

Multi-stage variable selection method for efficiency evaluation with DEA models and panel data

Patricia A. Iñiguez
Facultad de Ciencias Económicas
Universidad Nacional de Río Cuarto
piniguez@fce.unrc.edu.ar

Juan M. Gallardo
Facultad de Ciencias Económicas
Universidad Nacional de Río Cuarto
jgallardo@fce.unrc.edu.ar

Mariana Arburua
Facultad de Ciencias Económicas
Universidad Nacional de Río Cuarto
marburua@fce.unrc.edu.ar

Pablo Pagano
Facultad de Ciencias Económicas
Universidad Nacional de Río Cuarto
ppagano@fce.unrc.edu.ar

Abstract

A variable selection method based on a multi-attribute indicator for efficiency evaluation of units observed in more than one period is presented. The proposed search algorithm allows choosing a common set of variables for the T periods; controlling the number of selected variables, without the evaluator in each iteration analyzing the added value of the variables to be added; operating directly with the original variables, without reducing dimensionality by composing artificial variables and; deriving a collection of empirical models. The application is illustrated with the case of the efficiency evaluation of provincial public management hospitals in the province of Córdoba, Argentina.

Keywords: Data Envelopment Analysis, Variables Selection, Panel Data.

MSC Subject classifications: 90-08, 90C08, 90B50.

1. Introducción

La metodología de estimación de fronteras no paramétricas permite evaluar la eficiencia relativa de las unidades de producción, aún en aquellos casos en que se utilicen múltiples inputs y outputs. De los métodos desarrollados, el que registra mayor uso y avances, es el Análisis Envolvente de Datos (DEA¹). En efecto, Cooper, Seiford y Zhu, 2011 muestran que DEA es uno de los métodos más aplicados para analizar la eficiencia y la productividad de varios sistemas: empresas, bancos, hospitales, industrias, regiones, países, etc. Las cuatro décadas posteriores a Charnes, Cooper y Rhodes, 1978 han sido muy provechosas para el método DEA tanto por los muchos desarrollos teóricos, como por los estudios empíricos que se registran en innumerables artículos. No obstante, esta interesante evolución de los últimos años, una cuestión de estudio aún vigente en la literatura es el problema de la selección de los inputs y outputs a considerar en el modelo empírico.

Este interés persistente obedece a que al ser DEA un enfoque no paramétrico, surge el problema de la discriminación cuando, por ejemplo, existe un número relativamente grande de variables en comparación con las Decision Making Units (DMUs) (Adler y Yazhemsky, 2010). Tener un número bajo de unidades en evaluación en comparación con el número de inputs y outputs conduce a encontrar una gran proporción de unidades eficientes o con índices de eficiencia con valores optimistas. Esta clase de problemas, puede conducir a que la evaluación de eficiencia no cumpla con su propósito.

Lo cierto es que, en la práctica, generalmente, no se dispone de muestras grandes y los investigadores procuran analizar conjuntos de observaciones relativamente pequeños, de allí la necesidad de hallar metodologías para la selección de un conjunto de variables que mejoren la discriminación del modelo empírico.

Una variedad de métodos se ha propuesto, en particular para datos de corte transversal y, si bien, se han hecho diversas comparaciones a través de pruebas empíricas y/o simulación Monte Carlo (Adler y Yazhemsky, 2010, Nataraja y Johnson, 2011 y Eskelinen, 2017), todos poseen ventajas y desventajas, por lo que a ninguno se le puede atribuir la característica de inmejorable. Adicionalmente, la cuestión de elegir un conjunto de inputs y outputs, se dificulta cuando la evaluación de eficiencia comprende más de un período de tiempo. Aunque resulta factible seleccionar para cada uno de ellos un conjunto distinto de variables, ello no resulta razonable si se pretende comparar los resultados de los diferentes períodos o medir la evolución y cambio, por ejemplo, a través del índice de Malmquist. Esta circunstancia obliga a aplicar algún método que ayude a seleccionar un conjunto de inputs y outputs común para todos los períodos en los que se pretenda evaluar la eficiencia.

De la diversidad de métodos que se hayan en la literatura, relativamente pocos -González-Araya, Campo Hernández y Acosta Espejo, 2013 y Peyrache, Rose y Sicilia, 2020-, abordan el problema de la selección de variables con DMUs observadas en más de un período.

El primer método propuesto por González-Araya, Campo Hernández y Acosta Espejo, 2013 parte del desarrollado por Soares De Mello et al., 2004 para datos de corte transversal para seleccionar, en función de la frecuencia relativa de elección de los inputs u outputs en cada período, un conjunto de variables común para todos los períodos. Si bien, en cada año, se consideran dos criterios -el mejor ajuste a la frontera y la máxima discriminación-, en la conformación del conjunto común solo se tiene en cuenta la frecuencia relativa y, se retoma la máxima eficiencia media, en caso de empate.

En el segundo método presentado por los mencionados autores se aplica el formulado por Soares De Mello et al., 2004 una sola vez a la totalidad de los datos del panel.

¹ Del inglés Data Envelopment Analysis

La propuesta de Peyrache, Rose y Sicilia, 2020 define dos versiones del programa lineal entero mixto que permiten seleccionar un conjunto común de inputs y outputs. En la primera, se maximiza la eficiencia agregada de todas las observaciones en todos los períodos y, en la segunda, la eficiencia de la DMU promedio.

Con el propósito de facilitar la elección de un conjunto común de inputs y outputs para todos los períodos que integran el panel de datos que considere tanto el mejor ajuste a la frontera, como la máxima discriminación, la menor variación interanual y diversos escenarios, se propone un método de varias etapas integrando diversos métodos basados en técnicas de decisión multicriterio para datos de corte transversal - Soares De Mello et al., 2004, Castro Senra et al., 2007 -, y los propuestos por González-Araya, Campo Hernández y Acosta Espejo, 2013 . Se define un indicador multiatributo que, a pesos iguales, agrega la eficiencia media de los T períodos, la media de los desvíos de la eficiencia de las N DMUs en los T períodos, la media de DMUs eficientes considerando los T períodos y el desvío de DMUs eficientes de los T períodos.

Con esta propuesta se busca contribuir con un procedimiento de selección de variables que, considerando otros atributos, además, de la eficiencia media o agregada, evalúe una colección de modelos empíricos o escenarios que brinde al decisor o investigador la posibilidad de elegir aquel modelo empírico con mayor sentido conceptual y práctico.

Así, en este artículo se presenta el Método de Selección de Variables para Datos de Panel² con Base en un Indicador Multiatributo y se ilustra la aplicación con el caso de la evaluación de eficiencia de los hospitales de gestión pública provincial de la provincia de Córdoba de Argentina.

A continuación, en la sección 2 se describen brevemente los modelos básicos del DEA, luego, en la sección 3, sin aspirar a ser exhaustiva, se resume una revisión de la literatura. En la sección 4, se reseñan brevemente las propuestas de Soares De Mello et al., 2004, Castro Senra et al., 2007 y de González-Araya, Campo Hernández y Acosta Espejo, 2013, para luego exponer, en la sección 5, el Método de Selección de Variables para Datos de Panel con Base en un Indicador Multiatributo, sintetizando la aplicación empírica en la sección 6, para presentar las principales conclusiones, en la sección 7.

2. El Análisis Envolvente de Datos

Como se comentó al inicio, uno de los principales enfoques para analizar la eficiencia se conoce como el Análisis Envolvente de Datos (DEA). Este enfoque fue desarrollado por dos prestigiosos investigadores de la Investigación Operativa, Abraham Charnes y William Cooper, quienes junto a Eduardo Rhodes, publican el artículo fundamental, Charnes, Cooper y Rhodes, 1978. Este trabajo se inspira en el importante escrito, aunque relativamente desconocido hasta ese momento, de un destacado economista, Farrell, 1957. En efecto, dicho autor, propuso considerar como referente eficiente la mejor práctica observada de entre la muestra de unidades productoras objeto de estudio y, de este modo, calcular los índices de eficiencia de cada una por comparación con la/s que presenta/n un mejor comportamiento. Así, se obtiene una medida de eficiencia que tiene un carácter relativo, es decir, depende de la muestra.

El objetivo del DEA es comparar el desempeño de las DMUs, que utilizan los mismos inputs y producen los mismos outputs, entre sí. El DEA calcula un índice de eficiencia para cada una de

² Se habla de datos de panel cuando se tienen observaciones repetidas a lo largo del tiempo para una muestra de unidades individuales. Puede decirse que para una variable x_{it} se tienen $i= 1, \dots, N$ individuos observados a lo largo de $t= 1, \dots, T$ períodos de tiempo. (Arellano, 1992)

las DMUs, de modo que las DMUs con un indicador igual a uno se encuentran en la denominada frontera eficiente. Aquellas que se encuentran fuera de la frontera de eficiencia se consideran de menor desempeño y, un análisis adicional, facilita la información respecto de lo que pueden hacer para mejorar su eficiencia.

El método DEA generaliza la medida de la eficiencia para un único output e input de una DMU, medida por el cociente output/input formulando un programa fraccional que transforma los múltiples outputs e inputs para cada DMU en un único output e input virtual.

La eficiencia técnica relativa de una DMU se calcula mediante el cociente entre la suma ponderada de outputs y la suma ponderada de inputs, donde los pesos para los outputs e inputs de cada DMU son seleccionados de modo que la medida de la eficiencia responda al criterio de eficiencia de Pareto, sujeto a la restricción de que ninguna DMU puede tener una eficiencia relativa superior a la unidad.

DEA obtiene la frontera de producción a partir de las mejores prácticas observadas, es decir, en caso de una medida producto orientada, el máximo output alcanzable por cualquier DMU, dado un nivel de input.

La propuesta de Charnes, Cooper y Rhodes, 1978 dio lugar al Modelo Básico conocido en la literatura como Modelo CCR³ o modelo de retornos constantes a escala y se diseñó para estimar índices de eficiencia técnica. Luego, Banker, Charnes y W. Cooper, 1984 plantean un modelo lineal conocido como Modelo BCC⁴ o modelo de retornos variables a escala. A continuación, se presentan los modelos lineales de la envolvente (CCR: (1) y BCC: (2)) que permiten estimar eficiencia técnica producto orientada.

$$\begin{aligned}
 & \max \varphi \\
 & \text{sujeto a :} \\
 & \varphi y_o - Y_\mu \leq 0 \\
 & X_\mu \leq x_o \\
 & \mu \geq 0
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 & \max \varphi_B \\
 & \text{sujeto a :} \\
 & \varphi_B y_o - Y_\mu \leq 0 \\
 & X_\mu \leq x_o \\
 & \varepsilon \mu \leq 1 \\
 & \mu \geq 0
 \end{aligned} \tag{2}$$

Donde $x = (x_1, x_2, \dots, x_M) \in \mathbb{R}_+^M$ es el vector de M inputs; $y = (y_1, y_2, \dots, y_S) \in \mathbb{R}_+^S$ es el vector de S outputs; X e Y las matrices de los M inputs y S outputs de las N DMUs observadas; μ es el vector de escalares; ε es un vector de N componentes iguales a 1, y φ y φ_B los índices de eficiencia técnica.

³ En referencia al apellido de sus autores: Charnes, Cooper y Rhodes.

⁴ En referencia al apellido de sus autores: Banker, Charnes y Cooper.

3. Antecedentes

El interés por mejorar la discriminación de los modelos empíricos de evaluación de eficiencia, como se comentó al inicio, sigue siendo una cuestión vigente en la literatura. En este marco, diversos enfoques se han divulgado en los últimos treinta años, los que pueden reunirse en dos grandes grupos: a) los que seleccionan solo un subconjunto de las variables originales; b) los que utilizan una medida agregada.

Dentro del grupo a) se hallan aquellos que se basan en algún tipo de decisión heurística o juicio de valor acerca de qué variables son relevantes para el problema dado (Allen et al., 1997; Golany y Roll, 1989); pero también, se reúnen aquellos que utilizan técnicas estadísticas con la idea de comparar la contribución que, las potenciales variables a ser incluidas en el modelo, tienen sobre la eficiencia estimada (contribución a la eficiencia) (Banker, 1996, Simar y Wilson, 2001, Pastor, Ruiz y Sirvent, 2002, Jenkins y Anderson, 2003, Ruggiero, 2005, Wagner y Shimshak, 2007). Asimismo, se incluyen en este grupo, los métodos de selección desarrollados a partir de la propuesta de Lins y Moreira, 1999, que combinan la relación causal, con el potencial discriminatorio del modelo, dando lugar a procedimientos de selección que consideran múltiples criterios. Así, las variables candidatas son alternativas y como criterios se consideran el mejor ajuste a la frontera y la máxima discriminación (Soares De Mello et al., 2004, Castro Senra et al., 2007; González-Araya y Valdés, 2009 y González-Araya, Campo Hernández y Acosta Espejo, 2013).

Se suma a los anteriores, un cuarto subconjunto de métodos, que utilizan un enfoque DEA a través de programas lineales o cuadráticos enteros mixtos, desarrollados en los últimos tres años. En general, en estas propuestas, las variables binarias se introducen para seleccionar inputs y outputs mientras que las funciones objetivo y restricciones se formulan de acuerdo a un modelo DEA (Limleamthong y Guillén-Gosálbez, 2018, Benítez-Peña, Bogetoft y Morales, 2019 y Peyrache, Rose y Sicilia, 2020).

Por otra parte, los métodos que pueden reunirse en el grupo b), proponen reducir la dimensionalidad sustituyendo las variables originales por una medida agregada. La agregación puede hacerse basándose en algún tipo de decisión heurística o juicios de valor (usando ponderaciones ad-hoc) o usando algunos criterios estadísticos para elegir la combinación óptima (lineal) de los inputs u outputs (Friedman y Sinuany-Stern, 1998 y Morita y Avkiran, 2009). En este contexto, el enfoque más popular es la formulación DEA-ACP propuesta por Ueda y Hoshiai, 1997 y Adler y Golany, 2001 que se basa en el uso del análisis de componentes principales para reducir el número de inputs y outputs sustituyéndolos por componentes principales.

De los procedimientos de selección de variables mencionados anteriormente solo dos proporcionan formulaciones para datos de panel. González-Araya, Campo Hernández y Acosta Espejo, 2013 presentan tres métodos, dos de ellos basados en la propuesta de Soares De Mello et al., 2004 y, el tercero, parte del método presentado por González-Araya y Valdés, 2009. Finalmente, Peyrache, Rose y Sicilia, 2020 incluyen para seleccionar un conjunto común de variables para todas las observaciones en todos los períodos de tiempo que conforman el panel de datos, dos versiones del programa lineal entero mixto propuesto para datos de corte transversal.

4. Breve Reseña de los Métodos de Selección de Variables Combinados en la Propuesta

El Método Multicriterio para la Selección de Variables (MMSV) de Soares De Mello et al., 2004 y los propuestos por Castro Senra et al., 2007 nominados Método Multicriterio Combinatorio Inicial para

la Selección de Variables (MMCISV) y Método Multicriterio Combinatorio por Escenarios para la Selección de Variables (MMCESV), así como los presentados por González-Araya, Campo Hernández y Acosta Espejo, 2013, Método 1 y 2 de selección de variables en modelos DEA usando series de tiempo (M1SVST y M2SVST, respectivamente), se apoyan en un procedimiento de selección que considera múltiples criterios. Así, las variables candidatas son alternativas y como criterios se consideran el mejor ajuste a la frontera y la máxima discriminación.

En todos los casos se requiere un par inicial input-output. Este puede ser elección del decisor o , como proponen Castro Senra et al., 2007, escogido siguiendo un proceso similar al realizado en el MMSV considerando como conjunto de alternativas todos los posibles pares input-output.

Para seleccionar una de las variables candidatas se evalúa cada alternativa en los criterios antes indicados. El ajuste a la frontera es medido a través de la eficiencia media para cada variable a adicionar y su normalización da lugar al indicador S_{EF} , el que alcanza el valor 1 en la eficiencia media máxima y 0 en la eficiencia media mínima. La máxima discriminación es medida por el número de DMUs eficientes, el que se pretende minimizar. La normalización de este criterio genera el indicador S_{DIS} que alcanza el valor 1 para el menor número de DMUs en la frontera y 0 para el mayor. Para seleccionar la variable se agregan, a idéntico peso, los dos criterios mediante una suma ponderada de S_{EF} y S_{DIS} , $S = \alpha S_{EF} + (1 - \alpha) S_{DIS}$.

MMSV, MMCISV, M1SVST y M2SVST incorporan variables mientras el conjunto de inputs y outputs no exceda el tercio de las DMUs. En tanto que MMCESV no se detiene cuando el número de variables alcance el tercio de las DMUs. Se incluyen todas las variables para luego comparar distintos escenarios con diferentes números de variables.

5. Método de Selección de Variables para Datos de Panel con Base en un Indicador Multiatributo (MSVDPIM)

El procedimiento de selección de variables para evaluación de eficiencia mediante un modelo de la metodología DEA con datos de panel, que se plantea, busca elegir un conjunto de inputs y outputs común para todos los períodos que integran el panel, evaluando diversos escenarios a través de un indicador multiatributo que valora y pondera los criterios, mejor ajuste a la frontera, máxima discriminación y menor variación interanual.

El método propuesto se conforma de tres etapas. En la primera se define el par inicial. Éste puede, ser elegido por el investigador con cualquier criterio, tal como se propone en el MMSV; ser integrado por un par considerado relevante para la evaluación (Peyrache, Rose y Sicilia, 2020) o definirse siguiendo el criterio del MMCISV. En la segunda, se definen los diferentes escenarios a comparar. El primero de ellos, tomando la idea del M2SVST, surge de aplicar el MMCESV una sola vez a la totalidad de los datos del panel; los siguientes, se obtienen de aplicar el mismo método en cada período que integra el panel de datos y; el último, se conforma a partir de las frecuencias relativas, siguiendo el M1SVST. En la tercera etapa se estima la eficiencia con cada uno de los escenarios seleccionados para cada período del panel y se define el mejor escenario, mediante el indicador S_d que pondera cuatro atributos. El procedimiento se detalla en la Figura 1.

El indicador S_d , pondera a igual peso o a los que el investigador considere adecuados, cuatro atributos normalizados⁵ correspondientes a cada uno de los escenarios: la eficiencia media de los T períodos (

⁵ El procedimiento de normalización es el aplicado en el MMSV

S_{EFT}), la media de los desvíos de la eficiencia de las N DMUs en los T períodos (S_{VI1}), la media de DMUs eficientes considerando los T períodos (S_{DIST}) y el desvío de DMUs eficientes de los T períodos (S_{VI2}). El único atributo que se pretende maximizar es la eficiencia media, los restantes, se minimizan, para así seleccionar el escenario que mejor discrimine y ajuste a la frontera, al tiempo que menor variación interanual registre.



Fig. 1: Algoritmo método MSVDPIM

El algoritmo de selección de variables por agregación propuesto, es un método de búsqueda, que en cada iteración elige la mejor alternativa, por tanto, con la excepción de la determinación del escenario $T+2$, no es necesario que el investigador defina cuál es el número de variables en que se detiene la búsqueda. Se incluyen todas las variables para luego comparar distintos escenarios con diferentes números de variables.

6. Aplicación Empírica

Para ilustrar la aplicación del MSVDPIM se considera el caso de la evaluación de eficiencia de los hospitales de gestión pública provincial de la provincia de Córdoba en los años 2015, 2016 y 2017.

A partir de una revisión de los indicadores de insumo, proceso, producto e impacto que la literatura latinoamericana recomienda para el desempeño de los centros hospitalarios y los datos primarios que se obtienen del Registro Federal de Establecimientos Hospitalarios y de la unidad ejecutora central del Programa SUMAR⁶, ambas, dependencias del Ministerio de Salud de la Nación; el presupuesto del Ministerio de Salud de la provincia de Córdoba y el CENSO 2010, se logra estimar para veintidós (22) establecimientos de salud con internación general (hospitales) de gestión pública provincial, diecisiete (17) indicadores agrupados en: los vinculados al insumo cama, los relacionados al insumo personal, los referidos a las diversas prestaciones y los correspondientes al grado de atención de la población potencial beneficiaria de los servicios, según se describe en el la Figura 2.

ID	Descripción	Tipo	2015		2016		2017	
			Media	Desvío	Media	Desvío	Media	Desvío
I.1	Cantidad de camas disponibles por 1000 habitantes	Input	0.9645	0.9479	0.9645	0.9479	0.9645	0.9479
I.3	Porcentaje de camas generales	Input	0.8501	0.1833	0.8501	0.1833	0.8501	0.1833
I.6	Proporción de personal administrativo	Input	0.0899	0.0284	0.0918	0.0325	0.0906	0.0336
I.7	Proporción de personal conducción	Input	0.2245	0.0782	0.2331	0.0791	0.2385	0.0790
I.8	Proporción de personal profesional	Input	0.6158	0.0614	0.6087	0.0600	0.6069	0.0573
O.1	Razón de exámenes ecográficos por consulta	Output	0.2155	0.2684	0.3944	0.3885	0.3405	0.2783
O.2	Razón de Rx por consulta	Output	0.3863	0.3970	0.7894	0.8024	0.7341	0.5703
O.3	Razón de prácticas de laboratorio por consulta	Output	1.6737	2.2475	2.0889	2.0273	1.7710	1.7122
O.4	Proporción consultas clínica médica	Output	0.2888	0.1554	0.2233	0.1772	0.1767	0.1228
O.5	Proporción consultas ginecología y obstetricia	Output	0.2486	0.1145	0.2596	0.1308	0.2069	0.1099
O.8	Proporción consultas pediatría	Output	0.3700	0.1858	0.3709	0.2126	0.3824	0.1894
O.9	Proporción de actividades de prevención y promoción	Output	0.0464	0.0460	0.0369	0.0247	0.0333	0.0232
O.10	Índice de cobertura	Output	0.0652	0.0537	0.0756	0.0556	0.1034	0.0665
O.11	Intensidad de uso	Output	4.7878	1.5914	4.0223	1.3404	4.1730	1.5509
O.12	Promedio de Rx por paciente año	Output	1.4635	0.2444	1.3205	0.2336	1.3475	0.2487
O.13	Porcentaje de partos por cesárea	Output	0.4186	0.3029	0.4463	0.2874	0.4601	0.3157
O.14	Porcentaje de partos vaginales	Output	0.3087	0.2457	0.3719	0.2616	0.4036	0.3027

Fig. 2: Indicadores/VARIABLES

De este modo, se consigue conformar un panel de datos balanceado con 22 observaciones y 17 variables estimadas en los años 2015, 2016 y 2017. Sobre este conjunto de inputs-outputs se aplica el procedimiento descrito en la Figura 1, considerando diferentes pares input-output iniciales. En el primer caso el par inicial lo seleccionan los investigadores y, en el segundo, se escoge para todo el

⁶ El Programa SUMAR brinda cobertura a la población materno-infantil, niños, adolescentes y adultos hasta los 64 años sin obra social. Los datos se suministran en el marco de un Convenio de Colaboración y Confidencialidad por el que se acuerda que "Las ideas y opiniones expuestas en esta publicación son de responsabilidad exclusiva de su/s autor/es y no necesariamente reflejan la opinión del Programa SUMAR, ni del Ministerio de Salud de la Nación"

panel y en cada período t el que posea mayor valor S entre los 84 pares posibles.

Para construir y evaluar los escenarios 1, 2, ..., $T+1$ y $T+2$, el nivel de eficiencia se determina mediante el modelo (2). Se escoge esta formulación porque la mayoría de los indicadores constituyen ratios y, además, es un modelo invariante ante transformaciones afines de las variables. De este modo, si se utiliza el mismo conjunto de datos en diferentes unidades puede obtenerse la misma selección de variables (Peyrache, Rose y Sicilia, 2020).

Para el primer caso, los investigadores acuerdan seleccionar el par inicial compuesto por I.1 y O.10, no solo por ser el par con mayor correlación lineal input-output en los tres períodos, sino porque, además, se considera que una de las principales funciones del hospital de gestión pública, en Argentina, es brindar atención de salud a la población que no posee cobertura. A partir de dicho par inicial se conforman los cinco escenarios que se presentan en la Figura 3.

Como podrá observarse los Escenarios 1, 4 y 5 son idénticos. Por tal motivo quedan a comparar solo tres conjuntos de variables para definir el modelo empírico. De esos tres, dados dos diferentes conjuntos de pesos, el mejor es el Escenario 3, integrado solo por tres variables I.1, O.10 y O.12 y le sigue, aunque con un mayor número promedio de DMUs eficientes, el Escenario 1, el 4 o el 5.

En el segundo caso, y con el fin de ilustrar los resultados a los que se arriban en caso que el investigador no desee, no pueda o le resulte indiferente elegir el par input-output inicial, el mismo, se escoge tal como se realiza en el MMCISV.

Así, el par con el que se inicia el proceso de búsqueda de cada escenario puede ser diferente. El detalle de los pares iniciales y de los escenarios hallados se presentan en la Figura 4.

Atributos \ Variables	Escenario 1	Escenario 2	Escenario 3	Escenario 4	Escenario 5
	Considerando Todo el Panel Datos	Considerando Datos Año 2015	Considerando Datos Año 2016	Considerando Datos Año 2017	Mayor Frecuencia Relativa
	I.1-O.5-O.10-O.12	I.1-O.5-O.10	I.1-O.10-O.12	I.1-O.5-O.10-O.12	I.1-O.5-O.10-O.12
Eficiencia Media	1,1514	1,4682	1,2737	1,1514	1,1514
Media de los Desvíos de la Eficiencia	0,0938	0,2541	0,0894	0,0938	0,0938
Media DMUs Eficientes	7,3333	5,0000	5,6667	7,3333	7,3333
Desvío DMUs Eficientes	1,5275	0,0000	2,0817	1,5275	1,5275
Atributos Normalizados	I.1-O.5-O.10-O.12	I.1-O.5-O.10	I.1-O.10-O.12	I.1-O.5-O.10-O.12	I.1-O.5-O.10-O.12
SEFT	1,0000	0,0000	0,6139	1,0000	1,0000
SVI1	0,9733	0,0000	1,0000	0,9733	0,9733
SDIST	0,0000	1,0000	0,7143	0,0000	0,0000
SVI2	0,2662	1,0000	0,0000	0,2662	0,2662
Sd ($w_i=0,25$)	0,5599	0,5000	0,5821	0,5599	0,5599
Sd ($w_1=w_3=1/3, w_2=w_4=1/6$)	0,5399	0,5000	0,6094	0,5399	0,5399
Mínimos					
Máximos					

Fig. 3: Escenarios con el par inicial I.1-O.10

	Escenario 1	Escenario 2	Escenario 3	Escenario 4	Escenario 5
	Considerando Todo el Panel Datos	Considerando Datos Año 2015	Considerando Datos Año 2016	Considerando Datos Año 2017	Mayor Frecuencia Relativa
Par Inicial	1.8-0.12	1.7-0.11	1.7-0.12	1.6-0.9	N/C
Atributos / Variables	1.8-0.4-0.5-0.10-0.12-0.13	1.7-0.11-0.12-0.13	1.7-1.8-0.5-0.10-0.12	1.6-0.4-0.9-0.11	1.6-1.7-1.8-0.5-0.11-0.12-0.13
Eficiencia Media	1,0851	1,1698	1,1100	1,4237	1,0304
Media de los Desvíos de la Eficiencia	0,0535	0,0773	0,0870	0,2734	0,0328
Media DMUs Eficientes	9,3333	6,6667	8,0000	5,3333	17,0000
Desvío DMUs Eficientes	1,5275	1,5275	1,0000	2,0817	1,0000
Atributos Normalizados	1.8-0.4-0.5-0.10-0.12-0.13	1.7-0.11-0.12-0.13	1.7-1.8-0.5-0.10-0.12	1.6-0.4-0.9-0.11	1.6-1.7-1.8-0.5-0.11-0.12-0.13
S_{EFT}	0,8611	0,6455	0,7976	0,0000	1,0000
S_{VII}	0,9139	0,8149	0,7747	0,0000	1,0000
S_{DIST}	0,6571	0,8857	0,7714	1,0000	0,0000
S_{VII}	0,5123	0,5123	1,0000	0,0000	1,0000
S_d ($w_1=0,25$)	0,7361	0,7146	0,8359	0,2500	0,7500
S_d ($w_1=w_3=1/3, w_2=w_4=1/6$)	0,7438	0,7316	0,8188	0,3333	0,6667
Mínimos					
Máximos					

Fig. 4: Escenarios con par inicial de mayor S

Como puede observarse al seleccionar el par inicial a partir de S , el insumo cama pierde participación y el índice de cobertura, participa solo en algunos escenarios. Evidentemente el indicador O.12, es el que más favorece la productividad de estos establecimientos y cualquiera sea el punto de partida, integra el escenario con mayor S_d . En esta oportunidad el Escenario 3.

7. Conclusiones

En este trabajo se proporciona una solución, basada en un indicador multiatributo, para un problema vigente en la literatura del DEA: la selección de las variables que se incluirán en el modelo cuando la evaluación se realiza con datos de panel.

Si bien, no se exponen los cuadros con los valores de cada una de las etapas que permiten seleccionar los T+2 escenarios, la aplicación empírica muestra que es factible utilizar este algoritmo cuando se tiene un número bajo de unidades en evaluación en comparación con el número de inputs y outputs. No obstante, si el investigador lo considera oportuno, también, resulta de aplicación cuando la muestra sea lo suficientemente grande.

El algoritmo planteado permite, elegir un conjunto común de variables para los T períodos; controlar el número de variables seleccionadas, sin que el evaluador en cada iteración analice el valor agregado de las variables a adicionar; operar directamente con las variables originales, sin reducir la dimensionalidad componiendo variables artificiales que pueden ser de difícil interpretación y; derivar una colección de modelos empíricos, dado que se examinan T+2 escenarios.

También, es de notar que todos los conjuntos de variables se ajustan a las propuestas que se hallan en la literatura en relación al número de variables respecto del número de observaciones.

Finalmente, se debe señalar que a futuro queda por concluir el desarrollo de una rutina en el entorno libre **R**, que conceda acceso libre a todos los investigadores y la comparación de los resultados que arroja esta propuesta con los que surgirían de aplicar las otras, para datos de panel, hallados en la literatura.

Acerca de los Autores



Patricia Alejandra Iñiguez es Profesora Titular de Investigación Operativa en la Facultad de Ciencias Económicas de la UNRC. Integra Comités Científico de diversos congresos nacionales e internacionales y es evaluador de artículos para revistas y libros de nivel nacional e internacional. Dirige proyectos de investigación vinculados al estudio de la eficiencia. Posee publicaciones en congresos nacionales e internacionales y en revistas.



Juan Manuel Gallardo es doctorando del Doctorado en Desarrollo Territorial y Contador Público (Facultad de Ciencias Económicas – UNRC); Jefe de Trabajos Prácticos en la Facultad de Ciencias Económicas – UNRC. Es integrante de proyectos de investigación vinculados al estudio de la eficiencia. Posee diversas publicaciones nacionales e internacionales que abordan el estudio de la eficiencia del sector público.



Mariana Arburua es Maestría en Procesos Educativos Mediados por Tecnologías. Es Profesora Adjunta de Investigación Operativa en la Facultad de Ciencias Económicas de la UNRC. Actualmente codirige proyectos de investigación vinculados al estudio de la eficiencia. Posee diversas publicaciones nacionales e internacionales que abordan el estudio de la eficiencia del sector público.



Pablo Pagano es Diplomado Superior en Investigación Contable. Esta cursando la Especialidad en Métodos Cuantitativos para la gestión y Análisis de Datos en Organizaciones en la FCE de la UBA. Es ayudante de Primera de Investigación Operativa en la Facultad de Ciencias Económicas de la Universidad Nacional de Río Cuarto. Es integrante de proyectos de investigación vinculados al estudio de la eficiencia.

Referencias

- Adler, N. y B. Golany (2001). «Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe». En: *European Journal of Operational Research* 132.2, págs. 260-273.
- Adler, N. y E. Yazhemsky (2010). «Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction». En: *European Journal of Operational Research* 202.1, págs. 273-284.
- Allen, R., A. Athanassopoulos, R. Dyson y E. Thanassoulis (1997). «Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions». En: *Annals of operations research* 73, págs. 13-34.
- Arellano, Manuel (1992). *Introducción al análisis econométrico con datos de panel*. Banco de España. Servicio de Estudios.
- Banker, R. (1996). «Hypothesis tests using data envelopment analysis». En: *Journal of productivity analysis* 7.2, págs. 139-159.
- Banker, R., A. Charnes y W. Cooper (1984). «Some models for estimating technical and scale inefficiencies in data envelopment analysis». En: *Management science* 30.9, págs. 1078-1092.
- Benítez-Peña, S., P. Bogetoft y D. Morales (2019). «Feature selection in data envelopment analysis: A mathematical optimization approach». En: *Omega* 96, pág. 102068.
- Castro Senra, L., L. Nanci, J. Soares De Mello y L. Angulo Meza (2007). «Estudo sobre métodos de seleção de variáveis em dea». En: *Pesquisa Operacional* 27.2, págs. 191-207.
- Charnes, A., W Cooper y E. Rhodes (1978). «Measuring the efficiency of decision-making units». En: *European journal of operational research* 2.6, págs. 429-444.
- Cooper, W, L. Seiford y J. Zhu (2011). *Handbook on data envelopment analysis*. Springer Science & Business Media.
- Eskelinen, J. (2017). «Comparison of variable selection techniques for data envelopment analysis in a retail bank». En: *European Journal of Operational Research* 259.2, págs. 778-788.
- Farrell, M. (1957). «The measurement of productive efficiency». En: *Journal of the Royal Statistical Society: Series A (General)* 120.3, págs. 253-281.
- Friedman, L. y Z. Sinuany-Stern (1998). «Combining ranking scales and selecting variables in the DEA context: The case of industrial branches». En: *Computers & Operations Research* 25.9, págs. 781-791.
- Golany, B. e Y. Roll (1989). «An application procedure for DEA». En: *Omega* 17.3, págs. 237-250.
- González-Araya, M., G. Campo Hernández y L. Acosta Espejo (2013). «Selección de variables en modelos DEA usando series de tiempo». En: *Anais do XLV Simpósio Brasileiro de Pesquisa Operacional*, págs. 792-801.
- González-Araya, M. y N. Valdés (2009). «Método de selección de variables para mejorar la discriminación en el análisis de eficiencia aplicando modelos DEA». En: *Revista Ingeniería Industrial* 8.2, págs. 45-56.
- Jenkins, L. y M. Anderson (2003). «A multivariate statistical approach to reducing the number of variables in data envelopment analysis». En: *European Journal of Operational Research* 147.1, págs. 51-61.
- Limleamthong, P. y G. Guillén-Gosálbez (2018). «Mixed-integer programming approach for dimensionality reduction in data envelopment analysis: application to the sustainability assessment of technologies and solvents». En: *Industrial & Engineering Chemistry Research* 57.30, págs. 9866-9878.
- Lins, M. y M. Moreira (1999). «Método IO stepwise para seleção de variáveis em modelos de análise envoltória de dados». En: *Pesquisa Operacional* 19.1, págs. 39-50.
-

-
- Morita, H. y N. Avkiran (2009). «Selecting inputs and outputs in data envelopment analysis by designing statistical experiments». En: *Journal of the Operations Research society of Japan* 52.2, págs. 163-173.
- Nataraja, N. y A. Johnson (2011). «Guidelines for using variable selection techniques in data envelopment analysis». En: *European Journal of Operational Research* 215.3, págs. 662-669.
- Pastor, J., J. Ruiz e I. Sirvent (2002). «A statistical test for nested radial DEA models». En: *Operations Research* 50.4, págs. 728-735.
- Peyrache, A., C. Rose y G. Sicilia (2020). «Variable selection in data envelopment analysis». En: *European Journal of Operational Research* 282.2, págs. 644-659.
- Ruggiero, J. (2005). «Impact assessment of input omission on DEA». En: *International Journal of Information Technology & Decision Making* 4.3, págs. 359-368.
- Simar, L. y P. Wilson (2001). «Testing restrictions in nonparametric efficiency models». En: *Communications in Statistics-Simulation and Computation* 30.1, págs. 159-184.
- Soares De Mello, J., E. Gomes, L. Angulo Meza y M Lins (2004). «Selección de Variables para el incremento del poder de discriminación de los modelos DEA.» En: *Revista EPIO, Escuela de Perfeccionamiento en Investigacion Operativa* 24.1, págs. 40-52.
- Ueda, T. e Y. Hoshiai (1997). «Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs». En: *Journal of the Operations Research Society of Japan* 40.4, págs. 466-478.
- Wagner, J. y D. Shimshak (2007). «Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives». En: *European journal of operational research* 180.1, págs. 57-67.
-