A financial health classification of companies using multivariate ordering

Beatriz Bivar Roseiro

Instituto Superior Técnico, Lisboa, Portugal

October 2020

Abstract

The financial health of companies has proven to be an important factor regarding the analysis of financial performance and has been specially decisive on companies participation in the export market. The use of a financial dimension to understand international trade has grown increasingly over the last few years, considering that export is an indicator of business competitiveness.

The present work, done in collaboration with Tekever and AICEP, aimed to conduct a financial health classification of companies using a multivariate ordering methodology. Multivariate statistical methods were studied in order to perform sorting and grouping on the data, having in consideration financial measures and ratios obtained from financial statements.

Ordering methods, such as the reduced ordering here employed, allow the creation of a ranking, making it possible to analyse the firms considering their position in relation to others. With the present work, it was possible to explore the application of an effective multivariate distance metric that establishes an order and detects anomalies, differentiating from the financial rules technique and going beyond the assignment of observations to classes. Results were compared with the financial rules and raised questions regarding the fluctuation of financial results.

Keywords: Financial Health, Principal Components Analysis, Multivariate Ordering, Clustering

1 Introduction

In order to improve the quality of the service and keep contributing to the expansion of the national export base, the main current focus of the government business entity AICEP (aicep Portugal Global - Trade & Investment Agency) is in its digital transformation, with the first step being the new technological platform *"Portugal Exporta"* (AICEP 2019). AICEP's digital platform aims to help companies boost their exports, diversify markets and bring more companies to export on a recurring basis.

According to AICEP's criteria, the companies are segmented by Financial Health, Innovation, Export Volume and Export Intensity. Financial Health is the classification characteristic to consider in the present work.

There is not a precise way to determine the financial health of a company, but it can be achieved by using financial measures and ratios, obtained from financial statements such as the balance sheet and the income statement.

Financial ratios can help evaluate the overall financial condition of an organization as they compare the magnitudes of accounts in financial statements and can help to classify firms according to similarity of the structure of their financial statements. Usually, a company's financial health is evaluated with the objective of determining its worth, therefore, it consists on a specific diagnosis for a particular company. There is not a single metric that can identify the overall financial health of a company, but a combination of them can help us get closer to it.

Several studies have been used to analyse the financial health of companies, such as, the Z-score technique used by Altman (1968) or compositional data analysis (CoDa) by Linares-Mustarós, Coenders, and Vives-Mestres (2018). The ordering of multivariate data examined by (Barnett 1976) lead to sub-ordering principles that are key for this work as they *"enter into a variety of multivariate analysis procedures ... made of some form of data ordering or ranking"*. Also, D'Esposito and Ragozini (2008) proposed a new reduced ordering procedure for multivariate observations based on the idea of "worst best" direction, which goes from the "worst" performing units to the "best" ones.

The present work, in collaboration with Tekever and AICEP, aims to contribute for the effectiveness of *"Portugal Exporta"* platform, in order to guarantee that it delivers customized services to thousands of companies, agreeing with AICEP's mission to increase Portugal's competitiveness and reputation by supporting the structural investment and making companies international in scope, with special emphasis on small and medium-size companies. It is intended to apply multivariate statistical methods and multivariate ordering procedures on a database of companies with activity in Portugal considering financial statements metrics, in order to determine the financial health of companies.

2 Methodology

Multivariate statistical methods are very common when sorting and grouping and allow to create groups of similar objects based upon measured characteristics and under the required rules.

Measurements of several financial variables were used to develop a screening procedure that discriminates the financial position of the companies evaluated.

Besides the classical methods, robust statistics were also applied as it produces methods that are not unduly affected by outliers or other small departures from model assumptions.

2.1 Principal Component Analysis

Principal component analysis (PCA) is mainly used as a dimension reduction tool, but it can also contribute to the interpretation of the variables under study as it often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily result. The analysis of principal components is considered an unsupervised learning problem and can work as data preprocessing, serving as an intermediate step before applying other techniques.

The key idea behind PCA is to produce uncorrelated linear combinations of the original variables. This way, it is possible to get new uncorrelated variables, called principal components (PC), that retain the total variance of the original variables.

Although all the components are required to reproduce the total variability, often much of this variability can be accounted by a small number of principal components and thus reduce dimensionality while maintaing almost as much information as there is in the original variables.

The robust approach of principal components analysis aims to correct the sensitivity to outliers and is, generally, based upon two main strategies: using a robust covariance matrix Σ that can be generated by a robust estimator such as the Minimum Covariance Determinant (Rousseeuw 1985) or through Projection Pursuit (Friedman and Tukey 1974).

2.2 Ordering of multivariate data

To understand the companies positions financially, one can compare the businesses based on the performance metrics, having under consideration the industries of the companies, and create a ranking. Ranking proves to be a very effective procedure as it isolates outliers by properly weighting each ranked multivariate sample.

For univariate data, outliers detection is made simply by looking at their extreme value relative to the rest of the values. On the other hand, for multivariate data it is necessary to adapt the appropriate sub-ordering principle in order to express the extremeness of observations. Barnett (1976) discussed several sub-ordering techniques such as: marginal (M-ordering), conditional (C-ordering), reduced (R-ordering) and partial ordering (P-ordering). In this work, the focus will be on the concepts and methods that support reduced and partial ordering.

In reduced ordering or R-ordering, each multivariate observation is reduced to a single value as it implies the use of some generalized distance measure from a single fixed point (usually the center of the distribution). The aim of R-ordering is to reduce the multivariate ranking to a simple ranking operation of a set of transformed values and with this create an overall order on the original multivariate samples.

The reduction function most frequently employed in R-ordering is the generalized Mahalanobis distance. The Mahalanobis distance considers the correlation between variables and measures the distance of an observation to the center of the distribution (usually the center is the mean).

On the other hand, in partial ordering or P-ordering, subsets of data are grouped together forming minimum convex hulls. This relates to convex hull peeling depth where the convex layers are the depth contours of the data. The convex hull of a data set points is often defined as the smallest convex polygon that encloses all of those points in the set.

2.3 Cluster Analysis

Cluster analysis is an exploratory analysis that looks for patterns in the data by grouping the observations in clusters. It finds homogeneous groups of objects such that objects within a group are more similar among themselves than objects of different groups.

Clustering allows to understand hidden structures in the data set as it suggests interesting hypothesis concerning relationships among variables or objects.

There is not a precise way of grouping observations given that there is often a great deal of particularity involved in the choice of a measure of proximity.

There are several clustering methods, including the Hierarchical and Partitioning methods. Hierarchical clustering involves creating clusters that have a predetermined ordering, with no fixed number of clusters in advance, and there are two major types of hierarchical techniques: divisive (assigning all of the n observations to a single cluster and then partition the cluster to two less similar clusters and so on) and agglomerative (assigning each observation to its own cluster and then compute the similarity between each of the clusters and joining the two most similar clusters, repeating this step until there is only a single cluster left). Partitional clustering, or non-hierarchical clustering, divides the data set into a predefined number of groups, so the number of clusters (k) is fixed and stays unchanged through the course of the algorithm.

2.4 Rules and thresholds for financial variables

One other approach to consider is the definition of rules and thresholds for the financial ratios and measures taken into consideration for this work. Here, it is necessary to look closely at the chosen financial variables and establish a criterion in order to create groups that aggregate companies according to it.

From all the indicators available in the "Guide to Support the Use of the Financial Self-Diagnosis Tool" (IAPMEI 2015), regarding the present work, there was special attention paid to ratios such as Liquidity (current ratio), Profitability (ROI - Return On Investment, ROE - Return On Equity, ROA - Return On Assets), Debt and Financial Autonomy.

As liquidity can be measured by using the current ratio, it tells the capacity of a firm repay its short-term liabilities with its current assets, which can have significant impact towards the financial health of the company. As a general rule, the threshold to consider for liquidity is the value one.

The debt ratio (debt/asset ratio) represents the proportion of debt used to finance a company's assets (leverage) and the threshold can be usually define at 0.5. A ratio below 0.5 means that most of the company's assets are financed through equity and a ratio greater than 0.5 means most of the company's assets are financed through debt.

The profitability ratios (ROI, ROE or ROA) measure the ability of a company in generating returns from its investments and the result is reported as a percentage rate of return, giving an indication of the relationship between the profit and the investment. There is not a right value to define a good or bad profitability result but usually a ROA of 5% is considered to be a good value as well as a ROE of 15%.

Besides these financial ratios, one can also say that a direct way of describing the success of a business is through its net income, as it is a clear indicator of a company's profit after deducting all of the expenses. For the net income, as it is calculated by subtracting expenses (taxes, cost of goods sold, operating expenses, etc.) from the sales revenue, we know that if it is positive (values > 0) it tells us the business is profitable at that given period.

The objective is then to explore these rules and combine them in order to reach a final score for each company that can tell its position in terms of financial health.

The application of rules in the financial measures was already considered by Tekever and AICEP in their previous work and is a strategy that, besides taking a business perspective, can be complemented with other methods (as the methods previously seen).

3 Preliminary analysis

The data used in this study was provided by AICEP and IES - Informação Empresarial Simplificada (Simplified Business Information) for the period of 2015-2018.

The need to consider firms within the same industry in order to avoid the misinterpretation of financial metrics results, led to the Portuguese Business Activity Code (CAE - Classificação Portuguesa das Atividades Económicas). This Code compiles the business activity areas of the companies, and each business activity has a correspondent code, so eventually companies can have one or more codes. The main information considered in this work focus the year of 2018, where the initial data set contained 19.306 companies that could be divided into 678 business activity codes - CAEs.

The chosen CAE to analyse was CAE 15201, that groups footwear producing companies. Therefore, for the year 2018 and CAE 15201, the panel includes a total of 289 observations (companies) identified by the number that refers to their corresponding row number in the data frame.

The 8 variables selected are main heading measures that can be found in the financial statements of companies and intend to represent their financial situation. Six of these variables have information that is distributed within the income statement (Total Sales, Total Supplies, Total Purchases, Employee Total Costs, Total Commodity Material Cost and Net Income), whereas the other two are balance sheet items (Total Assets and Total Equity).

Also, outliers proved to have great significance to this study as they do not always reflect errors or mistakes and can even be the most interesting observations. The observations that differ significantly from the others are the ones that might tell us more about their financial health, as they stand out for a reason. For instance, that reason can be the outstanding results of a company's financial performance, meaning that it excelled financially.

3.1 Principal Component Analysis

Principal Components Analysis (PCA) also works as a tool for data preprocessing before applying other techniques. Considering the robust principal components analysis based on the Minimum Covariance Determinant estimator (MCD), the percentages of variance contribution explained by each of the first two components indicate that they explain 86.8% of the variation in the data, which is an adequate percentage. Therefore, it is possible to retain the first 2 robust principal components.

4 Applications and discussion of results

An analysis of the Mahalanobis distances was conducted in order to create a rank of the data and understand the relation between the position of the companies in that rank and their financial health. The Mahalanobis distance approaches considered were the classical Mahalanobis distance using the mean as reference point and the covariance matrix of the original data and the Robust Mahalanobis distance using the Minimum Covariance Determinant (MCD) estimates of location and scatter. It was also conducted an ordering considering the classical and robust Mahalanobis distances on the first 2 robust principal components.

The classical and robust approaches of the Mahalanobis distance on the original data and on the first 2 robust principal components, although difficult to interpret, showed to be consistent regarding companies that occupy higher positions, enhancing the connection between the Mahalanobis distance and the detection of outliers. These outliers, i.e the companies in higher positions, were in agreement with reality by highlighting companies considered extremely successful. For the rest of the companies on the ascendant ordered list, the interpretation is more difficult to make, not only because they change positions according to the approach taken, but also because there are more financially similar companies in middle positions.

It was also considered the application of this ordering method to the financial ratios previously mentioned (liquidity, ROA and ROE), in order to understand the differences when considering different types of financial evaluation and, even though there are some agreements in some of the quartiles that divide the ordering, there is a great difference between both approaches meaning that the majority of the companies is differently positioned depending on the financial data considered.

Twenty four convex sets were obtained when conducting a convex hull considering the scores for the first 2 (robust) principal components that explain around 87% variability of the data, as visualization with 8 variables is almost impossible. The first and last convex hulls end up contributing to the interpretation of some relationships between companies as they represent the most extreme points that stand out from the others and the closest points that are very similar between them, respectively.

The agglomerative hierarchical clustering procedure chosen to divide the data into clusters was the Ward's method, as it provides a stronger clustering structure, using the Euclidean distance and considering 4 clusters after all considerations. Also, the k-means clustering was applied to the data set considering k=4, being the only similarity to the hierarchical method the grouping of the two most extremes companies in one cluster very distant from the other 3 clusters. Figures 1 and 2 show the representation of both methods for k=4.

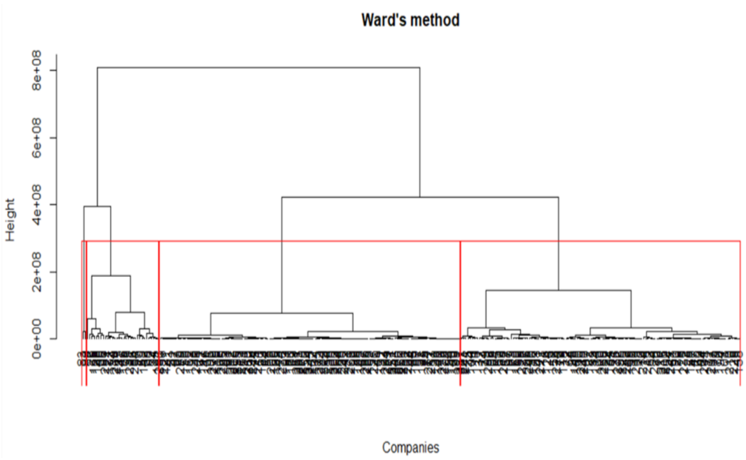

Figure 1: Ward's method dendrogram using Euclidean distance

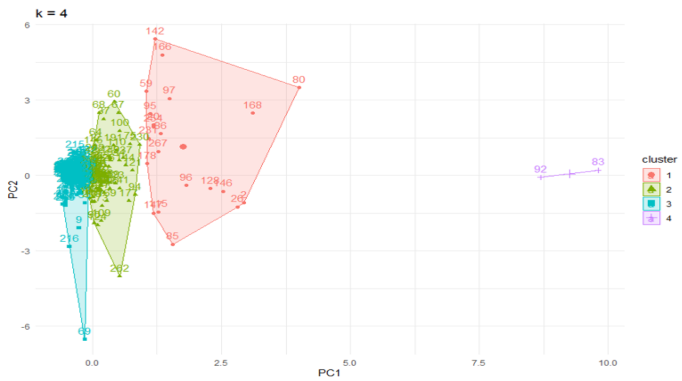

Figure 2: K-means for $k = 4$

The groups obtained from the clustering methods did not proved to be a reliable solution for the problem, raising some questions related to the optimal number of clusters and to the basis of division.

Finally, in order to complement the other methods and help understand the results, it was given focus to the net income and liquidity measures. The combination of both metrics allow to create some barriers and groups that help to locate the companies financially, considering the defined rules. With this, four groups were formed considering all alternatives when combining net income and liquidity, i.e, for net income above or below 0 and for liquidity above or below 1.

Therefore, we have Group I: companies with liquidity < 1 and net income < 0; Group II: companies with liquidity > 1 and net income < 0; Group III: companies with liquidity < 1 and net income > 0; Group IV: companies with liquidity > 1 and net income > 0 (due to the large number of companies present in this group, the firms are displayed in an ascending order according to their robust Mahalanobis distance, with the set divided into quartiles).

Table 1 is the contingency table between financial rules and the robust Mahalanobis distance method on the original data, defining its quartiles as groups.

Table 1: Robust Mahalanobis distance vs Financial rules

		Financial rules							
		I	**II**	**III**	**IV1**	**IV2**	**IV3**	**IV4**	**Total**
Robust Mahalanobis distance on original data	**1st Q**	2	7	0	26	22	9	6	72
	2nd Q	2	9	2	17	20	17	6	73
	3rd Q	7	12	1	10	13	18	11	72
	4th Q	3	11	1	5	3	14	35	72
	Total	14	39	4	58	58	58	58	289

The analysis of Table 1 is more difficult to do because the quartiles end up being very wide, but it is a good way to understand the distribution and create some kind of distinction between companies. There is a certain agreement between the quartiles and the sub-groups of IV, there is however, some contradictions. It is possible to see companies belonging to both rule I and 4th quartile and companies belonging to both rule IV4 and 1st quartile, being understood as examples of a loss that does not necessarily mean bad performance and of a profit that does not always mean a positive performance, respectively. This lead us to the problem of the fluctuation of financial results, which can be due to investments, loans, being a small business, an unexpected gain/loss, etc.

5 Conclusions

The results of the present work revealed that it is possible to create an alternative path to the classification of organisations according to their financial analysis, going beyond financial rules. To achieve it, multivariate statistical methods were applied on the data set consisting of financial statement variables.

The confrontation between the results obtained through the application of the multivariate ordering method and the financial rules considering net income and liquidity proved to be quite helpful, in the sense of understanding the ordering that had been obtained. It was specially enlightening for the cases where there was no agreement, and such findings lead to arguments related to the fluctuation of financial results and raised the question of whether a loss/profit means bad/good performance.

It can also be concluded that multivariate ordering allows to go further than assigning observations to categories or classes, and brings a new perspective to this kind of matter. Ranking the companies according to their Mahalanobis distances makes it possible to analyse them considering their position in relation to others, differentiating from the financial rules approach.

Even though the selected variables go according to the financial statements, it is important to realize that only through a comprehensive analysis of all the financial information can one make an informed decision on the financial health. One way to improve the knowledge had on the observations is to scrutinize the financial variables and understand the relationship and dependency between them. Additionally, more information such as company size, years of activity, investments and external events like Political, Economic, Social, Technological, Environmental and Competitive factors (PESTEC) should be considered, in order to achieve more solid results.

References

AICEP (2019). *Portugal Exporta*. URL: https://www.portugalexporta.pt/ (visited on 07/04/2019).

Altman, Edward I (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy". In: *The journal of finance* 23.4, pp. 589–609.

Barnett, Vic (1976). "The ordering of multivariate data". In: *Journal of the Royal Statistical Society: Series A (General)* 139.3, pp. 318–344.

D'Esposito, Maria Rosaria and Giancarlo Ragozini (2008). "A new R-ordering procedure to rank multivariate performances". In: *Quaderni di Statistica* 10, pp. 5–21.

Friedman, Jerome H and John W Tukey (1974). "A projection pursuit algorithm for exploratory data analysis". In: *IEEE Transactions on computers* 100.9, pp. 881–890.

IAPMEI (2015). *Guia de apoio à utilização da ferramenta de Autodiagnóstico Financeiro*. URL: https://www.iapmei.pt/PRODUTOS-E-SERVICOS/Revitalizacao-Transmissao/Revitalizacao-Empresarial/DOCS/GuiaADF.aspx (visited on 06/26/2019).

Linares-Mustarós, Salvador, Germà Coenders, and Marina Vives-Mestres (2018). "Financial performance and distress profiles. From classification according to financial ratios to compositional classification". In: *Advances in Accounting* 40, pp. 1–10.

Rousseeuw, Peter J (1985). "Multivariate estimation with high breakdown point". In: *Mathematical statistics and applications* 8.283-297, p. 37.