LUISS

Dipartimento
di Impresa e Management

Cattedra di Data Analysis for Business

Data Envelopment Analysis for different measures of efficiency in the retail industry

Prof. Iafrate Francesco
Relatore

Matricola n. 243811
Candidato

Anno accademico 2021/2022

"Efficiency is thus not a goal in itself. It is not something we want for its own sake, but rather because it helps us attain more of the things we value" (Stone 2012)

Index

1. Introduction

The principles of economic efficiency are based on the concept that resources are scarce. There are not sufficient resources to always ensure that all aspects of an economy function at their highest capacity, they must be distributed to meet the needs of the economy in an ideal way while also limiting the amount of waste produced. In more mathematical or scientific terms, economic efficiency signifies the level of performance that uses the least amount of inputs to achieve the highest amount of output. It often specifically comprises the capability of a specific application of effort to produce a specific outcome with a minimum amount or quantity of waste, expense, or unnecessary effort. Similarly, economists look at the amount of loss, referred to as waste, between pure efficiency and reality to see how efficiently an economy functions. Some terms that encompass phases of economic efficiency include productive efficiency, allocative efficiency, distributive efficiency, and Pareto efficiency. Productive efficiency of an industry requires that all firms operate using best-practice technological and managerial processes and that there is no further reallocation that bring more output with the same inputs and the same production technology. By improving these processes, an economy or business can extend its production possibility frontier outward, so that efficient production yields more output than previously. Its opposite can occur because the productive inputs physical capital and labor are underutilized, that is, some capital or labor is left sitting idle, or because these inputs are allocated in inappropriate combinations to the different industries that use them. Furthermore, a market can be said to have allocative efficiency if the price of a product that the market is supplying is equal to the marginal value consumers place on it and equals marginal cost. In other words, when every good or service is produced up to the point where one more unit provides a marginal benefit to consumers less than the marginal cost of producing it. Because productive resources are scarce, the resources must be allocated to various industries in just the right amounts, otherwise too much or too little output gets produced. At peak economic efficiency (when the economy is at productive and allocative efficiency), the welfare of one cannot be improved without subsequently lowering the welfare of another. This point is called Pareto efficiency.

Although a state of economic efficiency is essentially theoretical, a limit that can be approached but never reached, economists have long studied the efficiency of firms, industries, and entire economies. Measuring efficiency and identifying the sources of potential inefficiency are very important steps in improving the competitive position of the enterprises in their continuous development, sustainability, overall behavior in the current corporate environment. As an increasing number of people engaged in corporate performance evaluation from the middle of 1990s, it became a new management discipline. Indeed, performance evaluation, when applied properly, provides opportunity for management to find out which corporate activity ensures more revenue than cost (Neely, 2004). Organizations in particular use benchmarking in management to compare their processes to the best practices of others in a peer group of companies in the same industry or sector. The identification of the top businesses allows for the formulation of objectives in the best practice benchmarking process, allowing these organizations to learn from others and make strategies to improve some elements of their own performance. Performance evaluation helps businesses unleash creativity, improve reputation, and increase competitiveness, but also investors, especially private equity investors to measure the added value of their non-financial service. Within this new discipline, the contribution that data analytics brought in this decade was substantial.

Data analytics has become crucial in helping businesses optimize their performances. Taking it into account in their business models means that companies can strategically reduce costs by identifying more efficient ways of doing business. A company can also use data analytics to make better business decisions and help analyze customer trends and satisfaction, which can lead to new, and better, products and services. For instance, studies of benchmarking practices with Data Envelopment Analysis, a non-parametric linear programming technic that is finding more and more applications in the business analysis world, have identified numerous sources of inefficiency in some of the most profitable firms, firms that had served as benchmarks by reference to this (profitability) criterion, and this has provided a vehicle for identifying better benchmarks in many applied studies. (Seiford and Zhu 2011) Also clustering techniques, which are based upon some measure of resemblance or a measure of proximity, have heavily contributed to benchmarking. In fact, in cluster analysis, the problem is about finding the groupings whose populations are not known in advance, discovering "natural" clusters of the items based upon some internal criterion. Clustering problems cut across various disciplines and seem to be receiving greater attention. Entities to be grouped together may be traits, objects, persons, plants, groups, institutions, structures, fields, stars, cities, companies, variables, etc. Considerable attention is given to this matter in the behavioural and life sciences as it helps displaying hidden relations.

Despite the fact that the techniques to be described here may find useful application in other fields as well, in this work the fashion retail was used as an illustrative field of application. It is a type of retailing that includes selling clothing, apparel, and accessories. It includes groups of companies which part of the fashion supply chain that goes from the manufacturers to the consumer, offering fashion goods and services, through traditional seasonal spans and/or fast fashion timing, ranging from budget to designer price lines. To cater to the large, 3,000 billion, textile and garment industry new companies are entering the market daily. (Global Fashion industry statistics, 2022) With such a dynamic environment, it is worth to keep up with the changes of the last decades and more in general check the fashion chains health. The information used in the analysis is based on real past results that are released by the companies.

The objective of this work is to compare the more traditional way to assess firms' performances and efficiency from their accounting records with the results of Data Envelopment Analysis as well as with clustering labelling and it is organized as follows: in Section 1.2 methodology; Section 1.3 dataset; Section 2

Data Envelopment Analysis; Section 3 Clustering; Section 4 shows the modelling of the problem in R and results are presented and discussed. Lastly, the conclusions and are described in section 5.

1.1 Methodology

The most common way to describe the financial situation of a firm is Ratio Analysis. Ratio is a fraction whose numerator is the antecedent and denominator the consequent. It is simply an expression of one number in terms of another, but it allows stakeholders to make better sense of the accounts and deeper understand the current fiscal scenario of an entity. There are five categories of ratios used in financial statement analysis:

1. Liquidity ratios, which measure a firm's ability to meet cash needs as they arise;
2. Activity ratios, which measure the liquidity of specific assets and the efficiency of managing assets.
3. Leverage ratios, which measure the extent of a firm's financing with debt relative to equity and its ability to cover interest and other fixed charges;
4. Profitability ratios, which measure the overall performance of a firm and its efficiency in managing assets, liabilities, and equity;
5. Market value ratios, which bring in the stock price and give an idea of what investors, think about the firm and its future prospects.

Specific fnancial indicators can be attributed to the fundamental characteristics of activity, reflecting aspects of the effectiveness of companies and the risk of their financial strategies. For example, inventory levels for retail businesses are usually substantially higher than those for a manufacturing company. Costs for labor are likely to be higher than in retail. As a result of these differences, financial ratios tend to vary in importance among types of businesses and industries. Therefore, identifying the important ones for retail businesses involves analyzing which areas of performance are critical to success. As the Retail Owner's Institute reports, key retail industry average rations focus on aspects of activity, liquidity, and profitability. (Key Ratios Benchmarks, 2022) Indicators characterizing future cash flows or human capital were deliberately rejected for the reasons of data availability from the companies' public corporate reports; instead, the following relations will be used throughout this work:

Measures	Type	Description	Formulae
Day Sales Outstanding (DSO)	Activity	It is a measure of the average number of days that it takes a	$DSO = \frac{Trade\ receivables}{Sales} * 365$

		company to collect payment for a sale.	
Day Payable Outstanding (DPO)	Activity	It indicates the average time (in days) that a company takes to pay its bills and invoices to its trade creditors, which may include suppliers, vendors, or financiers.	$DPO = \frac{Trading\ Payables}{Cost\ of\ Goods\ for\ Sale} * 365$
Day Inventory Outstanding (DIO)	Activity	It is a financial ratio that indicates the average time in days that a company takes to turn its inventory, including goods that are a work in progress, into sales.	$DIO = \frac{Total\ inventories}{Cost\ of\ Goods\ for\ Sale} * 365$
Working Capital Conversion	Activity	It describes how many days it takes for a company to convert its working capital into revenue.	$WCR = Current\ Asset - Current\ Liabilities$ $WCRdays = \frac{WCR}{Sales} * 365$
EBITDA[1] to sales	Profitability	It is a financial estimator that compares gross revenue to earnings to determine a company's profitability. This metric represents the percentage of a company's earnings that remains after operating expenses.	$EBITDA\ to\ sales = \frac{EBITDA}{Sales}$
Return on sales (ROS)	Profitability	Return on sales (ROS) is a ratio used to evaluate a company's operational efficiency. This measure provides insight into how much profit is being produced per dollar of sales.	$ROS = \frac{EBIT^{2}}{Sales}$
Net Profit margin	Profitability	Profit margin represents net profit as a percentage of the revenue.	$Net\ profit\ margin = \frac{Net\ income}{Sales}$

[1] EBITDA is an abbreviation for "Earnings Before Interest, Taxes, Depreciation, And Amortization." Thus, it is calculated adding back these line items to net income, and so does include operating expenses such as the cost of goods sold (COGS) and selling, general, and administrative (SG&A) expenses.

[2] Similarly, EBIT stands for "Earning Before Interests and Taxes" and it is equal to EBITDA deducted from its depretiation and amortization components.

Return on Assets	Profitabilty	This financial ratio indicates how profitable a company is in relation to its total assets. Interest expenses are added back to the net income because this index is supposed to be independent from the financial structure of companies.	$ROA = \frac{Net\ income}{Equity}$
Return on Equity	Profitability	ROE is a metric of how well the company utilizes its equity to generate profits. It measures how many dollars of profit are generated for each dollar of shareholder's equity.	$ROE = \frac{Net\ income}{Equity}$
Sustainable Growth rate (SGR)	Liquidity	SGR is the rate at which a company can grow without having to borrow money to fund its growth.	$SGR = ROE * (1 - \frac{Retained\ Income}{Revenues})$
Quick Ratio	Liquidity	Quick ratio measures the ability of a company to use its near cash or quick assets to extinguish or retire its current liabilities immediately. It is defined as the ratio between quickly available or liquid assets and current liabilities.	$Quick\ ratio = \frac{Current\ asset - Inv - Prep\ expences}{Current\ Liabilities}$
Leverage Risk	Liquidity	This leverage ratio looks at how much capital comes in the form of debt (loans) or assesses the ability of a company to meet its financial obligations.	$Leverage = \frac{Debt}{Equity}$

Table 1.1 Financial ratios with their descriptions and formulas

Although this is far from being a comprehensive list, it should provide a good starting point for companies that wish to align managers around a common set of performance indicators. Certains highlight the degree of efficiency of a company in the management of its assets and other resources, others measures the ability to generate profits. It is important that assets and financial resources be allocated and used efficiently to avoid unnecessary expenses. For instance, there are several types of profit margin. The EBITDA/sales ratio can focus on the impact of direct operating costs while excluding the effects of the company's capital structure, tax exposure, and accounting quirks. EBITDA provides deeper insight into the operational efficiency of an organization based on only those costs management can control, so as return on sales (ROS). In everyday use, however, it usually refers to net profit margin, a company's bottom line after all other expenses, including taxes and one-off oddities, have been taken out of revenue.Comparing profits to revenue is a useful operational metric, but comparing them to the resources a company used to earn them displays the feasibility of that company's existence. Return on assets (ROA) is the simplest of such corporate bang-for-the-buck measures. Corporate management, analysts, and investors can use it to determine how efficiently a company employs its assets to generate a profit and it enables them to make necessary decisions about under-performing assets.

When employed correctly, ratio analysis throws light on many problems of the firm and also highlights some positives. Ratios are essentially whistleblowers, they draw the managements attention towards issues needing attention, however they have limitations. One of the biggest problems of financial indicators is the dimensional evaluation, as they might not show a proper picture on corporate performance to the management and shareholders (Abdoli et al., 2011). They mainly refer to the past and they are not able to point out more complex patterns. Therefore, ratio analysis metrics do not fully explain company performances. Meanwhile, DEA can complete the traditional financial ratio analysis, especially if the aim is to gain more information on operational and technical efficiency. The usefulness of DEA lies in its ability to estimate efficiency when multiple inputs are used to produce multiple outputs, without the need to specify distributions or functional forms. DEA is a fully nonparametric estimation method, meaning that is very flexible and can potentially be used to describe a wide variety of situations. The advantage of this type of analysis is that the aspects of financial performance are studied not in a sequential, but in a simultaneous way. As a comparison, some clustering techniques will be implemented to find some correspondences, analyzing the links between companies' clusters, DEA's results, and financial statement items.

1.2 Dataset

The dataset used comes from the Wharton Research Data Service. WRDS provides business intelligence, data analytics, and research platform to global institutions about Accounting, Banking, Economics, ESG, Finance, Healthcare, Insurance, Marketing, and Statistics, and enables historical analysis and insight into the latest innovations in academic research. It allows access to all of its data and includes support for many programming languages. Among its numerous sources there is Standard and Poor's, which releases the most comprehensive market and corporate financial databases. In particular, Compustat Fundamentals provides standardized North American and global financial statement and market data for over 80,000 active and inactive publicly traded companies that financial professionals have relied on for over 50 years. Compustat Global (comp.g_funda) and Compustat North America (comp.funda) are two databases of fundamental and market information on active and inactive publicly held companies around the world. As financial information for a business relies on the three primary financial statements, such data bank provides more than 300 annual Income Statement, Balance Sheet, Statement of Cash Flows, and supplemental data items. For most companies, annual history is available back to 1950 and quarterly history back to 1962 with monthly market history back to 1962. For the purpose of this work, it was used comp.g_funda.

The dataset considered starts with a sample of 106 firms distributed worldwide, represented through 443 variables and over a timespan of 11 years (from 2010 to 2021, for a total of more than 900 observations). Apart from companies like Hennes & Mauritz AB (H&M) or Only Corp, most of these companies remain rather unknown to the consumer audience, better known are the individual brands that are part of these global fashion companies. These firms have been chosen because they act in a constantly changing market. Even though some international retailers are of great importance, they would never dominate. This is because they generally consist of many chains and new banners/brands keep appearing very regularly. They are all branches of the fashion retail industry, in particular Clothing Store, Shoe Stores, and Jewelry, Luggage, and Leather Goods Stores. The footwear and clothing industries are similar in structure and share many of the characteristics of production and trade. Most of the countries that have emerged as successful producers and exporters of garments have also become important in footwear.

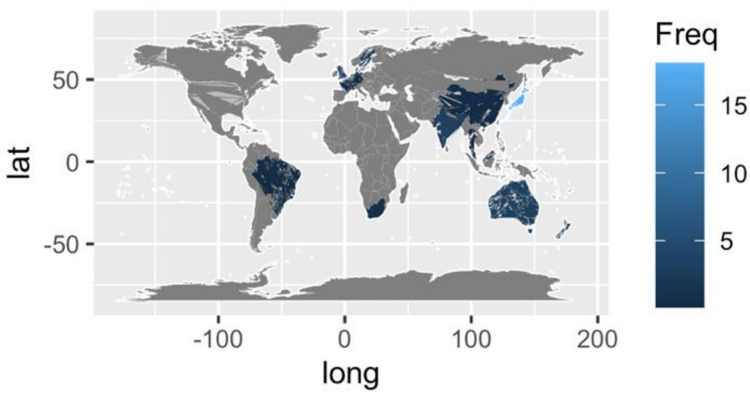
NAICS[3]	Titles	Revelant Markets
4481	**Clothing Stores**	This industry group comprises establishments primarily engaged in retailing new clothing.
448110	Men's Clothing Stores	
448120	Women's Clothing Stores	
448130	Children's and Infants' Clothing Stores	
448140	Family Clothing Stores	

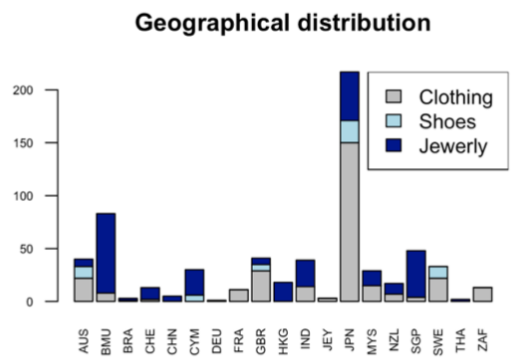
[3] The acronym NAICS is an abbreviation of the North American Industry Classification System. This system is the standard used by federal statistical agencies for classifying businesses for the purpose of collecting, analyzing, and publishing data related to the U.S. economy. NAICS industries are identified by a 6-digit code, the code accommodates all the sectors and allows flexibility in designating subsectors. The international NAICS agreement fixes only the first five digits of the code. The sixth digit, where used, identifies subdivisions of NAICS industries that accommodate user needs in individual countries. (SIX DIGIT NAICS CODES & TITLES n.d.)

448150	Clothing Accessories Stores	
448190	Other Clothing Stores	
4482	**Shoe Stores**	This industry comprises establishments primarily engaged in retailing all types of new footwear (except hosiery and specialty sports footwear, such as golf shoes, bowling shoes, and spiked shoes). Establishments primarily engaged in retailing new tennis shoes or sneakers are included in this industry.
448210	Shoe Stores	
4483	**Jewelry, Luggage, and Leather Goods Stores**	This industry group comprises establishments primarily engaged in retailing new jewelry (except costume jewelry); new sterling and plated silverware; new watches and clocks; and new luggage with or without a general line of new leather goods and accessories, such as hats, gloves, handbags, ties, and belts.
448310	Jewelry Stores	
448320	Luggage and Leather Goods Stores	

Table 1.2 The table shows the industry branches taken into consideration by their NAICS codes (SIX DIGIT NAICS CODES & TITLES n.d.)

The choice of a dataset with firms spread over the world is not a casualty. The geographical distribution of production in the clothing, and footwear industries has changed dramatically in the past 25 years resulting in sizeable employment losses in Europe and North America and important gains in Asia and other parts of the developing world. At present, more than 60 percent of world clothing exports are manufactured in developing countries. Asia is the major world supplier today, producing more than 32 percent of the world's clothing exports. (International Labour Organization 1996) As in the clothing and textile industries, footwear production has shifted largely to developing countries capable of producing large shares of the world's supply at far less cost. (International Labour Organization 1996)





Figures 1.1 and 1.2 Geographical distribution of firms displayed with a frequency map (on the left) and its relative barplot (on the right)

2. Data Envelopment Analysis

Data Envelopment Analysis is mathematical programming technique for evaluating the relative efficiency of a set of homogeneous entities or Decision Making Units (DMUs) which consume the same inputs (in different quantities) to produce the same outputs (in different quantities). Because it requires very few assumptions, DEA has also opened possibilities for use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in DMUs. Typically, the efficiency scores obtained by executing a DEA model allow us to classify the DMUs into two groups: efficient DMUs, which define the best practice frontier, and inefficient DMUs. Along with the measure, DEA also yields targets for performance, any gains realizable through changes in scale size and/or mix of resources used, identification of best practice and benchmark units.

Some 30 years ago Data Envelopment Analysis (DEA) (and frontier techniques in general) set out to overcome the problem that for actual firms one can never observe all the possible input-output combinations. The pivotal study "Measuring the efficiency of decision-making units" by Charnes, Cooper, and Rhodes (1978) described DEA as a ‘mathematical programming model applied to observational data that provides a new way of obtaining empirical estimates of relations - such as the production functions and/or efficient production possibility surfaces – that are cornerstones of modern economics. They built their CCR model on the concepts of Farrell (1957), who uses linear programming to estimate an empirical production technology frontier for the first time and set the ground for the actual implementations. At the time, the model relied on the economic concept of constant returns to scale, defined as inputs and output growing at a constant rate. After a few years, Banker et al. (1984) relaxed it by including the so-called convexity restriction. Thus, the resulting model, the BCC (or Banker, Charnes and Cooper’s DEA model), allowed the efficient frontier to exhibit variable returns to scale.

In the last two decades, there have been remarkable advances in both DEA methodologies and practical applications in many different fields (education, banking and finance, sustainability, arts and humanities, hospitals and healthcare, industrial sectors, agriculture, transportation, etc.). In fact, based on these basic DEA models, several extensions have been proposed in the DEA literature. For example, and without being exhaustive, the possibility of considering non-discretionary (or uncontrollable) inputs and/or outputs, the presence of categorical or ordinal inputs and/or outputs, imposing restrictions on the weights of inputs and/or outputs or taking into account the presence of undesirable factors. At the same time as these variations of the basic radial DEA models were emerging, methodological developments have also led to a wide variety of new DEA models. One of these first new DEA models is the additive model, which simultaneously allows input reductions and output increases. (Benitez, Coll-Serrano and Bolós 2021)

2.1 Scale assumption

The constant returns translate the firms' wish to see that their investments are generating consistent flows, but it might not be realistic because of mismanagement or external factors as imperfect competition or any types of economic, financial, or regulatory restrictions that makes unit not operating at an optimum scale. DEA is different depending on the model supporting scale assumptions. Generally two scale assumptions are applied: constant return to scale (CRS) and variable return to scale (VRS). The latter one includes both increasing and decreasing return to scale. CRS assumes that the output changes with the same ratio as the input, while VRS assumes that the return to scale can be increasing, constant or decreasing. Regarding return to scale the following options are possible in terms of efficiency:

- Changes occurred either in the input or in the output results a directly proportional change in the other. It is the constant return to scale, abridged CRS.
- Changes occurred in the input results in the larger scale of increase in the output. It is the increasing return to scale, abridged IRS.
- The increase of the input could also lead to proportionally lower increase of the output. It is the so called decreasing returns to scale, abridged DRS (Bogetoft and Otto, 2011).

The return to scale characteristics of an organization could depend on the nature of the industry, the size of the company, the way of operations and several other factors, which can limit the efficiency seeking strategies. For instance CRS assumption can only be used if the company's size is optimal and there is no perfect competition, there are no delivering, labour or financial, etc. limits. If the limits are existing, then applying VRS model scale efficiency and disturbing measurement problems can be avoided, which otherwise would lead to growth. Consequently, VRS model is the most popular type. Using CRS could be considered a bad choice for most of the companies, at the same time this model shows the efficiency the best and this indicator is present in the numerator of the scale efficiency as well.

2.2 BCC Model

Data envelopment analysis efficiency estimates are usually computed by solving a linear programming problem, i.e. a problem of maximizing or minimizing a linear function subject to system of linear constraints. The constraints may be equalities or inequalities. The linear function is called the objective function , of the form $f(x, y) = ax + by + c$. The solution set of the system of inequalities is the set of possible or feasible solution , which are of the form (x, y) and represents the problem's frontier (empirical production function, empirical production envelope and envelopment surface are all terms which are analogous to efficient frontier). If a linear programming problem can be optimized, an optimal value will occur at one of the vertices of the region representing the set of feasible solutions. DEA can be considered a quite flexible linear programming approach as it does not require the definition of an objective function that is valid for everyone and leaves to each decision-making unit the possibility of weighting the inputs and outputs in order to maximize its efficiency index.

The input-oriented model measures the ineffectiveness of the evaluated DMU from the perspective of input. It focuses on the degree to which the technical effectiveinputs should be reduced without reducing output. The output-oriented model measures the ineffectiveness of the evaluated DMU from the perspective of output. It focuses on the degree to which the technical effective outputs should be reducedwithout increasing input The extent of the increase. (Lai, Hongbo; Shi, Hao; Zhou, Yang, 2020). The two fundamental models are CCR model and BCC model which provide radial efficiency measures and can be either input- or output-oriented. The former is based on constant return scale (CRS) while the latter is based on variable return scale (VRS).

Assuming that there are n DMUs to be evaluated, each DMU consumes varying amounts of m different inputs to produce s different outputs. Specifically, DMU_j (possibly) consumes amount x_{ij} of input i and (possibly) produces amount y_{rj} of output r. We assume that $x_{ij} \geq 0$ and $y_{rj} \geq 0$ and further assume that each DMU has at least one positive input and one positive output value. If the constraint $\sum_{j=1}^{n} \lambda_j = 1$ is adjoined, the CCR turns into a BCC models. Here model with an output oriented objective:

$$\max\ \phi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$$

subject to

$$\begin{aligned} &\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{io} && i = 1,2,\dots,m; \\ &\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = \phi y_{ro} && r = 1,2,\dots,s; \\ &\lambda_j \geq 0. && j = 1,2,\dots,n. \end{aligned} \tag{2.1}$$

where ϕ represents the technical efficiency of entity j, ε is a non-Archimedean element smaller than any positive real number (This condition guarantees that solutions will be positive in these variables, see Arnold et al. (1998)), λ_j the associated weighting of outputs and inputs of entity j, and s^- s^+ the respective slacks, i.e., the leftover portions of inefficiencies.

After proportional reductions in inputs or increases in outputs, if a DMU cannot reach the efficiency frontier (to its efficient target), slacks are needed to push the DMU to the frontier. The efficiency frontier is the envelope representing "best performance" and is made up of the units in the data set which are most efficient in transforming their inputs into outputs. The units that determine the frontier will have an efficiency of 1 and will be defined as efficient. The remaining ones will have an efficiency index between 1 and +∞

inversely proportional to their distance from the border.

When comparing the CCR and the BCC efficiencies, the difference emerged from the constrain added to the BCC model, which makes DMUs be designated as efficient by the BCC model but not by the CCR model. Even when both models designate a DMUs as inefficient, the measures of inefficiency may differ.

2.3 Bootstrap

A main drawback of DEA is that it has no accommodation for noise or random error, as it uses a linear programming (nonstatistical) approach for the estimation of the frontier. The inefficiency scores derived from DEA and the envelopment surface are ‘calculated’ rather than statistically ‘estimated’. Hence DEA is not able to determine the accuracy of the efficiency estimates, or to provide a statistical foundation for the estimated frontier. Bootstrap DEA was introduced by Simar and Wilson (1998) mainly to allow extracting the sensitivity of efficiency scores which results from the distribution of (in)efficiency in the sample.

The basic idea of bootstrapping is that inference about a population from sample data (sample → population) can be modeled by resampling the sample data and performing inference about a sample from resampled data (resampled → sample). As the population is unknown, the true error in a sample statistic against its population value is unknown. In bootstrap-resamples, the 'population' is in fact the sample, and this is known; hence the quality of inference of the 'true' sample from resampled data (resampled → sample) is measurable. Indeed, its name is derived from the saying “pull oneself by one’s bootstraps”, often used as an exhortation to achieve success without external help. More formally, the bootstrap works by treating inference of the true probability distribution J, given the original data, as being analogous to an inference of the empirical distribution Ĵ, given the resampled data. The accuracy of inferences regarding Ĵ using the resampled data can be assessed because we know Ĵ. If Ĵ is a reasonable approximation to J, then the quality of inference on J can in turn be inferred. The purpose of using the bootstrapping approach in this case is two-fold: first, to obtain the bias corrected estimates and the confidence intervals of DEA-efficiency scores and second, to overcome the correlation problem of DEA-efficiency scores and to provide consistent inferences in explaining the determinants of retail industry efficiency. The method of Simar and Wilson (1998) for obtaining the bootstrapped DEA scores is technically consistent and comprises a valuable tool for implementing statistical inference on DEA. The outline of their proposed bootstrap procedure can be summarized in the following steps:

1. Use DEA to calculate efficiency scores.
2. Draw with replacement from the empirical distribution (ED) of efficiency scores. Simar and Wilson (1998) suggest that smoothing the ED provides more consistent results.

3. Divide the original efficient input levels by the pseudo-efficiency scores drawn from the (smoothed) empirical distribution to obtain a bootstrap set of pseudo-inputs.
4. Apply DEA using the new set of pseudo-inputs and the same set of outputs and calculate the bootstrapped efficiency scores.
5. Repeat steps 2-4 B times and use bootstrapped scores for statistical inference and hypothesis testing.

2.4 Window analysis

DEA window analysis is based on a dynamic perspective, regarding the same DMU in different period of time as entirely different DMUs. Moving average method is used to choose different reference set in order to determine the relative efficiency of each DMU. That is to say, when the set window slides once, the first period of each window will be deleted and a new period will be added at the same time. The benefit of this method is to describe the dynamic change of the efficiency of each DMU comprehensively, both horizontally and vertically. (Vargas Sánchez n.d.) More importantly, the number of DMU is increased in this method, hence, it enhances the discriminating power by increasing the number of DMUs when a limited number of DMUs is available. Window analysis in the assessment of influence on operational efficiencies after the establishment of branched hospitals). Nevertheless, this method does not consider the correlation structure of efficiencies and it does not use statistical technique to estimate efficiencies.

Consider a set of N (*n* = 1,…N) DMUs in T (t = 1,…T) period of time. Every DMU has r kinds of input and s kinds of output. Let $DMU_n^{\ t}$ denote the level of input or output for DMU n in t period of time, then input vector ($X_n^{\ t}$) (equation 2.2) and output vector ($Y_n^{\ t}$) will be presented as:

$$X_n^t = \begin{bmatrix} x_n^{1t} \\ \vdots \\ x_n^{rt} \end{bmatrix} \; Y_n^t = \begin{bmatrix} y_n^{1t} \\ \vdots \\ y_n^{st} \end{bmatrix} \tag{2.2}$$

Consider the window starts at the time point of *k* ($1 \le k \le T$), and the window width is *w* ($1 \le w \le T\text{-}k$), then input (equation 2.4) and output matrix (equation 2.5) of each window *kw* will be presented as :

$$X_{kw} = \begin{bmatrix} x_1^k & x_2^k & \cdots & x_N^k \\ x_1^{k+1} & x_2^{k+1} & \cdots & x_N^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{k+w} & x_2^{k+w} & \cdots & x_N^{k+w} \end{bmatrix} \quad Y_{kw} = \begin{bmatrix} y_1^k & y_2^k & \cdots & y_N^k \\ y_1^{k+1} & y_2^{k+1} & \cdots & y_N^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{k+w} & y_2^{k+w} & \cdots & y_N^{k+w} \end{bmatrix} \tag{2.3}$$

Substituting the above inputs and outputs of $DMU_n^{\ t}$ into relevant models will generate the results of DEA window analysis.

The window analysis technique represents one area for further research extending DEA. For example, the problem of choosing the width for a window (and the sensitivity of DEA solutions to window width) is currently determined by trial and error. Similarly, the theoretical implications of representing each DMU as if it were a different DMU for each period in the window remain to be worked out in full detail.

3. Clustering

Cluster analysis was originated in anthropology by Driver and Kroeber in 1932 and introduced to psychology by Joseph Zubin in 1938 and Robert Tryon in 1939 and famously used by Cattell beginning in 1943 for trait theory classification in personality psychology. Intuitively, clustering is a grouping of "similar" objects, where similarity is some predetermined function. More formally, given a set of *n* objects, the process of clustering partitions the set into unique subsets of objects such that each subset shares specific common characteristics. The common characteristics are usually specified in terms of some mathematical relation. Geometrically speaking, the objects can be viewed as points in some *d*-dimensional space. Clustering partitions these points into groups, where points in the same group are located near one another in space. The problem of cluster analysis is defined as follows. Suppose we have a sample of $x \in X \subset R^n$. It is required to divide it into non-intersecting subsets U_j , j = 1,...,k, with centers μ_j, such that

$$L = \min(\mu_j) \sum_{j=1}^{k} \sum_{x \in Uj} d(x, \mu_j) \tag{3.1}$$

is reached, where $d(x, \mu_j)$ is the distance metric value. The problem is nonconvex; thus, in general, L is a local minimum, and the result of clustering depends on the chosen measure d, method of normalization of x_i , i = 1, . . ., n, initial values of x fed to the input of the algorithm for solving (1), and the algorithm itself. In addition, the solution exists for any $0 < k \leq m$, where m is the number of elements in the sample, i.e., the number of clusters in the problem is undefined; so the problem has many "degrees of freedom", and a neat approach to its solution requires discussion of each parameter. However, by its unsupervised nature, there is no method for validating clustering results when the actual clusters of data are unknown.

The common approach of all the clustering techniques is to find cluster centers that will represent each group. A cluster center is a way to tell where the heart of each cluster is located, so that later when presented with an input vector, the system can tell which cluster this vector belongs to by measuring a similarity metric between the input vector and al the cluster centers, and determining which cluster is the nearest or most similar one. Therefore, a key issue of the clustering procedure is the number *k* of clusters obtained. In this work, two metrics will be used to test both the clustering algorithms. The first, the silhouette metric is defined for each sample element as $s_i = b_i - a_i$, max{a_i ,b_i} where a_i is the average distance from the element $x_i \in X$ to its cluster elements, b_i is the average distance to the elements of the nearest cluster. By construction, $x_i \in [-1, 1]$. If s_i = 1, then the element belongs to its cluster. If $s_i = -1$, then the element is definitely located in the wrong cluster. If $s_i = 0$, then the element is located on the boundary of at least two clusters. For generalized evaluation of the clustering quality, we use $s_j = s_i$, named the mean silhouette value over all cluster elements. A reasonable number of clusters is considered to be determined by the mean silhouette maximum. For the same reasons, when several clustering methods are used, the one with the maximum mean silhouette metric is recognized as the best one. Then, the elbow method, which is based on the comparative use of the total RMS distance $v_k = \sum k_j{=}1 \sum_{x_i \in X} (x_i - \mu_j)^2$ for various number of clusters (of the sum of within cluster variance with respect to the number of clusters). The sequence v_k decreases with respect to k, and the number of clusters is determined (as a rule, visually) as a transition from a large to a small change in the derivative of the resulting sequence.

There is no universal clustering algorithm. In our case, an a priori choice of algorithm is impossible; thus, we used 2 algorithms, the results of which will be compared in sections 4 and 5.

3.1 K-means

One of the major clustering approaches is based on the sum-of-squares criterion and on the algorithm that is today well-known under the name 'k-means'. The K-means clustering, or Hard C-means clustering, is an based on finding data clusters in a data set such that a cost function (or an objection function) of dissimilarity (or distance) measure is minimized. In most cases this dissimilarity measure is chosen as the Euclidean distance. A set of n vectors xj, j =1,,n, are to be partitioned into c groups. The cost function, based on the Euclidean distance between a vector x_k in group j and the corresponding cluster center c_i , can be defined by:

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \left(\sum_{k, \mathbf{x}_k \in G_i} \|\mathbf{x}_k - \mathbf{c}_i\|^2 \right) \tag{3.2}$$

where $J_i = \sum_{k, \mathbf{x}_k \in G_i} \|\mathbf{x}_k - \mathbf{c}_i\|^2$ is the cost function. The partitioned groups are defined by a c × n binary membership matrix U, where the element u_{ij} is 1 if the jth data point x j belongs to group i, and 0 otherwise. Once the cluster centers ci are fixed, the minimizing u_{ij} for Equation (3.3) can be derived as follows:

$$u_{ij} = \begin{cases} 1 \text{ if } \|\mathbf{x}_j - \mathbf{c}_i\|^2 \leq \|\mathbf{x}_j - \mathbf{c}_k\|^2 & \text{for each } k \neq i \\ 0 \text{ otherwise.} \end{cases} \tag{3.3}$$

which means that x_j belongs to group i if ci is the closest center among all centers. On the other hand, if the membership matrix is fixed, i.e. if u_{ij} is fixed, then the optimal center c_i that minimize equation (3.2) is the mean of all vectors in group i :

$$\mathbf{c}_i = \frac{1}{|\boldsymbol{G}_i|} \sum_{k,\mathbf{x}_k \in G_i} \mathbf{x}_k \tag{3.4}$$

where $|G_i|$ is the size of G_i ,or $|G_i| = \sum_{j=1}^{n} u_{ij}$.

The algorithm is presented with a data set x_i , i = 1, …, n; it then determines the cluster centers ci and the membership matrix U iteratively using the following steps:

1. Initialize the cluster center c_i, i=1, …, c. This is typically done by randomly selecting c points from among all of the data points;
2. Determine the membership matrix U by Equation (3.3);
3. Compute the cost function according to Equation (3.2). Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold;
4. Update the cluster centers according to Equation (3.4). Go to step 2;

3.2 Hierarchical clustering

The k-means algorithm gives what is sometimes called a simple or flat partition, because it just returns a single set of clusters, with no particular organization or structure within them. But it could easily be the case that some clusters could, themselves, be closely related to other clusters, and more distantly related to others. Here hierarchical clustering algorithms come to help.

There are two approaches to hierarchical clustering: we can go “from the bottom up”, grouping small clusters into larger ones, or “from the top down”, splitting big clusters into small ones. These are called divisive and agglomerative clusterings, respectively. The last algorithm is the one considered through this work, and it is very simple: it starts with each point in a cluster of its own until there is only one cluster and then it finds the closest pair of clusters and merge them together. It results in a cluster-merged tree called dendrogram. To turn this into a definite procedure, thought, there is to define how close the two clusters are, which is not the same as how close two data points are or how close two partitions are. Before any clustering is performed, it is required to determine the proximity matrix containing the distance between each point using a distance function (Euclidean etc.). Then, the matrix is updated to display the distance between each cluster. There are four methods for combining clusters in agglomerative (“bottom-up”) approach:

1. In single linkage hierarchical clustering, the distance between two clusters is defined as the shortest distance between two points in each cluster:

$$d(A,B) \equiv \min_{\vec{x} \in A, \vec{y} \in B} \left|\left|\vec{x} - \vec{y}\right|\right| \tag{3.5}$$

It is called “single link”, because it says clusters are close if they have even a single pair of close points a single “link”. This algorithm only wants separation and does not care about compactness or balance.

2. In complete linkage hierarchical clustering, the distance between two clusters is defined as the longest distance between two points in each cluster.

$$d(A,B) \equiv \max_{\vec{x} \in A, \vec{y} \in B} \left|\left|\vec{x} - \vec{y}\right|\right| \tag{3.6}$$

3. The average linkage method is a compromise between the single and complete linkage methods, which avoids the extremes of either large or tight compact clusters. the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster.

$$d(A,B) = \frac{T_{AB}}{N_A * N_B} \tag{3.7}$$

where T_{AB} is the sum of all pairwise distances between cluster A and cluster B. N_A and N_B are the sizes of the clusters A and B, respectively. At each stage of hierarchical clustering, the clusters A and B, for which d(A,B) is the minimum, are merged.

4. Instead of measuring the distance directly, the Ward’s Method, which is an alternative to single-link clustering, analyzes the variance of clusters. Ward’s method says that the distance between two clusters, A and B, is how much the sum of squares will increase when we merge them:

$$\begin{aligned}\Delta(A,B) &= \sum_{i \in A \cup B} ||\vec{x}_i - \vec{m}_{A\cup B}||^2 - \sum_{i \in A} ||\vec{x}_i - \vec{m}_A||^2 - \sum_{i \in B} ||\vec{x}_i - \vec{m}_B||^2 \\ &= \frac{n_A n_B}{n_A + n_B} \left|\left|\vec{m}_A - \vec{m}_B\right|\right|^2\end{aligned} \tag{3.8}$$

where $\vec{m}_J$ is the center of cluster j, and n_j is the number of points in it. Δ is called the merging cost of combining the clusters A and B. With hierarchical clustering, the sum of squares starts out at zero (because every point is in its own cluster) and then grows at the least pace possible.

4. Modelling with R

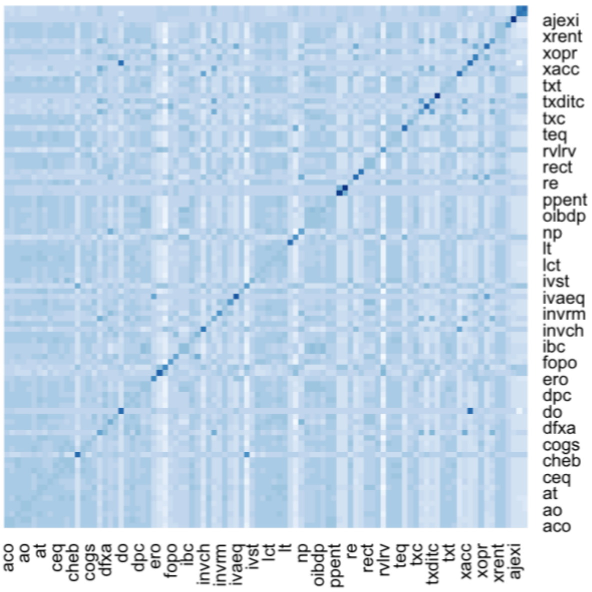
R is a programming language for statistical computing and graphics supported by the R Core Team and the R Foundation for Statistical Computing. The official R software environment is an open-source free environment within users have created packages to augment the functions available, and for this reason, it was the most suitable for calculations of this analysis. Here it follows the explanation of the work performed in R, structured with an introductory part, some variable selection models as well as the implementation of DEA and clustering.

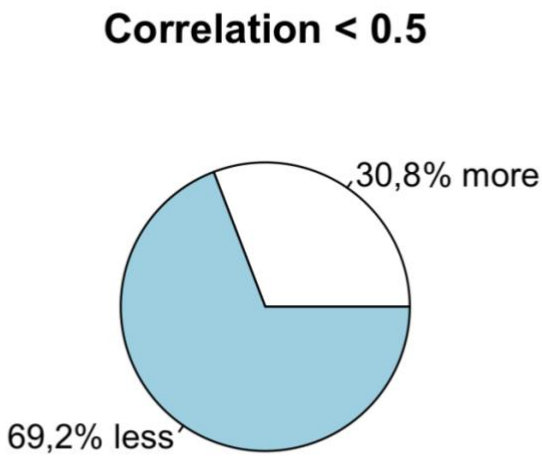
4.1 Import dataset

WRDS provides a direct interface for R access, allowing native querying of WRDS data right within Rstudio. All WRDS data is stored in a PostgreSQL database, and is available through R via a native R Postgres driver. Once being logged and connected to the server; it was possible to demand for the dataset of interest by typing a simple SQL query with the forehead mentioned characteristics (time span of 12 years and naics from table 1.2). At this point, compa_funda was downloaded in .csv format and ready to be accessed. Before to start, it was important to understand what kind of dataset was and, eventually, if there were some adjustments to do in order to optimize the access to the information available. Performing data cleaning means exactly that, to prepare data for subsequent analysis by removing or modifying incomplete, irrelevant, duplicated, improperly formatted, or missing information. In this case, compa_funda included more than 900 observation and 400 variables (see Appendix), mostly numeric data reported in the relative local currencies, with few observations expressed as date type or strings. Compa_funda was large and had a lot of information missing, more than half of the dataset columns were completely empty. They did not contain any information, so they were dropped as well as the few rows that remained incomplete, remaining with a 646x123 data frame. Then, by converting the categorical variables into factors, it was possible to visualize how many observations were present in each factors' levels. It was relevant to check how many firms per year remained after null values per row were removed (accounting for the 1.25% of total), ending up with mostly incomplete time series. There were no duplicates.

Going on with the numerical variables, a few outliers were identified, but still they were kept into account because within this data, which were revised and approved before being published, it was more probable, not certain though, that excessively high or low values just signaled a different business model rather than a miscalculation. For example, rental expenses (xrent) might change a lot depending on the number of facilities owned (machineries, buildings) as well as on the operating country average prices, while nonoperating income (nopi) might differ substantially especially if firms are diversified and do not operate only in the considered markets, same for the Net Cash Flow of Financing Activities - (fincf). Different firms have also different dividend policies, so Common Shares Outstanding - Issue (cshoi) and other equity invoices can reflect this fact. For completeness, for each compa_funda's variable median, mean, standard deviation and quartiles of the numerical variables was computed with the *summary()* function (see Appendix 2).

Besides that, also the correlation matrix highlights some peculiarities. The majority of the variables seemed to be independent with correlations value lower than 0.5, while 151 pairs were almost perfectly correlated, for instance showing the existence of some redundant information. Some correlations were direct, whereas other correlations were indirect in the sense that variable A is only correlated with variable D because A is related to B is related to C is related to D. It does not really come as a surprise, financial statements are built up with items that are just sums of other variables, but such intermediate values need to be discarded to avoid noise in the following models.





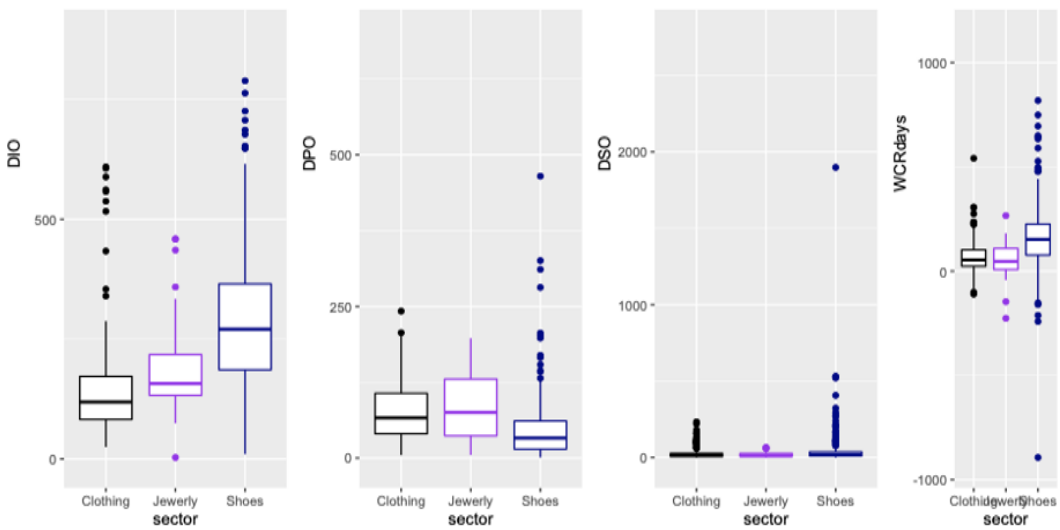
Figures 4.3 and 4.4 On the right, the correlation heatmap. On the left, a pie chart representing how many paired variables have a correlation of less than 0.5 (low)

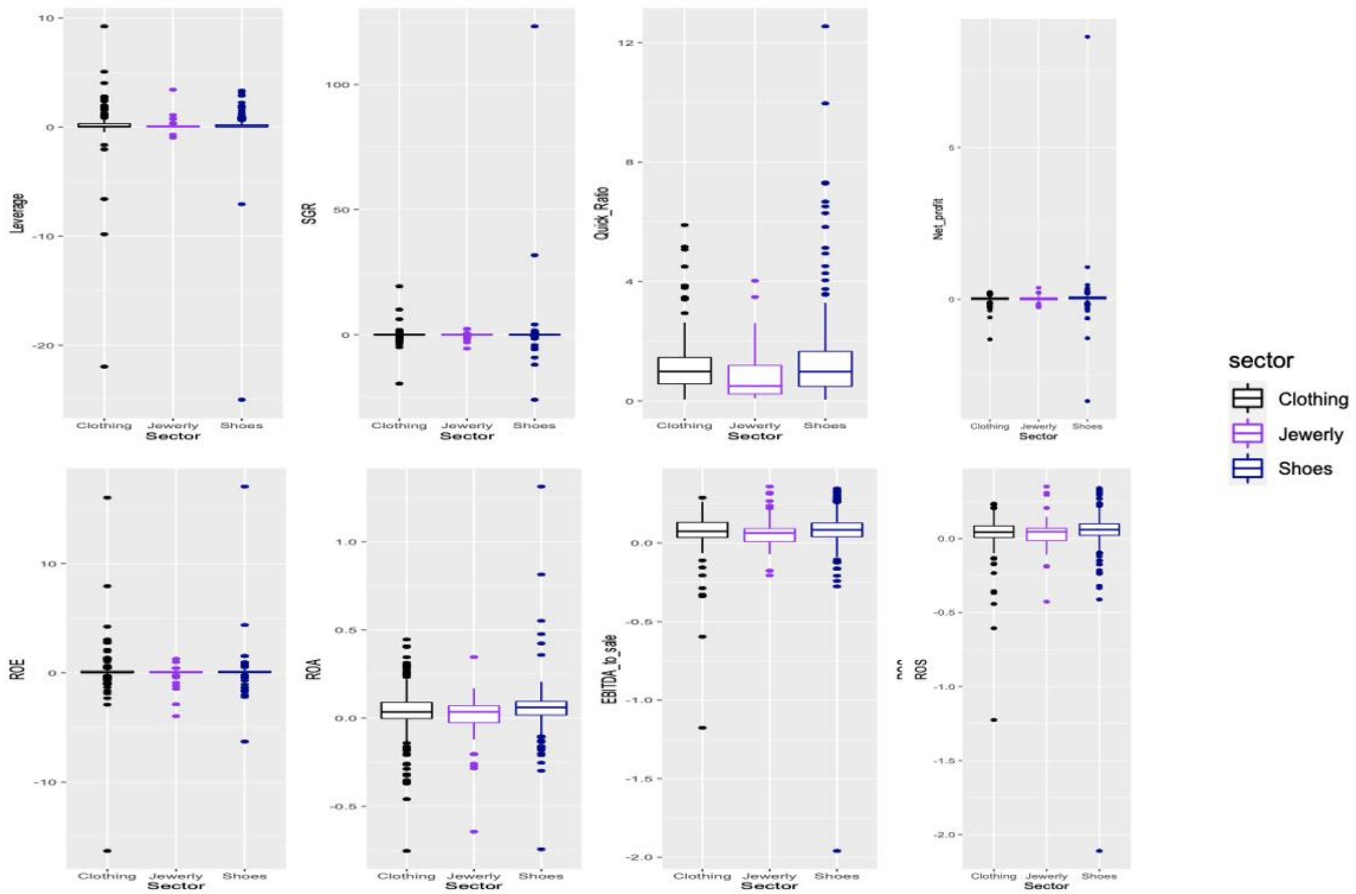
Nevertheless, in any practical correlation analysis it is important to identify and then focus on the direct correlations which really matter. It is also important to look for causality, to ensure that the selected correlations are real features of the underlying business rather than the result of coincidence. Therefore, a deeper insight will come ahead in the following section with a smaller number of pairs.

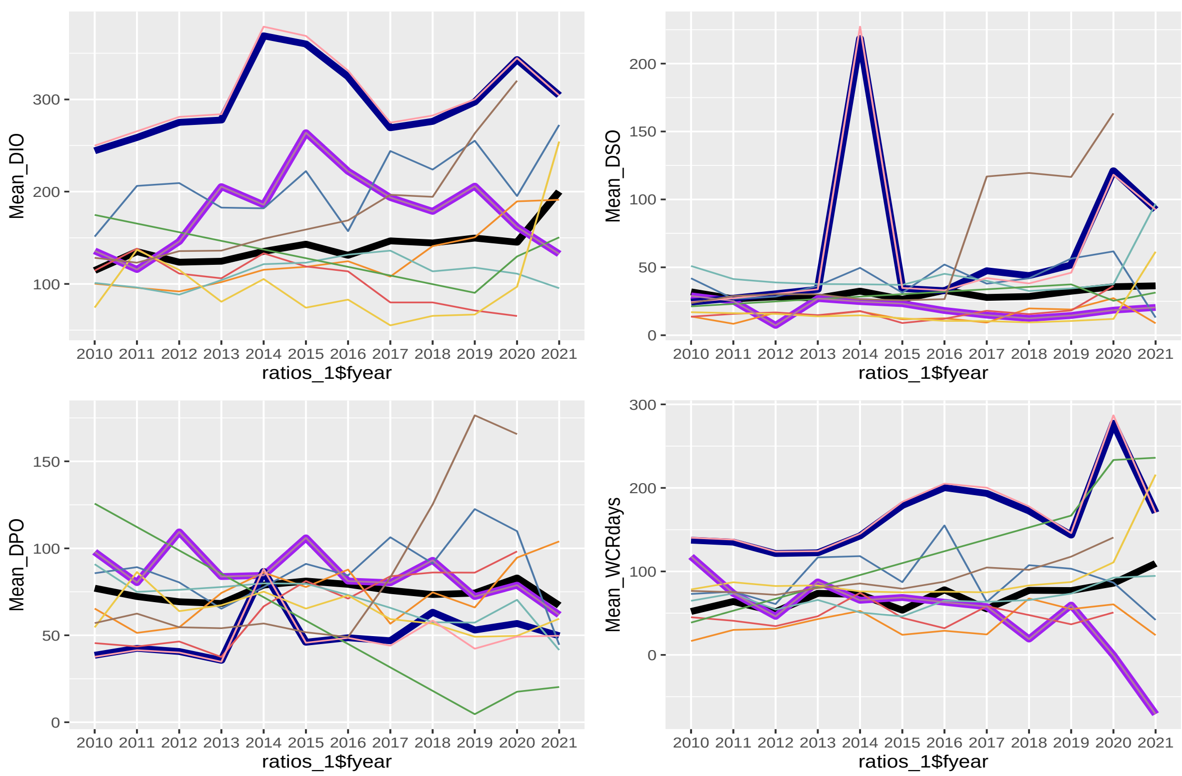
4.3 Ratios analysis

Following the example of the "Financial Performance Evaluation of agricultural enterprises with DEA Method", in which the authors chose some ratios that better translate the reality of the primary sector, for the purposes of this work, consideration has been limited to the parameters considered to be the most important for a general fashion retail company. Based on the data of the remaining 92 company, the financial ratios already discussed (see section 1.1) were computed and displayed through their values per sector and their average per year, as well as their correlations, so to simplify their interpretation.

By looking at the various plots, it was confirmed that the tree branches of the fashion retail could be compared using the same performance indicators. Boxplots showed that ratios did differ among Clothing, Jewerly, and Shoes retailing, especially when it came to efficiency measures like Day Inventory Outstanding, Day Payable Outstanding and days of Working Capital Requirements, despite that, their annual trends looked similar. On the other hand, profitability indicators were very close; EBITDA to sale and Return on Sales were respectively around 10% and 5% with a slight decreasing tendency until 2020, when Covid19 spread the most. The sanitary crisis had its impact on consumption worldwide, and some repercussion on the economy are still present at the time of this work, but overall organizations were able to produce greater earnings while keeping costs down through a time span of 10 years. This discrepancy between activity and profitability ratios could have happened for external circumstances (e.g. supply shortages) in which cash flow was impeded due to the value of inventory. This could have lead to increased financing costs to cover day-to-day cash needs. As a result, a high sales and inventory standing may raise questions about other aspects of a company’s health other than simply profits. Within the given sample, Jewerly and Clothing retail businesses seemed to be able to transform their resources in cash at a faster pace compared to shoes counterparts.




Leverage
SGR
Quick_Ratio
Net_profit
ROE
ROA
EBITDA_to_sale
ROS
Clothing
Jewerly
Shoes
Sector
sector
Clothing
Jewerly
Shoes


Mean_DIO
Mean_DSO
Mean_DPO
Mean_WCRdays
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021
ratios_1$fyear

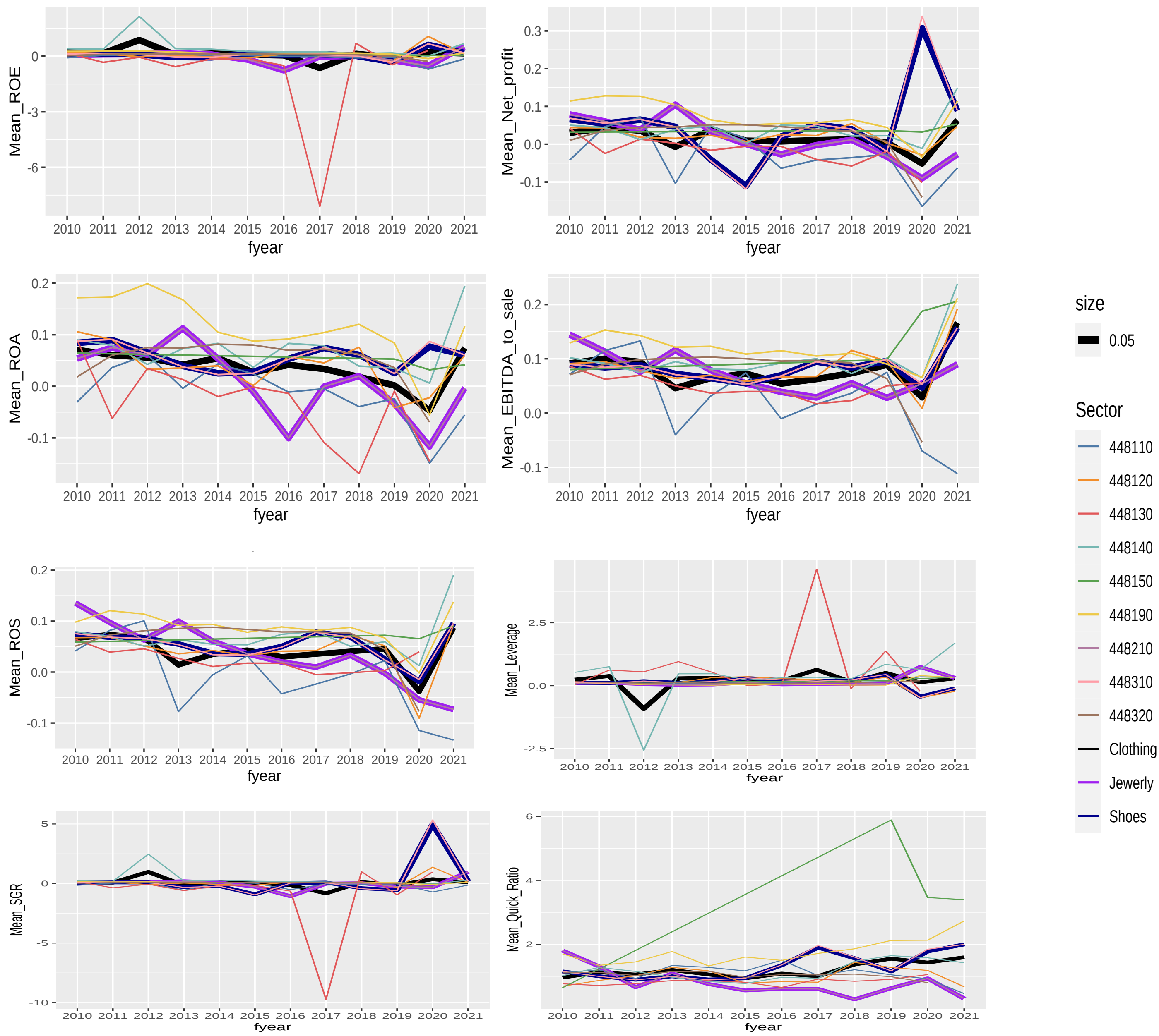


Figure 4.5 and 4.6 On the previous page, box plots of the financial ratios by their sectors. On the current page, the average value of each ratio along the considered timespan.

One can notice how a variable has changed over time but also how a variable has moved similarly to another just looking at more trends at the same time (Figure 4.6). EBITDA to sale, ROS, and Net profit moved together, except for the shoes sector, they looked the same just shifted. To obtain interpretable results afterwards, one should focus on as few ratios preferably uncorrelated as is possible, but forward, the variable selection method will have taken this in into account as well.

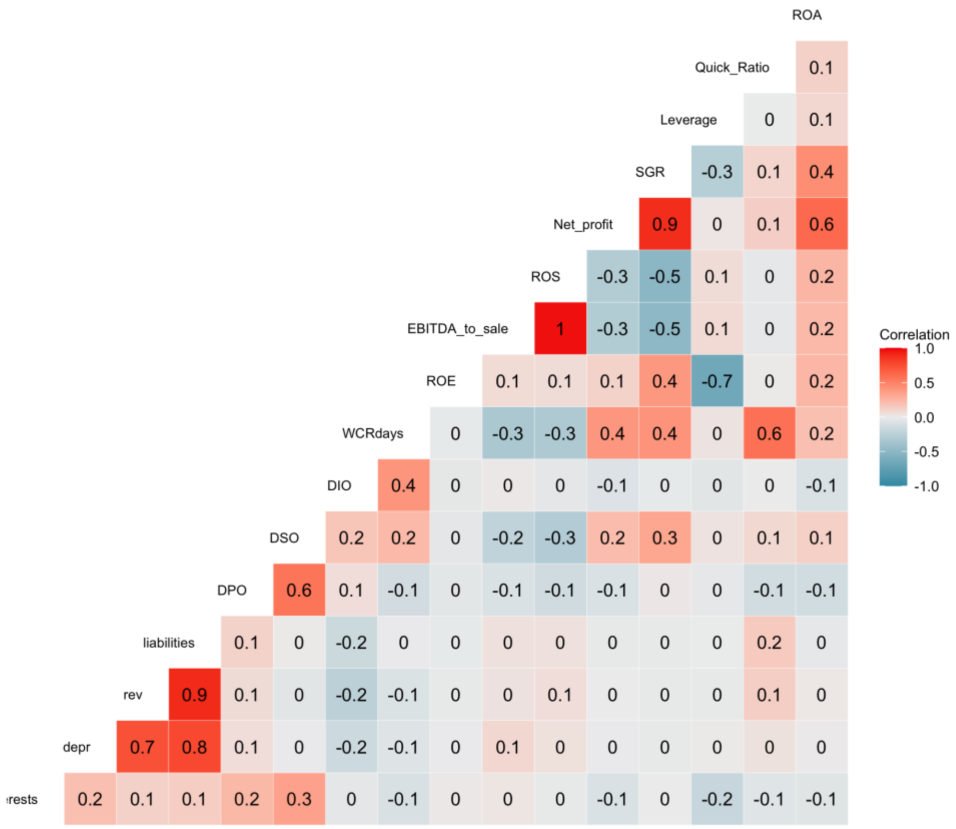


Figure 4.7 Correlation plot of the financial ratio of compa_funda4

While the perfect correlations between EBITDA to sales and ROS, SGR and Net Profit to sales were somehow expected by the way these ratios were formed there was some surprising and counterintuitive data. Return on Equity resulted having no correlation with any of the efficiency indicators, and a strong negative correlation with leverage that was seen also in the time plots. If by definition increasing debt with respect to equity (higher leverage) let to an increase in risk (associated to borrowing) for which shareholders should be paid accordingly (higher ROE), in reality leverage has a boomerang effect on profitability. Also, there was no correlation between leverage and liabilities at all. By contrast, liabilities found to be almost perfectly correlated with revenues, underling its indirect role in boosting sales and investing activities.

4.4 Data Envelopment Analysis

DEA provided opportunities for financial analysis by using data of 92 DMUs, which were compa_funda's companies, as input or output and whereby the units' general financial performance was evaluated using a complex indicator (score), which cannot be achieved by separate indicators gained from financial statements. The selection of inputs and outputs in Data Envelopment Analysis is regarded as an important step that is normally conducted before the DEA model is implemented. If one uses less than full information, they will lose some of the explanatory powers of the data, however variable selection methods are important for this work because DEA is a non-parametric approach and loses discriminatory power as the dimensionality of the production space increases. Many studies confirms (Łozowicka, 2014) that the suggested proportion for input and output versus decision making units is 1/3 because, as the number of inputs and outputs increases, the observations in the data set are projected in an increasing number of orthogonal directions and the Euclidean distance between the observations increases. This results in many observations lying on the frontier; thus, DEA loses its discriminatory power. When the condition of 1/3 is not satisfied it is advisable to increase the number of objects and/or remove some variables describing the objects. The confliction between the requirements of the practical conditions and traditional DEA methods lead to the situation that the selected data set is not suitable to apply traditional DEA methods to always occurs.

Before using DEA, some regressions were used in order to choose the variables correctly, helping to pick the most influential explanatory variable. Return on Assets (ROA) was used as outcome variable because it is a baseline that can be used to measure the profit contribution required from new investments, the remaining indicators were used as independent variables during regression.

4.4.1 Variables choice

Linear regression was primarily used to investigate the effect of variables on efficiency measurement, as one of the easiest ways to select variables consists of iteratively adding and removing predictors from the predictive model, to find the subset of variables in the data set resulting in the best performing model, that is a model that lowers prediction error. Another way to approach the problem was using models performing a regularization on the features. Indeed, LASSO was introduced trying to improve also the prediction accuracy and interpretability of regression models. But first, data needed to be standardized.

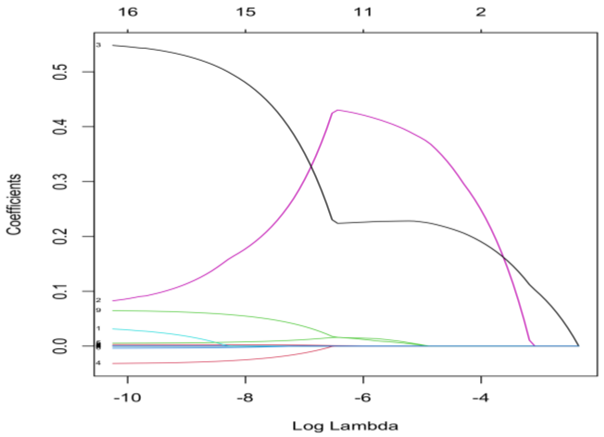
Even if one of the main advantages of DEA is that it is totally fine not to standardize data, as it compares different criteria even with different order of magnitudo (Epicoco, 2016), for linear regression algorithms is not the same. Generally speaking, standardization is useful when data has varied scales and the algorithm does make assumptions about data having a Gaussian distribution, such as linear regression, logistic regression, and linear discriminant analysis. Since R has a built-in function called *scale()* to standardize, it was applied on a copy of the dataset so that variables for DEA will have picked from the unscaled one. Then, the scaled data frame was initially split in a training data set (60%), so to have a set of examples used to fit the parameters of the variable selection models at stake. The first algorithms applied to this new set were forward, backward, and exhaustive search. Forward stepwise selection starts with an empty set of variables and adds variables to it, until some stopping criterion is met. Similarly, backward selection starts with a complete set of variables and then excludes variables from that set, again, until some stopping criterion was met. In exhaustive feature selection, the performance is evaluated against all possible combinations of the features in the dataset. The feature subset that yields best performance is selected, but at the cost of a higher computational level. Fortunately, variables were not too many and package *leaps()*, which performs an exhaustive search for the best subsets of the variables in x for predicting y in linear regression using an efficient branch-and-bound algorithm, came in help. Since the algorithm returned a best model of each size, the results did not depend on a penalty model for model size: it did not make any difference whether AIC, or BIC was used (Package 'leaps'). Within the inherited function *regsubsets()* it was possible to perform also forward and backward stepwise regression by changing the parameter "method" passing respectively the strings "forward" and "backward"; here below the results obtained on the test set.

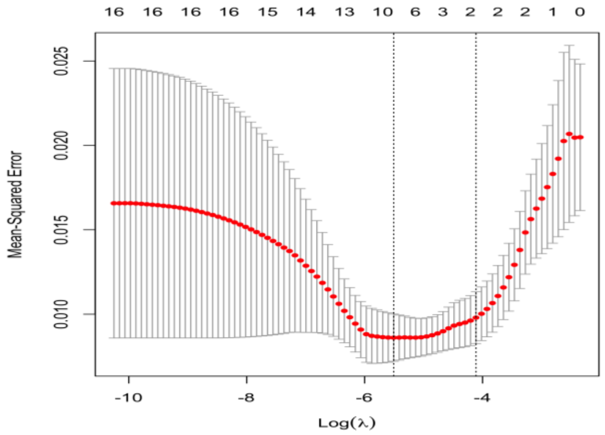
Coefficients	**Backward**	**Forward**	**Exhaustive**
(Intercept)	0.034272416***	2.791015e-02***	0.0177356410***
DSO	-	6.983332e-05	0.0001727219
ficCHE	-	-6.651368e-02*	-
ficFRA	-0.04139989	-	-0.0751621714*
ficJPN	-0.08898540***	-2.938116e-02***	-
ficMYS	-	-	0.0480347587
ficSGP	-0.02749350**	-	-
fyear2013	-	3.824107e-02	-
fyear2020	0.041399892	-4.548057e-02	-0.0399842205
Leverage	-	-	0.0114713134*
Net_profit	0.637730373***	2.492171e-01	0.6125164080***
ROE	0.055729273***	-	0.0596881211***
ROS	-	4.508299e-01	-
SectorJewerly	-	-4.108051e-02	-
SGR	-0.038048082	-	-0.036087091
Test Set Adj R^2	**0.7888**	**0.435**	**0.7888**
Mean Square Error (MSE)	**0.002875485**	**0.007722856**	**0.002874684**

Table 4.1 Backward, Forward, Exhaustive selection methods compared. The asterisks indicate the statistical significance of a variable inside the fitted mode.

From the table 4.1, it was seen that some of variables selected by the three methods were the same and had a very low p-value. Backward and Exhaustive model gave the closest results, differing just of a pair of variables, and with identical adjusted R^2 and MSE. As expected, of the EBITDA to sale, ROS, and Net Profit variables, which are all intermediate balances computed against the level of sales, only one was kept by the three selection methods. Of the previous discoveries about the impact 2020 were confirmed: among the possible dummies representing the fyear levels, it was the only widely confirmed to affect the return on assets.

On the other hand, within LASSO, a method for regularization is based on the L1 norm, the coefficients were shrunked accordingly to their relevance toward zero and for this reason it was used to perform a feature selection. A good way to visualize and understand better how LASSO actually worked for the purpose was by fitting a glmnet model (alpha = 1) for different values of lambda and then each value of lambda together with the value of the coefficient estimates. It became clear that higher lambda values are associated to more penalty and thus to more coefficients shrunked toward zero.





Figures 4.8 and 4.9 On the left, a LASSO model for different values of lambda and then each value of lambda together with the value of the coefficient estimates. On the right, a cross validation to pick the best lambda (min) with respect to the lowest MSE.

Variables	Coefficients
interests	1.889898e-05
depr	0.000000e+00
rev	0.000000e+00
liabilities	-2.834222e-08
DPO	-5.019529e-05
DSO	5.864348e-05
DIO	-4.137400e-05
WCRdays	4.637353e-05
ROE	1.060958e-02
EBITDA_to_sale	0.000000e+00
ROS	4.201498e-01

Net_profit	2.268642e-01
SGR	0.000000e+00
Leverage	1.546690e-02
Quick_Ratio	0.000000e+00
dev.ratio (R^2)	**0.6518918**
MSE	**0.008295255**

Table 4.2 LASSO coefficients at the best lambda. The coefficients that are meaningful for DEA model are: DSO, WCRdays, ROS, Net_profit, interests, Leverage, ROE, DPO, DIO, liabilities

According to LASSO, the most meaningful coefficients were Day Sales Outstanding, Day Inventory Outstanding, Day Payable Outstanding, and Return on Sales. Probably, those firms that had the lowest DSO, were the most inefficient, meaning that they might be unable to give an average time to their clients for illiquidity problems, and one should take into observation this variable with the other activity ratio as well. Other meaningful variables seemed to be Liabilities and Leverage, indicating that companies should also focus their attention on their level of debt in order to meet their obligation to third parties. Lastly, Concerning Net profit, which had a correlation of 0.6 with ROA, its coefficient was quite obvious in the sense that those firms that are able to generate more income are able to do so because they found better and more economic offers.

Finally, the series of inputs chosen to compute efficiencies is the one highlighted by the LASSO among the four variable selection methods applied. It was neither because of the measure of the R^2 nor the MSE, Thus, **INPUT = Interests, liabilities, DPO, DSO, DIO, ROE, ROS, WCRdays, Leverage and OUTPUT = ROA.** Net profit did not appear in the final input variables as the DEA algorithm cannot work properly with negative values and among the aforementioned variables Net profit had more negative observations, so ROS was selected instead. Also, the use of categorical variables could have been an important extension of the DEA, which could have improved the peer group construction process and incorporate "on-off" characteristics. Nevertheless, DEA will have performed on a yearly basis and the other relevant factors were discarded in the previous step, so they were not used.

4.4.2 Efficiency scores

DEA created a financial efficiency frontier for every year considered and financial efficiency score was assigned to all the analysed DMU to show if a DMU was stable, deteriorated or improved. It was also considered the possibility to apply the Malmquist Productivity Index (MPI), useful variation of the DEA that also considers time variations and can be decomposed into changes in efficiency and technology, but the dataset did not fit its assumption of continuity as some DMUs are missing in some years. Anyway, for this scope, the Benchmarking package was used iteratively. Within this package, Data Envelopment Analysis is supported under different technology assumptions (fdh, vrs, drs, crs, irs, add), and using different efficiency measures (input based, output based, hyperbolic graph, additive, super, directional). **In this case, the technology assumption was "vrs" and the model "output-oriented".** The efficiency in DEA was calculated by the LP method in the package lpSolveAPI. The estimates of efficiency scores in output-oriented models came out in a range from 1 to infinity but in order to be aligned with the commonly used efficiency range (0-1) it was enough to compute their inverse. Here a sample from the results.csv file:

Companies	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
FESTARIA HOLDINGS CO LTD	0.485720	0.414987	0.200342	0.070981	0.078227	0.126045	0.026171	0.076233	0	0.031871	0	NA
NEXT PLC	1	1	1	1	1	0.960360	1	0.814487	0.804183	1	0.774163	1
HENGDELI HOLDINGS LTD	0.536472	0.506098	0.486735	0.202426	0.387301	0.236536	0.155166	0	0.424181	0	0	1
TOMEI CONSOLIDATED BERHAD	0.855045	1	0.806136	0.271049	0.845070	0.709364	0.706016	0.825545	1	1	1	1
KAPPAHL AB	1	0.761054	0	0.984905	1	0.753379	0.683489	1	1	0.565115	NA	NA

FOSCHINI GROUP LTD	0.791665	0.636863	0.655860	0.373614	0.518353	0.565443	0.635067	0.266032	0.439018	0.300969	0	NA
STELUX HOLDINGS INTERNL LTD	0.687922	0.723149	0.509739	0.336479	0.422170	0	0	0	0	0	0	NA
AOKI HOLDINGS INC	0.343943	0.410684	0.380509	0.289398	0.292023	0.286190	0.281012	0.140066	0.212370	0.042983	0	NA
HENNES & MAURITZ AB	1	1	1	1	1	1	1	0.881469	0.596833	0.720459	0.044741	0.836866

Table 4.3 Efficiency scores of 9 among the ones that had the most complete records.

Investigating these scores revealed to be an interesting matter. DEA seemed to capture the dynamics of a healthy industry, in the sense that even the best performing companies might lose their status year by year (H&M after 2016), either by the hands of new entrants and innovative competitors or by fluctuation of supply and demand. When the pandemic crisis began, efficiencies scores were at their minimum for all DMUs in table 4.3, showing how susceptible retailers were to travel bans and shock in mass consumption. Moreover, the firms for which the algorithm gave a score of 0 probably faced negative ROS values. This interpretation found its basis on the ROS negative trends seen in preceding section but also on the obvious fact that without positive earnings, it is complicated to survive in very competitive markets. Nevertheless, this does not mean they could not have recovered nor that they could not in the future. For this reason, an average score was computed for the 92 DMUs, so to have an evaluation comprehensive of the physiological ups and downs firms might encounter. Excluding null values, if their average scores were higher than 0.50, then they were considered overall as efficient. This classification will have been useful further on in section 4.5.

Ultimately, slacks were calculated by the call of *dea()* using the option SLACK=TRUE and saved in separate spreadsheets, enabling doing deeper insights on the choices available for firms to react in such context.

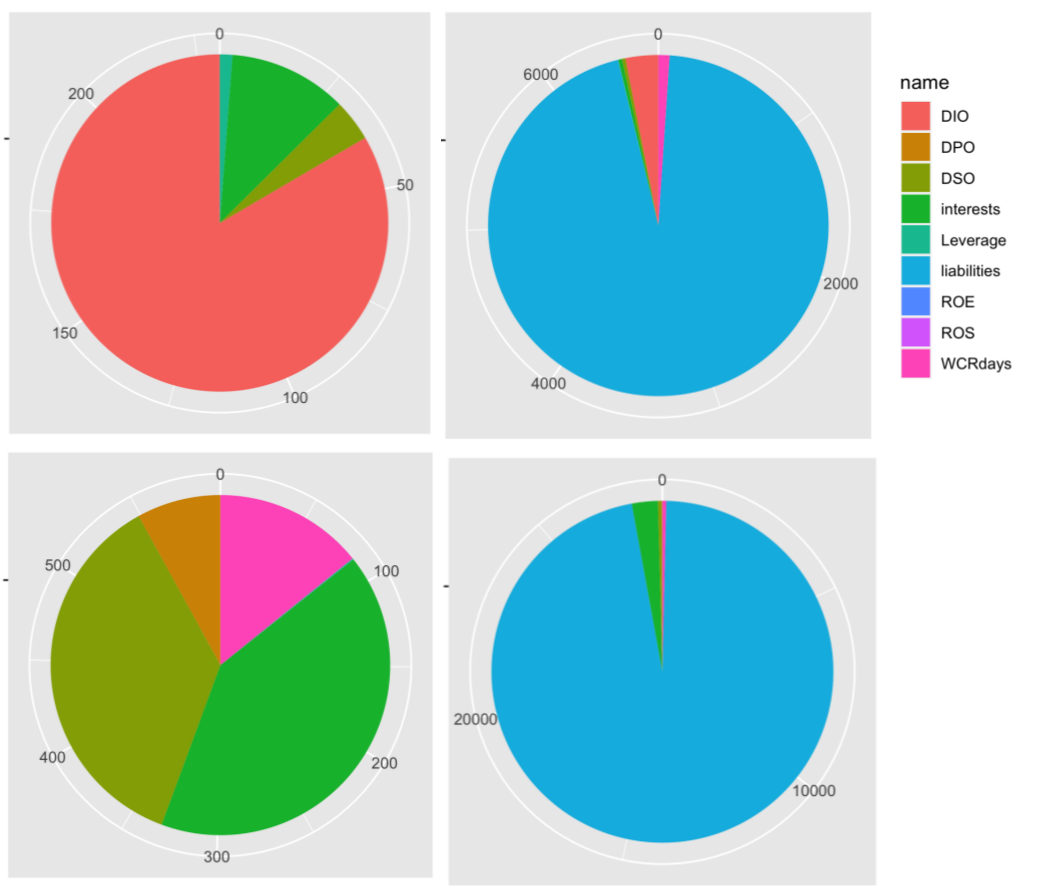


Figure 4.10 On the first row, slack values for Festaria Holding in 2010 and 2017. Within the same years, on the second row the slack values for Foschini Group LTD. Both firms were classified as inefficient.

For example, in 2013 Festaria Holdings and Foschini Group were both lacking a clear plan at operational level. By potentiating its inventory levels, the former would have probably fostered its sales, while by carefully giving more time to its clients to pay, the latter may have captured a larger clientele and consequently may have bargaining power to increase its average time to pay back suppliers. Instead, the data suggesting an increase in interest expences would be difficult to clear out without any references to the debt market in 2013. It may have been that it was rather cheap acquiring debt compared to other ways of financing at the time, or that even the most efficient firms were struggling for higher interests inherited by the Great Recession of 2008 maybe. After seven years, both pie charts showed a remarkable need for Festaria Holdings and Foschini Group to increase their liabilities, confirming that problems linked to bad activity benchmarks (DIO, DSO, DPO, WCRdays) made them end up searching for other sources of cash. Financial leverage offers an alternative way to increase profits and cash reserves by financing a portion of the business through loans or by issuing stock, but at the same time, it does not guarantee an improvement in profitability, as also the strong negative correlation score between ROE and leverage highlighted. Securing financial leverage may come at the cost of unfavorable interest rates and higher dividend payments for stockholders, which makes it more difficult to improve liquidity as well as profitability. Despite that, liabilities slack could be understood not properly as debt but more in general as investments in assets because, by the way ROA is built, "liabilities" is the only variable inside the model able to represent total assets among the all the selected benchmarks.

4.4.3 Bootstrap

The majority of DEA papers use the approach of Simar and Wilson (1998) or their more recently proposed method of confidence interval construction (Simar and Wilson, 2000) in order to test (usually) the hypothesis of whether: two firms from the same sample differ significantly in efficiency, two firms from different samples differ significantly in efficiency, and two samples have equal average efficiency.

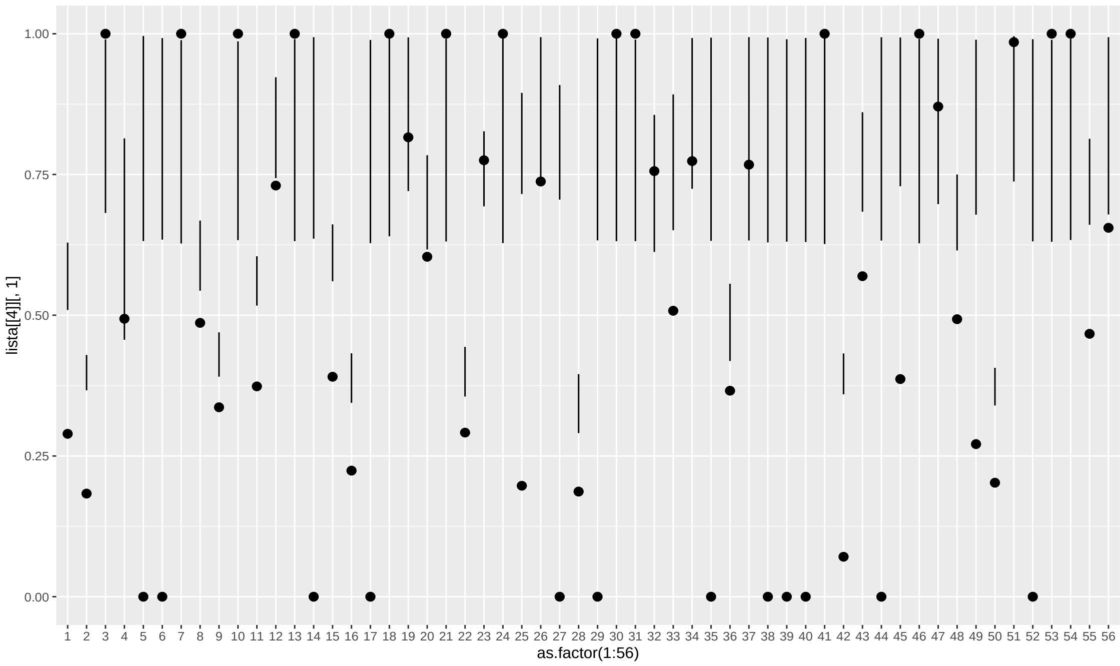


Figure 4.11 Bootstrapped confidence intervals for year 2013. The width of the confidence interval around the point estimate reveals the precision.

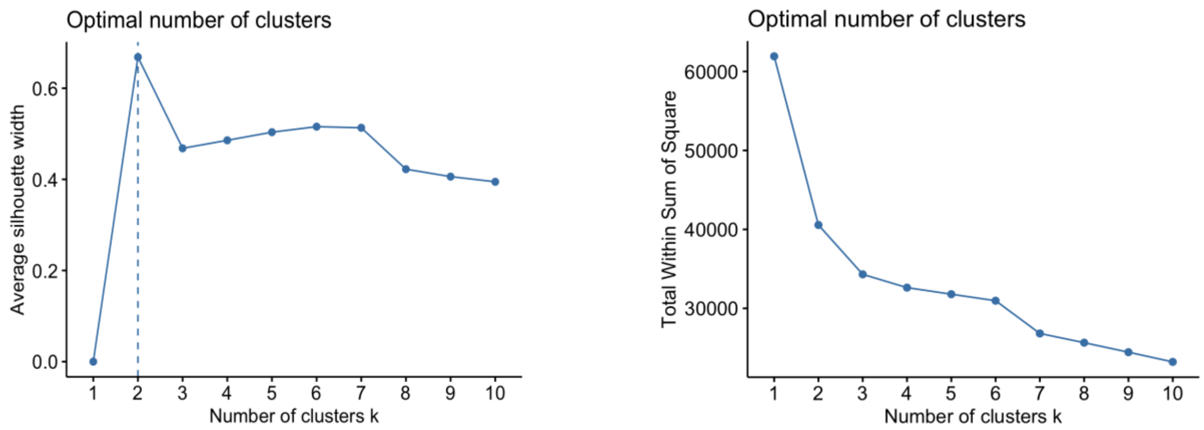
If we repeated DEA experiment 100 times (it was iterated 500 times), gathering 100 independent sets of firms, and computing a 95% CI for the mean difference each time, 95 of these confidence intervals would capture the population mean efficiencies. CIs helped to describe how uncertain the result of a non-statistical method estimated difference was, whether the estimate was a precise one or only a very "rough" one. If the range is narrow, the margin of error is small, and there is only a tiny range of plausible efficiency so that's a precise estimate. However, if the interval is wide, the margin of error is large, and the actual score is likely to fall somewhere within that more extensive range. That was the case one could see through the years (Figure

4.11) and there were also some estimations that were expected to be efficient but instead fell far out of the range.

It is true that the larger the sample size the smaller the variability in the bootstrap distribution, the narrower the interval the more precise the estimate is, and that unfortunately there were on average only 55 firms per year but using bootstrapping cannot improve point estimate. The quality of bootstrapping depends on the quality of collected data as well. If sample data is biased and does not represent population data well, the same will occur with bootstrap estimates.

4.5 Clustering

The given data were clustered in their entirety through the k-means method and the hierarchical method, with the aim of finding some correspondence within the efficiency scores and the groups that clustering could identify. The two algorithms have their differences, and it was worth to try them both. This was done considering the subset of the original data containing all the already standardized numerical variables; otherwise, the choice of units, e.g. different currencies, rather than the few outliers acknowledged before, for a particular variable would have greatly affect the dissimilarity measure obtained. It was first implemented the k-means method, for which it is crucial to define the number of clusters that best fit the dataset. So two kind of metric were used. The first, the silhouette metric, suggested that the optimal number of groups was two, as it can be seen from the graph (Figure 4.12). The second instead, the elbow method, was not as explicit as the other was, mainly because the right number of k is defined visually by looking at the graph. The point in which there is transition from a large to a small change in the derivative of the resulting sequence is the one to pick and could correspond either to 2 or 3.



Figures 4.12 and 4.13 Silhouette and elbow's results (respectively on the left and on the right) to choose the most suitable number of clusters. While it is well indicated from the graph on the left which one is the optimal

k in terms of largest average silhouette width, on the right is more a discretional matter, checking when there is no further improvement in increasing the number of k to have a lower total within sum of square value.

Then, the kmeans function allowed to access the results of the two clustering, the number of observations in each cluster, the centroid vectors (cluster means), the group in which each observation is allocated (clustering vector) and all the information about the intra and inter class variance of the clusters. The most useful information can be extracted from the intraclass variance and the interclass variance: in the first way we get how similar the observations are within the same group whereas from the second we get how far from each others the clusters are. There is no foolproof way to go on, but a metric of goodness for clustering is the BSS/TSS ratio, indicating that the cluster has the properties of internal cohesion and external separation. It should ideally approach 1, but in this case, it was a bit low and unsatisfying for both clustering fitting, anyway still higher with k = 3 (44.9 %).

Consequently, the best alternative was to fit a kmeans model the latter with 3 clusters. In order to study how companies and their financial ratios were represented inside the clusters during the years, many frequency tables were computed. It classified 90 out of 92 companies to belong to the first group, and the same happened with the benchmarks used for DEA in form of ranges. All these features belong mainly to the first cluster, maybe some other numerical variable can give more interesting results. By checking for revenues and total asset, it came out that the algorithm did have identified groups. Lower revenues and lower total assets belonged to the first cluster, higher revenues and higher total assets belonged to the second cluster the third class was around the average for both parameters (not reported here for compactness).

However, DEA efficiency distribution did not match with any of them. Since comparing results obtained from the previous analysis of the efficiency with what we got now was not possible, K means clustering comprehensive of the whole period was not significant with respect of this task. So, the hierarchical clustering was tried too. This other algorithm assigns iteratively the observations to the clusters accordingly to the dissimilarities and to the method of agglomeration used. It was decided to use other distance matrices, as well as three different clustering methods. At first, Ward and complete dendrograms were still built using the Euclidean distance. From figure 4. the difference between the two dendograms was quite evident. Since the two methods compared are very different and based on different concepts, we expected the two dendograms to have different shapes. The Ward's method tended to result in more balanced clusters, at least visually, compared to the complete linkage method. The optimal number of clusters were respectively 2 and 3, according to the dendograms larger height, but then thanks to the bidimensional projections it was possible to realize that even the Ward method dendrogram failed in its scope, as just very little pieces of information were embedded in the groups.

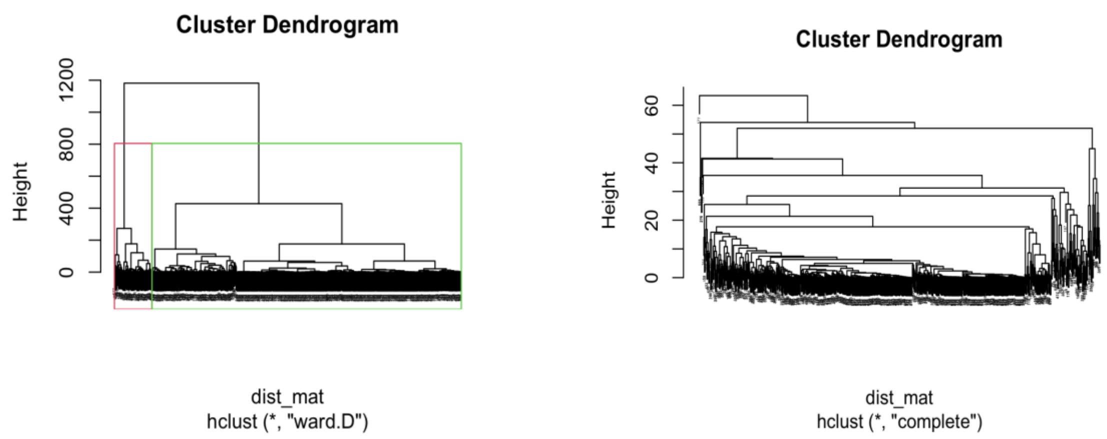


Figure 4.14 Ward and complete methods for dendrograms using the euclidean distance.

The previous results were disappointing, so a new distance matrix based on pairwise correlation between rows was applied, hoping for better results. Also in this case, two different methods were tried: average and complete. The steps to build the dendrograms were the same of the ones seen with euclidean distance and all showed 2 as optimal number of cluster. Last, clusters were projected into the 2D plane with the PCs and plotted as done previously, so to check for overlaps. To do this, information relative to the group of each single observation was extracted according to the clustering labelling which dendrograms suggested.

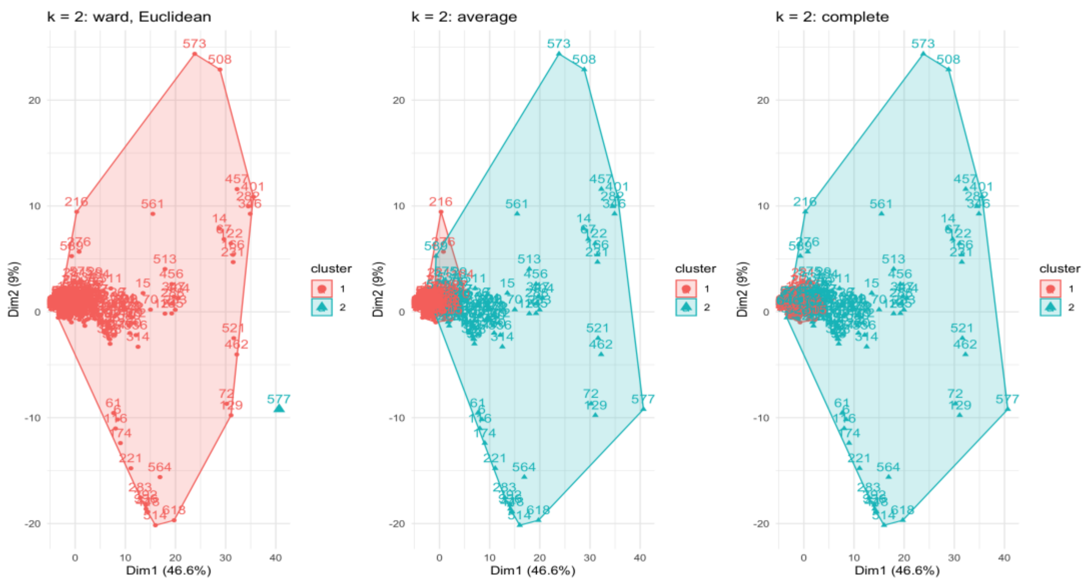


Figure 4.15 Bidimensional projections of clusters using different methods and distances

By investigating the plots in Figure 4.15, it appeared again that there was no as clear distinction between the two clusters generated by the average and complete linkage as there was with the k-means method clustering, but at least when compared to the Euclidean Ward, hierarchal algorithms with row correlation distance approximately distinguished the two groups. In addition, frequency table finally showed more similarities with companies' yearly efficiency levels built at the DEA stage. Here below a measure of accuracy:

	2010	*2011*	*2012*	*2013*	*2014*	*2015*	
Wrong guess	49	53	52	64	57	63	
Right guess	43	39	40	28	35	29	
	2016	2017	2018	2019	2020	2021	TOTAL
Wrong guess	53	60	55	56	59	61	682
Right guess	39	32	37	36	33	31	422

Table 4.4 Results from comparing complete linkage clustering labeling with the efficiencies (1 if the score was higher than 0.5, 0 otherwise) originated from the DEA

5. Conclusions

Financial ratios are presented in almost apologetic tones in most contemporary texts, and it appears though that their expected utility is very low. A financial ratio is held to be somewhat an ineffective predictor of financial difficulties if not framed in a broader context. The evidence bearing on this question is not overly abundant, but it appears that this general low opinion of the utility of financial ratios may have to be revised. Thanks to DEA and clustering techniques it was possible to investigate this field and doing an attempt to provide some answers. The key result of this research emphasizes the heterogeneity of companies, which is caused by their respective strategic preferences and by the constraints imposed by the environment, and that modern data analytics approaches are on the way to capture many of the different shades of this diversity.

For instance, Data Envelopment results lead to very practical interpretations, but it is important to understand that measures of efficiency across time may not be comparable. This is because DEA efficiency measures are relative to a frontier specific to a time period and that frontier may move over time. It is possible, though, to measure productivity change over time reflecting the combined effect of the change in a unit's efficiency over time and the movement of the frontiers against which those efficiencies have been measured.

Besides that, the process of data cleaning and variable selection aimed at finding the most complete observations and the most suitable parameters for the analysis, however, on the 92 firms considered, less than one tenth covered the entire 10 years timespan, and for sure this had affected the outcome of DEA. Unfortunately, there probably was a slight bias to the extent that only surviving firms over the period were included.

Hierarchical clustering using row correlations somehow intercepted a pattern like the efficiency classification previously made, the accuracy was not satisfactory but at least it could be improved. On average dendrogram the best number of clusters was two, while in k-means was assumed to be three. Of course, this is due the fact that the two methods uses different methodology to find clusters, but hierarchical algorithm seemed to be nearer the goal since the beginning.

It would be interesting to repeat this analysis in 10 years.

6. Appendix

Dataset's variables in alphabetic order:

	Variable Name	Type	Description
1.	CONM	string	Company Name (CONM)
2.	ISIN	string	International Security ID (ISIN)
3.	SEDOL	string	SEDOL (SEDOL)
4.	EXCHG	double	Stock Exchange Code (EXCHG)
5.	FYR	double	Fiscal Year-End (FYR)
6.	FIC	string	FIC -- ISO Country Code - Incorporation (FIC)
7.	ADD1	string	ADD1 -- Address Line 1 (ADD1)
8.	ADD2	string	ADD2 -- Address Line 2 (ADD2)
9.	ADDZIP	string	ADDZIP -- Postal Code (ADDZIP)
10.	BUSDESC	string	BUSDESC -- S&P Business Description (BUSDESC)
11.	CIK	string	CIK -- CIK Number (CIK)
12.	CITY	string	CITY -- City (CITY)
13.	CONML	string	CONML -- Company Legal Name (CONML)
14.	COUNTY	string	COUNTY -- County Code (COUNTY)
15.	DLRSN	string	DLRSN -- Research Co Reason for Deletion (DLRSN)
16.	EIN	string	EIN -- Employer Identification Number (EIN)
17.	FAX	string	FAX -- Fax Number (FAX)
18.	FYRC	double	FYRC -- Current Fiscal Year End Month (FYRC)
19.	GGROUP	string	GGROUP -- GIC Groups (GGROUP)
20.	GIND	string	GIND -- GIC Industries (GIND)
21.	GSECTOR	string	GSECTOR -- GIC Sectors (GSECTOR)
22.	GSUBIND	string	GSUBIND -- GIC Sub-Industries (GSUBIND)
23.	IDBFLAG	string	IDBFLAG -- International, Domestic, Both Indicator (IDBFLAG)
24.	INCORP	string	INCORP -- Current State/Province of Incorporation Code (INCORP)
25.	LOC	string	LOC -- Current ISO Country Code - Headquarters (LOC)
26.	NAICS	string	NAICS -- North American Industry Classification Code (NAICS)
27.	PHONE	string	PHONE -- Phone Number (PHONE)
28.	PRICAN	string	PRICAN -- Current Primary Issue Tag - Canada (PRICAN)
29.	PRIROW	string	PRIROW -- Primary Issue Tag - Rest of World (PRIROW)

30. PRIUSA	string	PRIUSA -- Current Primary Issue Tag - US (PRIUSA)
31. SIC	string	SIC -- Standard Industry Classification Code (SIC)
32. SPCINDCD	double	SPCINDCD -- S&P Industry Sector Code (SPCINDCD)
33. SPCSECCD	double	SPCSECCD -- S&P Economic Sector Code (SPCSECCD)
34. SPCSRC	string	SPCSRC -- S&P Quality Ranking - Current (SPCSRC)
35. STATE	string	STATE -- State/Province (STATE)
36. STKO	double	STKO -- Stock Ownership Code (STKO)
37. WEBURL	string	WEBURL -- Web URL (WEBURL)
38. DLDTE	date	DLDTE -- Research Company Deletion Date (DLDTE)
39. IPODATE	date	IPODATE -- Company Initial Public Offering Date (IPODATE)
40. ACCTSTD	string	ACCTSTD -- Accounting Standard (ACCTSTD)
41. ACQMETH	string	ACQMETH -- Acquisition Method (ACQMETH)
42. BSPR	string	BSPR -- Balance Sheet Presentation (BSPR)
43. COMPST	string	COMPST -- Comparability Status (COMPST)
44. CURCD	string	CURCD -- ISO Currency Code (CURCD)
45. FINAL	string	FINAL -- Final Indicator Flag (FINAL)
46. FYEAR	double	FYEAR -- Data Year - Fiscal (FYEAR)
47. ISMOD	double	ISMOD -- Income Statement Model Number (ISMOD)
48. PDDUR	double	PDDUR -- Period Duration (PDDUR)
49. SCF	double	SCF -- Cash Flow Format (SCF)
50. SRC	double	SRC -- Source Document (SRC)
51. STALT	string	STALT -- Status Alert (STALT)
52. UPD	double	UPD -- Update Code (UPD)
53. FDATE	date	FDATE -- Final Date (FDATE)
54. PDATE	date	PDATE -- Preliminary Date (PDATE)
55. ACCLI	double	ACCLI -- Accrued Liabilities - Increase/(Decrease) (ACCLI)
56. ACCO	double	ACCO -- Acceptances Outstanding (ACCO)
57. ACO	double	ACO -- Current Assets - Other - Total (ACO)
58. ACOFS	double	ACOFS -- Other Current Assets - Total - FS (Memo) (ACOFS)
59. ACOX	double	ACOX -- Current Assets - Other - Sundry (ACOX)
60. ACOXFS	double	ACOXFS -- Other Current Assets - FS (Memo) (ACOXFS)
61. ACQDISN	double	ACQDISN -- Acquisitions and Disposals - Net Cash Flow (ACQDISN)
62. ACQDISO	double	ACQDISO -- Acquisitions and Disposals - Other (ACQDISO)
63. ACT	double	ACT -- Current Assets - Total (ACT)

64. ADPAC	double	ADPAC -- Amortization of Deferred Policy Acquisition Costs (ADPAC)
65. AM	double	AM -- Amortization of Intangibles (AM)
66. AMDC	double	AMDC -- Amortization of Deferred Charges (AMDC)
67. AO	double	AO -- Assets - Other (AO)
68. AOLOCH	double	AOLOCH -- Assets and Liabilities - Other - Net Change (AOLOCH)
69. AOX	double	AOX -- Assets - Other - Sundry (AOX)
70. AP	double	AP -- Accounts Payable - Trade (AP)
71. APALCH	double	APALCH -- Accounts Payable and Accrued Liabilities - Increase/(Decrease) (APALCH)
72. APCH	double	APCH -- Accounts Payable - Increase (Decrease) (APCH)
73. APDPFS	double	APDPFS -- Customer Deposits Short Term - FS (Memo) (APDPFS)
74. APFS	double	APFS -- Trade Accounts Payable - FS (Memo) (APFS)
75. APO	double	APO -- Accounts Payable - Other (APO)
76. APOFS	double	APOFS -- Accounts Payable/Creditors - Other - FS (APOFS)
77. AQC	double	AQC -- Acquisitions (AQC)
78. ARTFS	double	ARTFS -- Accounts Receivable/Debtors - Total (ARTFS)
79. ASDIS	double	ASDIS -- Associated Undertakings - Disposal (ASDIS)
80. ASINV	double	ASINV -- Associated Undertakings - Investment (ASINV)
81. AT	double	AT -- Assets - Total (AT)
82. ATOCH	double	ATOCH -- Assets - Other - Change (ATOCH)
83. AU	string	AU -- Auditor (AU)
84. AUOP	string	AUOP -- Auditor Opinion (AUOP)
85. AUTXR	double	AUTXR -- Appropriations to Untaxed Reserves (AUTXR)
86. BCEF	double	BCEF -- Brokerage, Clearing and Exchange Fees (BCEF)
87. BCT	double	BCT -- Benefits and Claims - Total (Insurance) (BCT)
88. CA	double	CA -- Customers' Acceptance (CA)
89. CAPCST	double	CAPCST -- Capitalized Costs (CAPCST)
90. CAPFL	double	CAPFL -- Capital Element of Finance Lease Rental Payments (CAPFL)
91. CAPR1	double	CAPR1 -- Risk-Adjusted Capital Ratio - Tier 1 (CAPR1)
92. CAPR2	double	CAPR2 -- Risk-Adjusted Capital Ratio - Tier 2 (CAPR2)
93. CAPR3	double	CAPR3 -- Risk-Adjusted Capital Ratio - Combined (CAPR3)
94. CAPRT	double	CAPRT -- Risk-Adjusted Capital Ratio - Total (CAPRT)
95. CAPS	double	CAPS -- Capital Surplus/Share Premium Reserve (CAPS)
96. CAPX	double	CAPX -- Capital Expenditures (CAPX)
97. CAPXFI	double	CAPXFI -- Capital Expenditures and Financial Investment - Net Cash Flow (CAPXFI)

98. CEQ	double	CEQ -- Common/Ordinary Equity - Total (CEQ)
99. CFBD	double	CFBD -- Commissions and Fees - (Broker/Dealer) (CFBD)
100.CFERE	double	CFERE -- Commissions and Fees - (Real Estate) (CFERE)
101.CFLAOTH	double	CFLAOTH -- Cash Flow Adjustments - Other (CFLAOTH)
102.CFO	double	CFO -- Commissions and Fees - Other (CFO)
103.CFPDO	double	CFPDO -- Commissions and Fees Paid - Other (CFPDO)
104.CGA	double	CGA -- Capital Gains - After-Tax (CGA)
105.CH	double	CH -- Cash (CH)
106.CHE	double	CHE -- Cash and Short-Term Investments (CHE)
107.CHEB	double	CHEB -- Cash and Cash Equivalents at Beginning of Year (CHEB)
108.CHECH	double	CHECH -- Cash and Cash Equivalents - Increase/(Decrease) (CHECH)
109.CHEE	double	CHEE -- Cash and Cash Equivalents at End of Year (CHEE)
110.CHEFS	double	CHEFS -- Cash and Short Term Investments Total - FS (Memo) (CHEFS)
111.CHENFD	double	CHENFD -- Cash/Cash Equivalents/Net Funds - Increase/(Decrease) (CHENFD)
112.CHFS	double	CHFS -- Cash - FS (Memo) (CHFS)
113.CHS	double	CHS -- Cash and Deposits - Segregated (CHS)
114.CMP	double	CMP -- Commercial Paper (CMP)
115.COGS	double	COGS -- Cost of Goods Sold (COGS)
116.CRVNLI	double	CRVNLI -- Reserves for Claims (Losses) - Nonlife (Insurance) (CRVNLI)
117.CSHR	double	CSHR -- Common/Ordinary Shareholders (CSHR)
118.CSTK	double	CSTK -- Common/Ordinary Stock (Capital) (CSTK)
119.CUSTADV	double	CUSTADV -- Customer Advances (CUSTADV)
120.DBTB	double	DBTB -- Debt at Beginning of Year (DBTB)
121.DBTE	double	DBTE -- Debt at End of Year (DBTE)
122.DC	double	DC -- Deferred Charges (DC)
123.DCSFD	double	DCSFD -- Current Debt - Source of Funds (DCSFD)
124.DCUFD	double	DCUFD -- Current Debt - Use of Funds (DCUFD)
125.DD1	double	DD1 -- Long-Term Debt Due in One Year (DD1)
126.DD1FS	double	DD1FS -- Long Term Debt - Current Portion - FS (Memo) (DD1FS)
127.DFPAC	double	DFPAC -- Deferred Policy Acquisition Costs (DFPAC)
128.DFXA	double	DFXA -- Depreciation of Tangible Fixed Assets (DFXA)
129.DISPOCH	double	DISPOCH -- Disposals - Other - (Gain) Loss (DISPOCH)
130.DLC	double	DLC -- Debt in Current Liabilities - Total (DLC)
131.DLCCH	double	DLCCH -- Current Debt - Changes (DLCCH)

132.DLCFS	double	DLCFS -- Short Term Debt Total - FS (Memo) (DLCFS)
133.DLTIS	double	DLTIS -- Long-Term Debt - Issuance (DLTIS)
134.DLTR	double	DLTR -- Long-Term Debt - Reduction (DLTR)
135.DLTT	double	DLTT -- Long-Term Debt - Total (DLTT)
136.DO	double	DO -- Discontinued Operations (DO)
137.DOC	double	DOC -- Discontinued Operations (Cash Flow) (DOC)
138.DP	double	DP -- Depreciation and Amortization (DP)
139.DPACT	double	DPACT -- Depreciation, Depletion and Amortization (Accumulated) (DPACT)
140.DPC	double	DPC -- Depreciation and Amortization (Cash Flow) (DPC)
141.DPDC	double	DPDC -- Deposits - Demand - Customer (DPDC)
142.DPLTB	double	DPLTB -- Deposits - Long-Term Time - Bank (DPLTB)
143.DPSC	double	DPSC -- Deposits - Savings - Customer (DPSC)
144.DPSTB	double	DPSTB -- Deposits - Short-Term Demand - Bank (DPSTB)
145.DPTB	double	DPTB -- Deposits - Total - Banks (DPTB)
146.DPTC	double	DPTC -- Deposits - Total - Customer (DPTC)
147.DPTIC	double	DPTIC -- Deposits - Time - Customer (DPTIC)
148.DV	double	DV -- Cash Dividends (Cash Flow) (DV)
149.DVC	double	DVC -- Dividends Common/Ordinary (DVC)
150.DVP	double	DVP -- Dividends - Preferred/Preference (DVP)
151.DVPDP	double	DVPDP -- Dividends and Bonuses Paid Policyholders (DVPDP)
152.DVREC	double	DVREC -- Dividends Received (Cash Flow) (DVREC)
153.DVRRE	double	DVRRE -- Development Revenue (Real Estate) (DVRRE)
154.DVSCO	double	DVSCO -- Dividends - Share Capital - Other (DVSCO)
155.DVT	double	DVT -- Dividends - Total (DVT)
156.EA	double	EA -- Exchange Adjustments (Assets) (EA)
157.EBIT	double	EBIT -- Earnings Before Interest and Taxes (EBIT)
158.EBITDA	double	EBITDA -- Earnings Before Interest (EBITDA)
159.EIEA	double	EIEA -- Equity in Earnings - After-Tax (EIEA)
160.EIEAC	double	EIEAC -- Equity Interest in Earnings of Associated Companies (EIEAC)
161.EMP	double	EMP -- Employees (EMP)
162.EQDIVP	double	EQDIVP -- Equity Dividend Paid (EQDIVP)
163.ERO	double	ERO -- Equity Reserves - Other (ERO)
164.EXRE	double	EXRE -- Exchange Rate Effect (EXRE)
165.EXRES	double	EXRES -- Exchange Rate Effect - Source of Funds (EXRES)

166.EXREU	double	EXREU -- Exchange Rate Effect - Use of Funds (EXREU)
167.FATB	double	FATB -- Property, Plant, and Equipment - Buildings at Cost (FATB)
168.FATE	double	FATE -- Property, Plant, and Equipment - Machinery and Equipment at Cost (FATE)
169.FATL	double	FATL -- Property, Plant, and Equipment - Leases at Cost (FATL)
170.FATP	double	FATP -- Property, Plant, and Equipment - Land and Improvements at Cost (FATP)
171.FCA	double	FCA -- Foreign Exchange Income (Loss) (FCA)
172.FDFR	double	FDFR -- Federal Funds Purchased (FDFR)
173.FEA	double	FEA -- Foreign Exchange Assets (FEA)
174.FEL	double	FEL -- Foreign Exchange Liabilities (FEL)
175.FFS	double	FFS -- Federal Funds Sold (FFS)
176.FIAO	double	FIAO -- Financing Activities - Other (FIAO)
177.FINCF	double	FINCF -- Financing Activities - Net Cash Flow (FINCF)
178.FININC	double	FININC -- Financing Increase - Total (FININC)
179.FINLE	double	FINLE -- Finance Lease Increases (FINLE)
180.FINRE	double	FINRE -- Financing Repayments/Reductions - Total (FINRE)
181.FINVAO	double	FINVAO -- Funds from Investment and Finance Activities - Other (FINVAO)
182.FOPO	double	FOPO -- Funds from Operations - Other (FOPO)
183.FSRCO	double	FSRCO -- Sources of Funds - Other (FSRCO)
184.FSRCOPO	double	FSRCOPO -- Sources of Operating Funds - Other (FSRCOPO)
185.FSRCOPT	double	FSRCOPT -- Source of Funds From Operations - Total (FSRCOPT)
186.FSRCT	double	FSRCT -- Sources of Funds - Total (FSRCT)
187.FUSEO	double	FUSEO -- Uses of Funds - Other (FUSEO)
188.FUSET	double	FUSET -- Uses of Funds - Total (FUSET)
189.GDWL	double	GDWL -- Goodwill (GDWL)
190.IAEQ	double	IAEQ -- Investment Assets - Equity Securities (Insurance) (IAEQ)
191.IAFXI	double	IAFXI -- Investment Assets - Fixed Income Securities (Insurance) (IAFXI)
192.IALOI	double	IALOI -- Investment Assets - Loans - Other (Insurance) (IALOI)
193.IALTI	double	IALTI -- Investment Assets - Loans - Total (Insurance) (IALTI)
194.IAMLI	double	IAMLI -- Investment Assets - Mortgage Loans (Insurance) (IAMLI)
195.IAOI	double	IAOI -- Investment Assets - Other (Insurance) (IAOI)
196.IAPLI	double	IAPLI -- Investment Assets - Policy Loans (Insurance) (IAPLI)
197.IAREI	double	IAREI -- Investment Assets - Real Estate (Insurance) (IAREI)
198.IASSI	double	IASSI -- Investment Assets - Securities - Sundry (Insurance) (IASSI)
199.IASTI	double	IASTI -- Investment Assets - Securities - Total (Insurance) (IASTI)

200.IATI	double	IATI -- Investment Assets - Total (Insurance) (IATI)
201.IB	double	IB -- Income Before Extraordinary Items (IB)
202.IBC	double	IBC -- Income Before Extraordinary Items (Cash Flow) (IBC)
203.IBKI	double	IBKI -- Investment Banking Income (IBKI)
204.IBMII	double	IBMII -- Income before Extraordinary Items and Noncontrolling Interests (IBMII)
205.ICAPT	double	ICAPT -- Invested Capital - Total (ICAPT)
206.IDIIS	double	IDIIS -- Interest and Dividend Income - Investment Securities (IDIIS)
207.IDILB	double	IDILB -- Interest and Dividend Income - Loans/Claims/Advances - Banks (IDILB)
208.IDILC	double	IDILC -- Interest and Dividend Income - Loans/Claims/Advances - Customers (IDILC)
209.IDIS	double	IDIS -- Interest and Dividend Income - Sundry (IDIS)
210.IDIST	double	IDIST -- Interest and Dividend Income - Short-Term Investments (IDIST)
211.IDIT	double	IDIT -- Interest and Related Income - Total (IDIT)
212.IDITS	double	IDITS -- Interest and Dividend Income - Trading Securities (IDITS)
213.IIRE	double	IIRE -- Investment Income (Real Estate) (IIRE)
214.INITB	double	INITB -- Income - Non-interest - Total (Bank) (INITB)
215.INTAN	double	INTAN -- Intangible Assets - Total (INTAN)
216.INTAND	double	INTAND -- Intangible Assets - Disposal (INTAND)
217.INTANP	double	INTANP -- Intangible Assets - Purchase (INTANP)
218.INTC	double	INTC -- Interest Capitalized (INTC)
219.INTFACT	double	INTFACT -- Interest and Dividend Adjustments - Financing Activities (INTFACT)
220.INTFL	double	INTFL -- Interest Element of Finance Leases (INTFL)
221.INTIACT	double	INTIACT -- Interest and Dividend Adjustments - Investing Activities (INTIACT)
222.INTOACT	double	INTOACT -- Interest and Dividend Adjustments - Operating Activities (INTOACT)
223.INTPD	double	INTPD -- Interest Paid (INTPD)
224.INTPN	double	INTPN -- Interest Paid - Net (INTPN)
225.INTRC	double	INTRC -- Interest Received (INTRC)
226.INVCH	double	INVCH -- Inventory - Decrease (Increase) (INVCH)
227.INVDSP	double	INVDSP -- Investments - Disposal (INVDSP)
228.INVFG	double	INVFG -- Inventories - Finished Goods (INVFG)
229.INVO	double	INVO -- Inventories - Other (INVO)
230.INVRM	double	INVRM -- Inventories - Raw Materials (INVRM)
231.INVSVC	double	INVSVC -- Investments and Servicing of Finance - Net Cash Flow (INVSVC)
232.INVT	double	INVT -- Inventories - Total (INVT)
233.INVTFS	double	INVTFS -- Inventories - FS (Memo) (INVTFS)

234.INVWIP	double	INVWIP -- Inventories - Work In Process (INVWIP)
235.IOBD	double	IOBD -- Income - Other (Broker Dealer) (IOBD)
236.IOI	double	IOI -- Income - Other (Insurance) (IOI)
237.IORE	double	IORE -- Income - Other (Real Estate) (IORE)
238.IP	double	IP -- Investment Property (IP)
239.IPTI	double	IPTI -- Insurance Premiums - Total (Insurance) (IPTI)
240.ISGR	double	ISGR -- Investment Securities - Gain (Loss) - Realized (ISGR)
241.ISGT	double	ISGT -- Investment Securities - Gain (Loss) - Total (ISGT)
242.ISGU	double	ISGU -- Investment Securities - Gain (Loss) - Unrealized (ISGU)
243.ISOTH	double	ISOTH -- Investment Securities - Other (ISOTH)
244.IST	double	IST -- Investment Securities -Total (IST)
245.IVACO	double	IVACO -- Investing Activities - Other (IVACO)
246.IVAEQ	double	IVAEQ -- Investment and Advances - Equity (IVAEQ)
247.IVAO	double	IVAO -- Investment and Advances - Other (IVAO)
248.IVCH	double	IVCH -- Increase in Investments (IVCH)
249.IVGOD	double	IVGOD -- Investments Grants and Other Deductions (IVGOD)
250.IVI	double	IVI -- Investment Income - Total (Insurance) (IVI)
251.IVNCF	double	IVNCF -- Investing Activities - Net Cash Flow (IVNCF)
252.IVPT	double	IVPT -- Investments - Permanent - Total (IVPT)
253.IVST	double	IVST -- Short-Term Investments - Total (IVST)
254.IVSTCH	double	IVSTCH -- Short-Term Investments - Change (IVSTCH)
255.IVSTFS	double	IVSTFS -- Short Term Investments - FS (Memo) (IVSTFS)
256.LCABG	double	LCABG -- Loans/Claims/Advances - Banks and Government - Total (LCABG)
257.LCACL	double	LCACL -- Loans/Claims/Advances - Commercial (LCACL)
258.LCACR	double	LCACR -- Loans/Claims/Advances - Consumer (LCACR)
259.LCAG	double	LCAG -- Loans/Claims/Advances - Government (LCAG)
260.LCAL	double	LCAL -- Loans/Claims/Advances - Lease (LCAL)
261.LCALT	double	LCALT -- Loans/Claims/Advances - Long-Term (Banks) (LCALT)
262.LCAM	double	LCAM -- Loans/Claims/Advances - Mortgage (LCAM)
263.LCAO	double	LCAO -- Loans/Claims/Advances - Other (LCAO)
264.LCAST	double	LCAST -- Loans/Claims/Advances - Short-Term - Banks (LCAST)
265.LCAT	double	LCAT -- Loans/Claims/Advances - Total (LCAT)
266.LCO	double	LCO -- Current Liabilities - Other - Total (LCO)
267.LCOFS	double	LCOFS -- Other Current Liabilities - FS (Memo) (LCOFS)

268.LCOX	double	LCOX -- Current Liabilities - Other - Sundry (LCOX)
269.LCT	double	LCT -- Current Liabilities - Total (LCT)
270.LCTFS	double	LCTFS -- Other Current Liabilities - Total - FS (Memo) (LCTFS)
271.LCUACU	double	LCUACU -- Loans/Claims/Advances - Customer - Total (LCUACU)
272.LIQRESN	double	LIQRESN -- Management of Liquid Resources - Net Cash Flow (LIQRESN)
273.LIQRESO	double	LIQRESO -- Liquid Resources - Other Movements (LIQRESO)
274.LNDEP	double	LNDEP -- Loans and Deposits - (Increase) Decrease (LNDEP)
275.LNINC	double	LNINC -- Loan Increase/Additions (LNINC)
276.LNMD	double	LNMD -- Loans (Made)/Repaid (LNMD)
277.LNREP	double	LNREP -- Loan Repayments/Reductions (LNREP)
278.LO	double	LO -- Liabilities - Other - Total (LO)
279.LSE	double	LSE -- Liabilities and Stockholders Equity - Total (LSE)
280.LT	double	LT -- Liabilities - Total (LT)
281.LTDCH	double	LTDCH -- Long-Term Debt - Change (LTDCH)
282.LTDLCH	double	LTDLCH -- Long-Term Debt/Liabilities - Change (LTDLCH)
283.LTLO	double	LTLO -- Long-Term Liabilities - Other - Increase/(Decrease) (LTLO)
284.MIB	double	MIB -- Noncontrolling Interest (Balance Sheet) (MIB)
285.MIBN	double	MIBN -- Noncontrolling Interests - Nonredeemable - Balance Sheet (MIBN)
286.MIBT	double	MIBT -- Noncontrolling Interests - Total - Balance Sheet (MIBT)
287.MIC	double	MIC -- Noncontrolling Interest (Cash Flow) (MIC)
288.MII	double	MII -- Noncontrolling Interest (Income Account) (MII)
289.MISEQ	double	MISEQ -- Noncontrolling Interest In Stockholders Equity > Change (MISEQ)
290.MTL	double	MTL -- Loans From Securities Finance Companies for Margin Transactions (MTL)
291.NCFLIQ	double	NCFLIQ -- Net Cash Flow Before Management of Liquid Resources and Financing (NCFLIQ)
292.NEQMI	double	NEQMI -- Non-Equity and Noncontrolling Interest Dividends Paid (NEQMI)
293.NIO	double	NIO -- Net Items - Other (NIO)
294.NIT	double	NIT -- Net Item - Total (NIT)
295.NOASUB	double	NOASUB -- Net Overdrafts Acquired with Subsidiaries (NOASUB)
296.NOPI	double	NOPI -- Nonoperating Income (Expense) (NOPI)
297.NP	double	NP -- Notes Payable - Short-Term Borrowings (NP)
298.NPANL	double	NPANL -- Nonperforming Assets - Nonaccrual Loans (NPANL)
299.NPAORE	double	NPAORE -- Nonperforming Assets - Other Real Estate Owned (NPAORE)
300.NPARL	double	NPARL -- Nonperforming Assets - Restructured Loans (NPARL)
301.NPAT	double	NPAT -- Nonperforming Assets - Total (NPAT)

302.NPFS	double	NPFS -- Short Term Borrowings - FS (Memo) (NPFS)
303.OANCF	double	OANCF -- Operating Activities - Net Cash Flow (OANCF)
304.OANCFC	double	OANCFC -- Operating Activities - Net Cash Flow - Continuing Operations (OANCFC)
305.OANCFD	double	OANCFD -- Operating Activities - Net Cash Flow - Discontinued Operations (OANCFD)
306.OIADP	double	OIADP -- Operating Income After Depreciation (OIADP)
307.OIBDP	double	OIBDP -- Operating Income Before Depreciation (OIBDP)
308.ONBALB	double	ONBALB -- Other Net Balances at Beginning of Year (ONBALB)
309.ONBALE	double	ONBALE -- Other Net Balances at End of Year (ONBALE)
310.OPPRFT	double	OPPRFT -- Operating Profit (OPPRFT)
311.PACQP	double	PACQP -- Preacquisition Profits (PACQP)
312.PCL	double	PCL -- Provision - Credit Losses (Income Account) (PCL)
313.PI	double	PI -- Pretax Income (PI)
314.PLIACH	double	PLIACH -- Pension Liabilities - Change (PLIACH)
315.PPEGT	double	PPEGT -- Property, Plant and Equipment - Total (Gross) (PPEGT)
316.PPENT	double	PPENT -- Property, Plant and Equipment - Total (Net) (PPENT)
317.PRC	double	PRC -- Participation Rights Certificates (PRC)
318.PRODV	double	PRODV -- Proposed Dividends (PRODV)
319.PROSAI	double	PROSAI -- Proceeds From Sale of Fixed Assets and Sale of Investments (PROSAI)
320.PRSTKC	double	PRSTKC -- Purchase of Common and Preferred Stock (PRSTKC)
321.PRV	double	PRV -- Provisions (Cash Flow) (PRV)
322.PSFIX	double	PSFIX -- Proceeds From Sale of Fixed Assets (PSFIX)
323.PSTK	double	PSTK -- Preferred/Preference Stock (Capital) - Total (PSTK)
324.PSTKN	double	PSTKN -- Preferred/Preference Stock - Nonredeemable (PSTKN)
325.PSTKR	double	PSTKR -- Preferred/Preference Stock - Redeemable (PSTKR)
326.PTRAN	double	PTRAN -- Principal Transactions (PTRAN)
327.PURTSHR	double	PURTSHR -- Purchase of Treasury Shares (PURTSHR)
328.PVON	double	PVON -- Provisions - Other (Net) (PVON)
329.PVT	double	PVT -- Provisions - Total (PVT)
330.RADP	double	RADP -- Reinsurance Assets - Deposits and Other (Insurance) (RADP)
331.RAGR	double	RAGR -- Resale Agreements (RAGR)
332.RARI	double	RARI -- Reinsurance Assets - Receivable/Debtors (Insurance) (RARI)
333.RATI	double	RATI -- Reinsurance Assets - Total (Insurance) (RATI)
334.RAWMSM	double	RAWMSM -- Raw Materials, Supplies, and Merchandise (RAWMSM)
335.RCL	double	RCL -- Reserves for Credit Losses (Assets) (RCL)

336.RE	double	RE -- Retained Earnings (RE)
337.RECCH	double	RECCH -- Accounts Receivable - Decrease (Increase) (RECCH)
338.RECCO	double	RECCO -- Receivables - Current - Other (RECCO)
339.RECCOFS	double	RECCOFS -- Receivables - Other - FS (Memo) (RECCOFS)
340.RECT	double	RECT -- Receivables - Total (RECT)
341.RECTFS	double	RECTFS -- Receivables - Total - FS (Memo) (RECTFS)
342.RECTR	double	RECTR -- Receivables - Trade (RECTR)
343.RECTRFS	double	RECTRFS -- Receivables - Trade - FS (Memo) (RECTRFS)
344.REVT	double	REVT -- Revenue - Total (REVT)
345.RIS	double	RIS -- Revenue/Income - Sundry (RIS)
346.RLRI	double	RLRI -- Reinsurers" Liability for Reserves (Insurance) (RLRI)
347.RLT	double	RLT -- Reinsurance Liabilities - Total (RLT)
348.RPAG	double	RPAG -- Repurchase Agreements (RPAG)
349.RV	double	RV -- Reserves (RV)
350.RVBCI	double	RVBCI -- Reserves for Benefits - Life - Claims (Insurance) (RVBCI)
351.RVBPI	double	RVBPI -- Reserves for Benefits - Life - Policy (Insurance) (RVBPI)
352.RVBTI	double	RVBTI -- Reserves for Benefits - Life - Total (Insurance) (RVBTI)
353.RVEQT	double	RVEQT -- Equity Reserves - Total (RVEQT)
354.RVLRV	double	RVLRV -- Revaluation Reserve (RVLRV)
355.RVRI	double	RVRI -- Reserves - Reinsurance (Insurance) (RVRI)
356.RVSI	double	RVSI -- Reserves - Sundry (Insurance) (RVSI)
357.RVTI	double	RVTI -- Reserves - Total (RVTI)
358.RVUPI	double	RVUPI -- Reserves for Unearned Premiums (Insurance) (RVUPI)
359.RVUTX	double	RVUTX -- Reserves - Untaxed (RVUTX)
360.SAA	double	SAA -- Separate Account Assets (SAA)
361.SAL	double	SAL -- Separate Account Liabilities (SAL)
362.SALE	double	SALE -- Sales/Turnover (Net) (SALE)
363.SBDC	double	SBDC -- Securities Borrowed and Deposited by Customers (SBDC)
364.SC	double	SC -- Securities In Custody (SC)
365.SCO	double	SCO -- Share Capital - Other (SCO)
366.SEQ	double	SEQ -- Stockholders Equity - Parent (SEQ)
367.SHRCAP	double	SHRCAP -- Share Capital Transactions - Other (SHRCAP)
368.SIV	double	SIV -- Sale of Investments (SIV)
369.SPI	double	SPI -- Special Items (SPI)

370.SPPCH	double	SPPCH -- Sale of Fixed Assets - (Gain) Loss (SPPCH)
371.SPPIV	double	SPPIV -- Sale of Property, Plant and Equipment and Investments - Gain (Loss) (SPPIV)
372.SSNP	double	SSNP -- Securities Sold Not Yet Purchased (SSNP)
373.SSTK	double	SSTK -- Sale of Common and Preferred Stock (SSTK)
374.STBO	double	STBO -- Short-Term Borrowings - Other (STBO)
375.STFIXA	double	STFIXA -- Sale of Tangible Fixed Assets (STFIXA)
376.STINV	double	STINV -- Short Term Investments - (Increase)/Decrease (STINV)
377.STIO	double	STIO -- Short-Term Investments - Other (STIO)
378.STKCH	double	STKCH -- Change in Stocks (STKCH)
379.SUBDIS	double	SUBDIS -- Subsidiary Undertakings - Disposal (SUBDIS)
380.SUBPUR	double	SUBPUR -- Subsidiary Undertakings - Purchase (SUBPUR)
381.TDSG	double	TDSG -- Trading/Dealing Securities - Gain (Loss) (TDSG)
382.TDST	double	TDST -- Trading/Dealing Account Securities - Total (TDST)
383.TEQ	double	TEQ -- Stockholders Equity - Total (TEQ)
384.TRANSA	double	TRANSA -- Cumulative Translation Adjustment (TRANSA)
385.TSCA	double	TSCA -- Treasury Stock (Current Asset) (TSCA)
386.TSTK	double	TSTK -- Treasury Stock - Total (All Capital) (TSTK)
387.TSTLTA	double	TSTLTA -- Treasury Stock (Long-Term Asset) (TSTLTA)
388.TX	double	TX -- Taxation (TX)
389.TXC	double	TXC -- Income Taxes - Current (TXC)
390.TXDB	double	TXDB -- Deferred Taxes (Balance Sheet) (TXDB)
391.TXDC	double	TXDC -- Deferred Taxes (Cash Flow) (TXDC)
392.TXDI	double	TXDI -- Income Taxes - Deferred (TXDI)
393.TXDITC	double	TXDITC -- Deferred Taxes and Investment Tax Credit (TXDITC)
394.TXO	double	TXO -- Income Taxes - Other (TXO)
395.TXOP	double	TXOP -- Taxation - Operating Activities (TXOP)
396.TXP	double	TXP -- Income Taxes Payable (TXP)
397.TXPD	double	TXPD -- Income Taxes Paid (TXPD)
398.TXPFS	double	TXPFS -- Taxes Payable - Current - FS (Memo) (TXPFS)
399.TXT	double	TXT -- Income Taxes - Total (TXT)
400.TXW	double	TXW -- Excise Taxes (TXW)
401.UI	double	UI -- Unearned Income (UI)
402.UNL	double	UNL -- Unappropriated Net Loss (UNL)
403.UNNP	double	UNNP -- Unappropriated Net Profit (Stockholders" Equity) (UNNP)

404.VPAC	double	VPAC -- Investments - Permanent - Associated Companies (VPAC)
405.VPO	double	VPO -- Investments - Permanent - Other (VPO)
406.WCAP	double	WCAP -- Working Capital (Balance Sheet) (WCAP)
407.WCAPCH	double	WCAPCH -- Working Capital Change - Total (WCAPCH)
408.WCAPCHC	double	WCAPCHC -- Working Capital - Change (WCAPCHC)
409.WCAPOPC	double	WCAPOPC -- Working Capital/Net Operating Assets - Change (WCAPOPC)
410.WCAPS	double	WCAPS -- Working Capital Change - Source of Funds (WCAPS)
411.WCAPSA	double	WCAPSA -- Working Capital Change (Separate Account) (WCAPSA)
412.WCAPSU	double	WCAPSU -- Source and Use of Funds/Working Capital Adjustments - Other (WCAPSU)
413.WCAPT	double	WCAPT -- Working Capital/Cash/Net Funds Change - Total (WCAPT)
414.WCAPU	double	WCAPU -- Working Capital Change - Use of Funds (WCAPU)
415.XACC	double	XACC -- Accrued Expenses (XACC)
416.XACCFS	double	XACCFS -- Accrued Expenses & Deferred Income - FS (Memo) (XACCFS)
417.XAGO	double	XAGO -- Administrative and General Expense - Other (XAGO)
418.XAGT	double	XAGT -- Administrative and General Expense - Total (XAGT)
419.XCOM	double	XCOM -- Communications Expense (XCOM)
420.XCOMI	double	XCOMI -- Commissions Expense (Insurance) (XCOMI)
421.XDVRE	double	XDVRE -- Expense - Development (Real Estate) (XDVRE)
422.XEQO	double	XEQO -- Equipment and Occupancy Expense (XEQO)
423.XI	double	XI -- Extraordinary Items (XI)
424.XIDO	double	XIDO -- Extraordinary Items and Discontinued Operations (XIDO)
425.XIDOC	double	XIDOC -- Extraordinary Items and Discontinued Operations (Cash Flow) (XIDOC)
426.XINDB	double	XINDB -- Interest Expense - Deposits - Banks (XINDB)
427.XINDC	double	XINDC -- Interest Expense - Deposits - Customer (XINDC)
428.XINS	double	XINS -- Interest Expense - Sundry (XINS)
429.XINST	double	XINST -- Interest Expense - Short-Term Borrowings (XINST)
430.XINT	double	XINT -- Interest and Related Expense - Total (XINT)
431.XINTD	double	XINTD -- Interest Expense - Long-Term Debt (XINTD)
432.XIVI	double	XIVI -- Investment Expense (Insurance) (XIVI)
433.XIVRE	double	XIVRE -- Expense - Investment (Real Estate) (XIVRE)
434.XLR	double	XLR -- Staff Expense - Total (XLR)
435.XNITB	double	XNITB -- Expense - Noninterest - Total (Bank) (XNITB)
436.XOBD	double	XOBD -- Expense - Other (Broker/Dealer) (XOBD)
437.XOI	double	XOI -- Expenses - Other (Insurance) (XOI)

438.XOPR	double	XOPR -- Operating Expenses - Total (XOPR)
439.XOPRO	double	XOPRO -- Operating Expense - Other (XOPRO)
440.XORE	double	XORE -- Expense - Other (Real Estate) (XORE)
441.XPP	double	XPP -- Prepaid Expenses (XPP)
442.XPPFS	double	XPPFS -- Prepaid Expenses & Accrued Income - FS (Memo) (XPPFS)
443.XPR	double	XPR -- Pension and Retirement Expense (XPR)
444.XRD	double	XRD -- Research and Development Expense (XRD)
445.XRENT	double	XRENT -- Rental Expense (XRENT)
446.XS	double	XS -- Expense - Sundry (XS)
447.XSGA	double	XSGA -- Selling, General and Administrative Expense (XSGA)
448.XSTF	double	XSTF -- Staff Expense (Income Account) (XSTF)
449.XSTFO	double	XSTFO -- Staff Expense - Other (XSTFO)
450.XSTFWS	double	XSTFWS -- Staff Expense - Wages and Salaries (XSTFWS)
451.XT	double	XT -- Expense - Total (XT)
452.AJEXI	double	AJEXI -- Adjustment Factor (International Issue)-Cumulative by Ex-Date (AJEXI)
453.CSHOI	double	CSHOI -- Com Shares Outstanding - Issue (CSHOI)
454.CSHPRIA	double	CSHPRIA -- Common Shares Used to Calculate Earnings Per Share (Basic) - As Reported (CSHPRIA)
455.EPSEXCON	double	EPSEXCON -- Earnings Per Share (Basic) - Excluding Extraordinary Items - Consolidated (EPSEXCON)
456.EPSEXNC	double	EPSEXNC -- Earnings Per Share (Basic) - Excluding Extraordinary Items - Nonconsolidate (EPSEXNC)
457.EPSINCON	double	EPSINCON -- Earnings Per Share (Basic) - Including Extraordinary Items - Consolidated (EPSINCON)
458.EPSINNC	double	EPSINNC -- Earnings Per Share (Basic) - Including Extraordinary Items - Nonconsolidate (EPSINNC)
459.ICAPI	double	ICAPI -- Issued Capital (ICAPI)
460.NAICSH	string	NAICSH -- North America Industrial Classification System - Historical (NAICSH)
461.NICON	double	NICON -- Net Income (Loss) - Consolidated (NICON)
462.NINC	double	NINC -- Net Income (Loss) - Nonconsolidated (NINC)
463.PV	double	PV -- Par Value - Issue (PV)
464.SICH	double	SICH -- Standard Industrial Classification - Historical (SICH)
465.TSTKNI	double	TSTKNI -- Treasury Stock - Number of Common Shares - Issue (TSTKNI)

Frequencies Categorical Variables:

Variable	**N**	**Percent**
indfmt	646	
... INDL	646	100%
datafmt	646	

... HIST_STD	646	100%
consol	646	
... C	580	89.8%
... N	66	10.2%
popsrc	646	
... I	646	100%
acctstd	646	
... DI	384	59.4%
... DS	258	39.9%
... US	4	0.6%
bspr	646	
... GO	646	100%
curcd	646	
... AUD	46	7.1%
... BRL	3	0.5%
... CHF	2	0.3%
... CNY	17	2.6%
... EUR	23	3.6%
... GBP	44	6.8%
... HKD	115	17.8%
... INR	39	6%
... JPY	217	33.6%
... MYR	29	4.5%
... NZD	11	1.7%
... SEK	33	5.1%
... SGD	48	7.4%
... THB	2	0.3%
... USD	4	0.6%
... ZAR	13	2%
final	646	
... Y	646	100%
fyear	646	
... 2010	55	8.5%
... 2011	54	8.4%
... 2012	55	8.5%
... 2013	56	8.7%
... 2014	57	8.8%
... 2015	58	9%
... 2016	56	8.7%
... 2017	56	8.7%

... 2018	57	8.8%
... 2019	57	8.8%
... 2020	59	9.1%
... 2021	26	4%
fyr	646	
... 1	29	4.5%
... 2	103	15.9%
... 3	240	37.2%
... 4	6	0.9%
... 5	11	1.7%
... 6	75	11.6%
... 7	7	1.1%
... 8	43	6.7%
... 9	17	2.6%
... 10	7	1.1%
... 11	14	2.2%
... 12	94	14.6%
ismod	646	
... 1	646	100%
pddur	646	
... 11	1	0.2%
... 12	642	99.4%
... 13	1	0.2%
... 14	1	0.2%
... 15	1	0.2%
src	646	
... 3	6	0.9%
... 5	640	99.1%
iid	646	
... 01W	625	96.7%
... 02W	21	3.3%
curcdi	646	
... AUD	46	7.1%
... BRL	3	0.5%
... CHF	2	0.3%
... CNY	17	2.6%
... EUR	23	3.6%
... GBP	44	6.8%
... HKD	115	17.8%
... INR	39	6%

... JPY	217	33.6%
... MYR	29	4.5%
... NZD	11	1.7%
... SEK	33	5.1%
... SGD	48	7.4%
... THB	2	0.3%
... USD	4	0.6%
... ZAR	13	2%
conm	646	
... ACCENT GROUP LTD	11	1.7%
... ADASTRIA CO LTD	9	1.4%
... ALEXON GROUP PLC	1	0.2%
... AOKI HOLDINGS INC	11	1.7%
... AOYAMA TRADING CO LTD	11	1.7%
... ASIA COMMERCIAL HOLDINGS LTD	10	1.5%
... BALS CORP	1	0.2%
... BRAND CONCEPTS	4	0.6%
... CHARLES VOGELE HLDG AG	2	0.3%
... CHINA FORTUNE INVEST (HLDG)	8	1.2%
... CHIYODA CO LTD	9	1.4%
... CHOW SANG SANG HOLDINGS LTD	11	1.7%
... CIE FINANCIERE RICHEMONT AG	11	1.7%

... CITY CHIC COLLECTIVE LTD	11	1.7%
... CORTINA HOLDINGS LTD	11	1.7%
... COX CO LTD	11	1.7%
... CRG INCORPORATED BHD	3	0.5%
... D. P. ABHUSHAN LIM	4	0.6%
... DICKSON CONCEPTS (INTL) LTD	10	1.5%
... EDGARS CONSOLIDATED STORES	2	0.3%
... EMPEROR WATCH AND JEWELLERY	8	1.2%
... ETAM DEVELOPPEMENT SCA	5	0.8%
... FESTARIA HOLDINGS CO LTD	11	1.7%
... FOOTASYLUM LTD	1	0.2%
... FOSCHINI GROUP LTD	11	1.7%
... GARANT SCHUH PLUS MODE AG	1	0.2%
... GFOOT CO LTD	11	1.7%
... GROUPE JAJ	6	0.9%
... HARUYAMA HOLDINGS INC	9	1.4%
... HENGDELI HOLDINGS LTD	12	1.9%

... HENNES & MAURITZ AB	12	1.9%
... HONEYS HOLDINGS CO LTD	11	1.7%
... HONG KONG RESOURCES HLDGS CO	7	1.1%
... HOUR GLASS LTD	11	1.7%
... INCREDIBLE HOLDINGS LTD	4	0.6%
... JUBILEE ENTERPRISE PCL	2	0.3%
... KAPPAHL AB	10	1.5%
... KING FOOK HLDGS LTD	10	1.5%
... KONAKA CO LTD	11	1.7%
... LEYSEN JEWELLERY INC	4	0.6%
... LOVISA HOLDINGS LTD	7	1.1%
... LUK FOOK HLDGS	11	1.7%
... MAC HOUSE CO LTD	11	1.7%
... MAXI-CASH FINANCIAL SERV	3	0.5%
... MCLON JEWELLERY CO LTD	1	0.2%
... MICHAEL HILL INTL LTD	10	1.5%
... MOSAIC BRANDS LTD	11	1.7%
... MOSS BROS GROUP PLC	6	0.9%
... MOTHERCARE PLC	10	1.5%

... NEW ART HOLDINGS CO LTD	4	0.6%
... NEXT PLC	12	1.9%
... NICE CLAUP CO LTD	2	0.3%
... NISHIMATSUYA CHAIN CO LTD	11	1.7%
... ONLY CORP	4	0.6%
... ORIENTAL WATCH HLDGS LTD	11	1.7%
... PADINI HOLDINGS BHD	12	1.9%
... PALEMO HOLDINGS CO LTD	11	1.7%
... PROSPER ONE INTL HLDG CO LTD	4	0.6%
... PUMPKIN PATCH LTD	7	1.1%
... QUIZ PLC	3	0.5%
... RADHIKA JEWELTECH LTD	1	0.2%
... RAMSDENS HOLDINGS PLC	4	0.6%
... RIZZO GROUP AB (PUBL)	11	1.7%
... SACS BAR HOLDINGS INC	11	1.7%
... SECOND CHANCE PPTY LTD	8	1.2%
... SEKIDO CO LTD	3	0.5%
... SHIMAMURA CO LTD	5	0.8%
... SHOE ZONE PLC	5	0.8%

... SHREE GANESH JEWELLERY HOUSE	5	0.8%
... SIGNET JEWELERS LTD	4	0.6%
... SK JEWELLERY GROUP LTD	4	0.6%
... STELUX HOLDINGS INTERNL LTD	11	1.7%
... STYLIFE CORP	1	0.2%
... SUZUTAN CO LTD	1	0.2%
... TABIO CORP	11	1.7%
... TAKA JEWELLERY HOLDINGS LTD	7	1.1%
... TAKA-Q CO LTD	11	1.7%
... TASAKI SHINJU CO LTD	7	1.1%
... TATA HEALTH INTERNATIONAL	6	0.9%
... TCNS CLOTHING CO L	3	0.5%
... TOMEI CONSOLIDATED BERHAD	12	1.9%
... TRACK FIELD CO SA	1	0.2%
... TRENT LTD	11	1.7%
... TRIBHOVANDAS BHIMJ ZAVERI	10	1.5%
... TRINITY LTD	8	1.2%
... U.H. ZAVERI LTD	1	0.2%

... UNITED ARROWS LTD	8	1.2%
... VERITE CO LTD	10	1.5%
... VIVARA PARTICIPATES S A	2	0.3%
... WA INC	1	0.2%
... WATCHES OF SWITZERLAND GROUP	2	0.3%
... ZHULIAN CORP BERHAD	2	0.3%
costat	646	
... A	588	91%
... I	58	9%
fic	646	
... AUS	40	6.2%
... BMU	83	12.8%
... BRA	3	0.5%
... CHE	13	2%
... CHN	5	0.8%
... CYM	30	4.6%
... DEU	1	0.2%
... FRA	11	1.7%
... GBR	41	6.3%
... HKG	18	2.8%
... IND	39	6%
... JEY	3	0.5%
... JPN	217	33.6%
... MYS	29	4.5%
... NZL	17	2.6%
... SGP	48	7.4%
... SWE	33	5.1%
... THA	2	0.3%
... ZAF	13	2%
loc	646	
... AUS	50	7.7%
... BMU	4	0.6%
... BRA	3	0.5%
... CHE	13	2%

... CHN	5	0.8%
... DEU	1	0.2%
... FRA	11	1.7%
... GBR	44	6.8%
... HKG	127	19.7%
... IND	39	6%
... JPN	217	33.6%
... MYS	29	4.5%
... NZL	7	1.1%
... SGP	48	7.4%
... SWE	33	5.1%
... THA	2	0.3%
... ZAF	13	2%
naicsh	646	
... 448110	75	11.6%
... 448120	72	11.1%
... 448130	28	4.3%
... 448140	98	15.2%
... 448150	4	0.6%
... 448190	24	3.7%
... 448210	56	8.7%
... 448310	274	42.4%
... 448320	15	2.3%
au	646	
... 4	72	11.1%
... 5	80	12.4%
... 6	45	7%
... 7	70	10.8%
... 9	319	49.4%
... 11	43	6.7%
... 16	4	0.6%
... 17	4	0.6%
... 22	3	0.5%
... 24	6	0.9%
auop	646	
... 1	574	88.9%
... 2	6	0.9%
... 3	11	1.7%
... 4	55	8.5%
Sector	646	

... Clothing	301	46.6%
... Jewerly	56	8.7%
... Shoes	289	44.7%

Mean, median and quartile ranges of the final compa_funda data frame:

Variable	**N**	**Mean**	**Std. Dev.**	**Min**	**Pctl. 25**	**Pctl. 75**	**Max**
aco	646	0	1	-0.444	-0.441	-0.116	6.149
acox	646	0	1	-0.393	-0.393	-0.14	6.404
act	646	0	1	-0.457	-0.448	-0.059	7.521
ao	646	0	1	-0.459	-0.457	-0.055	4.816
aox	646	0	1	-0.451	-0.449	-0.06	4.858
ap	646	0	1	-0.48	-0.476	-0.059	4.98
at	646	0	1	-0.439	-0.432	-0.104	6.397
caps	646	0	1	-0.383	-0.383	-0.029	6.55
capx	646	0	1	-0.395	-0.391	-0.139	7.032
ceq	646	0	1	-0.7	-0.387	-0.101	8.253
ch	646	0	1	-0.435	-0.432	-0.155	5.398
che	646	0	1	-0.344	-0.342	-0.155	11.225
cheb	646	0	1	-0.433	-0.426	-0.153	7.282
chech	646	0	1	-19.274	-0.027	0.028	7.936
chee	646	0	1	-0.437	-0.43	-0.155	7.37
cogs	646	0	1	-0.449	-0.439	-0.07	9.022
cstk	646	0	1	-0.29	-0.284	-0.104	6.922
dc	646	0	1	-0.11	-0.11	-0.11	12.811
dfxa	646	0	1	-0.389	-0.383	-0.149	10.772
dlc	646	0	1	-0.371	-0.367	-0.043	12.972
dltt	646	0	1	-0.272	-0.272	-0.177	8.742
do	646	0	1	-16.425	-0.001	-0.001	11.198
dp	646	0	1	-0.384	-0.377	-0.154	10.243
dpact	646	0	1	-0.377	-0.374	-0.164	6.193
dpc	646	0	1	-0.38	-0.372	-0.152	10.532
ebit	646	0	1	-2.988	-0.365	-0.123	7.926
ebitda	646	0	1	-2.032	-0.404	-0.15	6.709
ero	646	0	1	-2.155	-0.331	-0.032	7.776
fca	646	0	1	-14.517	0.057	0.058	4.822
fincf	646	0	1	-8.878	0.072	0.255	8.047
fopo	646	0	1	-4.301	-0.252	-0.069	10.806
gdwl	646	0	1	-0.219	-0.219	-0.208	10.011
ib	646	0	1	-9.828	-0.241	-0.055	6.544

ibc	646	0	1	-9.74	-0.241	-0.058	6.514
icapt	646	0	1	-0.394	-0.388	-0.104	7.098
intan	646	0	1	-0.343	-0.342	-0.175	8.883
invch	646	0	1	-6.841	-0.061	0.176	6.051
invfg	646	0	1	-0.555	-0.541	0.087	5.038
invo	646	0	1	-0.344	-0.13	-0.13	9.799
invrm	646	0	1	-0.366	-0.366	-0.196	10.515
invt	646	0	1	-0.59	-0.572	0.063	4.917
invwip	646	0	1	-0.215	-0.215	-0.215	8.65
ivaeq	646	0	1	-0.131	-0.131	-0.131	11.795
ivao	646	0	1	-0.359	-0.359	-0.274	9.613
ivncf	646	0	1	-19.281	0.101	0.212	2.605
ivst	646	0	1	-0.139	-0.139	-0.135	15.695
lco	646	0	1	-0.454	-0.445	-0.129	6.502
lcox	646	0	1	-0.473	-0.436	-0.107	11.067
lct	646	0	1	-0.531	-0.521	-0.103	6.517
lo	646	0	1	-0.397	-0.397	-0.151	6.632
lse	646	0	1	-0.439	-0.432	-0.104	6.397
lt	646	0	1	-0.47	-0.461	-0.111	6.761
mii	646	0	1	-13.465	-0.039	-0.039	14.873
nopi	646	0	1	-14.772	0.101	0.171	4.699
np	646	0	1	-0.296	-0.296	-0.15	12.455
oancf	646	0	1	-2.551	-0.352	-0.165	7.551
oiadp	646	0	1	-2.988	-0.365	-0.123	7.926
oibdp	646	0	1	-2.032	-0.404	-0.15	6.709
pi	646	0	1	-5.743	-0.304	-0.093	8.12
ppegt	646	0	1	-0.358	-0.355	-0.2	5.467
ppent	646	0	1	-0.335	-0.332	-0.219	5.975
pstk	646	0	1	-0.066	-0.066	-0.066	19.385
pstkr	646	0	1	-0.042	-0.042	-0.042	25.373
re	646	0	1	-1.085	-0.359	-0.152	9.962
recch	646	0	1	-6.559	-0.035	0.001	16.001
recco	646	0	1	-0.254	-0.254	-0.196	9.552
rect	646	0	1	-0.285	-0.283	-0.138	8.433
rectr	646	0	1	-0.254	-0.252	-0.139	8.528
revt	646	0	1	-0.492	-0.483	-0.121	7.569
rvlrv	646	0	1	-8.147	0.131	0.131	0.181
sale	646	0	1	-0.492	-0.483	-0.121	7.569
seq	646	0	1	-0.7	-0.387	-0.101	8.253
teq	646	0	1	-0.693	-0.388	-0.104	8.193

transa	646	0	1	-15.699	-0.028	-0.028	10.48
tstk	646	0	1	-0.305	-0.305	-0.29	8.169
txc	646	0	1	-0.498	-0.382	-0.215	9.579
txdb	646	0	1	-0.241	-0.241	-0.198	9.371
txdi	646	0	1	-9.074	0.024	0.055	14.419
txditc	646	0	1	-0.241	-0.241	-0.198	9.371
txo	646	0	1	-5.943	-0.104	-0.104	10.386
txp	646	0	1	-0.33	-0.329	-0.196	8.717
txt	646	0	1	-2.371	-0.375	-0.187	9.832
wcap	646	0	1	-1.355	-0.356	-0.038	9.167
wcapopc	646	0	1	-6.256	-0.09	0.135	8.728
xacc	646	0	1	-0.228	-0.227	-0.133	11.225
xido	646	0	1	-16.425	-0.001	-0.001	11.198
xint	646	0	1	-0.271	-0.265	-0.13	11.611
xopr	646	0	1	-0.495	-0.485	-0.106	7.875
xopro	646	0	1	-3.437	-0.129	-0.109	17.718
xpp	646	0	1	-0.267	-0.267	-0.176	12.581
xrent	646	0	1	-0.535	-0.53	-0.012	5.021
xsga	646	0	1	-0.523	-0.516	-0.028	5.46
exchg	646	0	1	-2.171	-0.911	0.941	1.512
ajexi	646	0	1	-0.217	-0.09	-0.09	12.564
cshoi	646	0	1	-0.43	-0.408	0.011	11.759
cshpria	646	0	1	-0.479	-0.453	0.052	6.037

7. Bibliography

Stone, D. 2012. "Policy paradox: the art of political decision making."

Benitez, Rafael, Vicente Coll-Serrano, and Vicente J. Bolós. 2021. “deaR-Shiny: An Interactive Web App for Data Envelopment Analysis.” (MDPI).

Seiford, Lawrence M. , and Joe Zhu. 2011. “Data Envelopment Analysis: History, Models, and Interpretations.” *ResearchGate* 39.

Niranjan, Nataraja, and Andrew L Johnson. 2011. “Guidelines for using variable selection techniques in data envelopment analysis.” *European Journal of Operational Research.*

Fenyvesa, Veronika, Tibor Tarnóczi, and Kinga Zsidó. 2015. “Financial Performance Evaluation of agricultural enterprises with DEA Method.” *Procedia Economics and Finance* (Elsevier) 423 – 431.

Li, Yongjun, Xiao Shi, Min Yang, and Liang Liang. 2016. “Variable selection in data envelopment analysis via Akaike’s information criteria.” (Springer).

Rastogi, Shailesh. 2012. "Efficiency and Companies: An Empirical Study of Performance Measurement." (ResearchGate).

Tongying, Jia, and Yuan Huiyun. 2017. “The application of DEA (Data Envelopment Analysis) window analysis in the assessment of influence on operational efficiencies after the establishment of branched hospitals.” *BMC Health Services Research* 1-8.

Smriti , Tagdira Naznin, and Hasinur Rahaman Khan. n.d. “1 Efficiency Analysis of Manufacturing Firms Using Data Envelopment Analysis Technique.” *Journal of Data Science* 69-78.

Chong, Dazhi, and Hongwei Zhu. n.d. “Firm Clustering based on Financial Statements.”

Thanassoullis, Emmanuel, and Maria Conceiçao A Silva. 2018. “Measuring Efficiency Through Data Envelopment Analysis.” *Taylor & Francis* 37-41.

Asmild, Mette, Joseph C. Paradi, Aggarwall Vanita, and Claire Schaffnit. 2004. “Combining DEA Window Analysis with the Malmquist Index Approach in a Study of the Canadian Banking Industry.” *Journal of Productivity Analysis* 67-89.

Matada, Lucas, and Jehovaness Aikaeli. 2016. “Empirical Analysis of Efficiency of Community Banks in Tanzania.” *International Journal of Economics and Finance.*

Assaf, A, and K. M. Matawie. 2010. “Taylor & Francis Group.”

Tziogkidis, Panagiotis. 2012. “Bootstrap DEA and hypothesis testing.” *Cardiff Economics Working Papers.*

Dzuba, Sergey, and Denis Krylov. 2021. “Cluster Analysis of Financial Strategies of Companies.” *Mathematics.*

n.d. “SIX DIGIT NAICS CODES & TITLES.” https://www.naics.com/six-digit-naics/?code=44-45.

International Labour Organization. 1996. "Globalization Changes the Face of Textile, Clothing and Footwear Industries."

Lai, Hongbo, Hao Shi, and Yang Zhou. 2020. “Explanation of input-oriented BCC model.” *PLOS ONE.*

Vargas Sánchez, Jhon Jairo. n.d. “Malmquist Index with Time Series to Data Envelopment Analysis.”

Hammouda, Khaled, and Karray Fakhreddine. n.d. “A Comparative Study of Data Clustering Techniques.”

Lai, Hongbo; Shi, Hao; Zhou, Yang (2020): Explanation of input-oriented BCC model. PLOS ONE. Journal contribution.

Epicoco, N. (2016). Re: Do we need to standardize variable in DEA analysis? Retrieved from: https://www.researchgate.net/post/Do-we-need-to-standardize-variable-in-DEAanalysis/579089f74048548c5b4da5c1/citation/download.

Fashion United. 2022. Global Fashion industry statistics. [online] Available at: < https://fashionunited.com/global-fashion-industry-statistics > .

Łozowicka, Anna. (2014). About using Principal Component Analysis in Data Envelopment Analysis. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu.