -ésima.

²⁷ No obstante, **Farrell (1957:259)** defendió la elección en función de consideraciones técnicas y no conceptuales. Según él, si se analiza un caso con varios inputs y un sólo output lo más adecuado sería elegir una medida orientada al output ya que éste sería un escalar y no un vector.

²⁸ Según lo afirmado por **Seiford (1996:99)**.

²⁹ **Seiford (1996:101)** comenta que la aparición de rumores acerca de la existencia de trabajos con modelos estilo DEA durante los sesenta fueron corroborados por él analizando algunas ponencias del 39º Congreso Anual de la Asociación de Economía Agrícola del Oeste pero que, en cualquier caso, el modelo permaneció latente hasta **Charnes et alter (1978)**.

X_{ij} : input i -ésimo de la DMU j -ésima.

V_i, U_r : ponderaciones de inputs y outputs respectivamente (soluciones del programa).

Como puede observarse la función objetivo (h_o) es el ratio de eficiencia de la unidad comparada (la o) y su maximización está sujeta a que ningún ratio de eficiencia³⁰ supere la unidad. Obsérvese que el numerador de h_o es una suma ponderada de outputs y el denominador de inputs, lo cual implica que proporciona una medida de la cantidad de outputs generada por unidad de inputs de las DMUs (lo cual es un mero convencionalismo para evitar que la solución del programa se dispare a infinito – de hecho hay autores que limitan los ratios a un valor de cien en vez de uno - (**Boussofiene et alter (1991:2)** o **Dyson et alter (1990)**). El programa M.1. busca como soluciones los parámetros u y v que hagan que h_o alcance su máximo valor y, a la vez, hagan que el ratio de eficiencia de cualquier DMU no supere la unidad. Una particularidad importante del modelo es que es él mismo quien, en función de los datos, determina el valor de las ponderaciones y asigna el mismo valor para todas las DMUs. La ventaja evidente es el grado de flexibilidad que ello otorga y la no existencia de juicios de valor por parte del investigador³¹ referentes a la importancia relativa de cada variable en la determinación de la eficiencia de una DMU. La desventaja es que el programa puede asignar una ponderación nula o muy escasa a un determinado factor que, desde el punto de vista teórico, tenga una gran importancia en la eficiencia relativa de las DMUs. Al respecto ha habido varios intentos de solución³². Sin embargo todos implican la necesidad de incorporar información adicional acerca del funcionamiento de las DMUs que no ha podido ser obtenida para su utilización en esta tesis.

³⁰ Obsérvese que el numerador de h_o es una suma ponderada de outputs y el denominador de inputs, lo cual implica que proporciona una medida de la cantidad de outputs generada por unidad de inputs.

³¹ Esta afirmación, que es generalmente aceptada en la literatura sobre DEA, ha sido matizada, por ejemplo, por **Pedrajas et alter (1997: 220)** y **Salinas (1995:52-3)** quienes opinan que en realidad sí existen juicios de valor aun cuando sea el propio modelo quien determine libremente el valor de las ponderaciones pues, en ese caso, se estaría suponiendo implícitamente que ningún input o output posee una significatividad especial en la eficiencia de las DMUs. A la postre lo que persiguen los autores es justificar conceptualmente la inclusión de restricciones a las ponderaciones basadas en información adicional que se incorpore en el modelo.

³² Con respecto a este tema puede consultarse **Pedrajas et alter (1997)**. En este artículo se realiza un repaso a los diferentes métodos implementados y se propone otro basado en la acotación de los parámetros en función de información disponible.

La no linealidad del modelo, junto con la particularidad de que las soluciones del programa M.1. son infinitas³³, complicaba la resolución del problema. Así que los autores, tras proponer un ejemplo de aplicación del anterior programa fraccional al campo de la ingeniería de combustión, lo transformaron en un modelo lineal (**Charnes et al (1978:431)**). El procedimiento de transformación consistió en la consideración de diversas modificaciones del modelo M.1 mediante recíprocos duales. No obstante, parece más sencillo observar que del M.1 puede derivarse directamente un modelo lineal equivalente³⁴:

$$\begin{aligned}
 & \text{Max}_{u, v} \quad \sum_{r=1}^s U_r \cdot Y_{ro} \\
 & \text{S. A.} \\
 & \sum_{r=1}^s U_r \cdot Y_{rj} - \sum_{i=1}^m V_i \cdot X_{ij} \leq 0 \quad (\text{M.2})^{35} \\
 & \sum_{i=1}^m V_i \cdot X_{io} = 1 \\
 & U_r, V_i \geq 0, \forall r : 1 \dots s, \forall j : 1 \dots m
 \end{aligned}$$

³³ Ocurre que si el par de vectores (U, V) son solución de M.1. entonces (aU, aV) también es solución (siendo "a" un real positivo o nulo).

³⁴ La equivalencia entre el modelo fraccional y el modelo lineal no es difícil de entender. Las dos versiones existentes de un mismo modelo original resultan de la existencia de la posibilidad de maximizar un cociente, bien minimizando el denominador o bien maximizando el numerador (caeteris paribus). Así, el primer modelo lineal (M.2) pretende maximizar el numerador del modelo fraccional original manteniendo linealizadas las restricciones y constante el denominador. De este

modo, si en M.1 se denomina $(\sum_{r=1}^s U_r \cdot Y_{rj}) = a$ y $(\sum_{i=1}^m V_i \cdot X_{ij}) = b$ se obtendrá que $a/b \leq 1 \Rightarrow a \leq b \Rightarrow a -$

$b \leq 0$, lo cual es justo la primera restricción de M.2. La otra restricción de M.2 diferente a las de semipositividad de las

ponderaciones $(\sum_{i=1}^m V_i \cdot X_{io} = 1)$ es la que mantiene constante al denominador de la función objetivo de M.1 (de hecho

su igualación a la unidad es arbitraria puesto que otros autores prefieren su igualación a la centena (**Dyson et al (1990)**). Un razonamiento análogo puede realizarse en el modelo lineal M.2'.

³⁵ Si en vez de M.1 se considera su alternativo, M.1'

El mencionado modelo fraccional y original parte de la consideración de que la eficiencia de una organización se determina como el cociente entre su producción total y su consumo total. Como el caso más general es que existan múltiples inputs y outputs, cada uno de ellos tendrá que ir ponderado para componer una media de la producción y consumo total respectivamente. Dichas ponderaciones las proporciona el mismo modelo sin necesidad de información acerca de precios de insumos o productos. Una vez resuelto el programa y conocidas las ponderaciones (U_r y V_i) y el índice de eficiencia (h_o) para una entidad, el proceso debe repetirse para cada una de las unidades que se deseen evaluar³⁶.

$$\text{Min}_{u,v} \quad f_o = \frac{\sum_{i=1}^s V_i \cdot X_{io}}{\sum_{r=1}^m U_r \cdot Y_{ro}} \quad (\text{M.1})$$

S . A . :

$$\frac{\sum_{i=1}^s V_i \cdot X_{ij}}{\sum_{r=1}^m U_r \cdot Y_{rj}} \geq 1 \quad \forall j : 1 \dots n$$

$$U_r, V_i \geq 0 \quad \forall r : 1 \dots s \quad \forall i : 1 \dots m$$

Su conversión lineal sería:

$$\text{Min}_{u,v} \quad \sum_{i=1}^m V_i \cdot X_{io}$$

S . A .

$$\sum_{i=1}^m V_i \cdot X_{ij} - \sum_{r=1}^s U_r \cdot Y_{rj} \geq 0 \quad (\text{M.2})$$

$$\sum_{r=1}^s U_r \cdot Y_{ro} = 1$$

$$V_i, U_r \geq 0$$

³⁶ A primera vista las restricciones del programa pueden parecer sorprendentes puesto que, en función de la definición de eficiencia asumida, se estaría imponiendo que cualquier unidad de decisión (incluida la evaluada) fuera improductiva en el sentido de generar un output conjunto $(\sum_{r=1}^s U_r \cdot Y_{rj} \quad \forall j : 1 \dots n)$ inferior al input agregado utilizado

$(\sum_{i=1}^m V_i \cdot X_{ij} \quad \forall j : 1 \dots n)$. Sin embargo, hay dos motivos para que la restricción pueda ser válida. En primer lugar, el

output e input totales contemplados son sumas ponderadas por diferentes parámetros (el output por U_r y el input por V_i , con $r : 1 \dots s$ e $i : 1 \dots m$). En segundo lugar, existe una razón conceptual: la necesidad de acotar superiormente el ratio de eficiencia en un programa cuyo objetivo es la maximización de uno de esos ratios para una entidad. De este modo, cuando la entidad evaluada sea eficiente respecto a las otras, su ratio de eficiencia alcanzará el valor unitario y, en otro caso, otro nivel inferior. Sin

1.3.4. Modificaciones del modelo básico

Esta formulación original fue modificada más tarde por los propios **Charnes et alter (1979)** con el fin de que las ponderaciones alcanzaran valores estrictamente positivos y, así, evitar que la solución del programa no considerara a todos los factores y productos en el cálculo del índice de eficiencia al tiempo que se evitaba que el denominador del cociente de eficiencia fuera nulo y su valor no existiese³⁷.

La modificación de M.1 fue:

$$\begin{aligned}
 & \text{Max}_{u,v} \quad h_o = \frac{\sum_{r=1}^s U_r \cdot Y_{ro}}{\sum_{i=1}^m V_i \cdot X_{io}} \\
 & \text{S.A.} \\
 & \frac{\sum_{r=1}^s U_r \cdot Y_{rj}}{\sum_{i=1}^m V_i \cdot X_{ij}} \leq 1 \quad \forall j : 1 \dots n \quad (\text{M.1}'') \\
 & U_r, V_i \geq \varepsilon > 0 \quad \forall r, i
 \end{aligned}$$

donde ε es un número real positivo y pequeño (usualmente, en cálculos empíricos, 10^{-6} (**Norman y Stoker (1991:239)**)).

Linealizando tendríamos:

embargo, esto no es más que una condición necesaria pero no suficiente de eficiencia ya que, como se verá más adelante, existe otro requisito indispensable y adicional referente a las variables de holgura aún no comentadas.

³⁷ Parece existir cierto equívoco en la afirmación de **Boussofiene et alter (1991:1)** referente a que el modelo M.1 fue propuesto por **Charnes et alter (1978)** pues, en principio, no existe ninguna referencia a tal modificación en este último artículo. Habría que esperar hasta 1979 para que **Charnes et alter** introdujeran la restricción de positividad estricta de los parámetros u y v .

$$\begin{aligned}
 \text{Max}_{u,v} \quad h_o &= \sum_{r=1}^s U_r \cdot Y_{ro} \\
 \text{S.A.} \quad & \\
 \sum_{i=1}^m V_i \cdot X_{io} &= 1 & \text{(M.2'')} \\
 \sum_{r=1}^s U_r \cdot Y_{rj} - \sum_{i=1}^m V_i \cdot X_{ij} &\leq 0 \quad \forall j : 1 \dots n \\
 U_r, V_i &\geq \varepsilon > 0 \quad \forall r, i.
 \end{aligned}$$

Sin embargo, aunque este modelo ya era plenamente operativo, en general, no suele ser utilizado para obtener las medidas de eficiencia sino que se emplea su dual³⁸.

Asimismo, como se persigue el análisis de la eficiencia de un grupo de entidades de enseñanza pública que reciben un presupuesto anual y destinan éste a obtener los mejores resultados posibles, lo más lógico es utilizar el programa dual que mida la eficiencia por el lado de los outputs:

³⁸ Hay varios motivos que inducen a utilizar el dual en vez del primal. En primer lugar, por razones de operatividad y ahorro de tiempo. Como los duales tienen menos restricciones que los primales es más sencillo y corto calcular sus soluciones ((Boussofiane et al (1991:2)). Sin embargo, Green y Doyle (1997:70-1) encontraron que la búsqueda de soluciones con el primal es, contrariamente a lo difundido, más rápida. Aún así, se opta por utilizar en esta tesis el dual debido a dos motivos adicionales. Por un lado, los duales ofrecen una mejor interpretación de la eficiencia. Observando las m primeras restricciones del dual que interesa para el análisis de esta tesis, es posible concluir que el modelo determina la existencia de alguna entidad real o ficticia que consuma lo mismo o menos que la evaluada. En ese caso, se le asignaría un valor λ_j distinto de cero. Las s segundas restricciones tienen como finalidad verificar si existe alguna unidad real o no que produzca lo mismo o más que la evaluada. Por tanto, el modelo verifica que si existe otra unidad productiva que consuma lo mismo o menos y produzca lo mismo o más que la entidad cuya eficiencia relativa se desea conocer. Tal y como están presentados los duales de los modelos linealizados corregidos, su funcionamiento persigue la comparación de una unidad real o ficticia construida mediante los λ_j , de modo que la unidad evaluada ($j=0$) se califica de eficiente sólo cuando el ratio de eficiencia (φ^0) es igual a uno y las variables de holgura son todas nulas (ambas condiciones de modo conjunto constituyen la condición necesaria y suficiente de eficiencia demostrada por Charnes et al (1978:433) y Seiford y Thrall (1990:17). Finalmente, por otro lado, el dual ofrece una mejor información a la hora de elaborar estrategias de mejora puesto que es fácilmente identificable el grupo de comparación.

$$\begin{aligned}
 & \text{Max}_{\varphi_o, \lambda, S_{i-}, S_{r+}} \quad \varphi_o + \varepsilon \left(\sum_{i=1}^m S_{i-} + \sum_{r=1}^s S_{r+} \right) \\
 & \text{S.A.} \\
 & \sum_{j=1}^n \lambda_j \cdot X_{ij} + S_{i-} = X_{io} \quad \forall i : 1 \dots m \quad (\text{M.3}) \\
 & \sum_{j=1}^n \lambda_j \cdot Y_{rj} - S_{r+} = \varphi_o \cdot Y_{ro} \quad \forall r : 1 \dots s \\
 & \lambda_j, S_{i-}, S_{r+} \geq 0
 \end{aligned}$$

donde:

φ_o : parámetro que mide la eficiencia de la unidad evaluada.

λ_j : ponderaciones obtenidas como solución del programa. Expresan el peso que posee cada DMU dentro del grupo de comparación (*peer group*) de la DMU_o.

S_{i-} , S_{r+} : variables de holgura de inputs y outputs respectivamente. Transforman las restricciones de desigualdad en igualdades.

El programa M.3, para la unidad analizada (DMU_o), busca una combinación ponderada de unidades tal que, para cada input, su combinación de factores más la variable de holgura de los mismos sea igual que los insumos utilizados por la DMU_o (primera restricción) y, a la vez, que para cada output, la combinación ponderada del producto de las unidades menos la respectiva variable de holgura genere una proporción φ_o del output de la unidad comparada (segunda restricción).

De hecho, el programa M.3 procede del siguiente dual sin variables de holgura ni en las restricciones ni en la función objetivo³⁹:

³⁹ Si bien las variables de holgura tienen como finalidad transformar las restricciones en condiciones de igualdad, su utilización en la función objetivo tendría el fin de evitar máximos duales ((Norman y Stoker (1991:239)).

$$\begin{aligned}
 & \text{Max}_{\varphi_0, \lambda} \quad \varphi_0 \\
 & \text{S.A} \\
 & \sum_{j=1}^n \lambda_j \cdot X_{ij} \leq X_{i_0}, \forall i : 1 \dots m \\
 & \sum_{j=1}^n \lambda_j \cdot Y_{rj} \geq \varphi_0 \cdot Y_{r_0}, \forall r : 1 \dots s
 \end{aligned} \tag{M.3'}$$

En este último es más sencillo observar que se compara la DMU_0 con todas las DMU reales o ficticias que produzcan lo mismo o más que DMU_0 consumiendo menos o lo mismo respectivamente que la DMU_0 .

Así si φ_0 es igual a uno entonces la DMU_0 será eficiente⁴⁰ ya que el programa habrá buscado entre las DMUs reales y ficticias que produzcan lo mismo o más y usen lo mismo o menos que la DMU_0 y no habiendo encontrado ninguna no tendrá más remedio que considerar a la DMU_0 como eficiente dando al φ_0 el valor unitario. El hecho es que con el parámetro igual a uno y siendo DMU_0 eficiente sucede que $Y_{r_0} = Y_{r_0}$ (según la segunda restricción del programa) ya que sólo el λ_0 tendría valor positivo e igual a uno mientras que el resto lo tendría nulo. Por el mismo motivo $X_{i_0} = X_{i_0}$ según la restricción primera. Si por el contrario φ_0 fuera mayor que uno, DMU_0 no sería eficiente porque el programa habría encontrado otra DMU real (o no) que satisfaga las dos restricciones, o sea que produzca más o lo mismo que DMU_0 con iguales o menores inputs. Finalmente φ_0 no puede ser menor que uno porque siempre será posible para el programa asignar como valor mínimo el unitario a φ_0 ya que la DMU_0 siempre será plausible como solución al asignar un valor unitario al λ_0 y nulo al resto.

⁴⁰ Obsérvese que existe una exigencia adicional y es que las variables de holgura sean nulas. Ambos requisitos suponen conjuntamente la condición necesaria y suficiente de resultado de eficiencia según el programa.

Los comentarios realizados hasta ahora constituyen las pautas evolutivas básicas del modelo original. Sin embargo, DEA es un cuerpo de conceptos y métodos que, en su estructura básica, puede sintetizarse en cuatro tipos básicos: CCR⁴¹, BCC⁴², multiplicativos y aditivos⁴³. No obstante, sólo se comentarán aquellas extensiones que puedan tener alguna utilidad para el análisis posterior.

Modelo BCC (debido a **Banker et al. (1984:1084)** según **Boussofiene et alter (1991:12)** y **Pedraja y Salinas (1994:122)**): este modelo fue propuesto con el propósito de estimar la eficiencia puramente técnica eliminando la influencia que pudiera tener la existencia de economías de escala en la evaluación del ratio de eficiencia de las DMUs. El hecho es que la medida de eficiencia de una unidad está condicionada no sólo por la gestión de la misma sino también por la escala en la que opere y el modelo CCR suponía tácitamente la existencia de rendimientos constantes a escala (RKE) al definir las restricciones del modo que proponía⁴⁴. Esto significaba que todas las unidades se comparaban como si estuvieran sometidas a rendimientos constantes y no se contemplaba la posibilidad de existencia de ineficiencias debidas a las diferencias entre las escalas operativas en cada DMU.

El modo en que **Banker et al. (1984:1084)** proponen solucionar esa consideración implícita es añadir una restricción adicional al modelo CCR recogido en M.3. Dicha restricción es:

$$\sum_{j=1}^n \lambda_j = 1 \quad (\text{E.1.1})^{45}$$

⁴¹ Siglas de Charnes – Cooper - Rhodes.

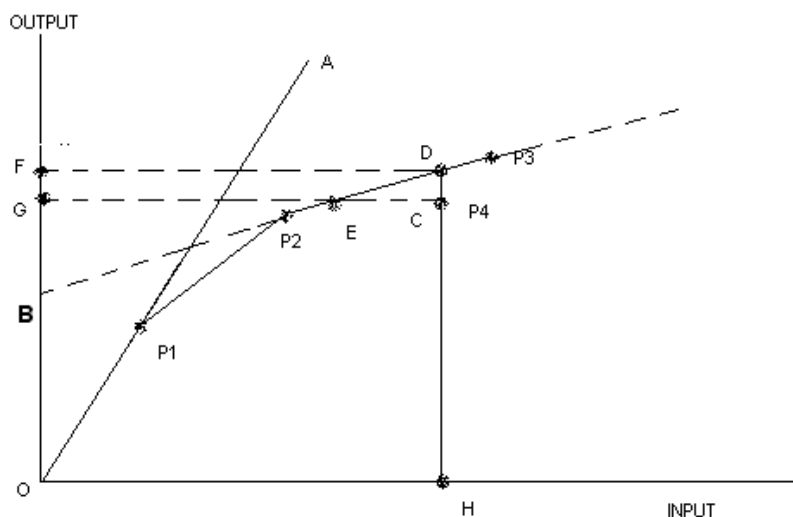
⁴² Siglas de Banker - Charnes – Cooper.

⁴³ A pesar de existir diversas variantes del modelo, todas han coincidido en ofrecer un carácter determinístico (no estocástico). No obstante, autores como **Banker (1996)** o **Simar (1996)** comenzaron a introducir procedimientos de tratamiento estocástico del modelo mediante la utilización de Procesos Generadores de Datos (**Simar (1996:177-9)**).

⁴⁴ **Banker et alter (1984:1081-3)** partiendo de la definición de función de distancia de **Shepard (1970:206)** y de cuatro supuestos sobre el conjunto de posibilidades de producción (T) (entre los que se encuentra el de no restricción del radio vector o de **RKE** del espacio de producción: si (X,Y) pertenece a T => (kX, kY) pertenece a T para todo k>0) establecen una equivalencia entre la medida de eficiencia de **Charnes et alter (1978)** y la medida de distancia antes mencionada.

⁴⁵ Es necesario destacar que una vez que se introducen en el modelo las economías de escala, los resultados de los índices de ineficiencia mediante una orientación al input no coinciden con los que se obtienen con una al output. El sentido en el que se defina la eficiencia técnica conlleva el modo de la proyección de la unidad evaluada. Si la orientación es al input aquella será

Gráfico 1.1.



Fuente: Banker et al (1984:1087) y elaboración propia.

El motivo de la misma según Norman y Stoker (1991: 97-109) se basa en la modificación de la frontera de posibilidades de producción al considerar rendimientos no constantes a escala. En el caso de los rendimientos constantes (RKE) dicha frontera está constituida por un radio vector que parte del origen y pasa por la entidad más eficiente de las evaluadas (línea OA en gráfico 1).

Pero este tipo de supuesto determina que la eficiencia de las DMUs sea medida respecto a P_1 pudiendo ocurrir que existan otras unidades igualmente eficientes en su

proyectada hacia atrás (hacia la frontera) en una magnitud $(\theta X, Y)$ y si la dirección es al output la unidad será proyectada hacia arriba en una cantidad $(X, \phi \cdot Y)$. En definitiva, la geometría implicada en cada modelo y orientación es diferente y los resultados de los índices pueden variar al introducir la restricción referente a las economías de escala (Seiford y Thrall (1990:20-26), Boussofiane et al (1991:12) y Charnes et al (1997:43)). En cualquier caso, si bien la condición de convexidad implica este resultado, también es cierto que una DMU será caracterizada como eficiente con el modelo BCC orientado al output si y sólo si también lo es con el BCC orientado al input Charnes et al. (1997: 36). No obstante, esta equivalencia ya no es cierta al comparar resultados obtenidos mediante la aplicación de modelos distintos, como CCR y BCC, por ejemplo. En concreto, aunque una DMU evaluada como eficiente con un CCR también lo será con un BCC, al revés, en general, no se dará la correspondencia (Charnes et al (1997:36 y 39)). Esto implica que la elección del programa a aplicar no es baladí y que, por tanto, tendrá que estar justificada por las particularidades y controlabilidad de inputs y/o outputs de cada caso (Boussofiane et al (1991:12)).

gestión (P_2 y P_3) pero que no operen en la misma escala que P_1 y, así, no obtengan tan buenos resultados de eficiencia. Al reconocer que con los RKE la medición de la eficiencia combina la escala con la que se opera y la gestión, es posible construir otra frontera de posibilidades de producción ($O-P_1-P_2-P_3$) que contenga a las DMU que posean el mejor comportamiento en el sentido de que, para cada unidad perteneciente a esa frontera, no existe otra con mejor ratio de eficiencia dentro de su escala de operación. En el gráfico 1 se supone que es $O-P_1-P_2-P_3$. En cualquier caso su posición deberá estar por debajo de OA ya que ésta era determinada como frontera de posibilidades de producción mediante un radio vector que pasa por el origen y la entidad con mayor productividad media (P_1).

Con estas consideraciones, y centrando la atención en la determinación del segmento P_2P_3 de la frontera eficiente de producción (desde un punto de vista orientado a la minimización en la utilización de inputs) la ecuación de dicho segmento que une P_2 y P_3 sería:

$$Y = b + aX \quad (\text{E.1.2})$$

donde a es la pendiente de la recta, b la ordenada en el origen, Y el nivel de output y X el de input.

El punto D , que está sobre la recta, usa los mismos inputs que P_4 así que su output valdrá:

$$Y = B + a \cdot X_{P_4} \quad (\text{E.1.3})$$

donde X_{P_4} son los inputs utilizados por la unidad representada en P_4 y B es la ordenada en el origen de la recta que contiene a los puntos P_2 y P_3 .

Como el punto F tiene el mismo output que el D por geometría se puede afirmar que:

$$\frac{GE}{GC} = \frac{BE}{BD} = \frac{BG}{Bf} = \frac{OG - OB}{OF - OB} = \frac{Y_{P_4} - B}{(a \cdot X_{P_4} + B) - B} = \frac{Y_{P_4} - B}{a \cdot X_{P_4}} \quad (\text{E.1.4})$$

observando que la medida de la eficiencia bajo **RVE** es similar a la de **RKE** diferenciándose sólo en el numerador. Así la forma matemática genérica sería:

$$\frac{\text{Suma ponderada de outputs} + \text{constante}}{\text{suma ponderada de inputs}} \quad (\text{E.1.5})$$

donde la constante es positiva, nula o negativa 0 cuando existen rendimientos decrecientes, constantes o crecientes a escala respectivamente.

Esquemáticamente, los tipos básicos de los modelos BCC son:

$$\begin{aligned} \text{Min}_{\theta, \lambda, S_+, S_-} \quad & \theta_o - \varepsilon \left(\sum_{i=1}^m S_{i-} + \sum_{r=1}^s S_{r+} \right) \\ \text{S.A} \quad & \\ \sum_{r=1}^s \lambda_j \cdot Y_{rj} - S_{r+} &= Y_{ro}, \forall r & (\text{M.4}) (\text{BCC}_p\text{-I})^{46} \\ \sum_{j=1}^n \lambda_j \cdot X_{ij} + S_{i-} &= \theta_o \cdot X_{io}, \forall i \\ \sum_{j=1}^n \lambda_j &= 1 \\ \lambda_j, S_{i-}, S_{r+} &\geq \varepsilon > 0, \forall j \end{aligned}$$

⁴⁶ El subíndice p hace referencia a "primal" y el D a "dual". Asimismo, la terminación I se refiere a la orientación al input y la O al output.

$$\text{Max}_{U, V} \quad \sum_{r=1}^s U_r \cdot Y_{ro} + u_o$$

S.A

$$\sum_{i=1}^m V_i \cdot X_{io} = 1, \forall i \quad (\text{M.4}') (\text{BCC}_D\text{-I})$$

$$\sum_{i=1}^m V_i \cdot X_{ij} - \sum_{r=1}^s U_r \cdot Y_{rj} - u_o \geq 0, \forall j : 1 \dots n$$

$$U_r, V_i \geq \varepsilon$$

u_o libre .

$$\text{Max}_{\theta, \lambda, S_+, S_-} \quad \phi_o + \varepsilon \left(\sum_{i=1}^m S_{i-} + \sum_{r=1}^s S_{r+} \right)$$

S.A

$$\sum_{r=1}^s \lambda_j \cdot Y_{rj} - S_{r+} = \phi_o Y_{ro}, \forall r \quad (\text{M.5}) (\text{BCC}_p\text{-O})$$

$$\sum_{j=1}^n \lambda_j \cdot X_{ij} + S_{i-} = X_{io}, \forall i$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda_j, S_{i-}, S_{r+} \geq \varepsilon > 0, \forall j$$

$$\text{Min}_{U, V, v_o} \quad \sum_{i=1}^m V_i \cdot X_{io} + v_o$$

S.A

$$\sum_{r=1}^s U_r \cdot Y_{ro} = 1, \forall r \quad (\text{M.5}') (\text{BCC}_D\text{-O})$$

$$\sum_{i=1}^m V_i \cdot X_{ij} - \sum_{r=1}^s U_r \cdot Y_{rj} + v_o \geq 0, \forall j : 1 \dots n$$

$$U_r, V_i \geq \varepsilon$$

v_o libre .

En cualquier caso, los modelos CCR y BCC no son los únicos tipos básicos que existen dentro de DEA. Además, existen otras dos clases: los aditivos y los multiplicativos.

Sin embargo estas dos últimas clases se obvian puesto que no serán utilizadas en esta tesis⁴⁷.

Estas versiones básicas de DEA, las comentadas hasta ahora, han ido incorporando otras modificaciones. Del conjunto de estas últimas⁴⁸ en esta investigación sólo se considerarán la introducción de variables no discrecionales en el modelo debido a que, como se argumentará en epígrafes posteriores es la única extensión que será necesaria para llevar a cabo el propósito marcado en la introducción⁴⁹.

Existen algunos tipos de productos cuya generación está fuertemente influida por factores no sujetos al control del gestor. Actividades como la agricultura o el turismo (con gran dependencia del clima) son un ejemplo típico. El caso que compete al objetivo que se persigue con este trabajo es también paradigmático. El resultado del proceso de enseñanza está indudablemente afectado por la calidad y estrategia docente pero también, tal y como se expone con mayor detalle posteriormente, existe un componente importante correspondiente al ambiente social, económico y familiar que contextualiza la vida del discente (**Boussofiane et al. (1991:3)**, **San Segundo (1991:29)** o **Gamer y Raundenbush (1991:258)**).

⁴⁷ En todo caso sí cabría apuntar que unos modelos se diferencian de otros, además de por su apariencia externa, por la forma de la envolvente que generan y no también por el modo de proyectar las unidades ineficientes hacia la frontera de eficiencia. De este modo, la elección de uno u otro depende del tipo de situación concreta que se analice.

⁴⁸ Fundamentalmente son: inclusión de variables no discrecionales, consideración de variables categóricas, incorporación de información vía restricciones a los parámetros y evaluación de la variación de la eficiencia a lo largo de varios períodos. Sin embargo, multitud de variaciones del modelo DEA han surgido en los últimos años debido al aumento de su utilización como herramienta de detección de ineficiencias. Entre esas variaciones cabe destacar la integración con otro tipo de técnicas (por ejemplo, **Retzlaff-Roberts y Puelz (1996)**), su modificación para la utilización bajo otro tipo de condiciones diferentes a las establecidas en las hipótesis básicas del modelo (v.g. **Gstach (1998)**), la implementación complementaria o suplementaria de DEA con otros modelos (v.g. **Sharma et al. (1997)**, **Bardhan (1995)**, **Bates (1997)** o **Bardhan et al. (1998)**), análisis de la reducción de la dimensión de las DMUs para estudiar su variación de eficiencia (v. g. **Ray y Mukherjee (1998)**), estudios acerca del efecto del tamaño muestral sobre los ratios de eficiencia (v. g. **Zhang (1998)**), modelos para el cálculo de la sensibilidad y estabilidad de los resultados de eficiencia (**Charnes, A. et al. (1996)** o **Banker, R.D. y Morey, R.C. (1994)**), variaciones destinadas al cómputo de la existencia de rendimientos a escala en contextos de múltiples soluciones (**Seiford, L. M. y Zhu, J. (1999)**) o , alteraciones de DEA con el fin de incluir análisis de inputs y outputs estocásticos (**Cooper, W et al. (1998)**).

⁴⁹ En realidad, tal y como se verá en capítulos posteriores, esta afirmación debe ser matizada. El estudio que se desea realizar de los centros educativos de secundaria de la provincia de Alicante necesitará considerar otras varias modificaciones de los modelos básicos que se han ido introduciendo con el tiempo. Sin embargo, se ha preferido mencionar aquí tan solo a las

De este modo, una unidad productiva será eficiente cuando, además de los requisitos ya mencionados con anterioridad, se enfrente a iguales o peores factores no controlables (**Banker y Morey (1986 (b):1613)**).

El modo en que las variables no controlables pueden ser consideradas es variado (**Muñiz (1998:460-2)**). Una primera forma es la propuesta por **Banker y Morey (1986b)** por la cual se modificarían las restricciones del modelo original. La clave del tratamiento de estas variables radicaría en que la información acerca de la cuantía en que podría reducirse un input no controlable (o aumentarse un output no controlable) no es significativa para el gestor de la unidad en evaluación (**Charnes et al (1997:50)**). De este modo se asume que inputs y outputs están divididos en dos subconjuntos: el de variables controlables (D) y el de no controlables (N)⁵⁰.

variantes básicas y dejar para epígrafes posteriores las alteraciones más avanzadas. El propósito es lograr que, de este modo, la exposición sea más clara.

⁵⁰ Esto es, siendo:

$$I=\{1,2\dots m\} = I_b \cup I_n, I_b \cap I_n = \emptyset$$

$$O=\{1,2\dots s\} = O_b \cup O_n, O_b \cap O_n = \emptyset$$

Donde: *I* hace referencia inputs, *D* a su carácter discrecional, *N* a su no discrecionalidad y *O* a outputs. Para el caso de un BCC-I se tendría:

$$\begin{aligned} & \text{Min}_{\lambda, \theta, S_{i-}, S_{r+}} \quad \theta - \varepsilon \left(\sum_{r \in OD} S_{r+} + \sum_{i \in ID} S_{i-} \right) \\ & \text{S.A.} \\ & \sum_{j=1}^n Y_{rj} \cdot \lambda_j - S_{r+} = Y_{r0} \quad \forall r : 1 \dots s \\ & \theta \cdot X_{io} - \sum_{j=1}^n X_{ij} \cdot \lambda_j + S_{i-} = \theta \cdot X_{i0} \quad \forall i \in ID \\ & \sum_{j=1}^n X_{ij} \cdot \lambda_j + S_{i-} = X_{i0} \quad \forall i \notin ID \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j, S_{r+}, S_{i-} \geq 0 \end{aligned} \tag{M.6}$$

Otra forma de tratar las variables no controlables es la propuesta por **Ray (1988 y 1991)** y después completada por **McCarty y Yaisawang (1993)**, **Lovell et alter (1997)** y **Chiligerian (1995)** mediante la cual primero se computaría la eficiencia mediante un modelo DEA sin considerar las variables no controlables y, en una segunda etapa, se regresaría la eficiencia obtenida frente a las variables no controlables, siendo los residuos de dicha regresión la eficiencia debida a la gestión puesto que representarían la parte de eficiencia no explicada por las variables no controlables.

Un tercer método es el propuesto por **Pastor (1994)** y mejorado por **Muñiz (1998:461-69)**. El método constaría de dos etapas. En la primera se aplicaría el método DEA con todos los outputs pero sólo con inputs exógenos y después, para las unidades ineficientes, se elevarían los outputs hasta que se consideraran eficientes, descontando así el efecto negativo que para esas unidades de producción tendría el entorno desfavorable. En la segunda etapa se volvería a aplicar DEA sobre las DMUs iniciales más las modificadas tomando todos los outputs y sólo los inputs controlables. **Muñiz (1998:467-68)** modificó el planteamiento de **Pastor (1994)** basándose en que la hipótesis de relación positiva entre inputs y outputs de DEA impedía que la propuesta de **Pastor (1994)** fuera válida. Partiendo de ese supuesto de relación positiva una unidad con buenas condiciones contextuales (y por tanto con inputs discrecionales elevados) tendría más posibilidades de ser evaluada como ineficiente y, por tanto, de que sus outputs se vieran corregidos al alza para convertirla en eficiente de cara a la implementación de la segunda etapa. Con lo cual una unidad con condiciones contextuales mejores que la media se vería beneficiada del incremento de output que, teóricamente, debería servir para compensarla de su peor situación inicial. Como solución **Muñiz (1998)** propone introducir las variables exógenas en el modelo DEA con relación negativa con el output⁵¹. Como ninguna de estas alternativas ha demostrado ofrecer

siendo θ el parámetro de eficiencia y λ_j las ponderaciones.

El modelo aquí expuesto es el ampliado por **Charnes et alter (1997:52)**. En realidad **Banker y Morey (1986:50)** propusieron un modelo bajo los mismos fundamentos pero sin ampliar la no discrecionalidad a los outputs. En definitiva se observa que el parámetro de eficiencia sólo afectaría a las restricciones de inputs pertenecientes al conjunto discrecional. La aplicación de la norma de modificación de las restricciones y función objetivo ante la existencia de variables no discrecionales es fácilmente extensible al resto de modelos DEA.

⁵¹ Recientemente **Muñiz (2000)** expuso una nueva variante basada en modificaciones de la propuesta de **Fried y Lovell (1996)**.

resultados más acertados que los obtenidos por el método de **Banker y Morey (1986b)**, en la parte empírica de esta tesis se empleará éste por ser el original y el más empleado en los diversos estudios consultados⁵².

Una variante adicional que se utilizará es la desarrollada por **Lovell y Pastor (1995:147-50)**, los cuales proponen un modelo DEA BCCp-O que incluye una modificación que convierte sus resultados en invariables respecto a las unidades en las que se midan las variables y frente a las traslaciones que se puedan efectuar a los inputs. Es decir, evita que la ordenación de la eficiencia de las DMUs dada por DEA sea susceptible de variación ante cambios en las unidades en que se midan los inputs y los outputs (adimensionalidad). Para que esto sea así el modelo a utilizar tendrá que tener la siguiente concreción:

$$\begin{aligned}
 & \text{Max}_{\lambda, \phi_0, S_{r+}, S_{i-}} \quad \phi_0 + \varepsilon \sum_{i=1}^m \left(\frac{S_{i-}}{\sigma_{i-}} \right) + \varepsilon \sum_{r=1}^s \left(\frac{S_{r+}}{\sigma_{r+}} \right) \\
 & \text{S.A} \\
 & \sum_{j \in \bigcup_{k=1}^{K'} Dk} \lambda_j \cdot Y_{rj} - S_{r+} = \phi_0 \cdot Y_{ro}, \forall r : 1 \dots s \\
 & \sum_{j \in \bigcup_{k=1}^{K'} Dk} \lambda_j \cdot X_{ij} + S_{i-} = X_{ro}, \forall i : 1 \dots m \\
 & \sum_{j \in \bigcup_{k=1}^{K'} Dk} \lambda_j = 1 \\
 & \lambda_j, S_{i-}, S_{r+} \geq 0
 \end{aligned} \tag{M.7}$$

donde σ_{i-} y σ_{r+} son las desviaciones típicas de los inputs y outputs respectivamente. El motivo de esta especificación del modelo radica en que la conversión de la medida de los inputs y outputs por cualquier multiplicador $\alpha > 0$ haría que las desviaciones típicas de los

⁵² Esta afirmación podrá corroborarse en el epígrafe 2.

datos fueran ahora $\alpha\sigma_{i-}$ y $\alpha\sigma_{r+}$ con lo cual en la función objetivo quedaría $\frac{\alpha S_{i-}}{\alpha\sigma_{i-}}$ y $\frac{\alpha S_{r+}}{\alpha\sigma_{r+}}$ lo

cual es exactamente lo que ya existía $\frac{S_{i-}}{\sigma_{i-}}$ y $\frac{S_{r+}}{\sigma_{r+}}$.

Por último, cabría realizar una breve exposición de las ventajas e inconvenientes que posee la aplicación de la técnica DEA.

Dentro de las ventajas cabe destacar:

- a) Se adapta al análisis de sectores que emplean en su proceso productivo múltiples inputs y generan varios outputs.
- b) Se ajusta a situaciones en los que los precios de factores y productos son desconocidos o difícilmente calculables, puesto que el propio DEA el que genera valores para esos precios.
- c) Tal y como se explicará en la parte empírica, ofrece una gran cantidad de información particularizada para cada DMU que puede ser empleada para establecer guías de actuación de cara a mejorar la eficiencia de las unidades ineficientes.
- d) No exige del conocimiento previo de la función de producción sino, únicamente, de un conjunto de combinaciones de factores que generen cantidades de output.

Por el lado de los inconvenientes cabe resaltar:

- a) La exigencia de la homogeneidad de las unidades sometidas a análisis, necesaria para evitar que la ineficiencias de los centros sean detectadas por causa de cualquier factor no uniforme y que queda centrada en dos aspectos:
 - a.1) Homogeneidad en la escala de producción (solventada por la modificación de **Banker y Morey (1986b)**).
 - a.2) Homogeneidad en el uso de inputs y outputs y en las circunstancias que constituyen el ámbito de actuación de las unidades.
- b) La flexibilidad de la elección de las ponderaciones, además de ventaja, puede ser considerada como un problema sobre la base de la existencia de la posibilidad de que la evaluación de eficiencia de alguna/s unidad/es algún/os

inputs o outputs reciban una ponderación nula y, por tanto, no sean contemplados en el proceso de cómputo. Ello podría implicar que variables trascendentales para el sector pasaran desapercibidas al establecer conclusiones encaminadas a mejorar el modo de producción de los centros ineficientes o, incluso, a basar las mismas en variables secundarias (no obstante, existen soluciones para estos casos ya comentadas con anterioridad en esta tesis).

- c) Es un método determinístico y, por tanto, supone que cualquier alejamiento de la frontera de una asignación de insumos y productos se deberá únicamente a un comportamiento ineficiente, no dando paso a la cabida de ineficiencia por motivos aleatorios.
- d) Es un modelo con el que se debe llevar especial cuidado al seleccionar las variables a incluir pues no existen tests adecuados para estimar si los resultados del análisis son estables o variarían significativamente con la utilización de otro tipo de variables. Ello obliga a realizar estudios de sensibilidad mediante diferentes especificaciones siempre y cuando los datos necesarios para ello sean accesibles.
- e) La fiabilidad de los resultados también depende de la relación existente entre el número de variables consideradas y el de unidades a analizar. Así, **Banker et aliter (1989)** establecían, a modo orientativo, el requisito de que el número de unidades analizadas sea mayor o igual a la suma de inputs y outputs para que el modelo tenga carácter discriminatorio (si bien otros autores como **Norman, M. Stoker (1991)** mencionan que veinte unidades sería suficientes sin hacer depender el número de la cantidad de variables o **Mancebón (1996:297)** recoge la recomendación de que el número de entidades analizadas sea al menos el triple de las variables relevantes introducidas en el modelo).

En conclusión, cabe pensar que es posible afirmar que la técnica DEA sea lo suficientemente apropiada para ser utilizada en el ámbito del sector público a tenor de las posibilidades que existen de minimizar sus inconvenientes, de las ventajas que ofrece sobre otro tipo de métodos de cuantificación de la eficiencia así como del uso que ha

recibido para tal fin en trabajos con similares características al que aquí se pretende desarrollar.

