



## SELECCIONES MATEMÁTICAS

Universidad Nacional de Trujillo

ISSN: 2411-1783 (Online)

2021; Vol. 8(2): 235-247.



### A non-parametric and multidirectional model in quantitative research

#### Un modelo no-paramétrico y multidireccional en investigación cuantitativa

Kelly Patricia Murillo<sup>Ⓜ</sup> and Eugénio Alexander Miguel Rocha<sup>Ⓜ</sup>

Received, Oct. 01, 2021

Accepted, Dec. 06, 2021



#### How to cite this article:

Murillo K, Miguel R E. A non-parametric and multidirectional model in quantitative research. *Selecciones Matemáticas*. 2021;8(2):235–247. <http://dx.doi.org/10.17268/sel.mat.2021.02.03>

#### Abstract

*In this work, a non-parametric and deterministic approach based on multidirectional efficiency analysis (MEA) is presented. The proposed model involves MEA with other important mathematical techniques in data analysis, such as the calculation of the NC value for the analysis of groups with different levels of efficiency. This model allows us to examine the factors that influence the behavior of decision-making units in different contexts such as business efficiency, educational quality, energy efficiency, circular economy, among others. Particularly in this work, we show the results obtained in 1787 Portuguese companies in the materials manufacturing sector (manufacture of wood and paper; manufacture of rubber products, plastics and other non-metallic mineral products; manufacture of common metals and metallic products, except machinery and equipment) in an eight-year period (2006-2013) of study. The results allow, not only a characterization of the financial structure of the sector and a diagnosis through indices that identify the strategic positioning of companies in terms of efficiency scores; but also a characterization of the most efficient units and an analysis of the variables that must be addressed differently, to obtain better results, in terms of economic performance.*

**Keywords** . Multidirectional efficiency analysis, NC value, dimensionality test, cluster analysis, fit model, manufacturing sector.

#### Resumen

*En este trabajo, es presentado un enfoque no paramétrico y determinístico basado en análisis de eficiencia multidireccional (MEA). El modelo propuesto envuelve MEA con otras técnicas matemáticas importantes en el análisis de datos, como el cálculo del valor NC para el análisis de grupos con diferentes niveles de eficiencia. Este modelo, permite examinar los factores que influyen en el comportamiento de unidades de decisión de diferentes contextos como eficiencia empresarial, calidad educativa, eficiencia energética, economía circular, entre otros. Particularmente en este trabajo, mostramos los resultados obtenidos en 1787 empresas portuguesas en el sector de fabricación de materiales (fabricación de madera y papel; fabricación de productos de caucho, plásticos y otros productos minerales no metálicos; fabricación de metales comunes y productos metálicos, excepto maquinaria y equipos) en un periodo de ocho años (2006-2013) de estudio. Los resultados permiten, no solo una caracterización de la estructura financiera del sector y un diagnóstico a través de índices que identifican el posicionamiento estratégico de las empresas en términos de puntajes de eficiencia; sino también una caracterización de las unidades más eficientes y un análisis de las variables que deben abordarse de manera diferente, para obtener mejores resultados, en términos de desempeño económico.*

**Palabras clave.** Análisis de eficiencia multidireccional, valor NC, prueba de dimensionalidad, análisis de conglomerados, modelo de ajuste, sector de fabricación.

\*Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIDMA), Departamento de Matemática, Universidade de Aveiro, Portugal. (kellymurillo@ua.pt).

†Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIDMA), Departamento de Matemática, Universidade de Aveiro, Portugal. (eugenio@ua.pt).

### Resumo

*Neste trabalho, é apresentada uma abordagem não paramétrica e determinística baseada na análise de eficiência multidireccional (MEA). O modelo proposto envolve o MEA com outras técnicas matemáticas importantes na análise de dados, como o cálculo do valor NC para a análise de grupos com diferentes níveis de eficiência. Este modelo permite examinar os fatores que influenciam o comportamento das unidades decisórias em diferentes contextos como a eficiência empresarial, a qualidade educacional, a eficiência energética, a economia circular, entre outros. Nomeadamente neste trabalho, apresentamos os resultados obtidos em 1787 empresas portuguesas do sector da transformação de materiais (fabrico de madeira e papel; fabrico de produtos de borracha, plásticos e outros produtos minerais não metálicos; fabrico de metais comuns e produtos metálicos, exceto máquinas e equipamentos) em um período de oito anos (2006-2013) de estudo. Os resultados permitem, não apenas uma caracterização da estrutura financeira do setor e um diagnóstico por meio de índices que identificam o posicionamento estratégico das empresas em termos de escores de eficiência; mas também uma caracterização das unidades mais eficientes, para obter melhores resultados, em termos de desempenho económico.*

**Palavras-chave.** Análise de eficiência multidireccional, valor NC, teste de dimensionalidade, análise de cluster, modelo de ajuste, setor de manufatura.

**1. Introducción.** La búsqueda de un modelo coherente, eficaz y completo para evaluar la eficiencia de unidades tomadoras de decisión (DMUs) en todos los sectores económicos, es cada vez mas comun y necesario en un mundo tan competitivo y en constante cambio industrial como el de la actualidad. En particular, cuando hablamos de eficiencia en la industria, existen algunos aspectos que es importante resaltar para identificar que tipo de análisis es el adecuado. El primero, relacionado con el uso de los insumos, se conoce como eficiencia técnica, que relaciona las variables inputs/insumos, utilizados por la empresa en razón de los outputs/productos, que se generan. El segundo aspecto relacionado con los costos de los factores productivos (inputs/insumos), se denomina eficiencia económica, este incluye consideraciones de minimización de dichos costos.

En este estudio, adoptamos un enfoque determinístico y no paramétrico, que combina el Análisis de Eficiencia Multidireccional-MEA (por sus siglas en ingles, Multidirectional Efficiency Analysis), propuesto en [6], con otras técnicas matemáticas. El modelo combina puntajes de eficiencia de las DMUs, tasas de eficiencia de la muestra, índices de ineficiencia de las variables, análisis de conglomerados, análisis de componentes principales con prueba de dimensionalidad, entre otras técnicas. Un análisis de grupos con diferentes niveles de eficiencia es realizado mediante el calculo del valor NC y es propuesto un adecuado ajuste del modelo (model fitting), que proporciona procedimientos para visualizar y comparar el desempeño de las variables.

Específicamente en este trabajo para mostrar la aplicabilidad del modelo, examinaremos el desempeño del sector de fabricación de materiales Portugues, en términos de puntajes de eficiencia, midiendo la forma en que se utilizan los insumos y se producen los productos, en un conjunto de empresas portuguesas, registradas de acuerdo a la clasificación NACE Rev.2 (Clasificación estadística de actividades económicas en la Comunidad Europea, sección C), durante el periodo 2006-2013.

En la literatura económica, la medición de la eficiencia de las empresas en sectores económicos, se ha fundamentado en el análisis de frontera paramétrica y no paramétrica. En el caso del sector de fabricación, los dos métodos más utilizados son el enfoque de análisis de regresión requiriendo métodos econométricos y el Análisis Envoltente de Datos-DEA (por sus siglas en ingles Data Envelopment Analysis); requiriendo en su desarrollo de programación lineal. Este último, permite resolver problemas de maximización simultánea de productos o minimización simultánea de insumos, construyendo una frontera de producción óptima y comparando cada unidad de observación frente al óptimo esperado. Sin embargo, debido a que el modelo DEA se basa en las contracciones radiales en todas las entradas y salidas indeseables; y expansiones radiales en todas las salidas deseables; no es el más adecuado para evaluar la eficiencia en la producción de materiales cuando se requiere establecer índices de ineficiencia en cada variable individualmente.

El modelo MEA, permite la reducción de entradas y la expansión de salidas, buscando una mejora potencial por separado en cada variable de entrada y cada variable de salida, lo que lo hace mas ventajoso en comparación a otros métodos. En la investigación de [12], por ejemplo, se demuestra que MEA proporciona una imagen de rendimiento mucho más sutil que DEA porque con ese modelo se consiguen evaluar los potenciales de mejora relativa específicas de los insumos. En [3], se demuestra que al usar MEA en el factor total del índice Malmquist en vez del enfoque convencional DEA, es posible obtener un análisis específico de variables, cambios en la productividad, así como sus componentes. Existen estudios científicos en diferentes contextos, donde son claramente detectados los beneficios de MEA. Por ejemplo, [4], en el cual el uso de MEA permite investigar cómo las reformas ferroviarias afectan las ineficiencias de los factores de costo específicos. En [10], MEA permite no sólo evaluar el nivel de eficiencia energética y ambiental

y la tendencia del sector de transporte de China, sino también investigar los patrones de eficiencia de 30 regiones. En [16], se detectaron tanto los niveles de eficiencia MEA integrados, como los patrones de eficiencia; los cuales están representados por la eficiencia MEA específica de la variable según cada tipo de emisión o descarga de contaminantes industriales de las ciudades principales chinasas.

El modelo que presentamos corresponde a la necesidad de establecer una metodología bien estructurada y completa, que permita evaluar la eficiencia desde varios ángulos (ratios, padrões, determinantes, índices, etc) y establecer claramente cuales son los factores que deben ser mejorados. Este modelo permite examinar los factores que influyen en el comportamiento de unidades de decisión en cualquier contexto, como saúde pública, eficiencia empresarial, calidad educativa, eficiencia energética, economía circular, entre otros.

El resto de este documento se presenta de la siguiente manera. En la siguiente sección, se discute el enfoque multidireccional propuesto y son presentadas las ideas principales del modelo MEA y las otras técnicas matemáticas aplicadas. En la sección 3, se examina la eficiencia del sector de fabricación de materiales en Portugal. En la Sección 5, se establecen los comentarios generales y observaciones finales.

**2. Enfoque multidireccional.** El modelo propuesto para el análisis de eficiencia técnica, sigue la estructura descrita en la Figura 2.1. La metodología utilizada en cada una de las etapas de la figura (selección de variables representativas, análisis de conglomerados y análisis de eficiencia), son descritas en detalle en las secciones siguientes. Comenzaremos por abordar la etapa mas extensa y base de este estudio, que es el análisis de eficiencia. Así, en la Sección 2.1, describimos matemáticamente como funciona el análisis de eficiencia multidireccional (MEA) y a seguir las etapas faltantes del diagrama, son presentadas en la Sección 2.2.

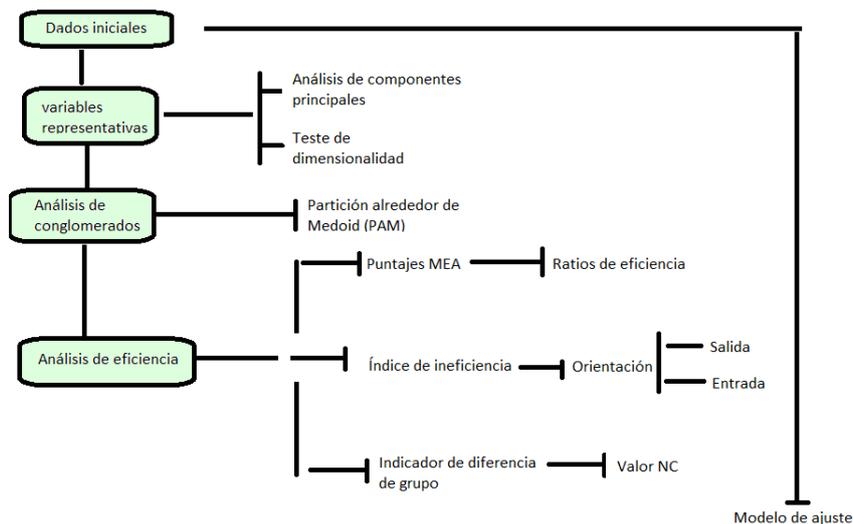


Figura 2.1: Estructura general del modelo aplicado

**2.1. Análisis de eficiencia.** Para examinar la eficiencia en este estudio, es usado un modelo no paramétrico y determinístico, que usa programación lineal. El análisis de eficiencia multidireccional, conocido como MEA por sus siglas en ingles (Mutidirectional Efficiency Analysis), fue propuesto inicialmente por [6]. Este enfoque es de gran importancia, desde que permite que las potenciales mejoras del modelo, sean realizadas de forma individual en cada variable.

A continuación, presentamos una descripción general del modelo MEA.

Sean  $I, J, T \in N$  y defina  $[m] = 1, \dots, m$ . Considere una dupla  $n = (e, t) \in ExT$ , identificando una empresa  $e \in E$ , y un año  $t \in T$ . Suponga que cualquier dupla  $n = n(e, t) \in N$  produce  $y_j(n)$ ,  $j \in [J]$  productos/salidas/ouputs, usando  $x_i(n)$ ,  $i \in [I]$  insumos/entradas/inputs; donde las primeras entradas  $1 < D \leq I$  son las llamadas entradas discretionales. Refiérase a los insumos que intervienen en el proceso de optimización. Por tanto,  $x(n) \in R^I$  es el vector de todas las entradas;  $y(n) \in R^J$  es el vector de todas las salidas y el conjunto de valores  $z(n) = (x(n); y(n))$ , es la base de datos para cada dupla  $n = n(e, t)$  dada.

Dada una dupla específica  $\bar{n} \equiv \bar{n}(e, t)$  y su conjunto de valores  $z(\bar{n}) = (x(\bar{n}); y(\bar{n}))$ . El procedimiento para obtener el puntaje MEA para  $z(\bar{n}) = (x(\bar{n}); y(\bar{n}))$  es establecido a seguir:

Primero se establece el conjunto en el cual estarán definidos los valores  $\lambda \in R^N$  considerados en los programas de optimización. Esto es, definir cuales son los rendimientos de escala considerados en el modelo: rendimientos de escala constante (CRS) o rendimientos de escala variable (VRS). Específicamente

este modelo definimos el conjunto  $\Lambda^n$ , con respecto a la variable rendimientos a escala (VRS) ([5]),

$$\Lambda^n = \left\{ \lambda \in \mathbb{R}^N : \sum_{n=1}^N \lambda_n = 1, \lambda_n \geq 0 \right\}. \tag{2.1}$$

A seguir se encuentra la solución para el programa de optimización lineal correspondiente a la minimización de los insumos:

$$\begin{aligned} &\text{Problem } P_m^\alpha(z, \bar{n}) : \\ &\min \alpha_m(\bar{n}) \text{ tal que} \\ &\sum_n \lambda_n x_m(n) \leq \alpha_m(\bar{n}), \\ &\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), i \in [I], i \neq m, \\ &\sum_n \lambda_n y_l(n) \leq y_l(\bar{n}), l \in [J], \end{aligned} \tag{2.2}$$

Luego se encuentra la solución para el programa de optimización lineal correspondiente a la maximización de las salidas:

$$\begin{aligned} &\text{Problem } P_j^\beta(z, \bar{n}) : \\ &\max \beta_j(\bar{n}) \text{ tal que} \\ &\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), i \in [I], \\ &\sum_n \lambda_n y_s(n) \leq \beta_j(\bar{n}), s \in [J], \\ &\sum_n \lambda_n y_l(n) \leq y_l(\bar{n}), l \in [J], l \neq j, \end{aligned} \tag{2.3}$$

Una vez obtenidas las soluciones optimas para los modelos anteriores, se define un tercer problema de optimización. Considerando  $\alpha_m^*(\bar{n})$  y  $\beta_j^*(\bar{n})$  soluciones optimas para los problemas  $P_{\alpha_m}(z, \bar{n})$  y  $P_{\beta_j}(z, \bar{n})$  respectivamente, se encuentra la solución para el siguiente programa de optimización lineal

$$\begin{aligned} &\text{Problem } P^\gamma(\alpha, \beta, z, \bar{n}) : \\ &\max \gamma(\bar{n}) \text{ tal que} \\ &\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}) - \gamma(\bar{n})(x_i(\bar{n}) - \alpha_i^*(\bar{n})), i \in [M], \\ &\sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), i \in [I] \setminus \{m\}, \\ &\sum_n \lambda_n y_l(n) \geq y_l(\bar{n}) + \gamma(\bar{n})(\beta_l^*(\bar{n}) - y_l(\bar{n})), l \in [L], \end{aligned} \tag{2.4}$$

De esta manera, calcular el puntaje MEA para una dupla especifica conlleva a abordar tres problemas de optimización, teniendo como punto de referencia ideal, el vector de salida MEA,

$$\zeta(\bar{n}) \doteq (\alpha_1^*(\bar{n}), \dots, \alpha_d^*(\bar{n}), \dots, x_I(\bar{n}), \beta_1^*(\bar{n}), \dots, \beta_J^*(\bar{n})). \tag{2.5}$$

En este sentido, el modelo MEA para una especifica observación  $z(\bar{n}) = (x(\bar{n}), y(\bar{n}))$  consiste de un total de  $(|D| + |J| + 1) \times N$  programas lineales, lo cual incluye un problema  $P_{\alpha_m}(z, \bar{n})$  para cada variable discrecional  $d \in D$ , un problema  $P_{\beta_j}(z, \bar{n})$  para cada una de las salidas  $j \in [J]$  y un problema  $P_\gamma(\alpha, \beta, z, \bar{n})$  con las soluciones optimas de los problemas  $P_{\alpha_m}(z, \bar{n})$  y  $P_{\beta_j}(z, \bar{n})$ . En otras palabras, el modelo MEA, consiste en encontrar soluciones:

$$\begin{aligned} &\alpha_m^*(\bar{n}) \in R^D \text{ tal que } \alpha_m^*(\bar{n}) \in P_{\alpha_m}(z, \bar{n}) \\ &\beta_j^*(\bar{n}) \in R^J \text{ tal que } \beta_j^*(\bar{n}) \in P_{\beta_j}(z, \bar{n}) \\ &\gamma(\bar{n}) \in R \text{ tal que } \gamma(\bar{n}) \in P_\gamma(\alpha, \beta, z, \bar{n}) \end{aligned} \tag{2.6}$$

Una vez que ya hemos presentado en detalle cada uno de los problemas de optimización que intervienen en el calculo del puntaje MEA, podemos abordar explícitamente su definición.

**Definición 2.1.** Dado un conjunto de datos  $Z = z(\bar{n}_N$  con  $z(\bar{n}) = (x(\bar{n}), y(\bar{n}))$ , el puntaje MEA denotado por  $MEA_z(\bar{n})$  es definido por medio de la siguiente ecuación,

$$MEA_z(\bar{n}) = \frac{\frac{1}{\gamma^*(\bar{n})} - \frac{1}{D} \sum_{i=1}^D \frac{x_i(\bar{n}) - \alpha_i^*(\bar{n})}{x_i(\bar{n})}}{\frac{1}{\gamma^*(\bar{n})} + \frac{1}{J} \sum_{j=1}^J \frac{\beta_j^*(\bar{n}) - y_j(\bar{n})}{y_j(\bar{n})}}, \quad (2.7)$$

donde,  $\alpha_i^*(\bar{n})$  y  $\beta_j^*(\bar{n})$  representan las soluciones optimas correspondientes a los problemas de optimización lineal  $P_{\alpha_m}(z, \bar{n})$  y  $P_{\beta_j}(z, \bar{n})$  respectivamente.

Observe que el puntaje MEA,  $MEA_z(\bar{n})$  es un valor en el intervalo  $[0, 1]$  y es obtenido por medio de la contribución direccional de cada entrada y salida, incluidas en el tercer problema. En efecto, para los insumos/entradas/inputs la contribución para cada unidad  $z(\bar{n})$  es dada por

$$\text{meff}_i(n) \doteq \frac{x_i(n) - \gamma(n)(x_i(n) - \alpha_i^*(n))}{x_i(n)} \chi_{[D]}(i), \quad (2.8)$$

para  $i \in [I]$ , donde  $\chi_{[D]}(i)$ , representa la function caracterisstica del conjunto D. Para los productos/salida-s/outputs, la contribución para cada unidad  $z(\bar{n})$  es dada por para  $j \in [J]$  por

$$\text{meff}_j(n) \doteq \frac{y_j(n)}{y_j(n) + \gamma(n)(\beta_j^*(n) - y_j(n))}, \quad (2.9)$$

Note, que  $\gamma(\bar{n})(x_i(\bar{n}) - \alpha_i^*(\bar{n}))$ , representa el exceso de entrada para todo  $i = 1, \dots, D$  y  $\gamma(\bar{n})(\beta_j^*(\bar{n}) - y_j(\bar{n}))$ , representa la insuficiencia de salida para todo  $j = 1, \dots, J$ .

Una característica muy interesante del modelo MEA, es la ineficiencia de cada variable de entrada, puede ser analizada individualmente. En efecto, con base en el exceso de entrada puede definirse el siguiente índice.

**Definición 2.2.** Dado un conjunto de datos  $Z = z(\bar{n}_N$  con  $z(\bar{n}) = (x(\bar{n}), y(\bar{n}))$ , el Índice de ineficiencia MEA, es definido como

$$INEF_i(\bar{n}) = \frac{\sum_{n=1}^N \gamma(\bar{n})(x_i(\bar{n}) - \alpha_i^*(\bar{n}))}{\sum_{n=1}^N x_i(\bar{n})} \quad (2.10)$$

para todo  $i \in [I]$  y dupla  $n \in N$ . Referimos el índice de ineficiencia  $INEF_i(\bar{n})$  para saber el número de veces en que cada entrada fue usada ineficientemente.

**2.2. Otras técnicas Matemáticas.** A continuación, presentamos brevemente las ideas principales de las técnicas usadas en las otras etapas del diagrama de la Figura 1.

*Selección sistemática de variables más significativas.* Un estudio de eficiencia coherente con los resultados, depende en gran medida de la relevancia de las variables consideradas en el estudio. Seleccionamos las variables mas relevantes para este estudio, utilizando el Análisis de Componentes Principales (PCA por sus siglas en ingles), acompañada de una prueba de dimensionalidad llamada test-dim. El análisis de PCA propuesto en [15], transforma una serie de variables correlacionadas en una serie de variables no correlacionadas, [1]. Una vez aplicando el PCA, se realiza una prueba de atenuación. Este permite probar el número de ejes en un análisis multivariante. El procedimiento se basa en el cálculo del coeficiente RV, [8].

*Análisis de conglomerados.* En los estudios de eficiencia, la importancia de aplicar el análisis de conglomerados o cluster analysis (CA, por sus siglas en ingles), se debe principalmente a que este muestra el grado de sensibilidad del puntaje de eficiencia de una unidad en particular a la presencia de otras unidades de la muestra que conforman la tecnología de referencia, [11]. En este estudio, se utiliza un algoritmo de agrupación espectral para obtener los conglomerados. El método utiliza Partition Around Medoid (PAM) basado en la distancia GDM2 (medida de distancia GDM para datos ordinales), [17], [14]. La principal ventaja de utilizar este método es que PAM determina el procedimiento de agrupación óptimo para el conjunto de datos. El análisis de conglomerados se realiza incrustando los datos en el subespacio de los vectores propios de una matriz de afinidad. El resultado del procesamiento de PAM contiene: el número de conglomerados encontrados para cada unidad y el número de cambios para cada unidad, con un número máximo de cinco variantes posibles.

*Indicador de grupo.* Analizar el comportamiento de grupos con diferentes niveles de eficiencia es un aspecto de relevante importancia en el análisis de datos. Las estadísticas del grupo eficiente con las del grupo no eficiente en este artículo, son comparadas utilizando una técnica desarrollada por [13], que se basa

en un coeficiente de intersección de distribución. Llamamos a este coeficiente valor-NC, para referirnos a las normales de distribución cruzadas. La idea es la siguiente.

Consideramos EFF como el conjunto de duplas eficiente, definido como

$$EFF \doteq \{n \equiv n(e, t) \in ExT, \text{ tal que } 0,6 \leq MEAz(n) \leq 1,0\}. \tag{2.11}$$

Definimos dos grupos de acuerdo con el coeficiente MEA:

-El grupo correspondiente a las duplas más eficientes, definido como

$$E_1 \doteq \{n \equiv n(e, t) \in ExT, \text{ tal que } 0,7 \leq MEAz(n) \leq 1,0\}. \tag{2.12}$$

-El grupo correspondiente a las duplas menos eficientes, definido como

$$E_0 \doteq \{n \equiv n(e, t) \in ExT, \text{ tal que } 0,0 \leq MEAz(n) \leq 0,4\}. \tag{2.13}$$

Es claro que  $E_1$  es un subconjunto de EFF y que  $E_0$  en la realidad corresponde a duplas no eficientes dada su puntuación muy baja. El valor-NC, se calcula como la intersección de las funciones gaussianas asociadas a las eficiencias mostradas en  $E_1$  y  $E_0$ . Cuanto mayor sea el valor de NC, menos común será el comportamiento de los dos grupos con respecto a las variables seleccionadas. De esta manera, es posible comparar el comportamiento de las variables de entrada y salida entre grupos con diferentes niveles de eficiencia y determinar desde el punto de vista su influencia en el modelo MEA.

*Modelo de ajuste.* Para comparar el comportamiento de las variables en el período de estudio, se aplica un modelo de ajuste/ fitting model convencional (mínimos cuadrados), minimizando la suma de distancias entre los puntos observados y la línea ajustada. Para aplicar el modelo procedemos de la siguiente forma: primero se seleccionan las variables de la muestra, que presentan un comportamiento más estable, durante todo el período de estudio. Para esto, consideramos todo el conjunto de datos inicial sin dividir por conglomerados, como se puede ver en el diagrama. Para continuar, se calculan los puntos de intersección entre cada dos rectas correlacionados con las variables seleccionadas en el paso anterior (variables mas estables). Los puntos de intersección establecen un orden relativo y se obtiene un gráfico que representa el comportamiento general de las variables en grupos a lo largo de los años. Na siguiente sección, puede verse cada paso em detalhe.

**3. Aplicación Empírica.** En esta sección examinamos la eficiencia técnica de empresas portuguesas en el sector de fabricación de materiales, durante el periodo 2006-2013. El conjunto de datos, recopilados de Amadeus (Bureau van Dijk), que llamaremos MATER, incluye información financiera y otras características de las empresas portuguesas que operan en el sector de la fabricación (NACE Rev.2 - Clasificación estadística de actividades económicas en la Comunidad Europea, sección C, códigos entre 1000 y 3399). Consideramos solo las empresas de fabricación de materiales con datos disponibles para todos los años en el período 2006-2013 (códigos NACE entre paréntesis). Específicamente los productos que son considerados en el estudio, son presentados en la Tabla 3.1.

Sector	NACE	Descripción códigos involucrados	Número empresas
MATER	1600 1899	Fabricación de productos de madera y papel e impresión;	1787
	2200 2599	Fabricación de productos de caucho y plásticos y otros productos minerales no metálicos;	
		Fabricación de metales básicos y productos elaborados de metal, excepto maquinaria y equipo.	

Tabla 3.1: Descripción del sector de fabricación estudiado

Para cada año y empresa, inicialmente extraemos de la base de datos Amadeus las siguientes 13 variables, descritas en la Tabla 3.2.

Las variables seleccionadas para el estudio son todas de carácter financiero. A continuación, son presentadas breves especificaciones en las definiciones de las variables para facilitar a el lector que no se encuentre familiarizado con los conceptos económicos utilizados, a establecer las diferencias en las variables seleccionadas. *CLIAB* requiere que se establezca dentro de los próximos doce meses o se espera que se establezca dentro de su ciclo operativo normal. *LTDEBT* incluye pagarés y bonos por pagar, alquiler a largo plazo, pensiones y otros beneficios. *PROFITM* muestra cuánto gana una empresa, teniendo en cuenta

los costos necesarios para producir sus bienes y servicios. *EBITM* indica la capacidad de una empresa para ser rentable y, en última instancia, para generar beneficios. *EBITDAM* se puede utilizar para analizar y comparar la rentabilidad entre empresas e industrias. *LIQR* se refiere a la capacidad de una empresa para movilizar activos y usarlos para pagar deuda, financiar operaciones actuales y reaccionar rápidamente ante las cambiantes condiciones comerciales. *SOLVR* se refiere al panorama crediticio general de una empresa y su capacidad para cumplir con las obligaciones a largo plazo y asegurar la financiación en el futuro. Está relacionado con la estructura general de capital de una empresa, su grado de apalancamiento financiero y el riesgo asociado con esa estructura.

Variable	Descripción
NE	número total de empleados incluidos en la nómina de la empresa;
CASH	cantidad de dinero en efectivo en bancos y en la mano de la empresa;
CAPITAL	capital social emitido (capital autorizado);
SALES	ventas, las ventas netas;
CLIAB	pasivos corrientes, la suma de los créditos, los acreedores y otros pasivos corrientes;
LTDEBT	la deuda a largo plazo (la deuda total de la empresa, debido a amortización superior a un año);
CASHFLOW	la suma de las ganancias por período y depreciación;
TASSETS	los activos totales (la suma de los activos fijos y bienes actuales);
PROFITM	las utilidades netas divididas por ventas;
EBITM	diferencia entre todos operativo los ingresos y todos los gastos operativos;
EBITDAM	suma de la utilidad de operación y depreciación;
LIQR	índice de liquidez (con efectivo e inversiones, dividido por los activos corrientes);
SOLVR	coeficiente de solvencia (después de impuestos de ganancias y la depreciación neta de la compañía, dividido por la cantidad de pasivos a largo plazo y corto plazo).

Tabla 3.2: Descripción de las variables

Note que el periodo de estudio considerado, puede ser dividido en tres períodos de tiempo: pre-crisis (2006-2008), pre troika (2009-2011) y troika (2012-2013), lo cual permite tener una visión de los efectos de la crisis financiera que se originó a finales del 2008 en toda Europa, sobre el sector de producción de materiales Portugueses.

**3.1. Análisis y resultados.** A continuación, abordamos la discusión de los resultados del estudio de la eficiencia técnica del sector MATER, siguiendo la estructura presentada en la Figura 1. Los resultados del algoritmo de CA, indican que los mejores conglomerados a utilizar en el estudio corresponden a las variables NE (conglomerado 1) y SALES (conglomerado 2). Las estadísticas descriptivas de todas las variables en el último año de estudio, se exhiben en la Tabla 3.3.

Los resultados de aplicar PCA al conjunto de datos mostraron las variables independientes y, por lo tanto, para la estimación de la eficiencia de las empresas, se consideraron las entradas: NE, TASSETS, LTDEBT, CLIAB y las salidas: PROFITM, LIQR, SOLVR, EBITM y EBITDAM.

*Tasas de eficiencia* Calculamos el puntaje MEA de cada empresa considerando las entradas y salidas definidas por la PCA. La Tabla 4 presenta información sobre los porcentajes de dos cantidades importantes: la Eficiencia total ( $EFF_{total} \equiv EFFT$ ), calculada como el coeficiente entre el número de duplas en EFF y el número total de duplas; y la Eficiencia superior ( $EFF_{superior} \equiv FULLEFF$ ), que representa los porcentajes de duplas con puntaje de eficiencia MEA igual a 1. Esto es, el porcentaje de duplas en el siguiente subconjunto de EFF

$$EFF \doteq \{n \equiv n(e, t) \in ExT, \text{ tal que } MEA_z(n) = 1, 0\}. \quad (3.1)$$

Las cantidades entre paréntesis representan la proporción de empresas por conglomerados en cada año (considerando 927 empresas para el conglomerado 1; y 860 empresas, para el conglomerado 2).

Variable	Conglomerado 1			Conglomerado 2		
	Suma	Media	D.Estandar	Suma	Media	D. Estandar
NE	7229	7,79	5,07	65537	76,205814	125,85
TASSETS	535781	577,97	554,77	14413213	16759,55	103598,01
LTDEBT	125204	135,06	166,07	3400138	3953,64	33693,594
CLIAB	208491	224,9	267,77	4592965	5340,65	22223,259
SFUNDS	178487	192,54	290,295	6012487	6991,26	49214,243
CASH	41395	44,65	81,87	1215737	1413,64	17660,423
CAPITAL	66392	71,62	126,28	2236836	2600,97	26605,622
PROFITM	-1901,36	-2,05	15,91	497,06	0,57	13,28
LIQR	1930,68	2,08	2,92	1102,08	1,28	1,05
PROFITM	-1901,36	-2,05	15,91	497,06	0,57	13,28
LIQR	1930,68	2,08	2,92	1102,08	1,28	1,05
SOLVR	27060,48	29,19	28,01	31659,89	36,81	20,24
EBITM	-113,66	-0,12	15,68	2573,4	2,99	12,15
EBITDA	6057,21	6,53	15,55	8145,8	9,47	12,17

Tabla 3.3: Estadísticas descriptivas, año 2013

Como podemos ver en la Tabla 3.4, el 2010 fue un año representativo, seguido por el 2007. En estos años, los EFFT totales tienen un porcentaje más alto en ambos clústeres (por encima del 74 %). En particular 2011 (conglomerado 2) también se caracterizó por un mayor porcentaje en FULLEFF; contrariamente al conglomerado 1. En términos generales el conglomerado 2, muestra un nivel de eficiencia superior al conglomerado 1.

AÑO	Conglomerado 1		Conglomerado 2	
	EFFT	FULLEFF	EFFT	FULLEFF
2006	72,6	3,78	78,3	4,19
2007	74,2	3,02	80,3	4,07
2008	72,8	2,59	75	3,95
2009	69,4	2,91	71,7	4,19
2010	77,3	3,78	83,7	5,7
2011	71,7	2,7	83,1	5,35
2012	66,1	2,16	76,3	4,19
2013	71,2	3,67	81,4	3,72

Tabla 3.4: Racios de eficiencia por conglomerados (%)

*Índice de ineficiencia* Los resultados detallados para el índice de ineficiencia de cada variable, se presentan en la Figura 3.1. El índice muestra la contribución de las variables en el cálculo del modelo.

Los porcentajes representan el número de veces que cada insumo se utilizó de manera ineficiente (insumos en exceso). La NE fue la variable menos utilizada de manera ineficiente en los dos conglomerados y LTDEBT fue la variable más utilizada de manera ineficiente. Curiosamente, el año 2012 fue el año más ineficiente en el conglomerado 1 y el menos ineficiente en el conglomerado defla mayoría de las variables en el conglomerado 2.

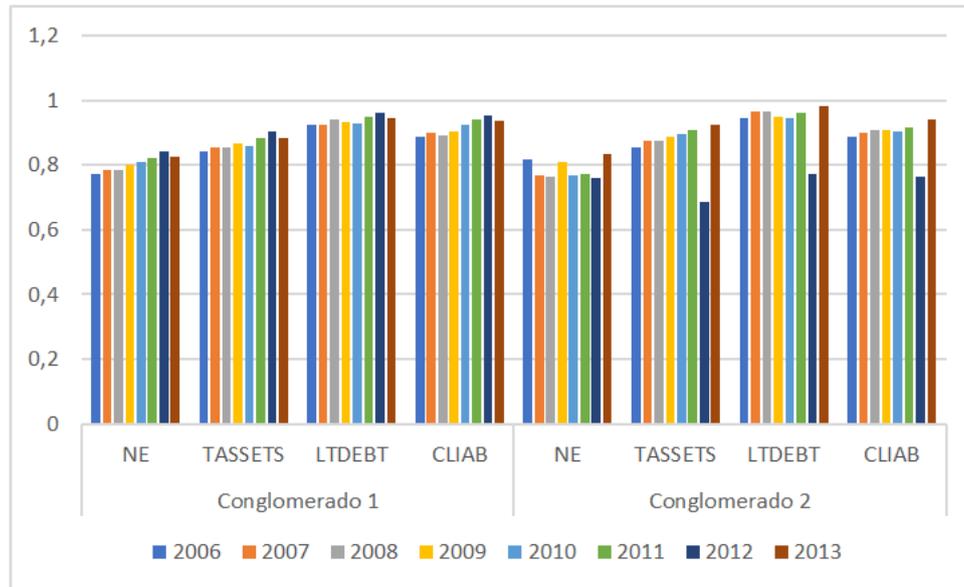


Figura 3.1: Indice de ineficiencia por conglomerados

Grupos con diferentes niveles de eficiencia. Determinamos los grupos  $E_1$  (duplas más eficientes) y  $E_0$  (duplas menos eficientes). En el conglomerado 1, el grupo  $E_1$  está conformado 13 duplas y el grupo  $E_0$  por 841. En el conglomerado 2, el grupo  $E_1$  está conformado 26 duplas y el grupo  $E_0$  por 612. A seguir calculamos el valor NC para cada variable. En la Figura 3.2, presentamos el valor NC para cada variable, correspondiente a cada conglomerado. Nótese que la diferencia entre los grupos de eficiencia e ineficiencia en los resultados de PROFITM es muy pequeña con respecto a los otros casos en especial en el conglomerado 2. La mayor diferencia entre los dos grupos se presenta en las variables CASHFLOW, CAPITAL y TASSETS, referente al conglomerado 1; y en las variables SOLVR, EBITDAM CASHFLOW para el conglomerado 2.

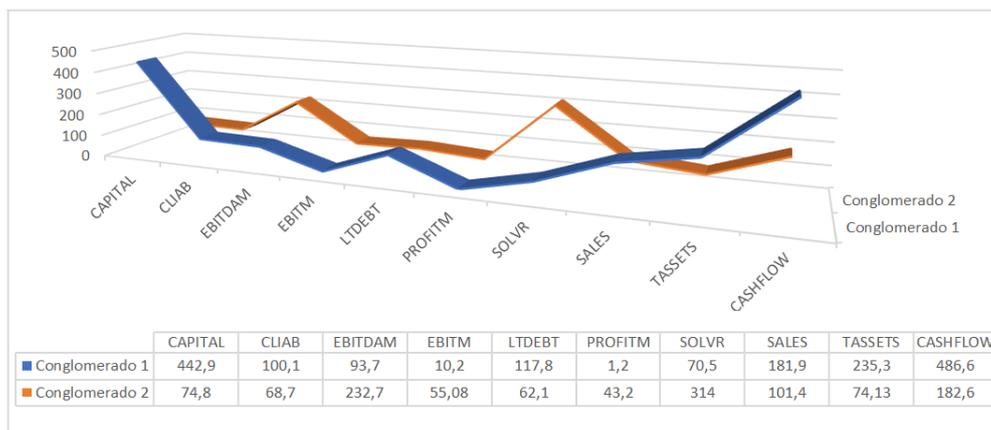


Figura 3.2: Valor NC para todas variables por conglomerados

Un ejemplo del comportamiento de cada variable frente al conglomerados 1, es representada en la Figura 3.3 para LTDEBT y SOLVR); y en Figura 3.4 para CLIAB y PROFITM, respectivamente. La línea azul representa la media del grupo mas eficiente  $E_1$  y la línea negra la menos eficiente  $E_0$ . Podemos ver que el grupo menos eficiente; adquiere un volumen mayor de deudas a largo (LTDEBT) y corto plazo (CLIAB) que el mas eficiente, durante todo el período de estudio.

Sin embargo, esto no implica que el grupo menos eficiente adquiera a lo largo de los años, un mayor lucro (PROFITM) y una mayor solvencia (SOLVR) que el grupo mas eficiente. Aunque si le permitió mantener una solvencia estavel, el lucro fue debilitándose con el pasar de los años.

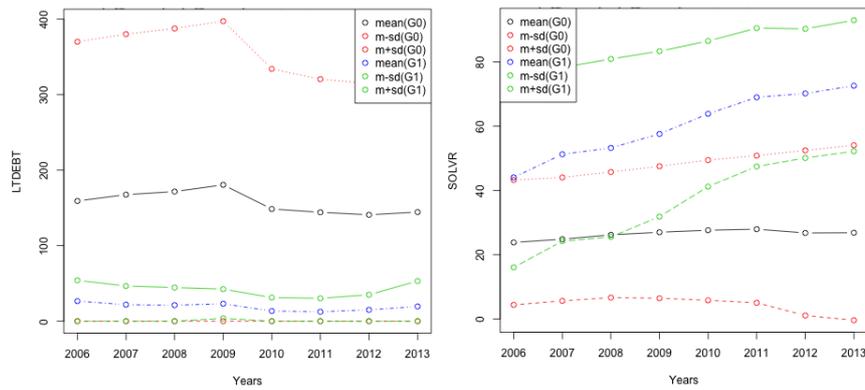


Figura 3.3: Comportamiento grupos de eficiencia LTDEBT (izquierda), SOLVR (derecha).

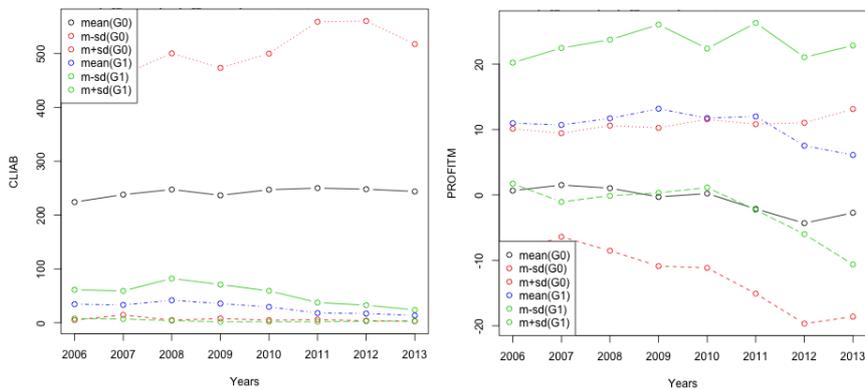


Figura 3.4: comportamiento grupos de eficiencia: CLIAB (izquierda), PROFITM (derecha).

*Caracterización del sector usando el ajuste de modelos* Para aplicar el ajuste del modelo, primero consideramos el conjunto de datos sin dividir por conglomerados, y seleccionamos las variables más relativamente estables durante todo el período de estudio: CAPITAL (para las entradas) y el ratio EBITM (para las salidas). Luego se genera la ecuación de la línea entre CAPITAL (resp. EBITM) y cada una de las otras entradas (resp. Salidas).

Para continuar, se calculan los puntos de intersección entre cada dos rectas correlacionados con CAPITAL (resp. EBITM), generando un orden relativo. Con esta información se generan, interesantes representaciones gráficas, que reflejan el comportamiento general de las variables en grupos, a lo largo de los años. Figura 3.5 muestra las variables relacionadas con CAPITAL. Figura 3.6 muestra las variables relacionadas con EBITM.

En cada gráfico se representan los valores mínimos y máximos para cada variable e intersecciones se consideran solo dentro de este rango. Usamos tres divisiones para la variable CAPITAL y cuatro divisiones para el EBITM. El orden relativo (columnas) representa el número de veces que una empresa cambia de conglomerado.

Para facilitar la descripción en el gráfico, en el caso CAPITAL, nos referimos a las grandes empresas, a las representadas con línea roja; empresas medianas con línea azul y pequeñas empresas con línea negra. Encontramos una disminución significativa de NE para todas las empresas en los años 2008-2010. Aumento de deudas a largo plazo de las pequeñas empresas durante todo el periodo de estudio, excepto en el 2009. Comportamiento muy similar para medianas y grandes empresas con relación a sus activos, durante todo el estudio. Las deudas a corto plazo fueron muy variantes para los tres tipos de empresas. Referente a la Figura 3.6, el ratio EBITM indica la capacidad de una empresa para ser rentable y, en última instancia, para generar beneficios, razón por la cual, resulta bastante informativo estudiar, el comportamiento de las variables relacionadas con este ratio.

En términos generales las empresas disminuyeron su solvencia notablemente en el año 2010-2011 y mantuvieron en grande variabilidad durante todo el periodo de estudio en los indicadores EBITDAM, LIQR y PROFITM. El estudio se centra en el período (2006-2013) en el que Portugal se vio directamente

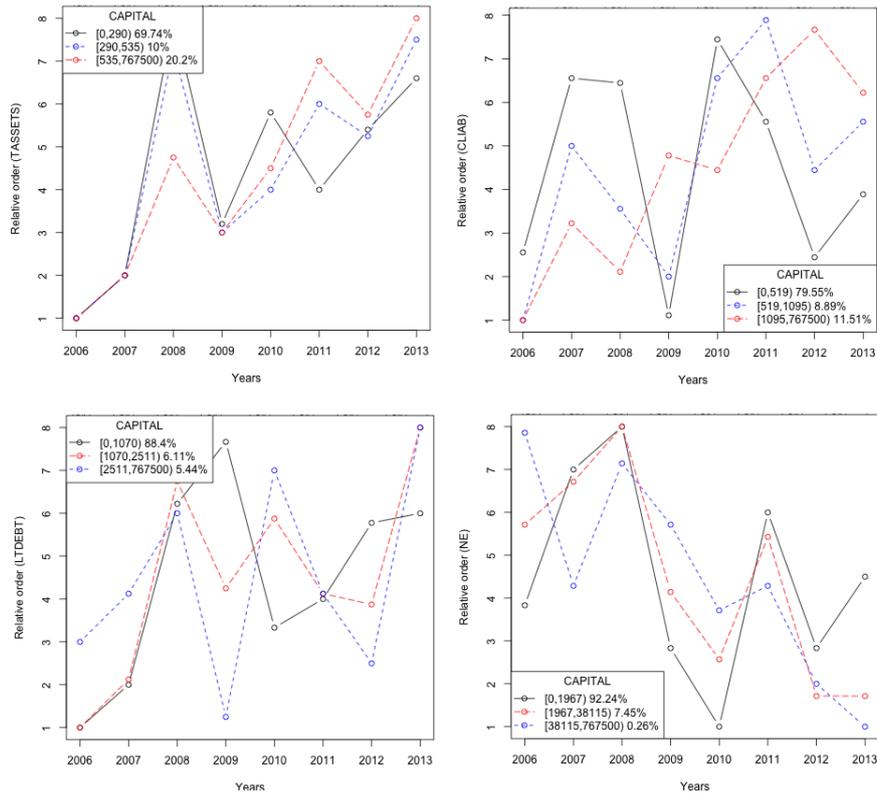


Figura 3.5: Orden relativo CAPITAL

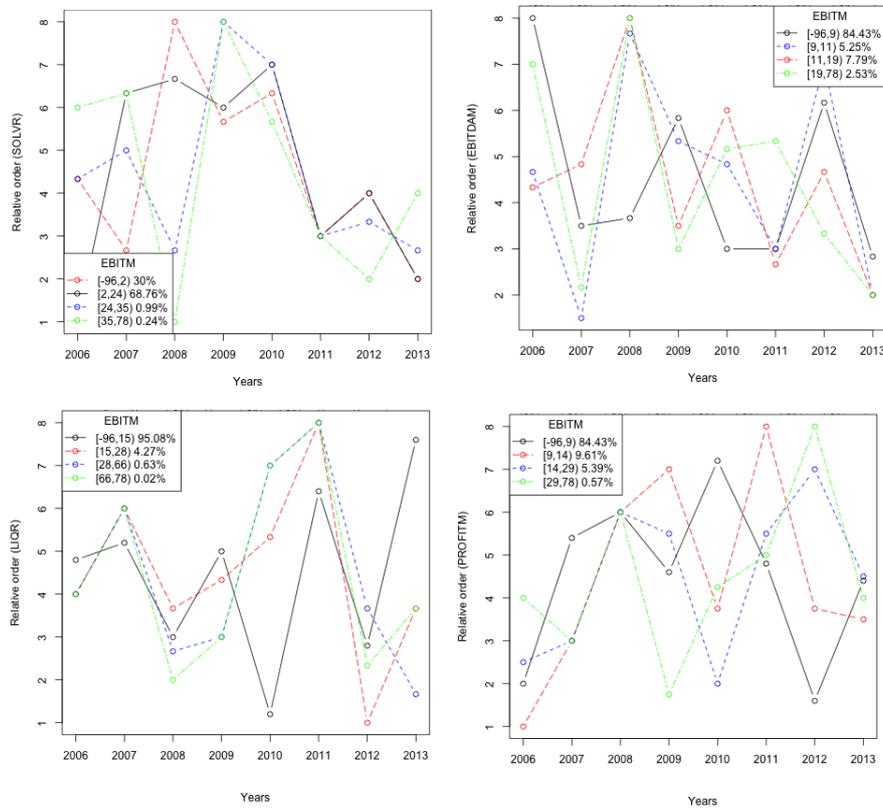


Figura 3.6: Orden relativo EBITM

afectado por la crisis financiera mundial: pre-crisis (2006-2008), pre troika (2009-2011) y troika (2012-2013). En este sentido, las consecuencias de la crisis se ven directamente reflejadas en la eficiencia de las empresas. En cuanto a los niveles de eficiencia, las grandes empresas fueron las más afectadas por la crisis en términos del desempeño de las variables año tras año; por el contrario, las pequeñas empresas mostraron mayor estabilidad en varias de las variables estudiadas.

**4. Conclusiones.** Este estudio presenta un modelo no paramétrico e determinístico que combina el Análisis de Eficiencia Multidireccional con otras técnicas matemáticas como análisis de componentes principales, teste de dimensionalidad, análisis de aglomerados, indicador de grupo y modelo de ajuste, para estudiar la eficiencia de unidades de decisión en diferentes contextos. En particular, se examina el desempeño del sector de fabricación de materiales (MATER) en Portugal, durante el período 2006-2013. El enfoque multidireccional presentado permite establecer las diferencias entre los patrones de eficiencia del sector considerando diferentes indicadores: ratios de eficiencia, índices de ineficiencia y valores de NC entre grupos con diferentes niveles de eficiencia. El análisis ofrece también una comparación directa entre variables, mediante el ajuste del modelo.

Con el fin de lograr una mejor caracterización de la eficiencia del sector MATER en términos económicos, el estudio incluye importantes indicadores del desempeño económico de las empresas, tales como rentabilidad, liquidez, solvencia, cobertura de intereses, entre otros. Observaciones importantes son establecidas sobre el sector MATER. En primer lugar, se muestra que es posible establecer características significativas entre las empresas, cuando se agrupan en conglomerados, en cuanto al número de empleados y ventas. En particular, el año 2010 fue un buen año para el sector MATER, relativo a la superación de las dificultades generadas por la crisis, ya que el porcentaje de eficiencia total superó el 77% en ambos conglomerados. El número de empleados fue la variable menos utilizada de manera ineficiente por los dos conglomerados, seguida por TASETS (conglomerado 1) e CLIAB (conglomerado 2). La mayor diferencia entre los grupos con diferentes niveles de eficiencia, se observa en las variables CASHFLOW para el conglomerado 1 (con valor NC 486,6) y SOLVR para el conglomerado 2 (con valor NC 314). El grupo de empresas menos eficiente aumentó la adquisición de deudas a largo y corto plazo (LTDEBT y CLIAB), durante todo el período de estudio, sin embargo no es posible establecer una evidencia de que esto implique la capacidad de las empresas para crecer más rápido. En este sentido sería necesario un estudio futuro con un número de periodo mas amplio para evidenciar las secuencias de las deudas adquiridas.

El modelo de desempeño utilizado en este estudio permite probar si las empresas bajo evaluación pueden reducir sus insumos mientras mantienen sus resultados en los niveles actuales. Los resultados son importantes en la determinación de las fallas que bloquean la mejora de la calidad de las empresas y en el establecimiento de estrategias para fortalecer el desempeño del sector. Dado que el sector MATER hace parte de los sectores que genera más empleo en Portugal, un trabajo futuro interesante sería comparar la eficiencia de este sector con otros países de características similares, estableciendo una clasificación comparativa y considerando un período de tiempo más extenso.

**Agradecimientos.** Los autores contaron con el apoyo del Centro de Investigación y Desarrollo en Matemáticas y Aplicaciones (CIDMA) a través de la Fundación Portuguesa para la Ciencia y la Tecnología (FCT - Fundação para a Ciência e a Tecnologia), referencias UIDB/04106/2020 y UIDP/4106/2020. Murillo también contó con el apoyo de Fondos Nacionales (OE), a través de FCT, IP, en el ámbito del contrato marco previsto en los números 4, 5 y 6 del artículo 23, del Decreto Ley 57/2016, de 29 de agosto, modificado, por la Ley 57/2017, de 19 de julio.

#### ORCID and License

Kelly Patricia Murillo <https://orcid.org/0000-0003-1479-8757>

Eugénio Alexander Miguel Rocha <https://orcid.org/0000-0003-3628-6795>

This work is licensed under the [Creative Commons - Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

## Referencias

- [1] Abdi H, Williams LJ. Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics; 2010; 2(4): 433–59. doi:10.1002/wics.101.
- [2] Abood S. Quality of improvement initiative in nursing homes. Am J Nurs [Internet]. [Consultado 22 Nov 2012]; 2002; 102(6). Disponible en: <http://www.nursingworld.org>.
- [3] Asmild M, Baležentis T, Hougaard JL. Multi-directional Productivity Change: MEA Malmquist, Journal of Productivity Analysis, 2016; 46: 109–119.
- [4] Asmild M, Holvad T, Hougaard J, Kronborg D. Railway reforms: do they influence operating efficiency? Transportation, 2009; 36(5): 617-638.
- [5] Banker RD, Charnes A, Cooper W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, 2009; 30(9), 1078–92. doi:10.1287/mnsc.30.9.1078.

- [6] Bogetoft P, Hougaard JL. Efficiency Evaluations Based on Potential (Non-proportional) Improvements. *Journal of Productivity Analysis*, 1999; 12(3):233–47. doi:10.1023/A:1007848222681.
- [7] Dawes J, González-Parra G, Rowley J. Enhancing the customer experience: contributions from information technology, *J Business Res.* 2005; 36(5):350-7.
- [8] Dray S. On the Number of Principal Components: A Test of Dimensionality Based on Measurements of Similarity Between Matrices. *Computational Statistics and Data Analysis*, 2008; 52(4): 2228–37. doi:10.1016/j.csda.2007.07.015
- [9] Evangelista A, Ortiz A, Ríos-Soto K, Urdapilleta A. USA the fast food nation: Obesity as an epidemic. *Los Alamos National Laboratory*; 2004.
- [10] Gongbing B, Pingchun W, Feng Y, Liang L, Energy and Environmental Efficiency of China's Transportation Sector: A Multidirectional Analysis Approach. *Mathematical Problems in Engineering*, 2014; 1-12. Article ID 539596. doi:10.1155/2014/539596
- [11] Hirschberg JG, Lye JN. Conglomeradoing in a Data Envelopment Analysis Using Bootstrapped Efficiency Scores (Department of Economics - Working Papers Series No. 800). The University of Melbourne. 2001.
- [12] Hougaard JL, Kronborg D, Overgård, C. Improvement Potential in Danish Elderly Care. *Health Care Management Science*, 2004; 7(3): 225–235.
- [13] Inman HF, Bradley Jr EL. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities, *Communications in Statistics, Theory and Methods*, 1989; 18(10): 3851-3874.
- [14] Karun K, Isaac E. Cognitive Analysis on k-means Conglomeradoing Algorithm and its Variants. *International Journal of Advanced Research in Computer and Communication Engineering*, 2013; 2(4):1875–80.
- [15] Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901; 2(11): 559–72. doi:10.1080/14786440109462720.
- [16] Wang K, Yu S, Li MJ, Wei Y. Multi-directional efficiency analysis-based regional industrial environmental performance evaluation of China. *Natural Hazards*. 2015.
- [17] Walesiak M. Multivariate Statistical Analysis in Marketing Research (Research Papers No. 654). *Wroclaw University of Economics Supplement*, 1993; 75(2): 273-299. <http://dx.doi.org/10.1007/s11069-014-1097-4>
- [18] Romero S, Moreno FJ, Rodríguez IM. *Linear Partial Differential Equations for Engineers and Scientists*. 2th. ed. Boca Raton: Chapman & Hall/CRC; 2002.
- [19] Thomas D, Weederma M, Fuemmeler B, Martin C, Dhurandhar N, Bredlau C, Bouchard C. Dynamic model predicting overweight, obesity, and extreme obesity prevalence trends. *Obesity*. 2014; 22(2):590-597.