

Factors Influencing the Economic Behavior of the Food, Beverages and Tobacco Industry: A Case Study for Portuguese Enterprises

KELLY P. MURILLO^a EUGÉNIO M. ROCHA^b

Received: 11.03.2020; Revised: 19.06.2020; Accepted: 01.10.2020

In today's world, it is increasingly important to conduct economic and financial analyzes of enterprises in all sectors to determine strengths, identify weaknesses and adopt strategies that allow them to be at the highest competitive level. In particular, the food sector plays an essential role in the economy of any country, representing a significant contribution to gross domestic product, total employment, and disposable income of households. In this work, we adopt a methodology for measuring efficiency based on the multidirectional efficiency analysis and other mathematical techniques (the calculation of the normal distribution intersection coefficient (NC value), analysis of clusters and principal components, and model fitting) in order to examine the factors that influence the performance of Portuguese enterprises in the food, beverages and tobacco industry for the period of 2006-2013. The results show a characterization of the financial structure of the sector and diagnosis through indexes that identify the strategic positioning of the enterprises in terms of efficiency scores. In addition, we also show that an analysis of the variables that must be approached differently to obtain better results regarding economic performance. Although there is an increase in credit with the acquisition of long-term debts, there is no evidence that this implies the ability of enterprises to grow faster, which affects profitability.


JEL codes: C14, C38, D22

Keywords: Multidirectional efficiency analysis, Clustering analysis, NC-value, Portuguese food industry

1 Introduction

The food products, beverage and tobacco (FBT) sector is of great importance in the economic growth of any country. The enterprises of the sector manufacture essential products for the community. Additionally, the sector has considerable interaction with the other sectors of the economy by providing the products that they use as intermediate inputs.

^a Corresponding Author. Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, Campus Universitario de Santiago, Aveiro, Portugal. email: kellymurillo@ua.pt  <https://orcid.org/0000-0003-1479-8757>

^b Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro. email: eugenio@ua.pt  <https://orcid.org/0000-0003-3628-6795>

The FBT sector is the manufacturing industry that contributes the most to the economy of the European Union (EU). This is not surprising since over 280,000 small and medium-sized enterprises are responsible for almost 50% of the sector's turnover, and the socioeconomic status of the sector in any country is related to the performance of all the enterprises involved. Particularly, Portugal was one of the EU countries whose economy was affected by the global financial crises in 2008/09. Based on information compiled in the Central Balance-Sheet Database of the Central Bank of Portugal, the FBT sector represented 13%, 14%, and 16% of the manufacturing industry in total employment, the number of enterprises, and turnover, respectively (Central Bank of Portugal, 2011).

A detailed study of the FBT sector in Portugal is of great importance from several points of view. Given the food component immersed in the FBT sector, it establishes an essential primary sector for the society, and therefore it has a direct influence on the economic development of the country. The current situation that the world is facing generated by the COVID 19 pandemic shows once again that the development of the FBT sector is very significant for the society at any moment, especially in the face of a crisis, whether it is financial or public health. We analyze the period of 2006-2013, in which Portugal was going through the financial crisis generated in Europe in mid-2008, by using a sample covering more than two thousand Portuguese firms.

FBT is one of the sectors that generate the most employment in the world, it is important to have a detailed characterization of the financial structure of the sector and diagnosis through indexes that identify the strategic positioning of companies in terms of efficiency scores. An analysis of the variables that must be approached differently, to obtain better results in terms of economic performance, is also needed. Although we are focusing on the case of Portugal, we believe that our study shed some light to understand the evolution of the financial situation of the FBT sector in other countries.

The aim of this work is to evaluate the efficiency of the Portuguese companies operating in the FBT sector between 2006 and 2013 by developing a model that combines the Multidirectional Efficiency Analysis (MEA), introduced by Bogetoft & Hougaard (1999), with other mathematical techniques described below. The analysis of efficiency through the MEA model has some well-known advantages over other models. MEA evaluation includes desirable output expansion and input contraction at the same time. For this reason, it is possible to measure the improvement potential for each variable separately. In this study, we consider an input-oriented model which minimizes inputs while satisfying at least the given output levels with variable returns to scale (Banker et al., 1984). The inputs used to get the MEA score (definition 1) are identified as an inefficiency index (definition 2). In this sense, it is possible to know if we use the inputs appropriately. This is the most significant characteristic of MEA: the ability to examine the misuse of inputs (when the analysis involves an input-oriented model) or to identify the wasted output (when the analysis involves an output-oriented model).

Eliminating any redundancy in the data is important in efficiency studies. The Principal Component Analysis (PCA) with a dimensionality test is applied to choose the more appropriate variables (Hotelling, 1933). PCA allows the reduction of the number of variables by keeping most of the original variability of the data. The results of approaches to estimate dimensionality are very sensitive to the level of correlations between the variables and the number of observations (Peres-Neto et al., 2005). In this study, the *test-dim* approach of Dray (2008) is used to estimate the dimensionality using singular value decomposition.

A clustering algorithm is used in the present analysis to capture the natural structure of the data and get the best size-based sample partition. The Partition Around Medoid (PAM) is based on the Generalized Distance Measure on ordinal scale (GDM2) distance (Walesiak, M., 1993; Karun & Isaac, 2013).

The so-called NC-value index, which allows the comparison between efficient and inefficient groups, is another measure used in the present analysis. This method involves the calculation of a normal distribution intersection coefficient that measures the overlapping of Gaussian distribution functions. This methodological approach was initially addressed in Murillo, Rocha, & Ramalho (2018). The authors proposed a reduction of the dimensionality of the MEA model if these are correlated with other input or output variables that have a relatively stable behavior. In addition, two procedures to visualize and make comparisons between long-run efficiencies are included: comparisons between enterprises with respect to specific variables using model fitting and comparisons between groups with different levels of efficiency based on the calculation of the NC-value.

In the efficiency evaluation of Decision-Making Units (DMUs), the Data Envelopment Analysis (DEA) has been popularly applied to a broad variety of industries such as software engineering (Asmild et al., 2006), banking and insurance (Kaffash & Marra, 2017), sports (Bhat et al., 2019) and e-commerce (Wen et al., 2003). However, MEA has been a valid alternative in the last years. Gongbing et al. (2014) presents a DEA model with a multidirectional approach for the study of the Chinese transport industry. Wang et al. (2013) uses the MEA approach for an evaluation of the Chinese industrial sector. Murillo & Rocha (2018) studies the impact of the austerity measures (troika) on the manufacturing enterprises in Portugal, using a model based on MEA.¹ However, there are few studies in the literature about the efficiency of the Portuguese FBT sector. In Central Bank of Portugal (2011), the solvency of enterprises of the FBT sector is evaluated by using two indicators by comparing to EBITDA (earnings before interest, taxes, depreciation, and amortization). Machado (2017) presents a general characterization of the Portuguese food and beverage sector, in the midst of a political and economic crisis. EU-MERCI (2016) presents a characterization of the FBT sector during the period 2011-2016 for five countries: Portugal, Turkey, France, Czech Republic and Spain. In Interreg Central Europe (2017), the main characteristics of the European food and beverage industry are given. The analysis shows that the food sector presents the potential to improve socio-economic levels. Most studies on efficiency use the DEA model, in which the units are limited to radial restrictions on the variables considered. In this study, we use the MEA model since this method allows improvements in each variable, reflecting a more detailed analysis of efficiency levels based on the behavior and particular influence of each variable. This makes MEA the most suitable model to analyze the impact of efficiency on enterprises.

The remainder of this paper is laid out as follows. In the next section, a general overview of the FBT sector is given and the data sample is presented. Section 3 introduces the MEA approach and the more important aspects of the mathematical techniques are discussed. In Section 4, the main results are discussed. In Section 5, the concluding remarks are given. We present further details of the MEA model in Appendix A, and the correlation coefficients in Appendix B.

¹ See also the work of Asmild et al. (2009) using the MEA method.

2 Development of the Portuguese FBT Sector and the Characterization of the Data

The FBT sector is the largest manufacturing industry in the EU. Referring to the current economic trends, the sector is generating the highest turnover (15%), value added (12.8%) and employment (15%), superior to other important manufacturing sectors (Interreg Central Europe, 2017).

The FBT industry constitutes a sector of great economic impact in Portugal for two aspects: firstly, it contributes most to the Portuguese economy in terms of turnover and gross value added; and secondly, the sector has contributed to the trade balance with a growth rate of exports higher than that of imports. Since 2010, the FBT sector presents a performance higher than other national manufacturing sectors. Despite the crisis, the FBT sector was responsible for over 108,000 jobs in 2016, becoming one sector that generates the most employment in Portugal. The performance of the FBT sector has also been outstanding in international markets; the exports of the sector reached 8.95% of the total exports (EU-MERCI, 2016).

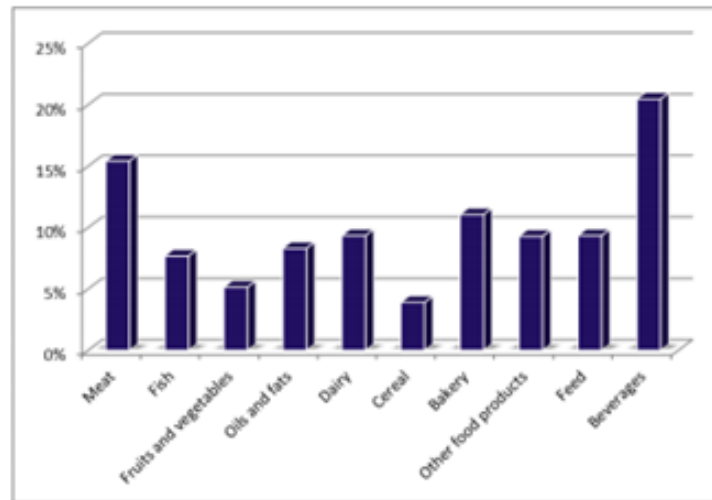


Figure 1: Turnover by Sector in Portugal in 2016
 Source: EU-MERCI (2016)

The beverage industry and meat preparation and preservation are the products of the FBT sector with higher turnover (20% and 15%, respectively), as depicted in Figure 1. Although the sector performed better than the total manufacturing industry in general, it experienced losses in the period of 2012-2103. For this reason, we are interested in studying the performance of enterprises during the years 2006 to 2013 and discover the factors that led to this negative growth.

The Amadeus (Bureau van Dijk) database considers information of enterprises in section C (Manufacturing) of NACE Rev. 2 (Statistical classification of economic activities in the European Community). Specifically, we consider 2,092 Portuguese enterprises of the FBT sector, between 2006 and 2013 inclusive. The data set contains only the enterprises common in all years. We extracted the following firm-level variables from the Amadeus database.

NE	Number of employees
CASH	Cash
CAPITAL	Capital stock
TASSETS	Total assets
LTDEBT	Long term debt
PROFITM	Profit margin (net income divided by revenues or net profits divided by sales)
LIQR	Liquidity ratio (the sum of cash and marketable securities divided by current assets)
SOLVR	Solvency ratio (the sum of post-tax profit and depreciation divided by the quantity of long-term and short-term liabilities)
SALES	Sales
EBITM	EBIT margin (the difference between all operating revenues and all operating expenses (gross profit-other operating))
EBITDAM	EBITDA margin (the sum of operating profit and depreciation)
CLIAB	Current liabilities (the sum of loans, credits and other current liabilities)
CASHFLOW	Cash flow (the sum of profit and depreciation)

During the 2006-2013 period, the FBT sector in Portugal maintained lower efficiency compared to some manufacturing sectors such as chemical products, pharmaceutical products, and transport equipment. However, in the same period, the FBT sector has a level of efficiency more similar to other sectors, including the manufacture of textiles, paper products and manufacture of computer, electronic and optical products (Murillo, Rocha, & Ramalho, 2018). Note that the previous period (2006-2013) involves the years in which the economic crisis in the world directly affected Portugal at the end of 2009. The sample period involves three significant sub-periods in the Portuguese economy: before the crisis (2006-2008), before the intervention of the troika (2009-2011) and during the troika (2012-2013). It can be expected that the consequences of the crisis over the FBT sector may have been reflected in the efficiency of their enterprises.

3 Methodology

In the next subsections, we provide the details of the model and the techniques used in this work. To extract the most information from the data, we consider the analysis with the structure shown in Figure 2. The methodology presented below is determined by four steps:

- 1) The determination of the relevant variables by using PCA and dimensionality test (Section 3.1);
- 2) Identification of clusters that best determine the classification of the data (Section 3.2);
- 3) The analysis of the MEA efficiency scores (Definition 1, Section 3.3) under three approaches:
 - The ratios of efficiency (see Section 3.3);
 - The input inefficiency index (Definition 2, Section 3.3);
 - Levels of efficiency by groups by calculating the NC-value (Section 3.4);
- 4) The characterization of the variables separately by using the method based on model fitting by least squares (Section 4.4).

To continue, crucial aspects of each step will be discussed. The data processing and each step of analyses are done in R program.

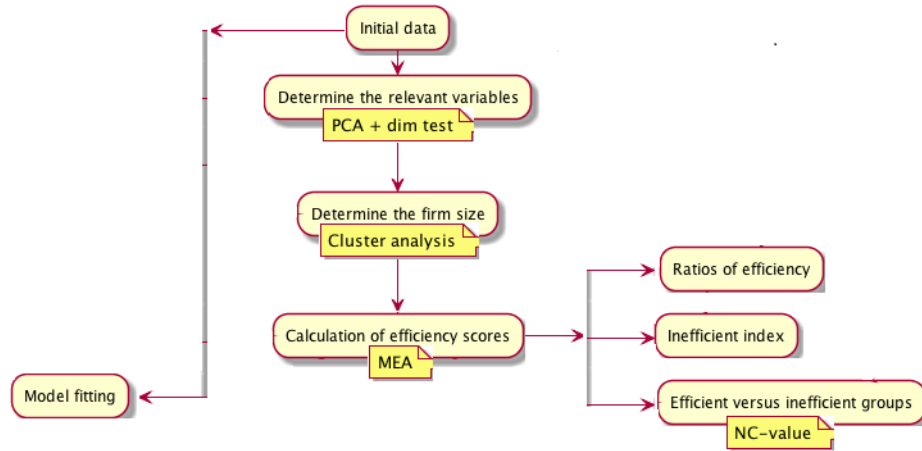


Figure 2: Structure of the Applied Methodology

3.1 Choosing (mathematically) meaningful variables

Techniques to obtain a better understanding of the relations and mathematical significances of the variables will be applied. We will select the better input-output variables for the study using the Principal Component Analysis (PCA), with the test of dimensionality *test-dim*. The PCA analysis proposed in Pearson (1901) transforms several correlated variables into various uncorrelated variables (Abdi & Williams, 2010). The PCA is often used as a first step to replace the original variables with the first major components. Then a key step is the correct choice of the number of axes to be retained. If the number of axes is not estimated correctly, it may introduce noise or loss of information into the analysis. In the literature, we find several approaches to estimate dimensionalities such as intuitively plausible ad-hoc rules, approaches that make distribution assumptions, and methods using computationally intensive procedures.² The method used here involves the calculation of the singular value decomposition of a data matrix, after mean centering (subtracting each data value from its variable's measured mean so that its empirical mean (average) is zero) the data for each attribute (Chen et al., 2011). Once the PCA has been applied, the *test-dim* is performed, which is based on the computation of the relationship between two sets of variables measured by the RV coefficient (Dray, 2008). We briefly explain it in what follows.

Let X be a table with the measurements of p centered variables (columns) for n individuals (rows). Consider the covariance matrix $(1/n)X^tX$, where X^t is the transpose of X . Then the PCA is based on its diagonalization. Set the singular value decomposition of

$$X^* = (1/\sqrt{n})X = UDV^t$$

where $D = \text{diag}(1, d_2, \dots, d_r)$ is a diagonal matrix with the r non-null singular values sorted in decreasing order; the column vectors in $U(n \times r)$ and V are orthonormal and satisfy $U = V^tV = I_r$, where I_r is the identity matrix of dimension r . Considering $X_m^\Delta = \sum_{i=1}^m X_i$ as the best approximation of X in the sense of least squares (minimization of $\|X - X_m^\Delta\|^2$),

² Jackson (1993), Ferré (1995) and Jolliffe (2002) provide some examples of the alternative methods.

we have

$$X = \sum_{j=1}^i X_j + R_{(i+1)}$$

where $X_j = d_j u_j v_j^t$ and R_i represents the residuals (Good, 1969).

In the next step, the test of dimensionality in PCA can be restated as a problem of the approximation of X . In this sense, we need to know if an element $\{X\}_i$ adds relevant information to the decomposition $X_{\{i-1\}}^{\Delta}$ of rank $i - 1$. The test is based on the similarity between X_i and R_i . If the i th dimension adds relevant information, the two configurations are close and their similarity can be measured by various coefficients such as RV, COI or RLS (Dray, 2008). This study considers the RV coefficient which is a measurement of the closeness between two configurations. The first configuration corresponds to the representation of the individuals in the one-dimensional space formed by the i^{th} principal axis. The second one is the configuration of individuals in the $(r - i + 1)$ -dimensional space formed by the last $(r - i + 1)$ principal axes (Escoufier, 1973). The corresponding dimensional RV statistic defined by

$$DIM_{RV}(X_i, R_i) = \frac{tr(X_i^t R_i R_i^t X_i)}{\sqrt{tr(X_i^t X_i X_i^t X_i) tr(R_i^t R_i R_i^t R_i)}} = \frac{\lambda_i}{\sqrt{\sum_{j=1}^r \lambda_j^2}} \quad (1)$$

The procedure to test the significance of the RVDIM is based on the permutation of values within each column of the original table X (Dray, 2008).

3.2 The Choice of Clusters

The cluster analysis is utilized for identifying structures within the data. Cluster analysis of a multivariate data set aims to divide a large data set into significant subgroups according to a set of specific characteristics. There are numerous methods available to classify objects based on their (dis)similarities with a variety of specific methods and algorithms (Bergman & Magnusson, 2001). There are four grouping algorithms.

- 1) Connectivity models: based on the idea that the closest data points in the data space exhibit more similarity to each other than the data points that are further away. An example of this model is the hierarchical grouping algorithm.
- 2) Centroid models whose notion of similarity is derived from the proximity of a data point to the centroid of the groups. An example of this model is the K-Means grouping algorithm.
- 3) Distribution models are based on the notion of how likely it is that all data points in the group belong to the same distribution. An example is the expectation of maximization algorithm.
- 4) Density models, which search the data space for areas of varying density of data points in the data space. Examples of these models are DBSCAN and OPTIC (Kim et al., 2019).

The method most used is the agglomerative hierarchical cluster analysis which has different rules on how to decide which of two clusters are more similar. The most appropriate cluster analysis method depends on the specific situation (Gordon, 1981). Methods that are generally tested with excellent results include the average linkage cluster analysis, and the

k-means relocation analysis (Morey et al., 1983).

In efficiency studies, the Cluster Analysis (CA) shows the degree of sensitivity of the efficiency score for a particular enterprise to the presence of the others in the sample that makes up the reference technology. CA can also compare the efficiency score of each enterprise to the other scores (Hirschberg & Lye, 2001). Including the size of the enterprises as a cluster, by using the formal definition of the European Commission (recommendation 2003/361/EC) is common in calculating the enterprise's efficiency. In this sense, many studies include the enterprises' size as a quantitative variable (e.g., sales) or as a categorical variable (e.g., micro, small, medium and large enterprises). Therefore, it is often assumed that the effects of size on the capital structure of the enterprises may vary (Ramalho & da Silva, 2009).

In this study, in contrast with the conventional division by size, a spectral grouping algorithm is used to get the clusters. The method uses Partition Around Medoid (PAM) based on the distance GDM2 (GDM distance measure for ordinal data) (Walesiak, M., 1993; Karun & Isaac, 2013). The PAM algorithm developed by Kaufman & Rousseeuw (1987) works with Medoids (an entity of the dataset that represents the group in which it is inserted). The algorithm partitions the dataset of n objects into k clusters, where both the dataset and the number k are inputs of the algorithm. PAM uses the following model to solve the problem.

$$\begin{aligned}
 F(x) &= \min \sum_{i=1}^n \sum_{j=1}^n d(i, j) z_{ij} \\
 &\quad s.t. \\
 &\quad \sum_{i=1}^n z_{ij} = 1, \quad j = 1, 2, \dots, n \\
 &\quad z_{ij} \leq y_i, \quad i, j = 1, 2, \dots, n \\
 &\quad \sum_{i=1}^n y_i = k, \quad k = \text{number of clusters} \\
 &\quad y_i, z_{ij} \in 0, 1, \quad i, j = 1, 2, \dots, n,
 \end{aligned} \tag{2}$$

where $F(x)$ is the objective function to minimize, $d(i, j)$ is the dissimilarity measurement between the entities i and j , z_{ij} is a variable that ensures that only the dissimilarity between entities from the same cluster will be computed in the main function, and y_i is defined as a binary variable equal to 1 if and only if the object i is selected as a representative object. In a general analysis, the algorithm proceeds in two phases: the construction phase and the exchange phase. The first phase comprises three parts: choosing k entities to become the medoids, calculating the dissimilarity matrix, and assigning each entity to its closest medoid. The second phase comprises two parts, which are circumstantial. First, if any of the entities in each group lowers the average dissimilarity coefficient, the entity that lowers this coefficient is selected as the medoid for this group. Second, if at least one medoid has changed, each entity is assigned its closest medoid. The major advantage of using this method is that PAM determines the optimal clustering procedure for the data set. The result of the algorithm contains the number of clusters found for each enterprise and the number of changes for each enterprise (mean and deviation standard) with a maximal number of five variants.

3.3 Measuring the MEA Efficiency Scores

For measuring the efficiency of the Portuguese FBT sector, we used the non-parametric deterministic MEA introduced by [Bogetoft & Hougaard \(1999\)](#). To continue, a general description of the MEA model is given and [Appendix A](#) provides further details.

Denote the set of enterprises and the set of years by E and T , respectively. Let $n = (e, t)$ be a tuple identifying the enterprise $e \in E$ and the year $t \in T$. Suppose that any given tuple $n \in N$ produces J outputs $y_j(n), j \in [J]$, using I inputs $x_i(n), i \in [I]$. We consider the first inputs $d \in [D], 1 < D \leq I$ being discretionary variables. The technical efficiency of each enterprise will be measured by calculating the MEA score.

Definition 1 (MEA score): Given a data set $z(n) = (x(n), y(n))$, where $x(n) \in RI$ is the vector of all the inputs $x_i(n)$, and $y(n) \in RJ$ is the vector of all the outputs $y_j(n)$, with $n \in N$; the MEA score of each $n \in N$ is obtained by

$$MEA_Z(n) = \frac{\frac{1}{\gamma^*(n)} - \frac{1}{D} \sum_{i=1}^D \frac{x_i(n) - \alpha_i^*(n)}{x_i(n)}}{\frac{1}{\gamma^*(n)} - \frac{1}{J} \sum_{j=1}^J \frac{\beta_j^*(n) - y_j(n)}{y_j(n)}} \quad (3)$$

where $\lambda \in \wedge^N$ and $\alpha_i^*(n), \beta_j^*(n)$ and $\gamma^*(n)$ correspond to the optimal solutions to the problems $P_i^\alpha(z, n), P_j^\beta(z, n)$ and $P^\gamma(z, n, \alpha^*, \beta^*)$.

3.3.1 Defining the Ratios of Efficiency

We consider fully efficient enterprises with $MEA_Z(n) = 1$, and null efficient enterprises with $MEA_Z(n) = 0$. Therefore, $MEA_Z(n)$ varies between 0 and 1. Consider EFF as a subset of units such as $0.6 \leq MEA_Z(n) \leq 1.0$. From the scores obtained, two ratios are generated to characterize the database such as the total efficiency (EFFT), which is the ratio of the number of efficient companies to the total number of companies, and Fully efficient (FULLEFF), which is the percentages of companies with $MEA_Z(n) = 1$.

Before continuing, an important aspect of the proposed model should be emphasized to ensure correct application and interpretation. In this study, we use an input-oriented model to test whether the enterprises under evaluation can reduce their inputs while maintaining the outputs at their current levels.

3.3.2 Calculating the Inefficiency Input Index

Using the MEA model allows us to analyze the inefficiency of the three input variables used in this study individually. One of the significant advantages of MEA is that it allows estimating the level of influence of each variable individually on the model. By following [Bogetoft & Otto \(2011\)](#), we introduce definition 2 to compute the number of times in which each input was used in an inefficient way.

Definition 2 (Inefficiency index): Given a data set $z(n) = (x(n), y(n))$ with $x(n) \in RI$ and $y(n) \in RJ$, and the MEA score of each tuple $n \in N$ (equation 3), the inefficiency index for each input $x_i(n)$ is obtained by

$$R_i(n) = \frac{\sum_{n=1}^N \gamma(n)(x_i(n) - \alpha_i^*(n))}{\sum_{n=1}^N x_i(n)}, \quad \text{for } i \in [I] \ \& \ n \in N \quad (4)$$

3.3.3 Comparing Groups with Different Levels of Efficiency

In order to compare the behavior of input and output variables between two groups, namely G_0 and G_1 , with different levels of efficiency, we use the NC-value. This indicator is defined to measure the overlap of Gaussian distribution functions of the groups (Inman & Bradley, 1989). This technique identifies the best practices presenting the variables that contribute to the difference between the two groups. The NC-value calculation has been applied in different contexts, such as the study of the efficiency of the Portuguese manufacturing firms during the financial crisis and the intensive and non-intensive energy sectors (Murillo, Rocha, & Ramalho, 2018; Murillo, Rocha, & Pardo, 2018), respectively.

The procedure requires defining which enterprises are in each group and then the indicator will make a comparison of the behavior of each group for each variable. The mean and the standard deviation of the variable in each group generate a normal distribution. Then, we compute the NC-value between the normal distributions of the groups for each of data subset, variable, and year of interest. In mathematical terms, the idea is the following. Let μ_1^t and μ_2^t be the average and σ_1^t and σ_2^t be the standard deviation of the groups F1 and F2 in the period $t \in T$, respectively. Assume $\mu_1^t < \mu_2^t$, without loss of generality. If $X_1 \sim N(\mu_1; \sigma_1)$ and $X_2 \sim N(\mu_2; \sigma_2)$ then

$$\begin{aligned} NC - value &= \frac{1}{|T|} \sum_{t \in T} (P(X_1 > c^t) + P(X_2 > c^t)) \\ &= \frac{1}{|T|} \sum_{t \in T} \left(1 - \frac{1}{2} \operatorname{erf} \left(\frac{c^t - \mu_1^t}{\sqrt{2}\sigma_1^t} \right) + \frac{1}{2} \operatorname{erf} \left(\frac{c^t - \mu_2^t}{\sqrt{2}\sigma_2^t} \right) \right) \end{aligned} \quad (5)$$

where c^t denotes the intersection point at which the probability density functions meet and erf stands for the error function. The point c^t is calculated as follows.

$$c^t = \frac{\mu_2^t(\sigma_1^t)^2 - \sigma_2^t \left(\mu_1^t(\sigma_2^t) + \sigma_1^t \sqrt{(\mu_1^t - \mu_2^t)^2 + 2((\sigma_1^t)^2 - (\sigma_2^t)^2) \log \left(\frac{\sigma_1^t}{\sigma_2^t} \right)} \right)}{(\sigma_1^t)^2 - (\sigma_2^t)^2} \quad (6)$$

The smaller the NC-value, the less common the behavior of the two groups regarding the selected variables. A higher NC-value means greater intersection between the behavior of the groups G_0 and G_1 .

3.4 Measuring Goodness of Fit

The macroeconomic variables defined every year in a country determine the level of overall economic activity in the country affecting positively or negatively the performance of enterprises. Therefore, it is important to study the behavior of enterprises over the years regarding its variables. In this sense, we consider it important to apply model fitting in our data. This allows us to have more tools to analyze the data set.

For a given data set, the model fit finds the parameters that minimize the difference between the original data and a predicted data set. To compare the behavior of the variables in the study period, the conventional method of fitting (least squares) will be applied. This method is characterized by minimizing the sum of distances between the observed points and the fitted line. Using this fitting procedure for affine functions, we proceed as follows. We calculate the p-values for the fittings that are obtained when relating each variable with

one of the input variables or output variables. Thus, suppose, there exist inputs $x_i(n), i \in [I]$ and outputs $y_j(n), j \in [J]$, such that

$$x_i(n) \approx a_i(t)C_i(n) + b_i(t) \wedge y_j(n) \approx c_j(t)E_j(n) + d_j(t) \tag{7}$$

where $C_i(n)$ and $E_j(n)$ are the value of selected input and output variables, respectively. The values $a_i(t)$ and $c_j(t)$ represent the coefficients of the relationship between each variable with the two selected variables, respectively. The values $b_i(t)$ and $d_j(t)$ are the intercepts of these relations.

4 Analysis and Results

4.1 Selecting More Relevant Variables

The results of PCA and the dimensionality test showed that nine variables should be considered in the estimation of efficiency: TASSETS, NE, CLIAB, LTDEBT, LIQR, PROFITM, SOLVR, EBITDAM and EBITM. Before applying the PCA, it was necessary to check if this method could be used in the database. For this, we apply the KMO (Kaiser-Meyer-Olkin) statistical test. The KMO indicates the proportion of the variance in the data (Figueiredo Filho & Silva Junior, 2010). This measure is represented by an index that assesses the adequacy of the factor analysis which is calculated by

$$KMO = \frac{\sum \sum_{j \neq k} r_{jk}^2}{\sum \sum_{j \neq k} r_{jk}^2 + \sum \sum_{j \neq k} q_{jk}^2} \tag{8}$$

where r_{jk}^2 is the square of the elements of the original correlation matrix outside the diagonal, and the term q_{jk}^2 is the square of the partial correlations between variables. The KMO value closer to 1, the more suitable it is for applying the final analysis. According to the criteria of Hair et al. (2006), the higher the level of KMO measurement is better and having 0.50 as the minimum level of suitability. A KMO Measure of Sampling Adequacy (MSA) takes values between 0.57 and 0.71 (Table 1) and shows that there is a mean correlation and the correlation between the variables increases through time.

Table 1: KMO Test Results

Year	Overall MSA	Year	Overall MSA
2006	0.57	2010	0.67
2007	0.58	2011	0.70
2008	0.64	2012	0.70
2009	0.67	2013	0.71

4.2 Cluster Analysis

The algorithm results of CA (Section 3.2) show that the best clusters to be considered in the study correspond to NE (cluster 1) and SALES (cluster 2). Recall that the algorithm ensures the best stability of the clusters for a long time. The descriptive statistics of the selected variables are presented in Table 2.

Table 2: Descriptive Statistics (2006)

Variable	Cluster 1			Cluster 2		
	Sum	Mean	St. dev.	Sum	Mean	St. dev.
NE	25,038.00	91.70	127.30	2,900.00	10.30	7.30
TASSETS	4,814,257.00	17,634.60	52,586.60	288,890.00	1,031.70	2,427.20
LTDEBT	708,951.00	2,596.80	4,910.90	95,818.00	342.20	1,127.20
CLIAB	2,269,499.00	8,313.10	32,475.60	110,293.00	393.90	931.90
PROFITM	407.70	1.40	8.00	-349.00	-1.20	13.60
LIQR	274.80	1.00	0.70	394.40	1.40	2.30
SOLVR	9,444.60	34.50	16.40	6,819.10	24.30	21.70
EBITM	923.90	3.30	7.30	343.10	1.20	12.90
EBITDAM	2,447.70	8.90	7.70	2,845.40	10.10	12.50

Source: Authors' calculations from Eurostat (2006-2013)

4.3 MEA Efficiency Score

4.3.1 Efficiency Ratios

We calculated the MEA score of each enterprise by considering the variables selected by the PCA. Table 3 presents information about the percentages of two quantities: Total Efficiency (EFFT) and Full efficient (FULLEFF) (Section 3.3). Recall that EFF is the subset of unities such as $0.6 \leq MEA_Z(n) \leq 1.0$ (equation 3). These values give us a general idea of the performance of the different enterprises over the years.

Table 3: Efficiency Ratios by Clusters

Year	Cluster 1		Cluster 2	
	EFFT	FULLEF	EFFT	FULLEF
2006	67	7.3	63.2	7.8
2007	71.4	7.6	63.6	11.4
2008	68.9	8.7	63.6	11.8
2009	74.4	8.1	66.8	7.14
2010	84.6	8.4	76.8	10.7
2011	77.7	9.5	70.7	8.93
2012	74	6.9	64.6	6.07
2013	75.5	9.9	63.6	8.93

As we can see in Table 3, 2010 was the year with the highest percentage of EFFT in both clusters (by above 84% and 76%, respectively). For cluster 2, the highest FULEFF was observed in 2013. Contrarily, the highest FULEFF values for cluster 2 were observed in 2007 and 2008.

4.3.2 Inefficiency Index

Detailed results of the MEA directional analysis with respect to clusters are presented in Figure 3. Note that each input involved in the MEA process can be improved individually. Figure 3 shows to what extent each variable could have been used better. Thus, the percentages in the figure show the number of years in which each variable was used in an inefficient way (equation 4). The NE is the variable less used inefficiently on the two clusters. LTDEBT and CLIAB are the variables more used inefficiently in cluster 1 and 2, respectively. Interestingly, most of the variables in cluster 2 have a lower percentage to be used inefficiently than cluster 1.

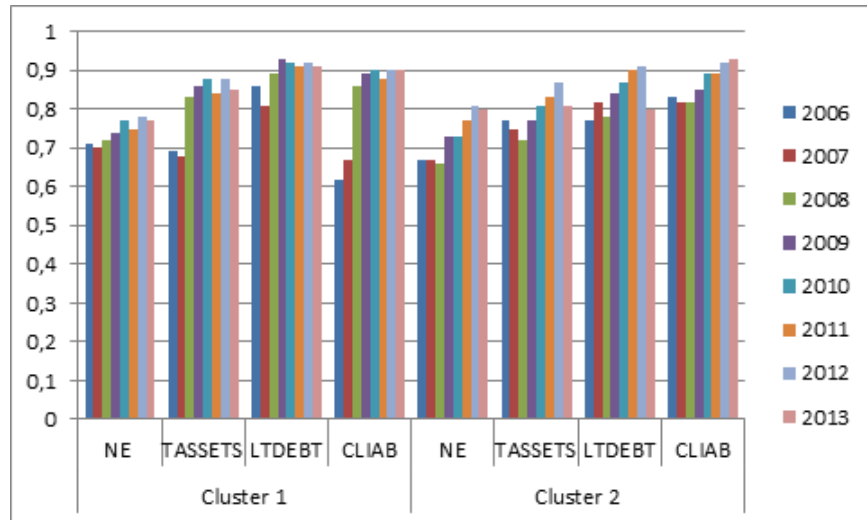


Figure 3: Inefficiency Index by Clusters

4.3.3 Groups with Different Levels of Efficiency

We calculate the NC-value for each variable (Section 3.3.3) by considering two groups with a different level of efficiency. The group G_1 corresponds to the more efficient enterprises (units such that $0.6 \leq MEAZ(n) \leq 1.0$). The group G_0 corresponds to the less efficient enterprises (units such that $0.0 \leq MEAZ(n) \leq 0.4$). Figures 4 and 5 represent the behavior of the two efficiency groups for the variables LTDEBT and LIQR. For the first variable, the NC-value is equal to 71.2 (cluster 1) and 322.5 (cluster 2). For the second variable, the NC-value is equal to 330.9 (cluster 1) and 220.5 (cluster 2).

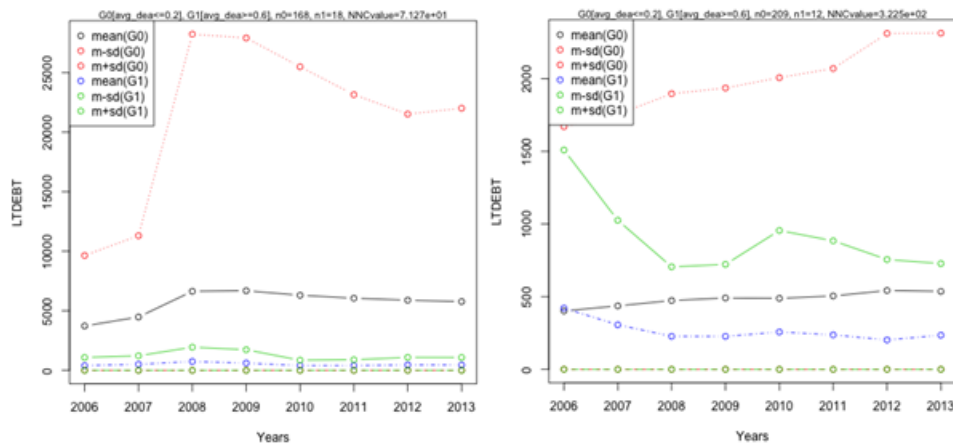


Figure 4: NC-value in the LTDEBT, for (a) cluster 1 (left side); (b) cluster 2 (right side)

The blue line represents the mean of the efficient group, the black line is the less efficient one, and the red and green lines represent the confidence intervals of standard deviation. We can see that the less efficient group gets more debt in both clusters during the entire period considered (Figure 4). For this reason, its liquidity is notably less than the efficient group, especially in 2008 (cluster 1) and 2010 (cluster 2), Figure 5.

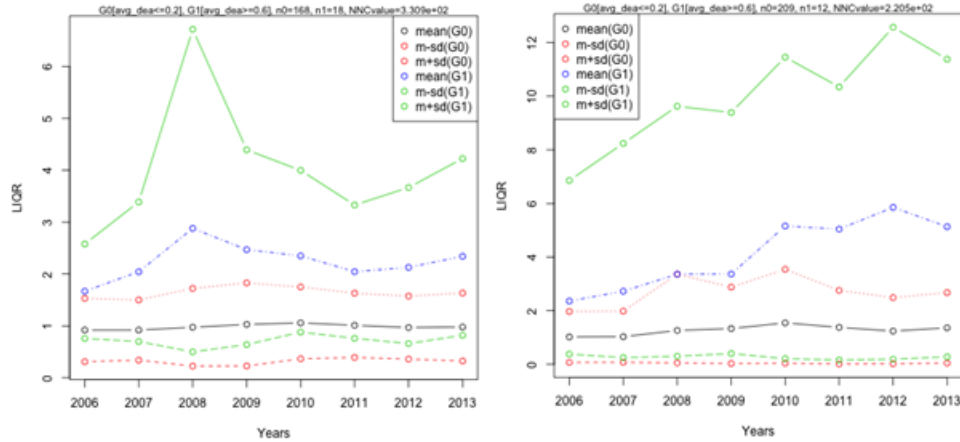


Figure 5: NC-value in the LIQR, for (a) cluster 1 (left side); (b) cluster 2 (right side)

In Table 4, we present the NC-value for each variable and cluster. Note that the difference between the efficient and inefficient group in the PROFITM results is very small compared to the other cases. The above is more noticeable in cluster 2. The bigger difference between the two groups is presented in EBITDAM for the cluster 1 (379.5) and in CAPITAL for the cluster 2 (520.1).

Table 4: NC-value for All Variables by Clusters

Variable	Cluster 1	Cluster 2	Variable	Cluster 1	Cluster 2
PROFITM	37.6	0.2	CASHFLOW	169.4	432.1
CLIAB	67.4	264.6	CAPITAL	182.5	520.1
LTDEBT	71.2	322.5	CASH	235.1	183.9
SALES	97.2	340.5	SOLVR	266.6	184.8
TASSETS	97.8	500.1	LIQR	330.8	220.5
EBITM	125.4	16.0	EBITDAM	379.5	190.6
NE	136.1	500.1			

4.4 Characterization of the FBT Sector by Using Model Fitting

Many problems in analyzing data involve describing how variables are related. A standard tool to measure the relationship between a pair of variables is the correlation. To get an idea of the relation between the variables of this study, we present the Pearson and Spearman correlation coefficients for the matrices for inputs and outputs in Tables B.1 and B.2, respectively, in Appendix B. The first correlation is used if variables are normally distributed, otherwise, the second one is used (Hauke & Kossowski, 2011). In these tables, the p-values measure how the variables are compatible.

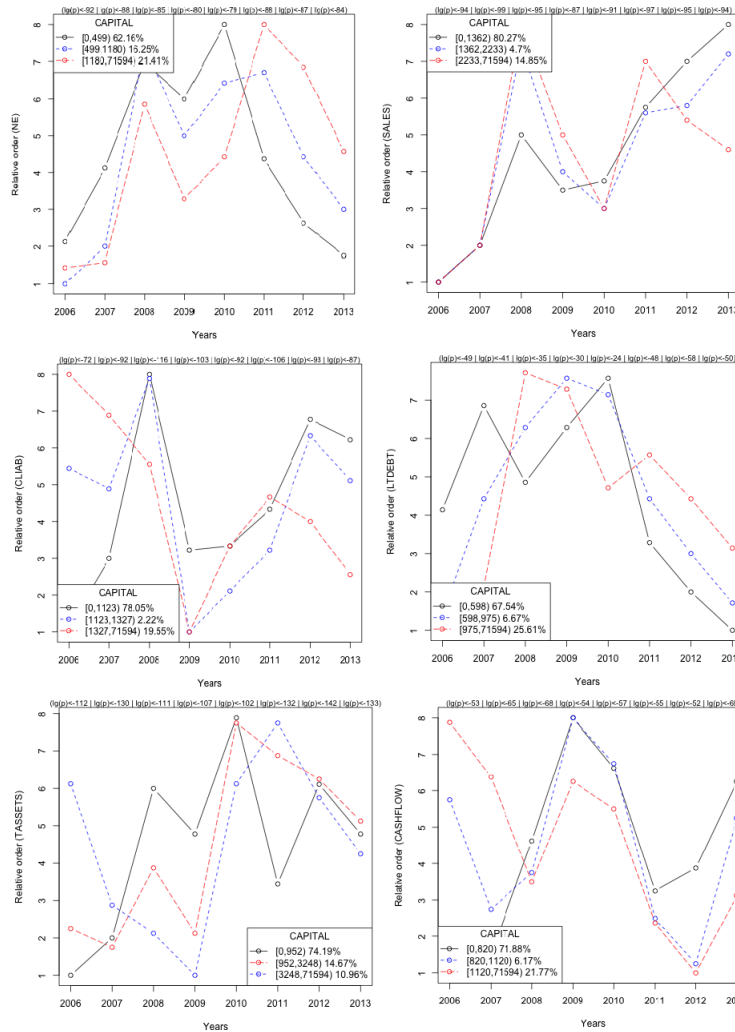


Figure 6: Relative order inputs: CAPITAL

To apply model fitting, we first consider the data set without dividing by clusters and select the variables that are relatively more stable between 2006 and 2013: CAPITAL (for the inputs) and the ratio EBITM (for the outputs). Then the equation of the line between CAPITAL (EBITM) and each one of the other inputs (outputs) is generated. The points of intersection between each two straight lines are calculated by generating a relative order (Murillo, Rocha, & Ramalho, 2018). By using this information, a relative order graph, which represents the general behavior of the variables in groups over the years, is obtained. Figure 6 is depicted for CAPITAL and Figure 7 shows the results for EBITM. Note that, in each graph, the minimum and maximum values for each variable are presented and the intersections are considered only within this range. We use three divisions for CAPITAL variable and four divisions for EBITM. The p-value is represented by log10p. The relative order (columns) represents the number of times in which an enterprise changes cluster.

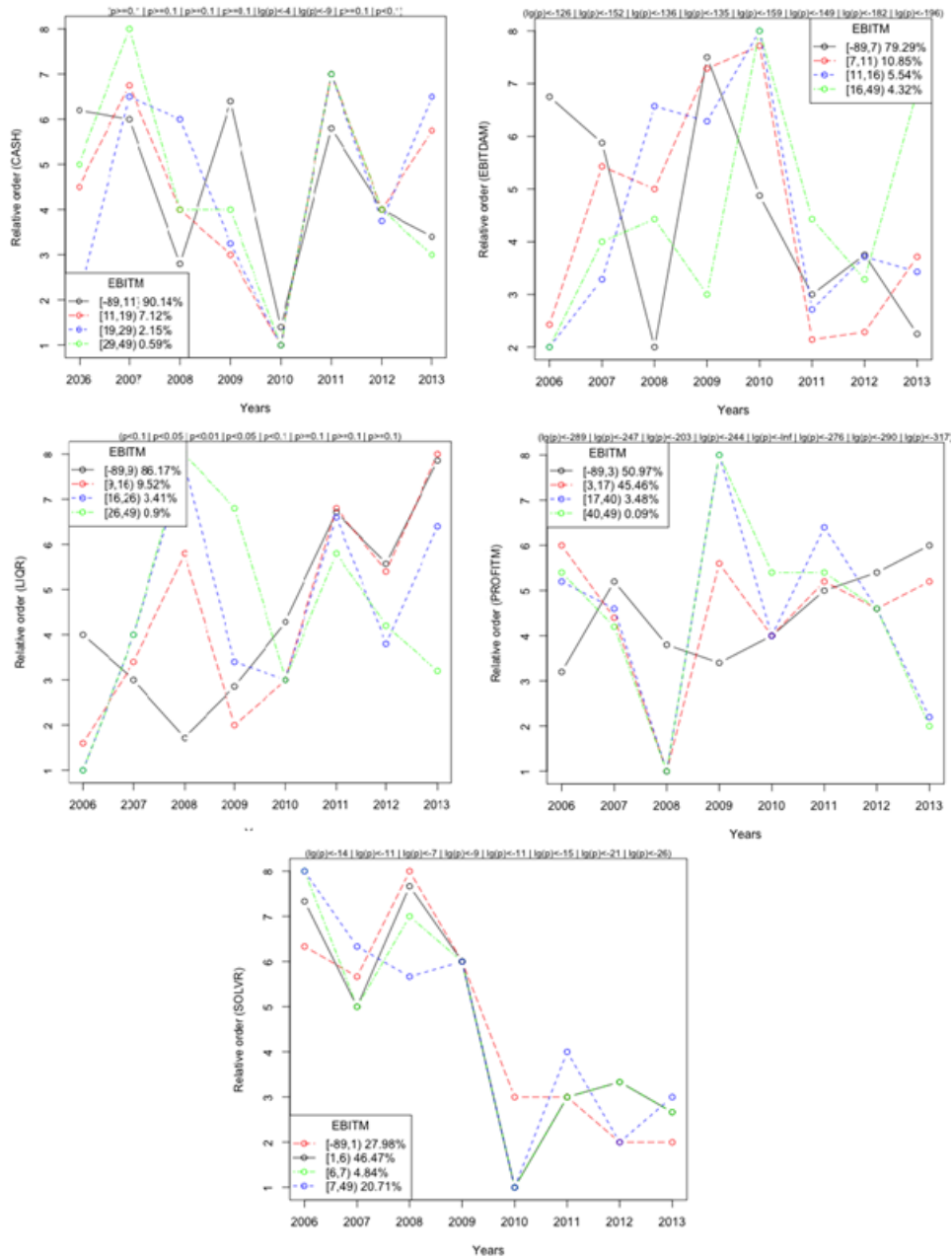


Figure 7: Relative order inputs: EBITM

In order to better understand the behavior of the enterprises and facilitate the description of the graphic, we adopted a color code. For the CAPITAL case (Figure 6), we represent the behavior of large enterprises by a red dashed line. Medium enterprises are represented by a blue line and small enterprises by a black line. Note that, this division does not correspond to the division of the European Commission (Section 3.2).

Interesting observations can be made from the relative order graphs. The most representative aspects of these variables during the study period are summarized as follows. In all the enterprises there was a decrease in the number of employees (NE) during the periods 2008-2009 and 2011-2013. However, small enterprises reflected an increase in their number of enterprises for the 2006-2010 period over other enterprises.

For all enterprises, SALES increased in 2007-2008 and decreased in 2009-2010. In the entire period, large enterprises had less CLIAB. Large enterprises acquired more LTDEBT in 2008 and 2011 than the other enterprises, while medium-sized enterprises had an increase in LTDEBT between 2007 and 2009. Between 2008 and 2010, small enterprises had higher TASSETS than other enterprises. Since EBITM indicates the ability of an enterprise to be profitable, it is important to study the behavior of the variables related to EBITM (Figure 7) during and after the crisis period. The enterprises increased the acquisition of debts during the period of 2007-2010. However, CASH decreased in 2010 for all enterprises. SOLVR decreased at the end of the 2010-2013 study period, and LIQR increased between 2010-2011 and 2012-2013.

5 Concluding Remarks

This work uses a nonparametric model that combines the Multidirectional Efficiency Analysis with other mathematical techniques to examine the performance of the Portuguese foods products, beverages and tobacco (FBT) sector in the 2006-2013 period. The study allowed the establishment of the differences between the efficiency patterns of the sector by considering different indicators: efficiency ratios, inefficiency indices, and NC-values between groups with different efficiency levels. The study also looks into a comparison between variables using model fitting.

First, we show that it is possible to establish significant characteristics among enterprises, when they are grouped into clusters, regarding employment and sales. Particularly, the year 2010 was a good year for the sector in overcoming the difficulties generated by the crisis as the percentage of total efficiency was above 60% in both clusters. All the enterprises show inefficiency of use of resources (input variables) between 0.62 and 0.93, for all years and clusters. The number of employees was the variable that was less used inefficiently by the two clusters.

At another stage of the analysis, two groups with different levels of efficiency are selected according to their MEA scores. The bigger difference between the groups is observed in the variables EBITDAM for the cluster 1 (with NC value 379.5) and CAPITAL for the cluster 2 (with NC value 520.1). The less efficient group increased the acquisition of debts (LTDEBT and CLIAB), during the entire study period. Although there is an increase in credit with the acquisition of long-term debts, there is no evidence that this implies the ability of enterprises to grow faster, which affects profitability. The liquidity of the less efficient group was notably less than the efficient group, especially in 2008 (2010) for cluster 1 (2). Regarding efficiency levels, it should be noted that, in most cases, large enterprises in Portugal were most affected by the crisis in terms of different variables. Large enterprises suffered more removals in the period 2008-2010. In contrast, small enterprises showed greater stability in several of the studied variables.

The study includes important indicators of the economic performance of the enterprises, such as profitability, liquidity, solvency, interest coverage and efficiency in getting a better characterization of the efficiency of the sector in economic terms. The study focuses on the period in which Portugal was directly affected by the world financial crisis of 2008/09. In this sense, it is expected that the consequences of the crisis on the FBT sector are reflected directly on the efficiency of the enterprises, especially in the indicators of financial performance. Although the enterprises increased the acquisition of debts during the 2007-2010 period to face the difficulties of the crisis, the results of the ratios are varied. Solvency decreased at the end of the 2010-2013 period, and liquidity increased between 2010 and 2011 and 2012 and 2013.

Given the impact the analysis mode may have on this type of study, using the most appropriate analysis model is vital. The performance model used in this study allows testing whether the enterprises under evaluation can reduce their inputs while maintaining their outputs at current levels. The results of this type of study are important in determining the failures that block the improvement of the quality of enterprises and establishing strategies to strengthen sector performance. Since the FBT sector is one sector that generates more employment in Portugal, an interesting future work would be to compare and benchmark the efficiency of this sector in countries of the European Union, establishing a comparative ranking and considering a more extended time period.

Acknowledgements

The authors were supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020. Murillo was also supported by national funds (OE), through FCT, I.P., in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law 57/2016, of August 29, changed by Law 57/2017, of July 19.

References

- Abdi, H., & Williams, L. J. (2010). Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–59. doi:10.1002/wics.101
- Asmild, M., Holvad, T., Hougaard, J. L., & Kronborg, D. (2009). Railway Reforms: Do They Influence Operating Efficiency? *Transportation*, 36(5), 617–38. doi:10.1007/s11116-009-9216-x
- Asmild, M., Paradi, J. C., & Kulkarni, A. (2006). Using Data Envelopment Analysis in Software Development Productivity Measurement. *Software Process Improvement and Practice*, 11(6), 561–72. doi:10.1002/spip.298
- Banker, R. D., Charnes, A., & Cooper, W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9), 1078–92. doi:10.1287/mnsc.30.9.1078
- Bergman, L., & Magnusson, D. (2001). Person-centered Research. In T. Cook & C. Ragin (Eds.), *Logic of Inquiry and Research Design* (Vol. 8 of the International Encyclopedia of the Social and Behavioral Sciences ed., pp. 11333–39). Oxford: Elsevier.

- Bhat, Z. U. H., Sultana, D., & Dar, Q. F. (2019). A Comprehensive Review of Data Envelopment Analysis (DEA) in Sports. *Journal of Sports Economics & Management*, 9(2), 82–109.
- Bogetoft, P., & Hougaard, J. L. (1999). Efficiency Evaluations Based on Potential (Non-proportional) Improvements. *Journal of Productivity Analysis*, 12(3), 233–47. doi:10.1023/A:1007848222681
- Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. Springer-Verlag New York. doi:10.1007/978-1-4419-7961-2
- Central Bank of Portugal. (2011). *Sectoral Analysis of Manufacture of Food Products* (Central Balance-Sheet Studies).
- Chen, X., Chen, C., & Jin, L. (2011). Principal Component Analysis in Anthropological Genetics. *Advances in Anthropology*, 1(2), 9–14. doi:10.4236/aa.2011.12002
- Dray, S. (2008). On the Number of Principal Components: A Test of Dimensionality Based on Measurements of Similarity Between Matrices. *Computational Statistics & Data Analysis*, 52(4), 2228–37. doi:10.1016/j.csda.2007.07.015
- Escoufier, Y. (1973). Le Traitement Des Variables Vectorielles. *Biometrics*, 29(4), 751–60.
- EU-MERCI. (2016). *Analysis of Food and Beverage sector in Different Countries* (Horizon 2020 Project Nr. 693845).
- Ferré, L. (1995). Selection of Components in Principal Component Analysis: A Comparison of Methods. *Computational Statistics & Data Analysis*, 19(6), 669–82. doi:10.1016/0167-9473(94)00020-J
- Figueiredo Filho, D. B. & Silva Junior, J. A. da. (2010). Visão Além do Alcance: Uma Introdução à Análise Fatorial. *Opinião Pública*, 16(06), 160 - 185. doi:10.1590/S0104-62762010000100007
- Gongbing, B., Pingchun, W., Feng, Y., & Liang, L. (2014). Energy and Environmental Efficiency of China's Transportation Sector: A Multidirectional Analysis Approach. *Mathematical Problems in Engineering*, 2014, 1–12. doi:10.1155/2014/539596
- Good, I. (1969). Some Applications of the Singular Decomposition of a Matrix. *Technometrics*, 11(4), 823–31. doi:10.1080/00401706.1969.10490741
- Gordon, A. D. (1981). *Classification*. Chapman and Hall, London.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. (2006). *Multivariate Data Analysis*. Pearson International Edition, New Jersey.
- Hauke, J., & Kossowski, T. (2011). Comparison of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae*, 30(2), 87–93. doi:10.2478/v10117-011-0021-1
- Hirschberg, J. G., & Lye, J. N. (2001). *Clustering in a Data Envelopment Analysis Using Bootstrapped Efficiency Scores* (Department of Economics - Working Papers Series No. 800). The University of Melbourne.
- Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *Journal Educational Psych*, 24(6), 417–41. doi:10.1037/h0071325

- Inman, H. F., & Bradley, E. L. (1989). 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*, 18(10), 3851–74. doi:10.1080/03610928908830127
- Interreg Central Europe. (2017). *Food Sector Related Knowledge Integration: Food Sector Global Market Trend Analysis* (I-CON project).
- Jackson, D. (1993). Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecology*, 74(8), 2204–14. doi:10.2307/1939574
- Jolliffe, I. (2002). *Principal Component Analysis*. Springer, Berlin.
- Kaffash, S., & Marra, M. (2017). Data Envelopment Analysis in Financial Services: A Citations Network Analysis of Banks, Insurance Companies and Money Market Funds. *Annals of Operations Research*, 253(1), 307–44. doi:10.1007/s10479-016-2294-1
- Karun, K., & Isaac, E. (2013). Cogitative Analysis on k-means Clustering Algorithm and its Variants. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(4), 1875–80.
- Kaufman, L., & Rousseeuw, P. J. (1987). Clustering by Means of Medoids. In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L_1 -Norm and Related Methods* (pp. 405–16). North-Holland.
- Kim, J. H., Choi, J. H., Yoo, K. H., Loh, W., & Nasridinov, A. (2019). A Fast Algorithm for Identifying Density-Based Clustering Structures Using a Constraint Graph. *Electronics*, 8(10), 1–23. doi:10.3390/electronics8101094
- Machado, D. M. (2017). *Portugal Food Processing Sector* (Gain Report, PT1102). Global Agricultural Information Network.
- Morey, L. C., Blashfield, R. K., & Skinner, H. A. (1983). A Comparison of Cluster Analysis Techniques Within a Sequential Validation Framework. *Multivariate Behavioral Research*, 18(3), 309–29. doi:10.1207/s15327906mbr1803_4
- Murillo, K. P., & Rocha, E. M. (2018). The Portuguese Manufacturing Sector During 2013–2016 After the Troika Austerity Measures. *World Journal of Applied Economics*, 4(1), 21–38. doi:10.22440/wjae.4.1.2
- Murillo, K. P., Rocha, E. M., & Pardo, C. I. (2018). Energy Production and CO₂ Emission Efficiency of Eight European Countries in the Manufacturing Area. *Journal of Management and Information Technology*, 13(1), 1–17. doi:10.24297/ijmit.v13i1.7427
- Murillo, K. P., Rocha, E. M., & Ramalho, J. S. (2018). About the Efficiency Behavior of the Portuguese Manufacturing Firms During the Financial Crisis. *Libertas Mathematica (new series)*, 38(1), 1–27.
- Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–72. doi:10.1080/14786440109462720
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How Many Principal Components? Stopping Rules for Determining the Number of Non-trivial Axes Revisited. *Computational Statistics & Data Analysis*, 49(4), 974–97. doi:10.1016/j.csda.2004.06.015
- Ramalho, J. J., & da Silva, J. V. (2009). A Two-part Fractional Regression Model for the Financial Leverage Decisions of Micro, Small, Medium and Large Firms. *Quantitative Finance*, 9(5), 621–36. doi:10.1080/14697680802448777

Walesiak, M. (1993). *Multivariate Statistical Analysis in Marketing Research* (Research Papers No. 654). Wroclaw University of Economics.

Wang, K., Wei, Y. M., & Zhang, X. (2013). Energy and Emissions Efficiency Patterns of Chinese Regions: A Multi-directional Efficiency Analysis. *Applied Energy*, 104, 105–116. doi:10.1016/j.apenergy.2012.11.039

Wen, H., Lim, B., & Lisa Huang, H. (2003). Measuring E-commerce Efficiency: A Data Envelopment Analysis (DEA) Approach. *Industrial Management & Data Systems*, 103(9), 703–10. doi:10.1108/02635570310506124

Appendices

Appendix A: Technical Details

In this section, some complementary details of the MEA model (Section 3.1) are presented.

Suppose any given tuple $n \in N$ under the conditions described in Section 3.1. Consider the first D inputs of the inputs set I , ($1 < D \leq I$) being discretionary (that means inputs that have a different value for each tuple and therefore inputs that into the optimization process, by calculating the MEA score), which will be represented by the indices d such that $1 < d < I$. Therefore $i \in [D]$ refers to the discretionary inputs, while $i \in [I] \setminus d$ refers to the non-discretionary inputs.

Regarding the variable returns to scale (VRS) (Banker et al., 1984), we define the set

$$\Lambda^N = \{\lambda \in R^N : \sum_{n=1}^N \lambda_n = 1\} \tag{A.1}$$

The MEA score for a specific observation $z(\bar{n}) = (x(\bar{n}), y(\bar{n}))$ with $\bar{n} \in N$ is found by solving the three optimization problems defined as

$$\begin{array}{ll}
 P_d^\alpha(z, \bar{n}) : & P_j^\beta(z, \bar{n}) : \\
 \min \alpha_d(\bar{n}) \text{ such that} & \max \beta_j(\bar{n}) \text{ such that} \\
 \sum_n \lambda_n x_d(n) \leq \alpha_d(\bar{n}) & \sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), i \in [I] \\
 \sum_n \lambda_n x_i(n) \leq x_i(\bar{n}), i \in [I] \setminus d & \sum_n \lambda_n y_j(n) \geq \beta_j(\bar{n}), j \in [J] \\
 \sum_n \lambda_n y_l(n) \geq y_l(\bar{n}), l \in [J] & \sum_n \lambda_n y_l(n) \geq y_l(\bar{n}), l \in [J] \setminus j
 \end{array} \tag{A.2} \tag{A.3}$$

$$\begin{aligned}
 & P^\gamma(\alpha, \beta, z, \bar{n}) : \\
 & \max \gamma(\bar{n}) \text{ such that} \\
 & \sum_n \lambda_n x_i(n) \leq x_i(\bar{n}) - \gamma(\bar{n})(x_i(\bar{n}) - \alpha_i(\bar{n})), i \in [M] \\
 & \sum_n \lambda_n y_j(n) \leq x_i(\bar{n}), i \in [I] \setminus -m \\
 & \sum_n \lambda_n y_l(n) \geq y_l(\bar{n}) + \gamma(\bar{n})(\beta_l^*(\bar{n}) - y_l(\bar{n})), l \in [J]
 \end{aligned} \tag{A.4}$$

where $\lambda \in \wedge^N$, $\alpha_i(n')$ and $\beta_i(n')$ are the optimal solutions to the problems $P_m^\alpha(z, \bar{n})$ and $P_j^\beta(z, \bar{n})$, respectively.

Note that, for the input $x_i(n), i \in [I]$ the contribution in $Z = \{-z(n)\}N$ is given by

$$\text{meff}_i(n) = \frac{x_i(n) - \gamma(n)(x_i(n) - \alpha_i(n))}{x_i(n)} \chi_{[D]}(i) \tag{A.5}$$

where $\chi_{[D]}(i) = 1$ if $i \in [D]$; otherwise, $\chi_{[D]}(i) = 0$.

In a similar way, the contribution for the outputs $j \in [J]$ is given by

$$\text{meff}_j(n) = \frac{y_j(n)}{y_j(n) + \gamma(n)(\beta_j(n) - y_j(n))} \tag{A.6}$$

Therefore, by the directional contributions $\text{meff}_i(n)$ and $\text{meff}_j(n)$, the definition of the MEA score (equation A.1), can be obtained. For further details, see [Murillo, Rocha, & Ramalho \(2018\)](#).

Appendix B: Correlation Coefficients

Table B.1: Correlation Matrix Pearson (with $p \leq 0.05$)

	NE	TASS.	CAPIT	LTDEBT	CLIAB	CASH	PROF.	LIQR	SOLVR	EBITM	SALES	CASHF.	EBITD.
NE	1	0.74	0.75	0.54	0.6	0.39	0.06	NA	0.07	0.06	0.73	0.65	NA
TASS.	0.74	1	0.77	0.41	0.94	0.4	0.07	NA	NA	0.07	0.76	0.85	0.06
CAPIT.	0.75	0.77	1	0.6	0.65	0.26	NA	NA	0.05	NA	0.63	0.58	NA
LTDEBT	0.54	0.41	0.6	1	0.24	0.06	NA	NA	NA	NA	0.36	NA	NA
CLIAB	0.6	0.94	0.65	0.24	1	0.2	0.05	NA	NA	0.06	0.68	0.8	0.04
CASH	0.39	0.4	0.26	0.06	0.2	1	0.08	NA	NA	0.05	0.39	0.52	NA
PROF.	0.06	0.07	NA	NA	0.05	0.08	1	NA	0.31	0.95	0.05	NA	0.74
LIQR	NA	NA	NA	NA	NA	NA	NA	1	0.27	NA	NA	NA	0.04
SOLVR	0.07	NA	0.05	NA	NA	NA	0.31	0.27	1	0.26	NA	NA	NA
EBITM	0.06	0.07	NA	NA	0.06	0.05	0.95	NA	0.26	1	0.05	NA	0.82
SALES	0.73	0.76	0.63	0.36	0.68	0.39	0.05	NA	NA	0.05	1	0.71	NA
CASHF.	0.65	0.85	0.58	NA	0.8	0.52	NA	NA	NA	NA	0.71	1	0.07
EBITD.	NA	0.06	NA	NA	0.04	NA	0.74	0.04	NA	0.82	NA	0.07	1

Note: All the correlations shown have $p \leq 0.05$ indicating a strong correlation. When such is not the case, we replaced the correlation value by “NA” (not applicable) indicating that it is insignificant at 5%.

Table B.2: Correlation Matrix Spearman, Pearson (with $p \leq 0.05$)

	NE	TASS.	CAPIT	LTDEBT	CLIAB	CASH	PROF.	LIQR	SOLVR	EBITM	SALES	CASHF.	EBITD.
NE	1	0.67	0.57	0.32	0.63	0.42	-0.05	NA	NA	NA	0.82	0.55	NA
TASS.	0.67	1	0.73	0.5	0.9	0.48	-0.06	NA	NA	0.04	0.86	0.69	0.07
CAPIT.	0.57	0.73	1	0.38	0.64	0.38	NA	7	NA	NA	0.69	0.5	-0.06
LTDEBT	0.32	0.5	0.38	1	0.36	NA	NA	0.07	-0.2	NA	0.4	0.341	NA
CLIAB	0.63	0.9	0.64	0.36	1	0.37	NA	-0.28	-0.23	NA	0.8	0.63	NA
CASH	0.42	0.48	0.38	NA	0.37	1	NA	0.38	0.29	NA	0.51	0.44	0.08
PROF.	-0.05	-0.06	NA	NA	NA	NA	1	0.26	0.33	0.91	NA	0.36	0.67
LIQR	NA	NA	0.07	0.07	-0.28	0.38	0.26	1	0.58	0.2	0.04	NA	NA
SOLVR	NA	NA	NA	-0.2	-0.23	0.29	0.33	0.58	1	0.23	NA	NA	NA
EBITM	NA	0.04	NA	NA	NA	NA	0.91	0.2	0.23	1	0.06	0.434	0.79
SALES	0.82	0.86	0.69	0.4	0.8	0.51	NA	0.04	NA	0.06	1	0.717	-0.04
CASHF.	0.55	0.69	0.5	0.34	0.63	0.44	0.36	NA	NA	0.43	0.71	1	0.47
EBITD.	NA	0.07	-0.06	NA	NA	0.08	0.67	NA	NA	0.79	-0.04	0.479	1

Note: See the previous table's note.