

El análisis envolvente de datos en la construcción de indicadores sintéticos. Una aplicación a las provincias españolas

MARTÍNEZ ROGET, F.**; MIGUEL DOMÍNGUEZ, J.C. DE* y MURIAS FERNÁNDEZ, P.*
**Departamento de Métodos Cuantitativos para la Economía y la Empresa; **Departamento de Economía Aplicada. Facultad de Ciencias Económicas y Empresariales. Universidad de Santiago de Compostela.*

Tfno.: 981563100; E-mail: eafimaro@usc.es; E-mail: eccharba@usc.es; E-mail: pmurias@usc.es;

RESUMEN

El objetivo de este trabajo es mostrar la utilidad del Análisis Envolvente de Datos (DEA) en la construcción de indicadores sintéticos que permiten la comparación regional. Para ello se propone su utilización para la comparación de las provincias españolas en relación con los problemas que más preocupan a los ciudadanos, lo que indirectamente, constituye un reflejo del nivel de bienestar de que se disfruta en cada una de ellas. Estos problemas son los recogidos mensualmente en el Barómetro de Opinión, la encuesta elaborada por el Centro de Investigaciones Sociológicas (CIS). A partir de una serie de indicadores parciales se estima un índice sintético para cada una de las provincias, utilizando el DEA como metodología de agregación y ponderación.

Palabras clave: Comparación provincial; DEA; Índice sintético.

DEA Construction of Composite Indicators. An Application to the Spanish Provinces

ABSTRACT

The purpose of this paper is to show the utility of Data Envelopment Analysis (DEA) in the construction of composite indicators which allow for regional comparisons. With this aim, its use is proposed in the comparison between the Spanish provinces in relation to the most worrying problems identified for Spanish citizens, which indirectly approximates every province's wellbeing. To identify those problems, the *Barómetro de Opinión* was used, a monthly survey carried out by the *Centro de Investigaciones Sociológicas (CIS)*. Basing on a series of sub-indicators a composite index is estimated for every province, using DEA as the methodological approach in the weighting and aggregation process.

Keywords: Regional comparison, DEA, Composite indicator.

Clasificación JEL: C14, R10, I30

Artículo recibido en enero de 2005 y aceptado para su publicación en julio de 2005.

Artículo disponible en versión electrónica en la página www.revista-eea.net, ref.: E-23308.

ISSN 1697-5731 (online) – ISSN 1133-3197 (print)

1. INTRODUCCIÓN

La construcción de indicadores sintéticos se ha convertido en los últimos años en una de las técnicas más utilizadas para la comparación regional. A finales de los sesenta y sobre todo en la década de los setenta, la visión tradicional que aproximaba el bienestar a partir de indicadores económicos o monetarios quedó superada por un movimiento que reclamaba la introducción de indicadores que abarcasen más ámbitos que el estrictamente económico (Nordhaus y Tobin, 1972; Sen 1979). A partir de este momento se asume que el bienestar es un concepto multidimensional, y que el aspecto económico no es más que uno de los componentes del mismo. Se inaugura así la corriente de los Indicadores Sociales, y se genera la necesidad de agregar la información procedente de múltiples componentes en un único índice o indicador sintético.

Un informe de la Comisión Europea apunta dos notas definitorias para un indicador sintético o compuesto. Se trata de un indicador construido a partir de otros que no tienen una unidad de medida común, y para los que no existe una forma obvia de ponderarlos (Joint Research Centre, 2002). Aunque son varios y complejos los pasos que deben darse en la construcción de un indicador de este tipo (OCDE, 2003), la propia definición identifica uno de los más problemáticos: la ponderación de los indicadores parciales que constituyen el índice. Sobre la complejidad de esta cuestión puede consultarse Cox et al. (1992) o Wall et al. (1995).

El problema de la ponderación de los indicadores parciales constituye el núcleo mismo de la construcción de un indicador sintético. Al mismo tiempo es el paso más controvertido de todo el proceso, ya que no existe ninguna metodología que sea generalmente aceptada. Por supuesto, las ponderaciones deben elegirse teniendo en cuenta el contexto teórico en el que se enmarca la construcción del indicador. Lo que ocurre es que, en la práctica totalidad de los casos, ese contexto no proporciona la información suficiente para seleccionar un conjunto de ponderaciones que resulten plenamente justificables. Por esa razón, la asignación de pesos a los distintos indicadores nunca puede realizarse de manera totalmente objetiva. Tradicionalmente se ha optado por asignar la misma ponderación a todas las componentes del índice, y de esta forma se calcula el índice de Desarrollo Humano (IDH) desde la ONU o el índice de bienestar económico (IEWB) propuesto para los países OCDE por Osberg y Sharpe (2002). El panorama actual presenta un importante número de propuestas, unas emplean algunos de los métodos estadísticos disponibles mientras que otras se denominan participativas porque se basan en la opinión emitida por expertos o por el público en general. Entre los enfoques metodológicos estadísticos destacan los que se basan en el uso de Componentes Principales o Análisis Factorial, el análisis de regresión o más recientemente el Análisis Envolvente de Datos. En el lado de las propuestas participativas se encuentran el método de la Opinión Pública (Public

Opinion) o el de consejo de expertos a través de técnicas como la *Budget Allocation*. A medio camino entre el enfoque estadístico y el participativo está el método de las Jerarquías Analíticas (AHP). Este trabajo se va a centrar en el Análisis Envolvente de Datos, porque es la técnica que se emplea en este caso, pero puede encontrarse una revisión de los distintos métodos en Joint Research Centre (2002), así como un análisis exhaustivo de los indicadores sintéticos más conocidos.

En España la experiencia más popular en la construcción de índices sintéticos es sin duda alguna, la promovida por la Fundación La Caixa. Esta entidad elabora anualmente un índice de bienestar social para las provincias españolas empleando un método estadístico basado en la corrección de correlaciones entre los distintos indicadores parciales (DP2). Este método, desarrollado inicialmente por Pena (1979) es ampliamente revisado en un trabajo posterior de Zarzosa (1996).

Sin embargo, en Quadrado y López-Bazo (1996) o Quadrado et al. (2001) pueden encontrarse índices compuestos basados en otros métodos estadísticos relacionados también con el cálculo de correlaciones pero en la esfera del Análisis de Componentes Principales. El primero de estos trabajos construye un índice de bienestar social mientras que en el segundo se analizan las disparidades regionales en la distribución de los centros de Educación Superior.

En este trabajo se presenta una aplicación del Análisis Envolvente de Datos a la construcción de un indicador sintético para las provincias españolas. Aunque esta técnica ha sido utilizada tradicionalmente para la estimación de la eficiencia relativa de un conjunto de unidades productivas, en los últimos tiempos se han desarrollado otras aplicaciones, entre las que se incluye su utilización para la ponderación de indicadores parciales.

En la siguiente sección se va a describir brevemente el funcionamiento de DEA, al mismo tiempo que se enumeran algunas de sus aplicaciones fuera del contexto de la eficiencia productiva. En particular, se revisan aquellas que están relacionadas con la construcción de índices sintéticos.

En el apartado 3 se plantea una aplicación de DEA para la estimación de un índice sintético para las provincias españolas. En este caso no se trata de comparar las provincias con respecto a unas componentes predeterminadas del bienestar económico o social sino de conocer cuál es su situación en relación con los problemas que más preocupan a los ciudadanos. Para ello se toma como referencia la encuesta que mensualmente elabora el Centro de Investigaciones Sociológicas (CIS) conocida como Barómetro (CIS, 2001).

En el apartado 4 se presentan algunos de los resultados obtenidos para las distintas provincias españolas y finalmente en una última sección se recogen las principales consideraciones a las que se ha llegado.

2. DEA EN LA CONSTRUCCIÓN DE INDICADORES SINTÉTICOS

El Análisis Envolvente de Datos fue inicialmente propuesto por Charnes, Cooper y Rhodes (1978) como una herramienta para estimar la eficiencia técnica de un conjunto de unidades productivas en contextos caracterizados por múltiples inputs y outputs y falta de información sobre los precios de los mismos. La técnica, que se basa en el uso de la programación lineal, ha sido ampliamente empleada en décadas posteriores en el contexto de la estimación de la eficiencia, como se puede comprobar en varias recopilaciones bibliográficas como la elaborada por Seiford (1996) o la más reciente de Tavares, (2002).

En términos generales DEA sugiere que, dado un proceso productivo caracterizado por rendimientos constantes a escala, en el que se utilizan p inputs (x_1, x_2, \dots, x_p) en la producción de q outputs (y_1, y_2, \dots, y_q) y para el que se cuenta con observaciones correspondientes a n unidades productivas, la eficiencia técnica de una determinada unidad 0 puede estimarse a través del siguiente programa lineal:

$$\max_{u_r, v_i} \frac{\sum_{r=1}^q u_r y_{r0}}{\sum_{i=1}^p v_i x_{i0}} \quad (1)$$

s.a.:

$$\frac{\sum_{r=1}^q u_r y_{rj}}{\sum_{i=1}^p v_i x_{ij}} \leq 1 \quad j = 1 \dots n \quad (2)$$

$$u_r, v_i \geq \varepsilon \quad (3)$$

$$r = 1 \dots q \quad i = 1 \dots p$$

donde ε representa un valor infinitesimalmente pequeño pero mayor que cero.

La idea que subyace a este planteamiento es la de maximizar una especie de índice de productividad total de factores (unidades de output producido por cada unidad de input empleada) para cada unidad. En el numerador del índice se resumen todos los outputs en un único output virtual, igual que en el denominador un único input virtual recoge todos los factores empleados en el proceso productivo. Sin embargo para esta agregación no se utiliza un sistema de precios convencional, por otra parte generalmente inexistente, sino un conjunto de ponderaciones $(u_1, \dots, u_q, v_1, \dots, v_p)$, cuyo valor es precisamente el que se pretende hallar, de tal forma que maximice la

ratio para cada unidad y al mismo tiempo haga que las ratios de las demás sean menores que uno.

Generalmente no se resuelve directamente este problema fraccional, sino que se utiliza alguna de sus variantes lineales para obtener la eficiencia estimada de la unidad 0, que será igual a 1 si ésta es técnicamente eficiente y menor que 1 en caso contrario. Teniendo en cuenta que las ponderaciones han sido estimadas de tal forma que otorguen el mayor valor posible a la función objetivo, si una determinada unidad no alcanza el valor 1 en esa función con sus pesos óptimos, tampoco podrá alcanzarlo con ningún otro conjunto de ponderaciones. Un problema similar al expuesto debe ser resuelto para cada una de las n unidades. Para una mayor descripción de DEA o de su funcionamiento en el campo de la estimación de la eficiencia puede consultarse en Charnes et al. (1994) y más brevemente en Boussofiane et al. (1991).

En el contexto descrito, DEA permite evaluar con qué grado de eficiencia se están convirtiendo múltiples inputs en múltiples outputs. Sin embargo, los modelos DEA no tienen por qué asumir necesariamente la relación de causalidad que existe entre inputs y outputs en el contexto productivo. La técnica es una potente herramienta de programación lineal que puede permitir otro tipo de análisis multidimensional fuera del ámbito productivo (ver Hashimoto, 1999), como por ejemplo la construcción de índices sintéticos a partir de una serie de indicadores simples. Así, si los inputs tradicionales son sustituidos por indicadores de carácter negativo (del tipo “cuanto más peor”) y los outputs por indicadores de carácter positivo (del tipo “cuanto más mejor”), el Análisis Envoltente ofrecerá una evaluación comprehensiva de las combinaciones de ambos tipos de indicadores. Dicha evaluación es comprehensiva porque tiene en cuenta toda la información recogida en los indicadores simples, y además es no uniforme, en la medida en que permite que cada región evaluada ofrezca ponderaciones distintas al mismo indicador.

La función objetivo (1) en el modelo expuesto más arriba representa el indicador sintético, cuyo valor se obtiene para cada región en el óptimo. Por lo tanto, al maximizar la expresión (1) para una determinada región se está buscando el conjunto de ponderaciones que otorgan el mayor valor posible del indicador sintético a esa región concreta. DEA resuelve de forma endógena el problema de la ponderación de los indicadores simples y además colocando a cada región en la posición más beneficiosa.

La restricción (2), que exige que con los pesos óptimos de una determinada región, todas las demás obtengan valores del indicador sintético menores o iguales que uno, introduce otro elemento característico de DEA: la idea del “benchmarking”. Los métodos más tradicionales de construcción de indicadores sintéticos obtienen un valor determinado para cada región y asumen implícitamente que las regiones se ordenan linealmente en función de esos valores. Sin embargo a través de la expresión (2) DEA permite comparar cada región con las regiones que presentan las mejores prácticas. Estas últimas obtienen el valor máximo de 1 para el indicador y todas las demás obtienen un valor inferior.

Aunque las aplicaciones DEA en la construcción de indicadores sintéticos son escasas, resultan cada vez más frecuentes. Hashimoto y Ishikawa (1993) y Hashimoto y Kodama (1997) han empleado DEA para la estimación de un índice de bienestar social y calidad de vida en Japón, mientras que Zhu (2001) hizo lo propio para comparar el nivel de vida que ofrecían un conjunto de ciudades. Storrie y Bjurek (2000) se sirven de esta misma técnica para construir un índice sintético que refleja el funcionamiento de los mercados de trabajo en la Unión Europea. Zaim, Färe y Grosskopf (2001) evalúan el bienestar individual en un conjunto de países también a través de DEA, mientras que Malhberg y Obersteiner (2001) utilizan el Análisis Envolvente para recalculan el popular IDH elaborado por la ONU. A nivel español, los autores de este trabajo han realizado una propuesta basada en DEA para la estimación de un índice de bienestar económico para las 50 provincias españolas (Murias et al., 2004).

Parece claro que una de los principales atractivos de la aplicación del Análisis Envolvente en este contexto radica en la ponderación endógena de los indicadores parciales, que permite un equilibrio entre objetividad y subjetividad que parece relevante en esta materia. Efectivamente, si bien la asignación de pesos ha de realizarse de manera objetiva, al mismo tiempo debería asumir y respetar la subjetividad con la que distintos grupos interpretan su bienestar. El Análisis Envolvente de Datos cumple esta doble premisa. Por un lado, constituye una herramienta objetiva porque no precisa de la asignación de ponderaciones a priori. Por otro lado, es enormemente flexible porque no exige que todas las unidades concedan la misma importancia a un mismo indicador parcial. Además, aunque esta flexibilidad puede parecer inicialmente excesiva, DEA permite graduarla a través de la introducción de restricciones adicionales sobre las ponderaciones absolutas o relativas. Así, aunque las distintas unidades no tienen por qué ponderar de igual forma un mismo indicador, es posible exigirles que la importancia de dicho indicador sea siempre mayor o menor que la de otro. Esta “libertad controlada” en la fijación de ponderaciones constituye a nuestro modo de ver la principal ventaja del índice sintético proporcionado por DEA.

En este trabajo se va a utilizar DEA para estimar un indicador sintético que permita resumir la situación de las distintas provincias españolas con respecto a los que los ciudadanos consideran como los principales problemas existentes en el país. Para el análisis se ha tomado como referencia el año 2001, para el que se han analizado los diferentes barómetros mensuales, cuyos resultados se exponen en el apartado siguiente. A partir de ahí se definen y priorizan los indicadores parciales que los describen y que serán la base para la construcción del índice sintético.

Es necesario recalcar una vez más, que la perspectiva de aplicación de DEA en este caso es distinta de la tradicional en el contexto de producción. No existe ningún tipo de relación orgánica entre indicadores positivos y negativos, a diferencia de lo que ocurre con inputs y outputs, y este hecho no entra en conflicto con las bases de la metodología. Es más, así como en el ámbito productivo no tendría sentido considerar un modelo sin inputs o un modelo sin outputs, sí podría ser empleada la técnica DEA

para obtener un indicador sintético a partir de un conjunto de indicadores exclusivamente positivos o exclusivamente negativos.

3. UN INDICADOR SINTÉTICO DEEA PARA LAS PROVINCIAS ESPAÑOLAS

Desde el año 1979, el Centro de Investigaciones Sociológicas (CIS) viene elaborando cada mes una encuesta (Barómetro), que pretende aproximar el estado de la opinión pública española en relación con la situación político-económica del país y con sus perspectivas de evolución. Aunque es frecuente que en el Barómetro se incluyan aspectos coyunturales que reflejan la actualidad social, económica y política de cada momento, muchas de las cuestiones planteadas se repiten mes a mes.

Una de estas cuestiones es la que solicita la identificación por parte del entrevistado de los tres principales problemas que existen en el país. Esta pregunta se formuló todos los meses del año 2001, con excepción de agosto, mes en el cual no se realiza el cuestionario y de octubre en el que la encuesta estuvo centrada en los ataques terroristas del 11 de septiembre. En la tabla 1 se recogen los problemas que más preocuparon a los ciudadanos en ese año ordenados de mayor a menor grado de preocupación.

Tabla 1: Principales problemas existentes en España

Terrorismo, ETA
Paro
Droga, alcoholismo
Inmigración
Inseguridad ciudadana
Problemas económicos
Problemas políticos
Problemas sociales
Sanidad

Fuente: elaboración propia a partir de los datos del CIS

El terrorismo de ETA, el paro, la droga y el alcohol, la inmigración, la delincuencia y la inseguridad ciudadana se han situado todos los meses entre las principales preocupaciones de los ciudadanos en el año 2001. En cambio otras cuestiones han preocupado profundamente a los encuestados en algunos meses concretos, pero su importancia relativa no se ha mantenido en el resto del año. Tal es el caso del “mal de las vacas locas” y otros problemas derivados de la relación entre ganadería y alimentación, que aparecía entre los principales problemas en el primer trimestre del 2001.

Cada uno de los problemas de la tabla 1 podría constituir una componente en la terminología de indicadores sociales (Zarzosa, 1996, pag. 65), que habría por lo tanto

que aproximar a través de uno o varios indicadores parciales. Sin embargo, tres de esos problemas no han sido explícitamente considerados en este trabajo. En primer lugar, se ha excluido la componente Terrorismo por varias razones. A nivel provincial no existe un indicador específico de este problema. Además, si existiera, su inclusión en el modelo podría introducir un grave elemento de heterogeneidad porque afectaría exclusivamente a un grupo reducido de provincias. En cualquier caso, las acciones terroristas también se incluyen en el apartado de Inseguridad Ciudadana, y en los indicadores que hemos seleccionado para representar esta componente, por lo que en cierta medida están presentes en el modelo especificado.

En segundo lugar, no se ha tenido en cuenta la componente Inmigración de manera explícita. La razón de nuevo es la inexistencia de un indicador adecuado para su medición. Obviamente, no consideramos que la población inmigrante constituya un problema *per se*, por lo tanto no tiene sentido emplear indicadores del volumen de inmigrantes. En realidad, suponemos que cuando los encuestados identifican la inmigración como un problema, se están refiriendo fundamentalmente a aspectos derivados de la no inserción social de algunos grupos de inmigrantes, como la delincuencia o el consumo de drogas. Lamentablemente, no existe ningún indicador que recoja el porcentaje de delincuencia común o tráfico de drogas generado por población inmigrante. De todas formas, estos aspectos están también incluidos en el modelo, dado que aparecen recogidos de manera general en la componente Inseguridad Ciudadana.

Finalmente, tampoco se incluye la componente Problemas Políticos, por la dificultad que nos planteaba su definición y la consecuente búsqueda de indicadores.

Para cada uno de los restantes problemas-componentes, se ha tratado de buscar algún indicador que permita describirlos. Esta tarea ha resultado muy difícil, debido a la limitación de los datos existentes a nivel provincial. Además, mientras que algunos problemas se aproximan totalmente a través de indicadores objetivos, otros sólo se han podido aproximar de forma limitada, ya que están más relacionados con las percepciones de los individuos.

La componente Paro se ha representado por el indicador tasa de paro registrado a nivel provincial. Con respecto a la componente Drogas y Alcohol, se ha considerado el indicador muertes violentas por síndrome de dependencia y psicosis alcohólica. Se ha intentado obtener algún indicador del consumo de drogas a nivel provincial, pero ha resultado imposible. La componente Inseguridad Ciudadana se ha aproximado a través de dos indicadores: el número de delitos y el de muertos en accidentes de tráfico, entendiendo por lo tanto la seguridad en un sentido amplio. Desgraciadamente no se ha podido recoger un aspecto más subjetivo, relacionado con la percepción concreta que los individuos tienen de su propia seguridad.

La cuarta componente considerada es la de Problemas Económicos. Para representarla se han incluido dos indicadores: el nivel de renta bruta disponible per capita y el porcentaje de hogares que registran dificultades para llegar a fin de mes. Este dato, recogido en la Encuesta Continua de Presupuestos Familiares elaborada para el

INE, está disponible a nivel autonómico, pero no provincial. En cualquier caso, y considerando la relevancia de dicha información para estimar los problemas económicos de las familias, se ha mantenido este indicador imputando a cada provincia el valor correspondiente a la comunidad autónoma a la que pertenece.

Al igual que la componente Problemas Políticos, la denominada Problemas Sociales puede dar lugar a dificultades en su interpretación. Sin embargo, por un lado, del análisis de las respuestas obtenidas a lo largo del año es posible descartar una serie de cuestiones que podrían ser incluidas dentro de esta categoría. Por otro, se ha considerado necesario reflejar dos problemas que están presentes en la mayoría de las sociedades desarrolladas y de forma más acentuada aún en la española. Estos dos problemas son el fuerte proceso de envejecimiento de la población y las dificultades existentes para conciliar la vida familiar y la laboral, que afecta sobre todo a la población femenina.

En relación con el primero de los problemas se ha utilizado el indicador número de plazas en residencias de la tercera edad por habitantes. Para aproximar el segundo, hubiese resultado interesante, considerar el número de plazas de guardería existentes en cada una de las provincias españolas. Dicha información que está disponible en el Anuario de Estadísticas Laborales y de Asuntos Sociales, 2002, no incluye datos del País Vasco ni de Navarra, por lo que finalmente no se ha incorporado.

Finalmente, para recoger la componente Sanidad se ha recurrido al número de camas en centros hospitalarios (régimen de internado), que por otro lado es un indicador tradicionalmente utilizado en la aproximación de esta componente.

Como resumen de lo expuesto hasta el momento, la tabla 2 recoge los ocho indicadores parciales clasificándolos en dos grupos. Por un lado aparecen los indicadores de carácter negativo, del tipo “cuanto más peor”, que estarán en el denominador de la función objetivo. Por otro lado están los de carácter positivo, del tipo “cuanto más mejor” que estarán en el denominador.

Tabla 2: Indicadores parciales

NEGATIVOS	POSITIVOS
X_1 : Tasa de paro (<i>Paro</i>) X_2 : Muertes violentas por síndrome de dependencia (<i>Alcohol</i>) X_3 : Delitos (<i>Delitos</i>) X_4 : Muertes por accidentes de tráfico (<i>Accid</i>) X_5 : Dificultad para llegar a fin de mes (<i>Dificultad</i>)	Y_1 : Renta disponible bruta (<i>Renta</i>) Y_2 : Plazas residencias (Plazas) Y_3 : Camas hospitales (<i>Camas</i>)

En relación al marco temporal ya se ha comentado que el trabajo estará referido al año 2001, aunque se deben hacer dos puntualizaciones. En primer lugar cuando estos

datos no estén disponibles para dicho año se considerarán los del año anterior. Esto ha ocurrido por ejemplo en los casos del número de camas en centros hospitalarios y del número de plazas en residencias de la tercera edad. En segundo lugar y para limitar el impacto de sucesos puntuales con una elevada incidencia en un determinado año en alguna provincia, en algunos indicadores se ha considerado la media de los datos correspondientes a los años 1999, 2000 y 2001. Esto se ha hecho con los delitos y con los accidentes de tráfico.

Los datos utilizados en el análisis se han obtenido fundamentalmente del Instituto Nacional de Estadística (Indicadores Sociales. Edición 2003 y Encuesta Continua de Presupuestos Familiares) y del Anuario Social de La Caixa (2003). De esta última fuente se consiguió el número de camas en centros hospitalarios y el número de plazas en residencias de la tercera edad existentes en cada provincia.

Es necesario puntualizar que los valores considerados para los indicadores se han relativizado por la población de cada provincia, a excepción de las plazas en residencias de la tercera edad para las que se ha considerado la población mayor de 65 años. Además, para evitar problemas de agregación relacionados con la escala de medida, se ha considerado la media nacional igual a 100 para cada indicador. De esta forma, todos los indicadores han sido expresados como números índices normalizados.

Con el objetivo de garantizar que la contribución de cada indicador parcial al índice sintético respete la importancia de la componente a la cual representa, se han impuesto restricciones adicionales sobre las aportaciones de cada indicador parcial al sintético. Así por ejemplo, dado que el paro es la respuesta con la mayor frecuencia media, se exige que la aportación al índice sintético correspondiente sea también la mayor. Midiendo la importancia de cada indicador en función de la frecuencia de respuesta, se ha obtenido una ordenación completa de todos ellos.

La formulación matemática de las restricciones es la siguiente:

$$IV_{1j} \geq IV_{2j} \quad (4)$$

$$IV_{2j} \geq IV_{3j} + IV_{4j} \quad (5)$$

$$IV_{3j} + IV_{4j} \geq IV_{5j} + OV_{1j} \quad (6)$$

$$IV_{5j} + OV_{1j} \geq OV_{2j} \quad (7)$$

$$OV_{2j} \geq OV_{3j} \quad (8)$$

$$IV_{3j} \geq IV_{4j} \quad (9)$$

$$OV_{1j} \geq IV_{5j} \quad (10)$$

$$\forall j, j=1...50$$

Donde:

$$OV_{kj} = \frac{u_k y_{kj}}{\sum_{r=1}^q u_r y_{rj}}$$

representa la aportación relativa del indicador positivo k

$$IV_{hj} = \frac{v_h x_{hj}}{\sum_{i=1}^p v_i x_{ij}}$$

representa la aportación relativa del indicador negativo h

4. RESULTADOS

Para obtener los valores del indicador sintético DEA-CIS se ha resuelto para cada una de las provincias una variante lineal del modelo general presentado en la sección anterior completado con las restricciones de prioridad señaladas. Aunque la incorporación de restricciones adicionales “endurece” la búsqueda de soluciones óptimas, en este caso no se ha presentado ningún problema de infactibilidad, es decir, todas las provincias han encontrado un sistema de ponderaciones que satisface conjuntamente todas las restricciones. En la tabla siguiente pueden observarse los valores que alcanza el indicador sintético para cada una de las provincias.

Tabla 3: Valores provinciales para el indicador sintético

<i>Provincia</i>	<i>Valor</i>	<i>Provincia</i>	<i>Valor</i>
A Coruña	0,3863	Las Palmas	0,4864
Álava	1,0000	León	0,5809
Albacete	0,4272	Lleida	0,6662
Alacant	0,4706	Lugo	0,4353
Almería	0,4770	Madrid	0,9375
Asturias	0,5253	Málaga	0,4279
Ávila	0,7023	Murcia	0,4471
Badajoz	0,3938	Navarra	0,8047
Balears	0,6542	Ourense	0,5531
Barcelona	0,8177	Palencia	0,8263
Burgos	0,6724	Pontevedra	0,3694
Cáceres	0,5993	Rioja, La	0,7327
Cádiz	0,3915	Salamanca	0,7360
Cantabria	0,6332	S.C. Tenerife	0,5307
Castelló	0,5879	Segovia	0,6447
Ciudad Real	0,4636	Sevilla	0,4435
Córdoba	0,4477	Soria	1,0000
Cuenca	0,5388	Tarragona	0,5910
Girona	0,9113	Teruel	0,7784
Granada	0,3470	Toledo	0,7156
Guadalajara	0,7726	Valencia	0,5560
Guipúzcoa	0,7902	Valladolid	0,7540
Huelva	0,5051	Vizcaya	0,7319
Huesca	0,8530	Zamora	0,5426
Jaén	0,4373	Zaragoza	0,7869

Lo primero que parece necesario destacar, es que sólo Álava y Soria alcanzan el valor unitario para el indicador. Por eso, únicamente de estas dos provincias puede decirse que presentan una buena situación general con respecto a los problemas destacados por los encuestados en el año 2001. Obviamente el número de unidades con índice unitario habría sido superior de no haber tenido en cuenta las restricciones, pero éstas se han considerado necesarias para asegurar la coherencia del modelo con los resultados de la encuesta. Además, la presencia de las restricciones impide que alguna unidad se aproveche de su buen comportamiento con respecto a un único indicador o conjunto de indicadores, asignándoles todo el peso a éstos e ignorando otros en el cálculo del indicador sintético.

Se podría pormenorizar un poco más en el análisis de los indicadores parciales de estas dos provincias. Alava tiene una tasa de paro del 8,8%, por debajo de la media nacional¹ que es del 12%. No es, sin embargo, en este indicador donde se manifiesta su superioridad de la provincia, pues hay otras como Soria, Castellón o Lleida donde la tasa de paro es menor. Álava destaca por encima de las demás provincias en el indicador parcial “Alcohol”, donde registra los valores más bajos de todas las provincias españolas. Los indicadores que sintetizan la inseguridad ciudadana (“Delitos” y “Accidentes”) no presentan un comportamiento similar. En cuanto al primero, los valores de Álava, aunque alejados de los registrados por algunas provincias andaluzas como Cádiz, Málaga o Granada, son bastante elevados. Algo similar ocurre con las cifras de muertos en accidentes de tráfico. La situación económica, sobre todo en cuanto a la R.B.D., es muy buena con respecto a las demás provincias y la situación socio-sanitaria reflejada por el número de plazas en residencias para la tercera edad y el número de camas en hospitales, también está muy por encima de la media española.

La buena situación de la provincia de Soria en el índice es reflejo, sobre todo de su reducida tasa de paro, la más baja de todas las provincias españolas y cuyo valor es inferior a la media nacional. Se podrían hacer en este caso algunas matizaciones, pues esta reducida tasa de paro puede estar condicionada por el intenso proceso migratorio registrado en la provincia, por su fuerte proceso de envejecimiento y en consecuencia de su pérdida de población, sobre todo la más joven. La provincia de Soria también destaca por el número de plazas en residencias de la tercera edad, donde ocupa los primeros lugares, sólo superada por Guadalajara y Palencia.

Las provincias de Madrid y Girona están también relativamente próximas al índice unitario y las siguientes mejor situadas son Huesca, Palencia, Barcelona, Navarra, Guipúzcoa y Zaragoza. A excepción de Madrid y Palencia, se trata de provincias del cuadrante nordeste de la península.

Madrid que es la provincia que está más próxima a alcanzar el valor unitario en el índice, presenta una situación relativamente buena en todos los indicadores, desta-

1. Entendiendo por media nacional la media de los valores registrados en las 50 provincias incluidas en el cálculo del indicador.

por habitante y que posee una de los menores ratios de plazas en Residencias, no resulta extraño que obtenga el valor más bajo en el índice.

Las dos variables más ponderadas en el índice (“Paro” y “Alcohol”) son también las dos variables en las cuales Pontevedra y A Coruña presentan los peores resultados. La situación de ambas provincias se ve agravada en cuanto al número de plazas en residencias para la tercera edad, indicador en el cual ocupan los últimos lugares.

La situación de Cádiz resulta especialmente mala en cuanto a la tasa de paro registrado y en cuanto al número de delitos. En ambos indicadores presenta los peores resultados de entre todas las provincias. La mala situación económica, indicada por el elevado paro, se refleja también en la RBDpc, cuya cifra es la segunda más baja de España, sólo por delante de Badajoz.

Hay un grupo de provincias cuya situación merece ser comentada por diferir su posición en el índice con la que ocuparían si se hiciese una ordenación en función de la renta. Podríamos hacer dos grupos en función de que dicha posición sea peor o mejor.

En la primera situación se encontrarían Guipúzcoa, Vizcaya, Baleares y Lleida. En el segundo grupo merecerían ser destacadas las provincias castellano-leonesas, Palencia y Salamanca y las castellano-manchegas, Guadalajara y Toledo.

La situación de Guipúzcoa vendría explicada sobre todo por el elevado número de delitos registrado en la provincia, mientras que en el caso de Vizcaya, su elevada tasa de paro, consecuencia de las transformaciones económicas originadas a raíz de la crisis industrial iniciada en los ochenta, sería el factor clave para explicar su relativa peor posición en el índice.

Baleares a pesar de ser una de las provincias españolas con RBDpc más elevada es también donde hay una mayor proporción de hogares que manifiestan tener dificultades para llegar a fin de mes. Lleida es una de las provincias españolas con un mayor índice de accidentes de tráfico, sólo por detrás de Cuenca, Huesca y Soria.

En cuanto a las provincias cuya posición es mejor en el índice que la que obtendrían si se ordenasen en función de la renta, Palencia destaca por ser la provincia española con un mayor número de plazas en residencias de la tercera edad en relación con este colectivo de personas y también por poseer el mayor número de camas en hospitales por habitante. Salamanca, además de disponer de una importante oferta de plazas en residencias para la tercera edad en comparación con la media española, obtiene buenos resultados en los indicadores de “Alcohol” y “Delitos”.

Guadalajara y Toledo con una RBDpc bastante por debajo de la media española tienen sin embargo, una mejor situación que la media en cuanto al paro registrado. A esto hay que unir que ambas provincias se encuentran entre las primeras en cuanto al indicador que recoge las plazas en residencias para la tercera edad y en el caso de Toledo, destaca por ser la tercera provincia española con una menor ratio de delitos por habitante.

A continuación, en la tabla 4 aparecen recogidas las contribuciones relativas de cada indicador parcial al valor del sintético. A través de su observación, es posible identificar los aspectos fuertes y débiles de cada una de las provincias y por lo tanto analizar también las razones del comportamiento diferencial de Cáceres y Ourense con respecto al resto de provincias de su comunidad autónoma.

Tabla 4: Aportaciones relativas de los indicadores parciales

	<i>Paro</i>	<i>Alcohol</i>	<i>Delitos</i>	<i>Accid</i>	<i>Dificultad</i>	<i>Renta</i>	<i>Camas</i>	<i>Plazas</i>
A Coruña	33,33	33,33	16,67	16,67	0,00	33,33	2,71	2,71
Álava	33,33	33,33	16,67	16,67	0,00	33,33	33,33	33,33
Albacete	29,04	29,04	14,52	14,52	12,87	16,17	0,00	26,55
Alacant	33,33	33,33	16,67	16,67	0,00	33,33	6,87	6,87
Almería	33,33	33,33	16,67	16,67	0,00	33,33	7,18	7,18
Asturias	28,57	28,57	14,29	14,29	14,29	14,29	19,64	19,64
Ávila	28,57	28,57	14,34	14,23	14,29	14,29	27,51	28,57
Badajoz	30,69	30,69	18,65	12,04	7,93	7,93	15,87	15,87
Balears	35,52	32,24	16,12	16,12	0,00	32,24	16,59	16,59
Barcelona	33,33	33,33	16,67	16,67	0,00	33,33	24,22	24,22
Burgos	31,61	31,61	15,80	15,80	5,18	26,43	20,40	20,40
Cáceres	29,34	29,34	16,56	12,78	11,99	11,99	23,99	23,99
Cádiz	33,33	33,33	16,67	16,67	0,00	33,33	3,29	3,29
Cantabria	28,57	28,57	14,29	14,29	14,29	14,29	25,11	25,11
Castelló	33,33	33,33	16,67	16,67	0,00	33,33	12,73	12,73
Ciudad Real	29,36	29,36	14,68	14,68	11,92	17,44	0,00	29,36
Córdoba	33,33	33,33	16,67	16,67	0,00	33,33	0,00	14,08
Cuenca	33,33	33,33	16,67	16,67	0,00	33,33	0,00	20,54
Girona	39,24	30,38	26,24	4,14	0,00	30,38	30,38	30,38
Granada	33,33	33,33	16,67	16,67	0,00	33,33	1,09	1,09
Guadalajara	29,54	29,54	16,51	13,03	11,37	18,17	29,54	29,54
Guipúzcoa	44,61	27,69	13,85	13,85	0,00	27,69	25,66	25,66
Huelva	33,33	33,33	26,58	6,76	0,00	33,33	8,59	8,59
Huesca	33,33	33,33	20,66	12,67	0,00	33,33	25,99	25,99
Jaén	33,33	33,33	16,67	16,67	0,00	33,33	6,86	6,86
Las Palmas	33,33	33,33	16,67	16,67	0,00	33,33	7,94	7,94
León	28,57	28,57	16,26	12,31	14,29	14,29	21,90	21,90
Lleida	33,33	33,33	16,67	16,67	0,00	33,33	16,64	16,64
Lugo	33,33	33,33	19,27	14,07	0,00	33,33	5,10	5,10
Madrid	33,33	33,33	16,67	16,67	0,00	33,33	30,21	30,21
Málaga	33,33	33,33	16,67	16,67	0,00	33,33	4,73	4,73
Murcia	33,33	33,33	16,67	16,67	0,00	33,33	5,69	5,69
Navarra	33,33	33,33	16,67	16,67	0,00	33,33	23,57	23,57
Ourense	33,33	33,33	21,92	11,41	0,00	33,33	10,99	10,99
Palencia	30,44	30,44	15,22	15,22	8,69	21,75	30,44	30,44
								.../...

Tabla 4: Aportaciones relativas de los indicadores parciales (cont.)

	<i>Paro</i>	<i>Alcohol</i>	<i>Delitos</i>	<i>Accid</i>	<i>Dificultad</i>	<i>Renta</i>	<i>Camas</i>	<i>Plazas</i>
Pontevedra	33,33	33,33	16,67	16,67	0,00	33,33	1,80	1,80
Rioja, La	35,18	32,41	16,20	16,20	0,00	32,41	20,43	20,43
Salamanca	28,93	28,93	21,07	7,86	13,20	15,73	28,93	28,93
S.C. Tenerife	33,33	33,33	16,67	16,67	0,00	33,33	9,87	9,87
Segovia	33,33	33,33	16,67	16,67	0,00	33,33	0,00	31,14
Sevilla	33,33	33,33	16,67	16,67	0,00	33,33	5,51	5,51
Soria	33,33	33,33	16,67	16,67	0,00	33,33	33,33	33,33
Tarragona	33,47	33,27	16,63	16,63	0,00	33,27	12,92	12,92
Teruel	33,33	33,33	20,49	12,84	0,00	33,33	22,25	22,25
Toledo	33,33	33,33	20,61	12,72	0,00	33,33	4,89	33,33
València	33,33	33,33	16,67	16,67	0,00	33,33	11,13	11,13
Valladolid	33,33	33,33	16,67	16,67	0,00	33,33	8,74	33,33
Vizcaya	33,33	33,33	16,67	16,67	0,00	33,33	19,93	19,93
Zamora	33,33	33,33	17,16	16,17	0,00	33,33	0,00	20,92
Zaragoza	40,60	29,70	14,85	14,85	0,00	29,70	24,50	24,50
	32,97	32,11	17,19	14,92	2,81	28,90	14,58	18,51

Por supuesto, los valores de los factores virtuales vienen determinados en gran medida por las restricciones que sobre ellos se ha impuesto. Por lo tanto, no supone ninguna sorpresa ver que la mayor contribución relativa promedio procede del indicador Paro, seguido del Alcohol, de la suma de Delitos y Accidentes... Sin embargo, sí resulta digno de mención el hecho de que muchas unidades saturan algunas de las restricciones impuestas. Así, es frecuente encontrarse con provincias que, obligadas a asignar a Paro un peso relativo superior o igual a Alcohol, optan por el igual. Lo mismo ocurre entre las parejas de indicadores Delitos-Accidentes y con mucha menos frecuencia entre Renta - Camas.

Otro hecho destacable es la elevada frecuencia con la que las distintas unidades asignan un valor cero al indicador Dificultad. A pesar de que esto implica que muchas unidades no lo están considerando en la construcción del índice, se ha decidido no imponer ninguna restricción a mayores sobre la aportación de este indicador ya que, como se comentó anteriormente, no permite discriminar entre provincias por estar definido únicamente a nivel autonómico.

Además, a partir de la tabla se puede observar que los mejores resultados de la provincia de Ourense con respecto al resto de Galicia se deben fundamentalmente a los indicadores Camas y Plazas. Algo similar ocurre en el caso de Cáceres y Badajoz.

En cualquier caso, y aunque el análisis de las contribuciones resulta muy útil para establecer en qué condiciones alcanzan el índice unitario las distintas provincias, es necesario reconocer una importante limitación. En problemas de optimización como

el que se ha planteado, es frecuente encontrarse con múltiples soluciones, y por lo tanto, las aportaciones no son únicas.

5. CONSIDERACIONES FINALES

El objetivo de este trabajo consistía en mostrar la utilidad del Análisis Envolvente de Datos para la construcción de índices sintéticos que faciliten la comparación regional. En concreto, DEA ofrece una alternativa para enfrentarse a la ponderación y agregación de los indicadores parciales a partir de los cuales se construye el índice sintético. La principal ventaja de esta técnica radica en el equilibrio que introduce entre los elementos objetivos y subjetivos del aspecto que se pretende aproximar (bienestar económico, bienestar social, etc.). DEA no utiliza la asignación de ponderaciones a priori y al mismo tiempo permite que las regiones en comparación concedan una importancia distinta a un mismo indicador parcial. Más aún, esta flexibilidad puede ser limitada añadiendo al modelo original un conjunto de restricciones adicionales sobre las ponderaciones.

Con la intención de mostrar la utilidad de DEA en un caso concreto, en el trabajo se planteó la construcción de un índice que recogiese la situación de las provincias españolas en relación con los problemas que los ciudadanos consideraron como más relevantes a lo largo del año 2001. Cada uno de esos problemas constituye una componente que intentó representarse a través de uno o varios indicadores parciales. La agregación de estos indicadores se realizó empleando un modelo DEA en el que se incluyeron una serie de restricciones adicionales. Estas restricciones demuestran la flexibilidad de la técnica y permiten garantizar que la contribución de cada indicador al índice sintético respete la importancia de la componente que representa.

Los resultados concretos obtenidos con la aplicación han sido considerados como secundarios. Por un lado este no era el objetivo central del trabajo y por otro la no disponibilidad de información limitó la elección de indicadores que representasen de forma fiable las componentes seleccionadas. En España el análisis regional está condicionado por la ausencia de una base de datos amplia, homogénea y fiable. Este problema se acentúa aún más cuando se pretende realizar comparaciones a nivel provincial para las que no se dispone de los indicadores necesarios o no existen en todos los casos.

Aún con las limitaciones señaladas en el párrafo anterior, el indicador obtenido, en la medida en que intenta recoger, a través de una serie de variables socioeconómicas, los problemas que más preocupaban a los ciudadanos en el año 2001 siguiendo la encuesta del CIS, puede servir como aproximación al bienestar de las provincias españolas.

Al incluir más variables que las estrictamente económicas los resultados difieren de los que se obtendrían de una ordenación provincial en función de la renta. Es

significativo el caso de las provincias de Segovia, Vizcaya, Baleares y LLeida cuya posición en el índice es bastante peor que la que obtendrían en una hipotética clasificación utilizando la renta. Por el contrario, otras provincias como las castellano-leonesas Soria, Palencia y Salamanca y las castellano-manchegas, Guadalajara y Toledo mejoran mucho en la clasificación si se incluyen más variables que la renta.

REFERENCIAS BIBLIOGRÁFICAS

- Boussofiene, A., R.G. Dyson y E. Thanassoulis (1991) Applied Data Envelopment Analysis. *European Journal of Operational Research*, 52, pp. 1-15.
- Centro de Investigaciones Sociológicas (CIS) (2001) Barómetros de Opinión, Madrid. Disponible en <http://www.cis.es>.
- Cox, D; R. Fitzpatrick, A. Fletcher, S. Gore, D. Spiegelhalter y D. Jones (1992) Quality-of-life assessment: can we keep it. Simple?, *Journal of the Royal Statistical Society*, 155 (3), pp. 353-393.
- Charnes, A., W.W. Cooper y E. Rhodes (1978) Measuring the efficiency on Decision Making Units, *European Journal of Operational Research*, 2, pp. 429-444.
- Charnes, A., W.W. Cooper, A. Lewin y L.M. Seiford (1994) *Data Envelopment Analysis: theory, methodology and application*, Boston: Kluwer Academic.
- Fundación La Caixa (2003) *Anuario Social de España* (Barcelona).
- Hashimoto, A. y M. Ishikawa (1993) Using DEA to evaluate the state of society as measured by multiple social indicators, *Socio-Economic Planning Sciences*, 27, pp. 257-268.
- Hashimoto, A. y M. Kodama (1997) Has livability of Japan gotten better for 1956-1990?: A DEA approach, *Social Indicators Research*, 40, pp. 359-373.
- Hashimoto, A. (1997) A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research*, 97, pp. 600-604.
- Hashimoto, A. (1999) Proposing non-uniform evaluation in social systems analysis, Discussion Paper Series, 827, Institute of Policy and Planning Science, University of Tsukuba.
- Instituto Nacional de Estadística: *Indicadores Sociales. Edición 2003 y Encuesta Continua de Presupuestos Familiares*. Madrid.
- Joint Research Centre (2002) *State of the art Report on Current Methodologies and Practices for Composite Indicator Development*. Informe preparado por M. Saisana y S. Tarantola del Applied Statistics Group. Institute for the Protection and Security of the Citizen, European Commission, junio 2002.
- Mahlberg, B. y M. Obersteiner (2001): *Remeasuring the HDI by Data Envelopment Analysis*, International Institute for Applied Systems Analysis Interim Report 01-069.

- Ministerio de Trabajo y Asuntos Sociales (2002) Anuario de estadísticas laborales y de asuntos sociales 2002, Madrid.
- Murias, P., F. Martínez y J. C. de Miguel (2005) An economic wellbeing index for the Spanish provinces: a Data Envelopment Analysis approach, próxima publicación en Social Indicators Research, Springer Science+Business Media B.V., Formerly Kluwer Academic Publishers B.V.
- OCDE (2003) Composite indicators of Country performance: a critical assessment. Informe preparado por Michael Freudenberg, marzo 2003.
- Osberg, L. y A. Sharpe (2002) An Index of Economic Well-being for Selected OECD Countries, Review of Income and Wealth, 48 (3), pp. 291-316.
- Pena, J.B. (1979) La distancia P: un método para la medición del nivel de bienestar, Revista Española de Economía, 4, pp. 50-91.
- Quadrado, L. y E. López-Bazo (1996) Macroeconomic variables and regional welfare. An application to the spanish regions, European Regional Science Association (Zurich, 36 Congreso Europeo).
- Quadrado, L., S. Loman y H. Folmer (2001) Multi-dimensional analysis of regional inequality: The case of higher educational facilities in Spain, Papers in Regional Science, 80 (2), pp. 189-209.
- Seiford, L. (1996) Data Envelopment Analysis: The Evolution of the State of the Art (1978-1995), Journal of Productivity Analysis, 7, pp. 99-137.
- Storrie, D. y H. Bjurek (2000) Benchmarking European labour market performance with efficiency frontier techniques, CELMS Discussion paper, Universidad de Göteborg.
- Tavares, G. (2002) A Bibliography of Data Envelopment Analysis (1978-2001), RUTCOR Research Report RRR 01-02 (New Jersey, Rutgers University).
- UNDP (2001) Human Development Report 2001. United Nations Development Program (UNDP). Disponible en <http://www.undp.org/hdr2001/spanish/>
- Zarzosa, P. (1996) Aproximación a la medición del Bienestar Social. Universidad de Valladolid. Valladolid.
- Wall R., K. Ostertag y N. Block (1995) Synopsis of selected indicator systems for sustainable development. Informe para el proyecto de investigación, 'Further development of. Indicator systems for reporting on the environment' Ministerio Federal de Medioambiente. Fraunhofer Institute for Systems and Innovation Research, Karlsruhe.
- Zaim, O., R. Färe y S. Grosskopf (2001) An economic approach to achievement and improvement indexes, Social Indicators Research, 56, pp. 91-118.
- Zhu, J. (2001) Multidimensional quality-of-life measure with an application to Fortune's best cities, Socio-Economic Planning Sciences, 35, pp. 263-284.