Theoretical and Applied Economics
Volume XIX (2012), No. 1(566), pp. 97-112

Using Quantitative Data Analysis Techniques for Bankruptcy Risk Estimation for Corporations

Ştefan Daniel ARMEANU
Bucharest Academy of Economic Studies
darmeanu@yahoo.com
Georgeta VINTILĂ
Bucharest Academy of Economic Studies
vintilageogeta@yahoo.fr
Maricica MOSCALU
Bucharest Academy of Economic Studies
mari.moscalu@yahoo.com
Maria-Oana FILIPESCU
Bucharest Academy of Economic Studies
oanadicea@yahoo.com
Paula LAZĂR
Bucharest Academy of Economic Studies
lazar_paula@yahoo.com

Abstract. *Diversification of methods and techniques for quantification and management of risk has led to the development of many mathematical models, a large part of which focused on measuring bankruptcy risk for businesses. In financial analysis there are many indicators which can be used to assess the risk of bankruptcy of enterprises but to make an assessment it is needed to reduce the number of indicators and this can be achieved through principal component, cluster and discriminant analyses techniques. In this context, the article aims to build a scoring function used to identify bankrupt companies, using a sample of companies listed on Bucharest Stock Exchange.*

Keywords: aggregate indicator; scoring function; principal components; bankruptcy risk; company.

JEL Codes: C81, D22, G30, G33.
REL Codes: 9B, 11Z.

1. Introduction

The first major model in both financial literature and practice belonged to E.I. Altman, who published in 1968 its original form, known as Z-score function. Although seemingly simple, this model took similar effect on the risk of bankruptcy prediction as the famous Black-Scholes model has had on the evaluation of derivatives.

The model suggested by Altman is based on discriminant analysis, which is used to develop models of classification and prediction of observations belonging to certain groups determined *a priori*. To this end, the discriminant analysis builds a classifier based on a set of observations and indicators characteristic for these observations. In the case of Altman model the set of observations is represented by a number of companies classified by the author in solvent and insolvent, and the considered indicators are certain financial ratios based upon the financial situation of companies is analyzed.

Z-score function proposed by Altman is actually an application of a linear classifier (Fisher type), with the following form:

$$Z(r_1, r_2, \ldots, r_n) = \alpha_0 + \alpha_1 \times r_1 + \alpha_2 \times r_2 + \ldots + \alpha_n \times r_n,$$

where:

$r_1, r_2, \ldots, r_n$ = the rates used for developing the classification model;

$\alpha_1, \alpha_2, \ldots, \alpha_n$ = the coefficients for each rate considered;

α_0 = the intercept for the classification function.

Based on the score of each company, it is performed the allocation to one of the two categories, namely the bankrupt companies or solvent companies. Also, based on the Z score it is estimated the probability of bankruptcy of the company.

Altman's original version of the model proposed in 1968 is as follows (Altman, 2002, p. 14):

$$Z = 1.2xr_1 + 1.4xr_2 + 3.3xr_3 + 0.6xr_4 + 1.0xr_5,$$

where

$r_i, i = \overline{1,5}$ are defined below:

$$r_1 = \frac{\text{Working capital}}{\text{Total assets}};$$

$$r_2 = \frac{\text{Retained earnings}}{\text{Total assets}};$$

$$r_3 = \frac{\text{Earnings before interest and taxes(EBIT)}}{\text{Total assets}};$$

$$r_4 = \frac{\text{Market value of equity}}{\text{Book value of total liabilities}};$$

$$r_5 = \frac{\text{Sales}}{\text{Total assets}}$$

Altman has defined three zones for classification of the companies:

- $Z > 2.99$: safe zone; the probability of bankruptcy is very low.
- $1.8 < Z < 2.99$: grey zone; the probability of bankruptcy is medium.
- $Z < 1.8$: distress zone; the probability of going bankrupt is high.

The original data sample consisted of 66 firms, publicly held manufacturers on the American market, half of which had filed for bankruptcy. Later on, Altman has re-estimated the model based on a date set of private companies, as follows:

$$Z' = 0.717xr_1 + 0.847xr_2 + 3.107xr_3 + 0.420xr_4 + 0.998xr_5,$$

where:

$$r_1 = \frac{\text{Working capital}}{\text{Total assets}};$$

$$r_2 = \frac{\text{Retained earnings}}{\text{Total assets}};$$

$$r_3 = \frac{\text{Earnings before interest and taxes(EBIT)}}{\text{Total assets}};$$

$$r_4 = \frac{\text{Market value of equity}}{\text{Book value of total liabilities}};$$

$$r_5 = \frac{\text{Sales}}{\text{Total assets}}$$

The zones of discrimination for Z' score are:

- $Z' > 2.9$: safe zone;
- $1.23 < Z' < 2.99$: grey zone;
- $Z' < 1.23$: distress zone.

There is a third version of Altman model, updated and extended, which has the benefit of usage for non-manufacturer industrials and emerging market credits:

$$Z'' = 6.56xr_1 + 3.26xr_2 + 3.72xr_3 + 1.05xr_4,$$

where:

$$r_1 = \frac{\text{Working capital}}{\text{Total assets}};$$

$$r_2 = \frac{\text{Retained earnings}}{\text{Total assets}};$$

$$r_3 = \frac{\text{Earnings before interest and taxes(EBIT)}}{\text{Total assets}};$$

$$r_4 = \frac{\text{Book value of equity}}{\text{Total liabilities}}$$

In this case the zones of discrimination are below:

- $Z'' > 2.6$: safe zone;
- $1.1 < Z'' < 2.6$: grey zone;
- $Z'' < 1.1$: distress zone with high risk of going bankrupt.

Altman's model, so used in financial practice, has found to be over 70% accurate in predicting bankruptcy (Stancu, 2007, p.787).

Another classification model, similar to Altman's, was developed by the economists J. Conan and M. Holder in 1979 and it is as follows:

$$CH = 0.24xr_1 + 0.22xr_2 + 0.16xr_3 - 0.87xr_4 - 0.1xr_5,$$

where:

$$\mathrm{r}_1 = \frac{\text{Gross operating surplus}}{\text{Total liabilities}};$$

$$r_2 = \frac{\text{Permanent capital}}{\text{Total assets}};$$

$$r_3 = \frac{\text{Working capital} - \text{Stock}}{\text{Total assets}};$$

$$r_4 = \frac{\text{Financial expenditures}}{\text{Net sales}};$$

$$r_5 = \frac{\text{Personnel expenditures}}{\text{Added value}}$$

According to Conan-Holder model, a CH value equal to -0.21 means a bankruptcy probability of 100%, a score of 0.068 indicates a probability of 50%, and the CH score of 0.164 implies a bankruptcy probability of 10%.

2. Principal components analysis

We now intend to develop a scoring function similar to Altman's on a sample of 60 Romanian companies listed on the Romanian stock exchange to highlight both their financial strength but also their ability to meet the obligations. This way we took into account a total of seven economic and financial indicators for the activity of the companies (*total assets* – Activ total, s*ales* – CA, *operating profit* - EBIT, *net cash flow from operating activities* - CF, *net profit* - PN, *total liabilities* – Datorii totale *and average market value* - CB.

First, we standardized the considered indicators. Table 1 shows the correlation matrix for the seven original variables. Obviously, the main diagonal elements of the matrix are equal to unity:

Table 1

The correlation matrix of the original variables

	Correlations (baza de date + indicatori2010.sta)						
Variable	Activ total	CA	Datorii totale	PN	EBIT	CB	CF
Activ total	1,000000	0,955671	0,904545	0,811757	0,883171	0,984581	0,864128
CA	0,955671	1,000000	0,967587	0,644614	0,741079	0,905970	0,803860
Datorii totale	0,904545	0,967587	1,000000	0,493520	0,608211	0,826175	0,686195
PN	0,811757	0,644614	0,493520	1,000000	0,990361	0,891671	0,839384
EBIT	0,883171	0,741079	0,608211	0,990361	1,000000	0,944330	0,876308
CB	0,984581	0,905970	0,826175	0,891671	0,944330	1,000000	0,899479
CF	0,864128	0,803860	0,686195	0,839384	0,876308	0,899479	1,000000

Source: own results.

The correlation matrix shows the close relationship existing between all seven variables considered, predicting a better representation of them in a substantially reduced number of new variables, principal components. The existence of strong correlations between the analyzed variables diminishes the individual significance of the latter, on the one hand, and highlights the existence of redundancy information, on the other hand: there is a significant amount of information dissipated in the connections between variables. In our approach, we propose to reduce the dimension of the initial causal space, and to remove redundancy information, and therefore we use principal component analysis method.

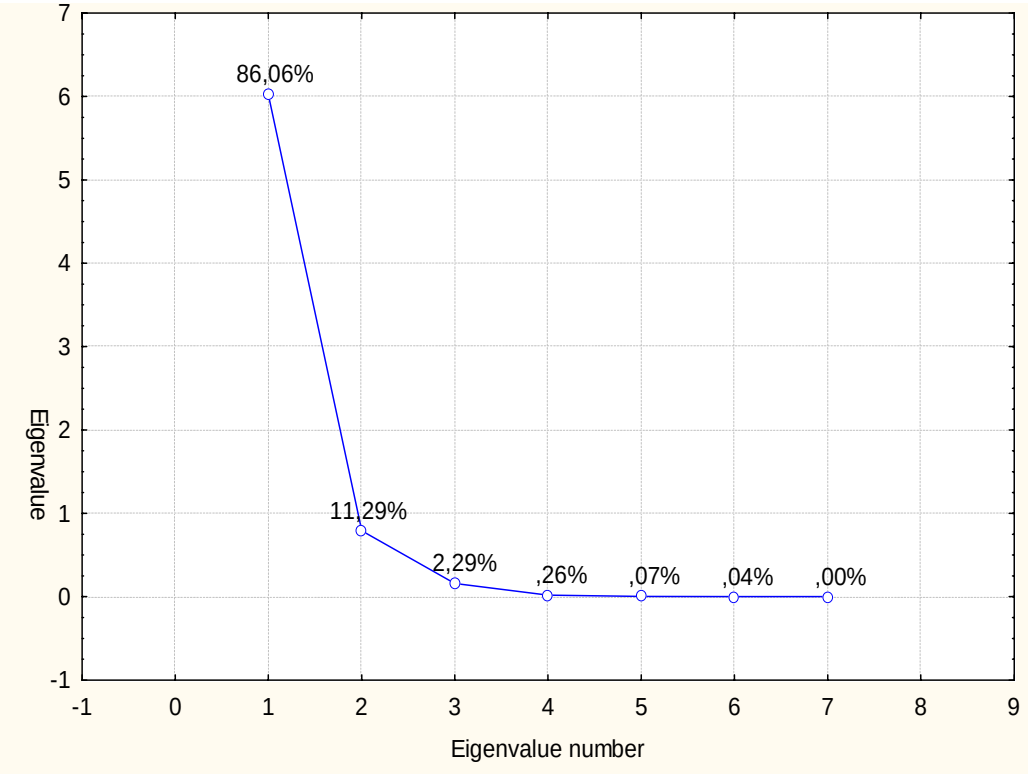
After the standardization of initial data, we present below the eigenvalues resulting from principal component analysis. It is worth mentioning that only the eigenvalues greater than unity are retained for only the principal components that have variance greater than the original standardized variables (mean zero and variance equal to 1) should be extracted, according to Kaiser's criterion. The results are presented in the Table 2.

Table 2

Eigenvalues of correlation matrix

Value number	Eigenvalues of correlation matrix, and related statistics (baza de date + indicatori2010.sta) Active variables only			
	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	6,024163	86,05947	6,024163	86,0595
2	0,790360	11,29086	6,814522	97,3503
3	0,160099	2,28713	6,974622	99,6375
4	0,018130	0,25900	6,992751	99,8964
5	0,004561	0,06516	6,997312	99,9616
6	0,002558	0,03654	6,999870	99,9981
7	0,000130	0,00185	7,000000	100,0000

Source: own results.



Source: own results.

Figure 1. *Eigenvalues of the correlation matrix*

Note that only the first new variable thus formed has its eigenvalue – which is interpreted in terms of variance or informational quantity – greater than unity, so it is retained only the first principal component. The same decision can be taken based on studying the graph in Figure 1.

Once determined the number of principal components retained in the analysis, further testing will proceed to the interpretation of principal components.

We will continue by computing the factor matrix for the single principal component resulted from the analysis. The factor matrix is very important factor in our analysis because its elements (also known as the *factor loadings*) are correlation coefficients between original variables and principal components. The formula for an element of this matrix is:

$$f_{ij} = \frac{\sqrt{\lambda_j}}{\sqrt{VAR(x_i)}} \cdot \beta_{ij}, i = 1, 2, \ldots, n \qquad j = 1, 2, \ldots, k$$

where k is the number of principal components retained in the analysis.

The previous formula gives the correlation coefficient between the original variable i and the principal component j. The relationship is based on demonstration of correlation coefficient definition. This can be argued by defining the correlation coefficient:

$$\rho_{x_i, z_j} = f_{ij} = \frac{COV(x_i, z_j)}{\sqrt{VAR(x_i)} \cdot \sqrt{VAR(x_j)}}$$

Transferring to a matrix, the previous equation can be written as:

$$F = VXW,$$

where X is the covariance matrix between the vectors x and w (the vectors of original variables and principal components) and V and W are diagonal matrix whose elements on the main diagonal are equal to the inverse of the original variables variance and, respectively, of the principal components variance, as it follows:

$$V = \begin{pmatrix} \frac{1}{\sqrt{VAR(x_1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{VAR(x_n)}} \end{pmatrix}$$

$$W = \begin{pmatrix} \frac{1}{\sqrt{VAR(z_1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{VAR(z_n)}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}$$

X matrix is computed from the mathematical definition of covariance:

$$X = COV(x,z) = E\left[(x - E(x)) \cdot (z - E(z))^t\right]$$

Let's assume now the simplifying hypothesis of centralizing the original variables and principal components as it is known that the centralization has no impact on the variance of a stochastic variable. Considering the previous equation, the main equation will be:

$$X = E(xz^t) = E\left(x(B^t x)^t\right) = E(xx^t B) = E(xx^t)B = \Sigma B$$

Replacing, we have:

$$F = VXW = \begin{pmatrix} \frac{1}{\sqrt{VAR(x_1)}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{VAR(x_n)}} \end{pmatrix} \Sigma B \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}$$

Assuming the form of the given matrix Σ and the configuration of the matrix A, it results that indeed a certain element of the matrix F is the correlation coefficient between the original variable i and the principal component j. Factor matrix is shown in Table 3.

Table 3

Factor Matrix

Variable	Factor-variable correlations (factor loadings) Factor 1
Activ total	-0,987380
CA	-0,927273
Datorii totale	-0,844738
PN	-0,875847
EBIT	-0,932918
CB	-0,995254
CF	-0,920787

Source: own results.

To be noted that the new principal component presents high negative correlations with all seven initial variables, of over 85%. Table 4 presents the coefficients of linear combinations that define the principal components (eigenvectors of the correlation matrix), from which we calculate the observations scores in the principal components space:

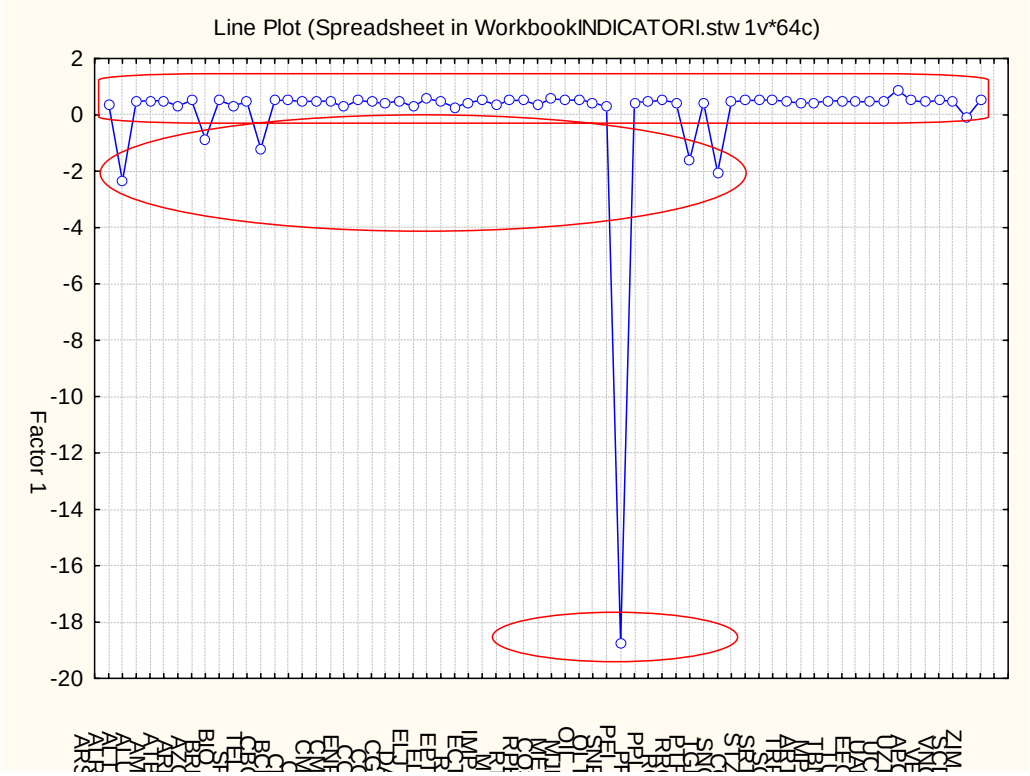
Table 4

Eigenvectors of the correlation matrix

Variable	Eigenvectors of correlation matrix Factor 1
Activ total	-0,402287
CA	-0,377798
Datorii totale	-0,344171
PN	-0,356845
EBIT	-0,380098
CB	-0,405495
CF	-0,375155

Source: own results.

The representation of the companies in the new space provided by the principal component is drawn in Figure 2.



Source: own results.

Figure 2. *Representation of companies in principal components' space*

It should be noted that SNP detaches from the other companies by the principal component, as the recorded values for all variables are significantly higher compared to other companies. Another category of companies is represented by LRA, AZO, TEL, RRC, TGN; they show high levels for all indicators, and the other companies in the third category have low and average indicators.

After the application of principal component analysis we identified one principal component that summarizes over 86% of the information generated by the initial indicators and thus we now identify an Altman model which has the following form:

$$Z = -0.402xr_1 - 0.377xr_2 - 0.344xr_3 - 0.356xr_4 - \\ - 0.38xr_5 - 0.405xr_6 - 0.375xr_7 ,$$

where $r_i, i = \overline{1,7}$ are defined below:

$r_1 = Total\ Assets$; $r_2 = Sales$; $r_3 = Total\ Liabilities$; $r_4 = Net\ \Pr ofit$;

$r_5 = EBIT$; $r_6 = Market\ Value$; $r_7 = CF$.

The analysis of the graphical representation of enterprises against the first principal component and the case scores led to the identification of three zones of classification:

- Z < -2.34: safe zone. The probability of bankruptcy very low.
- -2.34 < Z < -0.102: grey zone. Medium risk of going bankrupt.
- Z >-0.102: distress zone. High probability of bankruptcy.

3. Cluster analysis

Cluster analysis plays an important role in the unsupervised shape classification methods (also known as unsupervised learning methods). The purpose of cluster analysis is rank of data (cases, observations or forms) in significant and relevant structures from an informational point of view, known as classes, groups or clusters.

A key concept used in cluster analysis is therefore *the cluster*. A cluster is defined as a subset of the initial set of objects (observations) that has the property that the degree of dissimilarity between any two objects belonging to the cluster is less than the degree of dissimilarity between any object belonging to the cluster and any object that does not belong to that cluster.

It is worth mentioning a series of technical specifications. First, to evaluate the distance (dissimilarity) between objects (companies listed in Category I) or between clusters it will be used Manhattan distance. Manhattan distance, also called rectangular distance, "City-Block" distance or L1 norm, is

calculated as the sum of absolute values of differences of coordinates for two objects or two variables.

Secondly, we will use as an agglomerative hierarchical clustering method the Ward's classification method. This method is considered to be the most effective and powerful of all hierarchical clustering "algorithms" because it is the only one explicitly dealing with the issue of homogenization of classes i.e. minimizing within cluster variance: at each step, the pairs of clusters with minimum cluster distance are merged.

An important prerequisite of Ward's method is the decomposition of total variance in within cluster variance and between cluster variance, as follows:

$$\sigma_T^2 = \sigma_w^2 + \sigma_b^2,$$

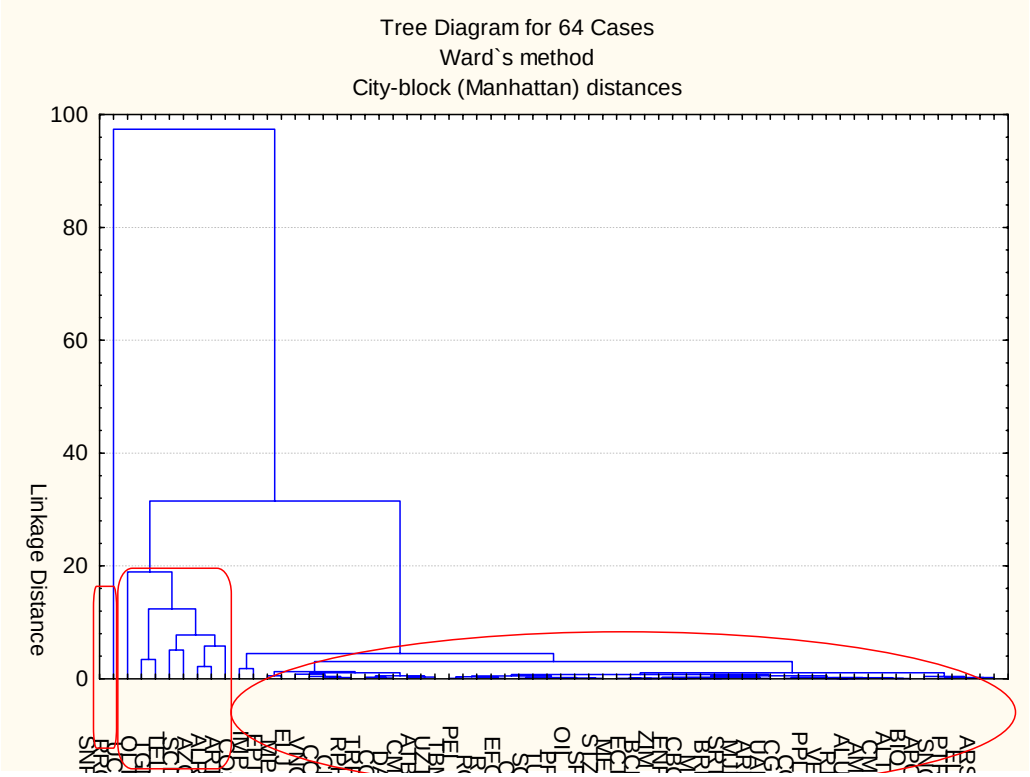
where σ_w^2 and σ_b^2 are within and between cluster variance.

Ward's method is based on the following reasoning: if at any step of the clustering process there are p groups $\{\omega_1, \omega_2, \ldots \omega_p\}$, and the total within-cluster variance is σ_w^2, the pair of clusters will be merged so that intra-cluster variance (which will mandatory be greater than the two individual intra-cluster variances, as the increase in the number of objects in the group makes the latter more heterogeneous, thus with higher variability), noted with $\tilde{\sigma}_w^2$, be the lowest possible, that is to be the solution for the following optimization problem:

$$\min\left(\tilde{\sigma}_w^2 - \sigma_w^2\right)$$

The argument of the optimization function is Ward distance indeed. Applying this technique on our own set of data provided the results presented in Figure 3.

A feature of agglomerative hierarchical techniques (including Ward's method) is to produce more cluster solutions, choosing one of them having to be made according to the objectives set out in the analysis. Selecting a cluster solution is achieved by drawing a parallel to the abscissa axis for different levels of linkage distance. Thus, considering a small level for the linkage distance, we get three clusters of companies (marked in red on the graph in Figure 2), the clusters being similar to those resulting from the PCA: First cluster - SNP; Second cluster - compared to the assignment from PCA, here we have added UCM, OLT and SCD; Third - cluster the rest.



Source: own results.

Figure 3. *Dendrogram for the 60 firms*

4. Discriminant analysis

Discriminant analysis implies using a set of methods, techniques and algorithms in order to determine those characteristics of objects that are the most relevant in terms of recognition of membership to certain classes of default (thus, we deal with a supervised shape classification technique) and to determine the most likely group to belong. Let's note that SNP was removed from the analysis because discrimination can not be done with classes containing a single element. Classes were considered those obtained in cluster analysis. A first result of discriminant analysis is presented in the Table 5.

Table 5

Model's worthiness and the discrimination power of each variable

	Discriminant Function Analysis Summary (data base + indexes 2010.sta) No. of vars in model: 7; Grouping: Clasa* (2 grps) Wilks' Lambda: ,15952 approx. F (7,55)=41,399 p<0,0000					
N=63	Wilks' Lambda	Partial Lambda	F-remove (1,55)	p-level	Toler.	1-Toler. (R-Sqr.)
Activ total	0,174561	0,913815	5,18724	0,026663	0,043724	0,956276
CA	0,278340	0,573099	40,96945	0,000000	0,082668	0,917332
Datorii totale	0,175662	0,908089	5,56677	0,021882	0,008899	0,991101
PN	0,160368	0,994686	0,29381	0,589978	0,000951	0,999049
EBIT	0,159590	0,999539	0,02539	0,873991	0,001363	0,998637
CB	0,180197	0,885235	7,13040	0,009946	0,124843	0,875158
CF	0,313932	0,508124	53,24131	0,000000	0,352962	0,647038

Source: own results.

First, it is observed that the overall discrimination is very strong, as indicated by the table header information: *Wilks's Lambda* statistic has a value of 0.15952 (the closer to zero the statistical value is, the higher the power of discrimination is; the closer to unity lambda is, the lower the discrimination power is), p-value is less than 10^{-4}. In Table 6 are presented the two classification functions.

Table 6

Classification function

	Classification Functions; grouping: Clasa* (data base + indexes 2010.sta)	
Variable	G_1:1 p=,12698	G_2:2 p=,87302
Activ total	26,8541	-36,2707
CA	-30,7169	12,2313
Datorii totale	49,8322	-0,7513
PN	65,1031	26,9284
EBIT	-47,9097	-35,8740
CB	-26,4144	16,5351
CF	14,3467	-4,9985
Constant	-13,8868	-2,4161

Source: own results.

The classification matrix (Table 7) shows the number of correct and incorrect predictions made by the classification model compared to the actual outcomes (target value) in the data.

Table 7

Classification matrix

	Classification Matrix Rows: Observed classifications Columns: Predicted classifications		
Group	Percent Correct	G_1:1 p=,12698	G_2:2 p=,87302
G_1:1	87,5000	7	1
G_2:2	100,0000	0	55
Total	98,4127	7	56

Source: own results.

The matrix shows that all companies classified in the second class after the cluster analysis have been allocated in the same class in discriminant analysis, only one company, SCD, belonging to class 1, was allocated to class 2 after discrimination.

We have obtained a percentage of correct classification of 98.41%.

5. Conclusions

As Heffernan points out (2005) bankruptcy risk predicting models developed based on discriminant analysis (such as Altman model and Conan-Holder model) can easily mislead because, firstly, they rely on historical data. Even if at the time of their development these models were reasonably accurate, their accuracy decreases over time if no action is taken to update the considered variables and/or to recalibrate the models. It is plausible to believe that the financial rates can change in time, even according to the market where they operate. It is necessary for banks to (re)test with a sufficiently high frequency discriminant models and to perform regular updates of risk models used in practice (Heffernan, 2005, p. 161).

A more difficult problem consists in the fact that the result required by the model is binary: either the debtor is solvent or not. In practice, there are several possible scenarios, such as delays in monthly repayments, failure to pay them, failure to pay fees or penalty interest and so on. Most times the debtor lets the bank know about its financial difficulties and the credit terms are renegotiated but discriminant analysis models used may not include the state of solvency, insolvency and restructuring simultaneously.

The suggested techniques of multivariate data analysis prove to be extremely useful when the research is done on a set of objects characterized by a large number of variables, which makes the study of causal dependencies and classification of objects to be difficult. This is our case, the object of the analysis consisting of companies listed on Bucharest Stock Exchange, for

which we considered a representative number of seven characteristics (total assets, net turnover, operating income - EBIT, net profit, net cash flows from operating activities, total liabilities and average market capitalization).

As we have seen, the seven individual variables are characterized by high levels of volatility, but are strongly interrelated, which means that in addition to the intrinsic information content of each variable, there is a significant amount of information dissipated into directly undetectable links between the variables. In this context, principal component analysis is a useful tool, because it can both synthesize information and eliminate duplication of information.

Applying the principal components method on our data set, we obtained a component that synthesizes approximately 86.10% of the information contained in the original causal space. Thus, the transition from seven variables to only one was performed in conditions of minimum information loss, of about 23%. The first principal component salvages 86% of the information in the original space and is strongly negatively correlated with all indicators considered, thus providing information on business volume, profitability of companies (both in the operation and overall activity level), on the market value of shares issued by companies. After considering the application of principal component analysis we identified one principal component that summarizes over 86% of the information generated by the initial indicators and we identify an scoring model that has the following form:

$$Z = -0.402xr_1 - 0.377xr_2 - 0.344xr_3 - 0.356xr_4 - \\ - 0.38xr_5 - 0.405xr_6 - 0.375xr_7 ,$$

The analysis of the graph representation of firms against the first principal component and the scores obtained by the firms allowed us to identify three zones used for their classification:

- $Z < -2.34$: safe zone. Probability of bankruptcy very low.
- $-2.34 < Z < -0.102$: grey zone. Medium risk of going bankrupt.
- $Z > -0.102$: risky zone. High probability of bankruptcy.

Applying the cluster and discriminant analysis helped us testing if the three zones identified by the scoring function are correct.

Acknowledgements

In this paper are disseminated a part of the research results obtained within the Exploratory Research Project, identified as PN II ID-PCE-2008-2, No. 1764, obtained through a CNCSIS competition and financed from governmental resources by the Executive Unit for Financing Superior Education and Academic Scientific Research, Development and Innovation (UEFISCDI).

References

Altman, E.I. (2002). *Managing the Commercial Lending Process*

Armeanu, D. (2005). *Evaluarea riscului activităţii financiare cu aplicaţii pe economia românească*, teză de doctorat

Damodaran, Aswath (2005). *Applied corporate finance*, 2nd edition, Wiley

Everitt, B., Landau, S., Leese, M. (2006). *Cluster Analysis*, Arnold, Fourth edition

Ghahramani, Z., *Unsupervised learning*,
http://www.inf.ed.ac.uk/teaching/courses/pmr/docs/ul.pdf

Gujarati (2005). *Basic econometrics*, McGraw-Hill

Heffernan, S. (2005). *Modern Banking*, Wiley

Helfert, E. (2001). *Financial analysis tools and techniques – a guide for managers*, The McGraw-Hill Companies

Moore, A. „K-Means and Hierarchical Clustering",
http://www.autonlab.org/tutorials/kmeans.html

Nicolae, M., „Utilizarea metodei componentelor principale la analiza „stării" de sănătate a societăţilor bancare", *Revista Oeconomica*, 2004, Bucureşti

Ruxanda, Gh. (2005). *Econometrie II*, suport de curs masterat Management Financiar şi Pieţe de Capital", Bucureşti

Ruxanda, Gh. (2007). *Analiza multidimensionala a datelor*, Master Baze de Date – Suport pentru Afaceri

Simar, L. (2004). *Applied Multivariate Statistical Analysis*, Springer

Spircu, L. (2004). *Tehnici de analiza datelor*, Academia de Studii Economice Bucureşti

Spircu, L. (2006). *Analiza Datelor: Aplicaţii Economice*, Editura ASE, Bucureşti

Stancu, I. (2007). *Finanţe*, Editura Economică, Bucureşti

Vintilă, G., Toroapă, M.G., „Building a Scoring Model for Bankruptcy Risk Prediction on Multiple Discriminant Analysis", The International Conference *Present issues of global economy,* 8th Edition, April 16th-17th, 2011, Annals of the "Ovidius" University, Economic Sciences Series Volume XI, Issue 1 /2011

Vintilă, G. (2005). *Gestiunea financiară a întreprinderii*, Editura Didactică şi Pedagogică, Bucureşti

"Cluster Analysis" -*http://www.statsoft.com/textbook/stathome.html?stcluan.html&1*

"Cluster Analysis – Statnotes, from North Carolina State University",
http://www2.chass.ncsu.edu/garson/PA765/cluster.htm

"Notes on Cluster Analysis", http://www.uic.edu/classes/idsc/ids472/clustering.htm

"A Tutorial on Clustering Algorithms. K-Means Clustering",
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html