

Comparación de rankings de eficiencia mediante análisis de componentes principales y DEA

Úrsula Faura Martínez

Departamento de Métodos Cuantitativos
Universidad de Murcia

Juan Cándido Gómez Gallego

Universidad Católica San Antonio, Murcia

María Concepción Pérez Cárceles

Universidad Católica San Antonio, Murcia

Juan Gómez García

Departamento de Métodos Cuantitativos
Universidad de Murcia

Resumen

Aplicando el Análisis Envolvente de Datos (DEA) y el método de Análisis de Componentes Principales (ACP) se comparan las ordenaciones de un conjunto de unidades productivas (DMU), de acuerdo a sus medidas de eficiencia. Se contrastan las condiciones de aplicación y la consistencia entre las diferentes ordenaciones. Los datos corresponden al sector de Cajas de Ahorros de España: 2002-2007.

El ranking de DMUs obtenido mediante la aplicación de ACP, es válido y fiable respecto al resultante con el análisis de supereficiencia y además evita el problema de infactibilidad que puede presentarse en la aplicación del DEA.

Palabras clave: DEA, Supereficiencia, Análisis de Componentes Principales, Rankings.

Clasificación AMS: 62-07, 62H25, 90C99

Comparison of rankings of efficiency through principal components analysis and DEA

Abstract

In this paper, Data Envelopment Analysis (DEA) and Principal Component Analysis (PCA) are used to rank a set of DMU, in order of the efficiency scores obtained. Application conditions and consistency are tested between different ranks. Data set is a sample of Spanish saving banks in the period 2002-2007.

The ranking of DMUs obtained by means of PCA's application is valid with respect to the resultant with the analysis of superefficiency and in addition the problem of unfeasibility that can appear in the application of the DEA is avoided.

Key words: DEA, Superefficiency, Principal component analysis, Rankings.

AMS classification: 62-07, 62H25, 90C99

1. Introducción

La evaluación de la eficiencia es un tema que, en los últimos años, ha suscitado un gran interés, debido fundamentalmente, al hecho de que en entornos competitivos es imprescindible utilizar eficientemente los recursos si se pretende mejorar la rentabilidad.

Una forma de estudiar la eficiencia es mediante el Análisis Envolvente de Datos, conocido como DEA (Data Envelopment Analysis). El DEA es una técnica no paramétrica para la medición de la eficiencia relativa de unidades organizacionales, en situaciones donde existen múltiples inputs y outputs.

Los orígenes de DEA se remontan a los años 70, cuando Charnes, Cooper y Rhodes (1978) desarrollaron la técnica. Desde su introducción, la investigación en DEA ha sido muy productiva¹, tanto en el ámbito teórico como en el aplicado, medición de procesos logísticos, selección de equipos industriales, regulación de servicios públicos, sector bancario, etc. En el ámbito económico, la eficiencia se concibe como una medida que pone en relación los medios empleados con los fines obtenidos, considerándose una determinada técnica, procedimiento o sistema de producción eficiente cuando para un determinado nivel de inputs es capaz de producir la máxima cantidad de output, o si para alcanzar un determinado output emplea la menor cantidad de inputs (Lovell, 1993).

Los modelos DEA parten de las cantidades de inputs empleadas y de outputs producidas por un conjunto de DMUs (Decision Making Units o Unidades de Toma de Decisiones), para determinar cuáles son las mejores opciones, comparando cada DMU con todas las posibles combinaciones lineales de todas las unidades de la muestra. El conjunto de DMUs eficientes forma lo que se denomina frontera eficiente, midiéndose la eficiencia

¹ En Cooper, Seiford y Tone (2006) se citan más de 2.000 trabajos en distintas áreas y en Emrouznejad, Parker y Tavares (2008) se pueden encontrar más de 4000 referencias, y en <http://www.deazone.com/bibliography>, se pueden añadir las referencias de trabajos realizados sobre este tema.

de cada unidad como distancia a la misma. De esta manera se pueden clasificar las distintas DMUs en función de su eficiencia.

El DEA obtiene para todas las DMUs eficientes, un valor de eficiencia igual a 1. Pero en el estudio de la eficiencia de las distintas unidades pueden aparecer varias DMUs eficientes o incluso varias DMUs ineficientes con el mismo valor de ineficiencia y en estas situaciones es difícil establecer un ranking completo y válido entre las distintas DMUs. Uno de los principales problemas de la técnica DEA es el hecho de que califica muchas unidades como eficientes, salvo en el caso de que la suma del número de inputs y outputs sea pequeña en relación con el número de observaciones. De hecho, las unidades muy especializadas pueden considerarse eficientes como consecuencia de un único input u output que, por otra parte puede ser de escasa importancia.

Para resolver este problema², se han desarrollado diversos métodos para clasificar las DMUs, entre los que se encuentran el uso de técnicas multivariantes (correlación canónica, análisis discriminante, análisis de componentes principales...) y los que emplean la técnica DEA (supereficiencia, eficiencia cruzada...). Unos trabajos complementan las técnicas multivariantes con el DEA y en otros tan solo se comparan las utilidades de unos frente a otros.

Nuestro estudio se centra en los modelos de Supereficiencia y en el Análisis de Componentes Principales, determinando de qué manera pueden complementarse ambas técnicas y cómo aprovechar las cualidades de una y otra. Comparamos ambas técnicas respecto a las ordenaciones que producen en el conjunto de DMUs evaluadas, valorando la consistencia de los rankings obtenidos en la aplicación en el sector de Cajas de Ahorros de España. En concreto, aplicamos la metodología desarrollada por Zhu (1998) para el ACP realizando algunas modificaciones sobre la misma.

2. Revisión de la literatura sobre estudios de ranking mediante análisis de componentes principales y supereficiencia

No todos los artículos que tratan la ordenación de las DMUs mencionan la existencia de ACP como metodología alternativa al DEA o viceversa. Investigadores como Adler y Golany (2001), Premachandra (2001), Serrano y Mar Molinero (2001), Vargas y Bricker (2000) y Zhu (1998) entre otros, conectaron DEA y ACP con la finalidad de llegar a una ordenación de las DMUs.

Diversos trabajos han considerado el uso de ACP como paso previo o en unión con el DEA. El primer trabajo en esta dirección es de Ueda y Hoshiai (1997) quienes aplicaron componentes principales como técnica para reducir el número de inputs y outputs y obtener así un mejor ranking de las DMUs.

Otros trabajos como Mancebón y Mar Molinero (2000), Bradley et al. (2001) y Nath (2001) usan de manera secuencial ambos métodos.

² En Adler et al. (2002) se puede encontrar una primera revisión de los distintos métodos de ranking en el contexto DEA.

Kardiyen y Örkü (2006) estudian la situación de 15 países miembros de la Unión Europea en 2002 con respecto a diversos aspectos financieros y posteriormente mediante simulación (realizan 1000 repeticiones para diferente número de inputs y outputs y diferente número de DMUs) comparan las ordenaciones entre ACP y DEA.

Azadeh et al. (2008) completan el estudio de localización de plantas solares mediante el DEA comparándolo con los resultados aplicando ACP, obteniendo una correlación entre ambos ranking de 0,621 con un nivel de significación de 0,01.

Serrano Cinca y Mar Molinero (2001) proponen una metodología utilizando conjuntamente DEA y ACP. Consideran todas las posibles combinaciones de inputs y outputs y obtienen las eficiencias para cada uno de las combinaciones (estudian así las fortalezas o debilidades de cada unidad con respecto a cada variable). Sobre la matriz de eficiencias obtenida aplican ACP. En esta misma línea, Vázquez Cueto (2008) utiliza DEA junto con ACP para obtener una ordenación de las universidades españolas. Combina diferentes modelos con distintos inputs y outputs y con las distintas eficiencias y modelos aplica ACP. Segovia González et al. (2009), estudian la eficiencia de una cartera de asegurados en el sector del automóvil, mediante estas técnicas. Ho y Wu (2009) aplican también esta metodología para estudiar la banca online y más recientemente Kordrostami et al. (2011) la adaptan para el estudio de la eficiencia de un conjunto de bancos iraníes.

3. Supereficiencia y análisis de componentes principales

3.1 Fundamentos

Uno de los métodos para ordenar las DMUs de acuerdo a su medida de eficiencia es el denominado como supereficiencia, formulado por Andersen y Petersen (1993) y perfeccionado por Wilson (1995).

Este procedimiento está íntimamente relacionado con el DEA, pues consiste en resolver un programa de programación lineal, similar al convencional en el DEA, donde la unidad evaluada se compara con una combinación lineal del resto de unidades eficientes.

La particularidad del método consiste en eliminar del programa original la restricción correspondiente a la DMU bajo estudio, con lo que los parámetros que se optimizan dejan de estar acotados por 1, y sus valores se alejan más de 1 cuanto más eficiente es la DMU evaluada.

Para una DMU eficiente la diferencia entre 1, su puntuación de eficiencia, y su puntuación de supereficiencia indica el empeoramiento que podría soportar sin dejar de ser eficiente. En la versión minimizadora de inputs, las entidades eficientes en el modelo convencional obtendrán una ratio igual o superior a la unidad, indicando el complementario a 1 de este valor el incremento de inputs que se podrían permitir estos centros sin dejar de ser eficientes.

Wilson (1995) se centró en la determinación de las DMUs eficientes atípicas, caracterizadas por un índice de supereficiencia muy bajo o muy alto según la

orientación, o no factible. El límite de este valor de supereficiencia es una cuestión abierta; Wilson (1995) propone como límite 0,6 y Mancebón (1996) lo eleva a 0,8, considerando que para que una DMU sea calificada de atípica su puntuación de supereficiencia debe estar muy alejada de 1 en las dos orientaciones del modelo.

El método de la supereficiencia tiene también algunos inconvenientes. El primero es el problema de la interpretación de la ordenación obtenida; puesto que para obtener la ordenación de supereficiencia las DMUs eficientes se evalúan con multiplicadores diferentes, no serían realmente comparables.

En segundo lugar, el método asigna a las DMUs “especializadas” una puntuación excesivamente alta. Este problema se puede evitar, o al menos corregir, completando la clasificación con el método de las regiones de seguridad, como indica Sueyoshi (1999).

En tercer lugar, el programa de supereficiencia puede no tener solución. Thrall (1996), Zhu (1996) y Seiford y Zhu (1999) estudian las condiciones bajo las cuales aparece esta circunstancia y Mehrabian et al. (1999) proponen modificaciones a la formulación del dual para asegurar la existencia de solución admisible.

En cuarto lugar, la supereficiencia no ordena el conjunto total de DMUs, solo las clasificadas como eficientes.

No obstante estos inconvenientes, el método de supereficiencia es uno de los procedimientos más utilizados para la ordenación de DMUs. Su aplicación es sencilla, solo requiere volver a aplicar el DEA con una restricción menos.

En este trabajo se ha optado por el modelo BCC que asume rendimientos variables a escala en la relación entre inputs y outputs. En la sección siguiente se presenta la formulación del modelo aplicado.

3.2 Modelo DEA-BCC

Supongamos una muestra de n DMUs tales que cada DMU_j ($j=1, 2, \dots, n$) produce s outputs, Y_{rj} ($r=1, 2, \dots, s$), utilizando m inputs, X_{ij} ($i=1, 2, \dots, m$). En DEA, la eficiencia técnica de cada una de las unidades, se define como el cociente entre la suma ponderada de los outputs y la suma ponderada de los inputs.

El modelo BCC output orientado en su formulación multiplicativa está dado por las siguientes ecuaciones:

$$\begin{aligned} \text{Min } h_0(u, v) &= \sum_{i=1}^m v_i x_{i0} + v \\ \text{s.a. } &\begin{cases} \sum_{r=1}^s \mu_r y_{r0} = 1 \\ \sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s \mu_r y_{rj} \geq 0 & j = 1, 2, \dots, n \\ u_r \geq 0 & r = 1, 2, \dots, s \\ v_i \geq 0 & i = 1, 2, \dots, m \\ v \text{ libre} \end{cases} \end{aligned}$$

El programa lineal selecciona las ponderaciones que minimizan el input virtual de la unidad evaluada, condicionadas a que su output virtual sea igual a la unidad, así como que la aplicación de dichas ponderaciones al resto de unidades de decisión evaluadas no permita que su input virtual exceda del output virtual. La unidad será eficiente si su input virtual es la unidad.

En la práctica, el cálculo de los índices de eficiencia resulta más sencillo si se utiliza la forma dual del modelo anterior, a través de la cual se construye una aproximación lineal por tramos a la verdadera frontera de producción³. La formulación dual es la siguiente:

$$\begin{aligned} \text{Max } \varphi & \\ \text{s.a. } &\begin{cases} \sum_{j=1}^n \lambda_j x_{ij} \leq x_{i0} & i = 1, 2, \dots, m \\ \sum_{j=1}^n \lambda_j y_{rj} \geq \varphi \cdot y_{r0} & r = 1, 2, \dots, s \\ \lambda_j \geq 0 & j = 1, 2, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \end{aligned}$$

En este caso, si $\varphi = 1$, la unidad evaluada se considera eficiente, pues no existe otra que produzca más o que consiga el mismo nivel de producción con menores recursos que ella.

El modelo anterior asigna a todas las unidades eficientes el mismo valor, por tanto no es útil a la hora de establecer un ranking. Por ello, Andersen y Petersen, con propósitos de establecer ordenaciones, formulan el modelo de supereficiencia. Tal modelo, que aplicamos en este trabajo, se expresa mediante las ecuaciones siguientes:

³ Esta preferencia por el modelo dual se explica por el importante esfuerzo computacional que requiere la resolución del modelo de programación lineal (Cooper et al., 2000).

$$\begin{aligned} & \text{Max } \varphi^{\text{super}} \\ & \text{s.a.} \left\{ \begin{array}{l} \sum_{j=1, j \neq 0}^n \lambda_j x_{ij} \leq x_{i0} \quad i = 1, 2, \dots, m \\ \sum_{j=1, j \neq 0}^n \lambda_j y_{rj} \geq \varphi^{\text{super}} \cdot y_{r0} \quad r = 1, 2, \dots, s \\ \lambda_j \geq 0 \quad j = 1, 2, \dots, n; j \neq 0 \\ \sum_{j=1, j \neq 0}^n \lambda_j = 1 \end{array} \right. \end{aligned}$$

3.3. Método de Zhu

Otro de los métodos utilizados para la ordenación de DMUs es el ACP. En este contexto, Zhu (1998) aplica el ACP definiendo para cada una de las DMUs el cociente entre el output y el input. Así, para cada DMU_j, se define el cociente entre el output r (y_{ij}) y el input i (x_{ij}):

$$d_{ir}^j = \frac{y_{rj}}{x_{ij}} \quad i=1, 2, \dots, m; r=1, 2, \dots, s \quad \text{con } j=1, 2, \dots, n.$$

El valor d_{ir}^j proporciona el cociente entre todos los outputs e inputs para cada DMU y cuanto mayor sea d_{ir}^j mejor será el rendimiento de la DMU_j en términos del output r y el input i. Se define $d_k^j = d_{ir}^j$, de forma que k=1 se corresponde a i=1, r=1; k=2 se corresponderá con i=1, r=2 y así sucesivamente, con k=1, 2, ..., p; p=mxs. Se construye una matriz $D = (d_1, d_2, \dots, d_p)_{n \times p}$, donde cada fila representará p ratios individuales de d_k^j para cada DMU y cada columna representará un ratio output/input específico, siendo

$$d_k = (d_k^1 \ d_k^2 \ \dots \ d_k^n)_{1 \times p} \quad k = 1, 2, \dots, p.$$

Mediante el ACP se obtienen nuevas variables independientes que son combinaciones lineales de d_1, d_2, \dots, d_p . Estas medidas son una combinación de las d_k^j , donde la importancia de cada componente principal está dada por los valores propios. Siguiendo a Zhu, a partir de la matriz de datos D:

1. Se calcula la matriz de correlación, R.
2. Se obtienen los valores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ $\left(\sum_{k=1}^p \lambda_k = p \right)$ y vectores propios I_1, I_2, \dots, I_p de R.
3. Se determinan las componentes principales, $CP_k = \sqrt{\lambda_k} I_k \quad k = 1, 2, \dots, p$.

4. Se escogen las primeras m componentes principales tales que $\frac{\sum_{k=1}^m \lambda_k}{p} > 0,9$.

Se calcula una combinación lineal de las m CP escogidas dada por $T_Z = \sum_{k=1}^m w_k CP_k$

siendo la ponderación el porcentaje de varianza explicada por la componente principal. Para determinar el signo de las ponderaciones, Zhu aplica la siguiente regla:

- si todos los elementos de la CP son negativos, la ponderación será negativa;
 - si son todos positivos, la ponderación será positiva;
 - si más de la mitad son negativos entonces la ponderación será negativa; en cualquier otro caso, positiva.
5. Para usar las puntuaciones de las componentes principales en el ranking, se trabaja con la matriz D pero tipificada, D_Z .
6. Para obtener las puntuaciones se utiliza $CP_{korden} = D_z \cdot T_Z$ y las DMUs se ordenan en función de los valores de estas puntuaciones.

3.4 Método de Premachandra

Esta metodología utilizada por Zhu es modificada por Premachandra (2001), al considerar que falta tener en cuenta el rendimiento de cada DMU respecto al total de variables d_{ir}^j , además de conocer cómo está cada DMU concreta cuando se compara con el resto de DMUs. Por ello propone las siguientes modificaciones:

- En lugar de trabajar con la matriz D construye otra matriz D' añadiendo otra variable cuyos elementos d'_{ij} para cada unidad es la suma de los elementos en las primeras columnas de la matriz D , es decir, $d'_{i,p+1} = \sum_{j=1}^p d_{ij}$ $i=1, 2, \dots, n$. La nueva variable añadida d'_{ij} tiene en cuenta el rendimiento de cada DMU con respecto al total de variables d_{ir}^j .
- Define una nueva matriz $D'' = (d''_{ij})$ dividiendo todos los elementos de cada columna de la matriz D' por el mínimo de su columna. De esta forma, cada elemento de la columna k de la matriz D'' indica cómo de bien está la DMU con respecto al input i y el output r cuando se compara con la peor DMU con respecto a las mismas variables.
- Emplea ACP sobre D'' .

4. Aplicación para el estudio de la banca en España

En esta aplicación se consideran 43 entidades de crédito, Cajas de Ahorros del sistema bancario español. La información utilizada es parte de la publicada por la Confederación Española de Cajas de Ahorros (CECA) para los años 2002, 2005 y 2007.

Las definiciones de los input y output se muestran a continuación, representando las tres áreas principales en el proceso de producción bancaria.

INPUT:

- Número de empleados: en miles de empleados
- Depósitos: miles de €
- Activo Material: miles de €

OUTPUT:

- Créditos: miles de €
- Cartera de Valores: miles de €

El análisis parte del enfoque de intermediación de la empresa bancaria y se basa en el concepto de eficiencia técnica con orientación hacia el output y rendimientos variables a escala.

En primer lugar se aplica para cada DMU el modelo BBC (Banker et al. 1984) y se obtienen los correspondientes coeficientes de eficiencia. Los resultados indican (Tabla 1), que 9 entidades operan en la frontera (BANCAJA, CAB, CAJASTUR, CAMPA, CAMPB, CAN, CAVA; CETE y CMAD).

Tabla 1

Eficiencia de las entidades bancarias en 2002 según distintos procedimientos

	$T_{Z,02}^1$	$T_{Z,02}^2$	$T_{Z,02}^3$	DEA	Super- eficiencia	$T_{P,02}^1$	$T_{P,02}^2$
BANCAJA	0,39130	0,22701	0,30915	1,00000	1,25282	0,97629	0,79522
BBK	-0,57364	-0,33717	-0,45540	0,74402	0,74402	-0,46332	-0,25634
CAB	1,41818	1,92732	1,67275	1,00000	1,42477	1,64727	1,86836
CABAD	-0,65068	-0,66465	-0,65767	0,68229	0,68229	-0,95373	-0,91995
CAG	0,96294	1,01095	0,98694	0,97297	0,97297	0,64763	0,63779
CAIXA	-0,11399	-0,00268	-0,05834	0,86521	0,86521	-0,09710	-0,07472
CAIXANOVA	0,35466	0,35805	0,35635	0,84688	0,84688	0,18340	0,17094
CAJASTUR	1,08650	0,64051	0,86351	1,00000	1,22100	0,66647	0,47505
CALR	-0,26483	-0,11834	-0,19158	0,88938	0,88938	0,12836	0,16088
CAM	-0,15207	-0,14650	-0,14929	0,89243	0,89243	0,27027	0,23286
CAMPA	0,93541	0,79110	0,86325	1,00000	1,09613	0,48285	0,34029
CAMPB	0,90320	0,91918	0,91119	1,00000	1,20305	0,25255	0,24292
CAMPE	0,16713	-0,06729	0,04992	0,77897	0,77897	-0,18210	-0,27591
CAN	0,30940	0,19918	0,25429	1,00000	1,04220	0,73205	0,71625
CAO	-0,79126	-0,77577	-0,78351	0,92780	0,92780	-0,65797	-0,72274
CAPG	-0,22819	-0,32677	-0,27748	0,76369	0,76369	-0,40559	-0,44747
CAS	0,76197	0,82933	0,79565	0,94125	0,94125	0,60367	0,59760
CASC	-0,25964	-0,10240	-0,18102	0,79808	0,79808	-0,10633	-0,00504
CASS	-0,71031	-0,37879	-0,54455	0,83280	0,83280	-0,37593	-0,14025
CAVA	0,95459	1,13746	1,04602	1,00000	1,01204	1,14174	1,23646
CC	-0,08119	0,11729	0,01805	0,90327	0,90327	0,26282	0,33514
CCM	0,25608	0,11097	0,18352	0,79214	0,79214	-0,14114	-0,20694
CDUERO	-0,43621	-0,45977	-0,44799	0,56934	0,56934	-0,73141	-0,63039
CECM	-0,63744	-0,64222	-0,63983	0,74409	0,74409	-0,66473	-0,66068
CEG	0,16898	-0,02615	0,07141	0,76432	0,76432	-0,10760	-0,16660
CEL	-0,52358	-0,19576	-0,35967	0,79122	0,79122	-0,51831	-0,35513
CEM	-0,24801	-0,32346	-0,28574	0,77504	0,77504	-0,31536	-0,35268
CEP	-0,71143	-0,53536	-0,62340	0,70500	0,70500	-0,65341	-0,52996
CES	-0,28551	-0,21525	-0,25038	0,82802	0,82802	-0,05782	-0,04754
CESP	-0,23938	-0,21991	-0,22965	0,68640	0,68640	-0,38329	-0,32571
CETA	-0,58004	-0,79882	-0,68943	0,68257	0,68257	-0,68129	-0,75213
CETE	0,77538	0,19163	0,48350	1,00000	1,06709	0,54091	0,26279
CGAC	0,38439	0,16509	0,27474	0,91615	0,91615	0,54789	0,43504
CGAG	-0,18276	-0,28648	-0,23462	0,84746	0,84746	-0,13956	-0,22558
CIAC	-0,58602	-0,50585	-0,54593	0,67302	0,67302	-0,61532	-0,52866
CM	-0,00096	0,08020	0,03962	0,81507	0,81507	0,11126	0,14812
CMAD	0,62139	0,81211	0,71675	1,00000	1,02959	0,82455	0,83867
COLONYA	-0,67501	-0,74434	-0,70968	0,69720	0,69720	-0,68949	-0,68884
CPAJ	-0,97878	-1,02416	-1,00147	0,82196	0,82196	-0,92278	-0,95898
CSUR	-0,32647	-0,33357	-0,33002	0,79396	0,79396	-0,10203	-0,11352
ÍBERCAJA	0,05732	-0,13417	-0,03842	0,80049	0,80049	0,16329	0,06634
SANOSTRA	-0,16933	-0,05728	-0,11330	0,84153	0,84153	0,07868	0,12605
UNICAJA	-0,10209	-0,09450	-0,09829	0,81413	0,81413	-0,29638	-0,30103

El problema para la ordenación está en el conjunto de DMUs con eficiencia igual a la unidad. Para resolverlo, de acuerdo con lo indicado en la metodología, se aplica el modelo de supereficiencia. Observando la quinta columna de la Tabla 1, se observa que no aparecen problemas de infactibilidad y se puede establecer el ranking de eficiencia de las entidades bancarias.

Para la aplicación del método de Análisis de Componentes Principales, se comprueba en primer lugar que aunque los datos originales no siguen una distribución normal, el cociente entre los outputs y los inputs sí es una distribución normal, por lo que puede aplicarse el ACP para estos cocientes. Se retienen tres componentes, aquellas cuya varianza total explicada es mayor al 90%. Los tres autovalores son: $\lambda_1 = 3,057$; $\lambda_2 = 1,753$; $\lambda_3 = 0,595$, que explican el 50,597%; 29,210% y el 9,910% de la varianza total, respectivamente. La Tabla 2, contiene la matriz de componentes que expresa las correlaciones entre las variables originales, d_i , y las componentes principales retenidas.

Siguiendo la metodología de Zhu, para obtener el valor T_Z , al tener la tercera componente igual número de valores positivos que negativos, y no existir ningún criterio a priori para escoger el signo, calculamos el valor de T_Z de dos formas:

$$T_{Z,02}^1 = 0,509 \cdot PC1 + 0,292 \cdot PC2 + 0,099 \cdot PC3$$

$$T_{Z,02}^2 = 0,509 \cdot PC1 + 0,292 \cdot PC2 - 0,099 \cdot PC3$$

Tabla 2

Matriz de componentes (Versión Zhu)

	1	2	3
d1	0,965	0,154	-0,150
d2	0,958	0,005	0,015
d3	0,911	0,242	0,305
d4	0,179	0,805	-0,479
d5	-0,536	0,591	-0,109
d6	-0,245	0,820	0,487

También cabe la posibilidad de escoger sólo dos componentes, utilizando el criterio de seleccionar sólo aquellas componentes que lleven asociado un valor propio mayor que 1:

$$T_{Z,02}^3 = 0,509 \cdot PC1 + 0,292 \cdot PC2$$

Con esta opción, incluso evitamos el problema de tener que decidir el signo positivo o negativo para la tercera componente.

Para poder comparar las distintas ordenaciones, se obtiene el test de correlación de Spearman (Tabla 3). Dado que las tres posibilidades están correlacionadas de manera muy similar con la ordenación realizada por el DEA, para escoger entre uno de ellos se debería comparar las eficiencias de cada método, no sólo las ordenaciones. Para ello, realizamos una transformación en las T_Z , restando a cada valor el mínimo de la columna⁴ (Tabla 1), de manera que tenemos valores siempre positivos y comparables con los del DEA.

⁴ Ver Nadimi y Jolai (2008).

Tabla 3

Correlación de Spearman y MSE entre eficiencia y $T_{Z,02}$

		$T_{Z,02}^1$	$T_{Z,02}^2$	$T_{Z,02}^3$
Coeficiente de correlación de Spearman	DEA	0,719	0,773	0,745
	SUPEREFICIENCIA	0,721	0,771	0,744
MSE	SUPEREFICIENCIA	0,367	0,356	0,351

Se puede utilizar el criterio MSE (Mean Square Error) para comparar además las eficiencias de los diferentes métodos (Nadimi y Jolai, 2008):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\text{eficiencia}_* - \text{eficiencia}_{DEA})^2$$

siendo eficiencia_* el valor obtenido para la eficiencia mediante el método * y eficiencia_{DEA} la obtenida mediante DEA. En función de los resultados es mejor escoger sólo aquellas componentes cuyo valor propio sea mayor que 1.

Para aplicar la variación de Premachandra, seleccionamos también tres componentes principales asociados a los autovalores $\lambda_1 = 3,383$; $\lambda_2 = 2,271$; $\lambda_3 = 0,741$, que explican el 48,325%; 32,442% y el 10,582% de la varianza total, respectivamente.

Tabla 4

Matriz de componentes (Versión Premachandra)

	1	2	3
d1	0,967	-0,174	-0,025
d2	0,882	-0,352	0,155
d3	0,896	-0,147	0,388
d4	0,502	0,785	-0,350
d5	-0,332	0,696	0,204
d6	-0,008	0,769	0,559
d7	0,711	0,636	-0,299

El valor de la eficiencia se obtendría como

$$T_{p,02}^1 = 0,483 \cdot PC1 + 0,324 \cdot PC2 + 0,105 \cdot PC3$$

También cabe la posibilidad de escoger sólo dos componentes, seleccionando sólo aquellas que lleven asociado un valor propio mayor que 1:

$$T_{p,02}^2 = 0,483 \cdot PC1 + 0,324 \cdot PC2$$

Realizados los cálculos correspondientes se obtienen los resultados que aparecen en la Tabla 1.

Para poder comparar las distintas ordenaciones, se realiza el test de correlación de Spearman y calculamos el MSE (Tabla 5). Sería preferible escoger también en este caso sólo las componentes con valores propios mayores o iguales a 1.

Tabla 5

Correlación de Spearman y MSE entre eficiencia y $T_{P,02}^1$

		$T_{P,02}^1$	$T_{P,02}^2$
Coeficiente de correlación de Spearman	DEA	0,860	0,856
	SUPEREFICIENCIA	0,843	0,839
MSE	SUPEREFICIENCIA	0,369	0,335

Aplicando el mismo procedimiento para los años 2005 y 2007 y seleccionando sólo aquellas componentes cuyo valor propio es mayor a 1, se obtienen los resultados que aparecen en la Tabla 6.

Tabla 6

Correlación de Spearman y MSE entre supereficiencia, T_Z y T_P : 2002-2007

		$T_{Z,05}^1$	$T_{P,05}^2$	$T_{Z,07}^1$	$T_{P,07}^2$
SUPEREFICIENCIA	Coeficiente de correlación	0,855	0,856	0,919	0,937
	MSE	0,432	0,810	0,459	0,753

En cuanto a la estabilidad temporal de las puntuaciones de eficiencia se muestra en la Tabla 7 que son más estables las ordenaciones de las entidades bancarias cuando se aplica el método de Premachandra y presenta menos estabilidad cuando se calculan medidas de supereficiencia. Siempre es mayor la estabilidad si comparamos el año 2005 y 2007.

Tabla 7

Evolución temporal del coeficiente de correlación de Spearman

	2002-2005	2005-2007	2002-2007
Supereficiencia	0,360 (*)	0,692 (**)	0,340 (*)
T_Z	0,554 (**)	0,874 (**)	0,537 (**)
T_P	0,707 (**)	0,889 (**)	0,664 (**)

* La correlación es significativa al nivel 0,05 (bilateral).

** La correlación es significativa al nivel 0,01 (bilateral).

5. Conclusiones

Para los datos empleados, podemos obtener las siguientes conclusiones:

1. La obtención de un ranking de DMUs mediante la aplicación de Análisis de Componentes Principales, método de Zhu o Premachandra, es válido y fiable comparado con el obtenido mediante el análisis de supereficiencia. Lo anterior es relevante, ya que se evita el problema de infactibilidad que puede presentarse en la aplicación del DEA.
2. En base a las correlaciones y valores estimados del MSE, sería recomendable el método de Zhu frente al propuesto por Premachandra.
3. La consideración de aquellas componentes principales asociadas a autovalores mayores o iguales a la unidad es suficiente para establecer una ordenación similar a la obtenida por el método de supereficiencia.
4. La estabilidad temporal de las puntuaciones de eficiencia es superior en las ordenaciones cuando se aplica el método de Premachandra y es menor cuando se calculan medidas de supereficiencia.

Referencias

- ADLER, N.; GOLANY, B. (2001), «Evaluation of deregulated airline networks using data envelopment analysis combined with principal analysis with application to Western Europe», *European Journal of Operational Research*, 132, 2, 18-21.
- ADLER, N; FRIEDMAN, L; SINUANY-STERN, Z (2002), «Review of ranking methods in the data envelopment analysis context», *European Journal of Operational Research*, Vol. 140, 249-265.
- ALIREZAEI, M R, AFSHARIAN, M (2007), «A complete ranking of DMUs using restrictions in DEA models», *Applied Mathematics and Computation* 189, 1550-1559.
- ANDERSEN, P; PETERSEN, N C (1993), «A Procedure for Ranking Efficient Units in Data Envelopment Analysis» *Management Science*, Vol. 39, 1993, 1261-1264.

- AZADEH, A.; GHADERI, S.F.; MAGHSOUDI, A. (2008), «Location optimization of solar plants by an integrated hierarchical DEA PCA approach», *Energy Policy*, 36, 3993-4004.
- BANKER, R.D.; CHARNES, A. Y COOPER, W.W. (1984): «Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis», *Management Science*, 30, 1078-1092.
- BRADLEY, S.; JOHNES, G.; MILLINGTON, J. (2001), «The effect of competition on the efficiency of secondary schools in England», *European Journal of Operational Research*, 135, 545-568.
- CHARNES, A; COOPER, W W Y RHODES, E L. (1978), «Measuring the efficiency of decision making units», *European Journal of Operational Research*, Vol. 2, 429-444.
- COOPER W.W.; SEIFORD L.M.; TONE, K. (2000), «Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software», Kluwer Academic Publishers, Boston.
- COOPER, W.L; SEIFORD. K; TONE, K. (2006), «Introduction to data envelopment analysis and its uses: with DEA-solver software and references», ed. Springer.
- EMROUZNEJAD, A, PARKER, B; G. TAVARES (2008), «Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature», *DEA Journal of Socio-Economics Planning Science*, 42(3), 151-157.
- HO, C-T.; WU, D. (2009), «Online banking performance evaluation using data envelopment analysis and principal component analysis», *Computers & Operational Research*, 36, 1835-1842.
- KARDÏYEN, F; H.H. ÖRKÇÜ (2006), «The comparison of principal component analysis and data envelopment analysis in ranking of decision making units», *G.U. Journal of Science*, 19 (2), 127-133.
- KORDROSTAMI, S; AMIRTEIMOORE, A; MASOUMZADEH, A. (2011), «Evaluating the efficiency of DMUs with PCA an application in real data set of Iranian banks», en *International Journal of Industrial Mathematics*, 3, 4, 251-258.
- LOVELL, C.A.K. (1993), «Production frontiers and productive efficiency», en: H.O. Fried, C.A.K. Lovell and S.S. Schmidt, Editores, *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford, 3-67.
- MANCEBÓN. M.J. (1996), «La evaluación de la eficiencia en los los centros educativos públicos», Tesis Doctoral, Universidad de Zaragoza.
- MANCEBON, M.J.; MAR MOLINERO, C. (2000), «Performance in primary school», *Journal of the Operational Research Society*, 51, 843-854.
- MEHRABIAN, S.; ALIREZAEI, M.R.M JAHANSHALOO, G.R. (1999), «A Complete efficiency ranking of decision making units in data envelopment analysis», *Computational Optimization and applications*, 14, 261-266.

- NADIMI, R.; JOLAI, F. (2008), «Joint use of factor analysis (FA) and data envelopment analysis (DEA) for ranking of data envelopment analysis», *International Journal of Mathematical, Physical and Engineering Sciences*, Vol.2, Nº 4, 218-222.
- NATH, P. (2001), «Efficiency analysis of Indian Banking Industry- An exploratory study», *APORS 2000*, The Fifth Conference of the Association of Asian-Pacific Operational Research Societies within IFORS, 5-7 July 2000, Singapore.
- PREMACHANDRA, I.M. (2001), «A note on DEA vs principal component analysis: An improvement to Joe Zhu's approach», *European Journal of Operational Research*, 132, 553-560.
- PREMACHANDRA, I.M.; BHABRA, G.S.; SUEYOSHI, T. (2009), «DEA as tool for banruptcy assesment: A comparative study whit logistic regression technique», *European Journal of Operational Research*, 193, 412-424.
- SEGOVIA GONZÁLEZ, M. M.; CONTRERAS RUBIO, I.; MAR MOLINERO, C. (2009), «Evaluación de la eficiencia de una cartera de asegurados en el sector del automóvil», *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 7, 57-76.
- SEIDORF, L.M.; ZHU, J. (1999), «Infeasibility of super efficiency data envelopment analysis models», *Information Systems and Operational Research*, 27, 174-187.
- SERRANO CINCA, C.; MAR MOLINERO, C. (2001), «Selecting DEA especifications and ranking units via PCA», *Discussion Papers in Management*, University of Southampton, Nº M01-3.
- SUEYOSHI, T. (1999), «Data envelopment analysis non-parametric ranking test and index measurement: slack-adjusted DEA and an application to Japanese agriculture cooperatives», *Omega, International Journal Management Science*, 27, 315-326.
- THRALL, R.M. (1996), «Duality, classification and slacks in DEA», *Annals Operational Research*, 66, pp 109-138.
- UEDA, T.; HOSHIAI, Y. (1997), «Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs», *Journal of the Operational Research*, Vol.40, Nº.4, 466-478.
- VARGAS, S. C.; BRICKER, D. (2000), «Combining DEA and Factor Analysis to improve the evaluation of academic departments given uncertainty about the output constructs», *Research paper*, Department of Engineering, University of Iowa, Iowa City, USA.
- VÁZQUEZ CUETO, M^a.J. (2008), «Medición de la eficiencia de las universidades públicas españolas. Combinación de la metodología DEA con PCA», XV Jornadas de Asepuma y III Encuentro Internacional.
- WILSON, P.W. (1995), «Protecting influential observations in Data Envelopment Analysis», *Journal of Productivity Analysis*, 4, 27-45.

- ZHU, J. (1996), «Robustness of the efficient DMUs in data envelopment analysis», *European Journal of Operational Research*, Vol. 90, 451-460.
- ZHU, J. (1998), «Data envelopment analysis vs. principal component analysis: an illustrative study of economic performance of Chinese cities», *European Journal of Operational Research*, 111, 1, 56-61.

