An assessment of the bankruptcy risk on the Romanian capital market

Daniel-Stefan Armeanu
dan, Sorin-Iulian Cioaca

Abstract

In this article, we use statistical models, such as Principal Component Analysis, Cluster Analysis Discriminant Analysis and Altman model to assess the bankruptcy risk on the Romanian capital market. Working on the financial data for the fiscal year 2012, we identify 3 groups of companies listed on the Bucharest Stock Exchange, based on their associated bankruptcy risk. The obtained results can be used by professionals and investors to build appropriate investment strategies, adding new insights on the Romanian capital market. Also, the presented method can function as an early-warning mechanism, helping the authorities adjust their regulatory and supervising tools.

Keywords: financial education; bankruptcy risk; cluster analysis; discriminant analysis; principal component analysis

1. Introduction

Pattern recognition theory developed mainly during the 1960s has direct uses in different fields, like finance, medicine, sociology etc., the main findings being used to develop optimization algorithms and to find patterns for data series. Faced with bankruptcy risks, the financial community became more interested in finding methods to estimate this type of risks, using as input variables some measures considered to assess a company’s financial stability that can be studied using pattern recognition theory. One of the oldest and most important estimation models for the bankruptcy risk, the Z-score function defined by E.I. Altman in 1968, is based on the bankruptcy risk assessment using the Discriminant Analysis, by considering 2 classes of firms (solvent and insolvent). These

* Sorin-Iulian Cioaca. Tel.: +4- 075-303-0409.
E-mail address: cioaca_sorin@yahoo.com
concepts have been further extended and refined, using Principal Component Analysis (PCA), Discriminant Analysis and Cluster Analysis, in order to identify the companies faced with a significant bankruptcy risk.

These concepts have been used on the companies listed on the Bucharest Stock Exchange, obtaining 3 sets of companies, with different levels of bankruptcy risk, using the financial data for the fiscal year 2012. The results have been validated by the evolution of some companies assessed as having a high bankruptcy risk, since, at the end of the third quarter of 2014, they are insolvent.

One contribution of this article to the financial literature dealing with the bankruptcy risk is that it analyses the Romanian capital market, an underdeveloped market that, with proper measures in place, has a great potential to grow. With this article, an assessment of the bankruptcy risk is provided, in order to foster the financial education of both prospective issuers (and their shareholders) and investors (especially locals), that have little knowledge of the advantage and mechanisms of the capital markets.

2. An overview of the main theoretical findings in pattern recognition theory

The pattern recognition theory identifies the common features of the variables from a large data set and simplify them using the PCA (Jolliffe, 2002), such that, using the Discriminant Analysis and Cluster Analysis, to refine the results, by finding the most important characteristics. Moreover, it uses optimization algorithms and computation of the eigenvalues for matrices, obtained by defining different metrics, in order to use them in finding analysis algorithms of the principal components.

De la Torre (2008) evaluates 4 methods of PCA, starting from the results obtained by Borga (1998), restating PCA algorithms, Partial Least-Square, Canonical Correlation Analysis and Multiple Linear Regression, by considering generalized eigen-value problems. Further research has been done by Yan, Xu, Zhang and Zhang (2007), using graph theory and defining a non-parametrical version for Linear Discriminant Analysis.

Cha (2012) defines and uses a new metric in optimization algorithms, named earth mover’s distance, and proves that is more efficient in the classification process than the classical metrics (as the Euclidian distance or city-block).

The concepts of pattern recognition theory have been used in almost all research fields. In finance, Pai and Lin (2005), as well as Chan, Wong and Lam (2000) use the pattern recognition theory to study the data series representing the prices of financial assets, in order to find the undervalued companies (and, therefore, to buy them) and overvalued companies (in order to sell them). Altman, Marco and Varreto (1985) analyse the financial stability of the companies form the Italian banking system. Blokh (2012) uses a classification method (by considering the uncertainty coefficient, instead of the classical similarity measures, like Euclidian distance) for the financial data series related to 25 companies listed on the Tel Aviv Stock Exchange, based on a non-linear measure for the correlation between the data series (a normalized mutual information, as a measure of the uncertainty coefficient), the resulted classification being more relevant to the classification process.

The Cluster Analysis aims to classify the variables into separate structures, that are relevant and have clear distinctive features, named classes (groups or clusters), characterized by the fact that the dissimilarity between every two components of one class is less than the one between one element from the class and one element outside the class. In order to apply Cluster Analysis, different metrics are used, as dissimilarity measure between the components of the initial space, the most known being Manhattan distance (or rectangular distance or City-block distance), that is calculated as the sum of the absolute values of the differences of the two variables coordinates.

In order to classify the elements of a set, we use Ward method, considered as being the most efficient method for hierarchical classification, as it leads to the minimization of the intra-cluster variability, as a result of reunion – at each step of the algorithm – of clusters for which the resulting cluster variability is less than the variability resulted from maintaining the respective clusters as distinct subsets. Such that, the total variance is decomposed in intra-cluster variance and inter-cluster variance: \( \sigma_I^2 = \sigma_W^2 + \sigma_B^2 \), where \( \sigma_W^2 \) and \( \sigma_B^2 \) are intra-variance and, respectively, inter-cluster variance.

At each step of the algorithm, two clusters will be unified such that the resulted intra-cluster variability (that will be greater than the individual intra-cluster variability, as by increasing the number of the elements of a class, the class variability will increase) is the smallest considering the total intra-variance \( \sigma_W^2 \) (by solving an optimization problem, the clusters are defined, starting form the initial variables).

Starting form the classification under by the Cluster Analysis, we use the Discriminant Analysis’s concepts to
find the characteristics of the initial space’s elements, that are the most relevant to find whether an element belongs to a specific cluster (subset) and, therefore, to identify the probability of each element belonging to the identified clusters.

A measure to find the variables’ importance in classification into the resulted clusters from the Cluster Analysis is the Wilks’ Lambda, a measure that can interpreted as follows: if the value is close to zero, the greater discrimination power a variable has (and vice-versa, when the value is close to 1, the discrimination power fades), and p-value is almost equal to zero.

3. Methodology

To assess the bankruptcy risk, we use the Z-score function defined by E.I. Altman in 1968:

\[ Z(r_1, r_2, ..., r_n) = \alpha_0 + \alpha_1 r_1 + \alpha_2 r_2 + ... + \alpha_n r_n, \]

where \( r_1, r_2, ..., r_n \) are the indicators considered in order to find the classification model, \( \alpha_1, \alpha_2, ..., \alpha_n \) are the indicators’ coefficients, and \( \alpha_0 \) is the free-term of the classification function.

By using the Z-score function, each company is allocated to one of the two classes defined by Altman (solvent and insolvent company), being also estimated the bankruptcy probability for each company.

Applying this model for a group of 66 companies listed on the American stock market, from which half filed for bankruptcy, Altman obtained the following Z-score function:

\[ Z = 1.2r_1 + 1.4r_2 + 3.3r_3 + 0.6r_4 + 1.0r_5, \]

where:

\[ r_1 = \frac{\text{Working capital}}{\text{Total assets}}; \quad r_2 = \frac{\text{Retained Earnings}}{\text{Market Capitalization}}; \quad r_3 = \frac{\text{EBIT}}{\text{Sales}}; \quad r_4 = \frac{\text{Net Cash Flows from Operating Activities}}{\text{Sales}}; \quad r_5 = \frac{\text{Net Profit}}{\text{Total Debt}}. \]

Moreover, Altman defined 3 areas of bankruptcy probability: low bankruptcy probably (when \( Z \) is greater than 2.99), medium bankruptcy probability (when \( Z \) is between 1.8 and 2.99) and high bankruptcy probability (when \( Z \) is less than 1.8).

4. The results related to the Romanian capital market

4.1 Applying the Principal Component Analysis

Starting from Altman’s model, we will apply these theoretical concepts for 68 (out of 74) companies listed on the Romanian capital market, namely those listed on the main venue, Bucharest Stock Exchange (at the first and second category), in order to assess their bankruptcy risk. We consider the financial data for the fiscal year 2012, prepared according with the International Financial Reporting Standards and published on the Bucharest Stock Exchange and their own web pages. In order to find the Z-score function for the Romanian capital market, we consider 7 financial measures for each company, that are relevant to finding their financial stability: Total Assets, Sales, Earnings before interest and Taxes (EBIT), Net Cash Flows from Operating Activities, Net Profit, Total Liabilities, Market Capitalization (year-end).

Table 1. Descriptive statistics for selected companies

| Variable                                      | Mean    | Minimum | Maximum  | Std. Dev. |
|-----------------------------------------------|---------|---------|----------|-----------|
| Total Assets                                  | 2438642 | 12602   | 48890590 | 8232994   |
| Sales                                         | 763157  | 3045    | 19122510 | 2794920   |
| EBIT                                          | 97049   | -573355 | 5067820  | 623823    |
| Net Cash Flows from Operating Activities      | 177263  | -263080 | 6664950  | 864943    |
| Net Profit                                   | 64208   | -547386 | 3850610  | 481078    |
| Total Debt                                   | 1533679 | 928     | 43128297 | 6343037   |
These indicators have been standardized, in order to eliminate the effects of the indicators’ extreme value for some companies, and we obtain the correlation matrix (instead of covariance matrix, whose components are directly influenced by the presence of high values for some indicators, like the market capitalization or sales):

Table 2. Correlation matrix for the 7 variables.

| Variable                        | Total Assets | Sales   | EBIT    | Net Cash Flows from Operating Activities | Net Profit | Total Debt | Market Capitalization |
|---------------------------------|--------------|---------|---------|-----------------------------------------|------------|------------|----------------------|
| Total Assets                    | 1.000000     | 0.595845| 0.499268| 0.797175                                | 0.491911   | 0.937198   | 0.717739             |
| Sales                           | 0.595845     | 1.000000| 0.733185| 0.810025                                | 0.697820   | 0.392971   | 0.812558             |
| EBIT                            | 0.499268     | 0.733185| 1.000000| 0.909351                                | 0.989505   | 0.214476   | 0.907424             |
| Net Cash Flows from Operating Activities | 0.797175 | 0.810025| 0.909351| 1.000000                                | 0.898284   | 0.578578   | 0.945286             |
| Net Profit                      | 0.491911     | 0.697820| 0.989505| 0.898284                                | 1.000000   | 0.209809   | 0.893010             |
| Total Debt                      | 0.937198     | 0.392971| 0.214476| 0.578578                                | 0.209809   | 1.000000   | 0.435326             |
| Market Capitalization           | 0.717739     | 0.812558| 0.907424| 0.945286                                | 0.893010   | 0.435326   | 1.000000             |

The correlation matrix shows strong relationships between 6 out of 7 variables (Total Assets, Sales, Net Cash Flows from Operating Activities, Net Profit, Total Debt and Market Capitalization), and the EBIT seems to be the least correlated with the other 6 measures (fact that is derived from the presence of some extreme values for the 68 companies). Since the initial causal space is characterized by the presence of redundances (as 6 out of 7 variables are strongly correlated), we use PCA to eliminate the redundant information that alters the conclusions that can be derived from the collected data.

Starting from the correlation matrix, we calculate the eigenvalues and, using the Kaiser criterion (that states that only the correlation matrix’s eigenvalues greater than unity qualify as principal components), we can identify the number of principal components, as shown in the following table:

Table 3. Eigenvalues of correlation matrix.

| Value number | Eigenvalue | %Total Variance | Cumulative Eigenvale | Cumulative % |
|--------------|------------|-----------------|----------------------|--------------|
| 1            | 5.758336   | 82.26194        | 5.758336             | 82.26194     |
| 2            | 0.763745   | 10.91064        | 6.522081             | 93.172658    |
| 3            | 0.458486   | 6.54981         | 6.980567             | 99.722425    |
| 4            | 0.019145   | 0.27350         | 6.999712             | 99.995937    |
| 5            | 0.000234   | 0.00335         | 6.999947             | 99.999284    |
| 6            | 0.000053   | 0.00076         | 7.000000             | 100.00000    |

As shown in the table, the first resulting variable has an eigenvalue greater than unity, so that can be used for PCA analysis. Moreover, the identified principal component explains more than 80% of the variations in the initial causal space, that leads to the conclusion that those can adequately be used to classify the initial variables. A similar
conclusion is derived from the following graph, using the slope criterion:

![Figure 1: Eigenvalues of correlation matrix](image)

After we find the number of principal components used for the analysis of the initial causal space, we can find the factor matrix, whose elements are the correlation coefficients between the original variables and the principal components (called factor loadings).

### Table 4. Factor matrix.

| Variable                                | Factor Loadings (Unrotate Extraction:Principal Component Analysis) Marked loadings are >.70 |
|-----------------------------------------|------------------------------------------------------------------------------------------------|
| Factor 1                                | Factor Loadings (Unrotate Extraction:Principal Component Analysis) Marked loadings are >.70 |
| Total Assets                            | -0.814565                                                                                     |
| Sales                                   | 0.775370                                                                                      |
| EBIT                                    | 0.988153                                                                                      |
| Net Cash Flows from Operating Activities | 0.945702                                                                                      |
| Net Profit                              | 0.894252                                                                                      |
| Total Debt                              | 0.891022                                                                                      |
| Market Capitalization                   | 0.927757                                                                                      |
| Expl. Var.                              | 0.927757                                                                                      |
| Prp. Totl.                              | 0.822619                                                                                      |

*Source: own calculation*

Analyzing this table, we find that the principal component has strong positive correlations with 5 out of the 7 initial variables and strong negative correlation with Total Assets and Total Debt.

In the next table there are the coefficients of the linear combination that defines the principal component (first column), that can be used to calculate the Z-score function for the observations from the principal component space:

### Table 5. Eigenvectors of correlation matrix.

| Variable                                | Eigenvectors of correlation matrix. Active variables only |
|-----------------------------------------|-----------------------------------------------------------|
| Total Assets                            | Factor 1 Factor 2 Factor 3 Factor 4 Factor 5 Factor 6    |
| Sales                                   | 0.339451 -0.663188 0.024039 -0.124313 -0.222773 0.214447 |
| EBIT                                    | 0.323117 -0.149028 -0.912109 0.144876 -0.116727 0.081677 |
| Net Cash Flows from Operating Activities | 0.411790 -0.032158 0.219814 0.170856 0.257209 0.827721 |
| Net Profit                              | 0.372659 -0.484320 0.193128 0.458110 0.387092 -0.472678 |
Using the factors derived from the PCA (with one principal component, that explains 82% of the information contained in the initial variables), we can write an Altman Z-score function for the Romanian capital market as being given by:

$$Z = -0.33r_1 + 0.32r_2 + 0.41r_3 + 0.37r_4 + 0.40r_5 - 0.38r_6 + 0.39r_7,$$

where $r_1$ is Total Assets, $r_2$ is Sales, $r_3$ is EBIT, $r_4$ is Net Cash Flows from Operating Activities, $r_5$ is Net profit, $r_6$ is Total Debt and $r_7$ is Market capitalization (all being normalized).

### 4.2 Applying cluster analysis

To apply the Cluster Analysis, we use Manhattan distance (also known as rectangular distance or City-Block distance) as a dissimilarity measure for the initial space’s components. In order to realize the hierarchical classification, we use the Ward method for the data set comprised of the 68 companies listed on the regulated market. Using the cluster analysis, we can identify 3 clusters, first with 3 components (SNP, BRD and TLV), the second with 6 components (RRC, FP, BCC, TEL, TGN and ALR) and the third cluster with the remaining 59 companies.

Using a methodology proposed by Armeanu (2012) and the Z-score function obtained by using PCA, we can define 3 regions for assessing the bankruptcy risk of a company:

- When the absolute values of Z-score is greater than 1.3 (safe zone), the associated probability of bankruptcy is small (low bankruptcy risk);
- When the value of Z-score is between 0.11 and 1.3 (unsafe zone), the company has a moderate bankruptcy risk;
- When the value of Z-score is between -1.3 and 0.11 (risky zone), the probability of bankruptcy is high.

### 4.3 Applying Discriminant Analysis

We start the Discriminant analysis considering the 3 classes derived from the Cluster Analysis and we will apply the Discriminant Analysis step by step (by adding a variable in the model at each iteration), the results being shown below:

| Source: own calculation |

---

| Table 6. Model worthiness and the discrimination power of each variable. |
| --- |
| N=68 | Discriminant Function Analysis Summary No. of variables in model: 7; Grouping: Clasa (3grps). |
| Wilks’ Lambda: 0.00494 approx. F(14,118) = 111.54 p <0.0000 |
| Wilks’ Lambda | Partial Lambda | F-remove (2,59) | p-value | Toler. | 1-Toler. (R-sqr) |
| Total Assets | 0.012031 | 0.410287 | 42.4096 | 0.000000 | 0.008562 | 0.991438 |
| Sales | 0.010685 | 0.461945 | 34.36035 | 0.000000 | 0.133509 | 0.866491 |
| EBIT | 0.009989 | 0.494162 | 30.19703 | 0.000000 | 0.007177 | 0.992823 |
| Net Cash Flows from Operating Activities | 0.005948 | 0.829875 | 6.04751 | 0.004082 | 0.028241 | 0.971759 |
| Net Profit | 0.007326 | 0.673786 | 14.28244 | 0.000009 | 0.017499 | 0.982501 |
| Total Debt | 0.010176 | 0.485084 | 31.31419 | 0.000000 | 0.0005827 | 0.994173 |
| Market Capitalization | 0.009936 | 0.496801 | 29.87988 | 0.000000 | 0.004114 | 0.995886 |
| Source: own calculation |
It can be found that, by adding each variable, the discrimination power of the model improves itself, as can be seen by the reduction of the Wilks’ Lambda at each step (to a value of 0.00494 by considering all the variables).

Using the Discriminant Analysis, the percent of the correctness is 98.5294%, influenced by the difference from the second cluster.

Table 7. Classification matrix.

| Class | Classification matrix. Rows: Observed classification | Columns: Predicted classifications | Percent Correct | 1 | 2 | 3 |
|-------|-------------------------------------------------------|-----------------------------------|-----------------|---|---|---|
|       |                                                       |                                   | p=.0441         |   |   |   |
| 1     |                                                       |                                   | 100.000000      | 3.000000 | 0.000000 | 0.000000 |
| 2     |                                                       |                                   | 83.333333       | 0.000000 | 5.000000 | 1.000000 |
| 3     |                                                       |                                   | 100.000000      | 0.000000 | 0.000000 | 59.000000 |
| Total |                                                       |                                   | 98.529400       | 3.000000 | 5.000000 | 60.000000 |

Source: own calculation

5. Conclusions

Using the PCA on the Romanian capital market and the financial data of the 68 listed companies, we can find a model (similar to the one developed by Altman) to assess the bankruptcy risk for the respective companies. In order to define the Z-score function, we use 7 financial indicators that are relevant for the companies’ stability, namely Total Assets, Sales, EBIT, Net Cash Flows from Operating Activities, Net Profit, Total Debt and Market Capitalization. Using the PCA, we find a principal component, that explains more than 82% of the variations of the initial space’s variables and, therefore, we can use it to classify the initial variables. Moreover, we define an Altman-type function, that can be used for assessing the bankruptcy risk.

Still we must be aware that this model can derive misleading conclusions, since it is based on historical data (not prospective), the possibility that the financial data might be subject of accounting shenanigans and creativity. Moreover, the binary-type classification for the companies (as solvent or insolvent) doesn’t consider the occurrence of some possible temporary difficulty in servicing debt (that does not necessarily means that the company is insolvent).

This article provides an overview of the Romanian capital market, by assessing the bankruptcy risk of the issuers listed on Bucharest Stock Exchange. This paper can be of interest to prospective issuers and investors looking for portfolio diversification, because assessing the bankruptcy risk companies face, they can put adequate investment strategies in place.

Acknowledgment

This work was cofinanced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2013-2020, project number POSDRU 159/1.5/S/142115 "Performance and excellence in doctoral and postdoctoral research in Romanian economics science domain”.

References

Altman, E., Marco, G., & Varetto, F. (1994). Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks (the Italian Experience), Journal of Banking and Finance, Vol.18, pp.505-529.
Altman, E.I. (1984). Managing the commercial lending process, Handbook for Banking Strategy, John Willey, New York
Armeanu, D., Vintila G., Moscalu M., Filipescu MO., & Lazar P. (2012). Using quantitative data analysis techniques for bankruptcy risk estimation for corporations, Theoretical and Applied Economics, Volume XIX (2012), No. 1(566), pp.97-112
Blokh, D. (2012). Clustering financial time series via information-theory analysis and rank statistics, *Journal of Pattern Recognition Research*, Vol. 7, pp. 106-115

Borga M. (1998). Doctoral thesis on “Learning multidimensional signal processing”, *Linkoping University, Sweden*, retrieved from http://liu.diva-portal.org/smash/get/diva2:302872/FULLTEXT01.pdf

Cha, HY (2012). Distance between distributions with special topologies of cost matrices, *Journal of Pattern Recognition Research*, Vol. 7, No. 1, pp. 1-25

Chan, MC., Wong CC., & Lam CC. (2000). Financial time series forecasting by neural network using conjugate gradient learning algorithm and multiple linear regression weight initialization, http://machine-learning.martinsewell.com/ann/ChWL00.pdf

De la Torre, F. (2008). A least-squares framework for Component Analysis, retrieved from http://www.cs.cmu.edu/~fiorre/uca_final_revision.pdf

Jolliffe, IT. (2012). *Principal Component Analysis*, Springer, New York

Pai PF., & Lin CS. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting, *Omega*, Vol. 33(6), pp. 497-505

Yan S., Xu D., Zhang B., Zhang HJ., Yang Q., & Lin S. (2007). Graph embedding and extensions: a general framework for dimensionality reduction, *Pattern Analysis and Machine Intelligence*, IEEE Transactions, Vol 29 (1), pp. 40-51