Making Predictions of the Profitability on the Financial Markets Using Discriminant Analysis

STEFAN ALEXANDRU IONESCU
Statistics and Mathematics Department
Romanian American University
16, Dumbraveni Street, 4th District, 040779-Bucharest
Romania
ionescusa@yahoo.com
CRISTIANA STEFANIA MURGOCI
Economics Department
Romanian American University
12, Doamna Elena Street, 1st District, Bucharest
Romania
cristianastefania@yahoo.com
CAMELIA MONICA GHEORGHE
Economics Department
Romanian American University
20, Oltenitei Street, Bl. 3E, Apt. 17, 4th District, Bucharest
Romania
camighe@yahoo.com
EMILIA IONESCU
Economics Department
Academy of Economic Studies
16, Dumbraveni Street, 4th District, 040779-Bucharest
Romania
emacont@yahoo.com

Abstract: - In this paper we tried to group in three classes the companies listed without interruption for 6 years from Bucharest Stock Exchange. We used cluster analysis, namely an iterative method of clustering, the k-means algorithm. Using data results, we have made tests for the three classes of prediction using discriminant analysis. Fisher s functions have helped us to make predictions on the affiliation of a new listed company on one of the 3 classes of risk. In this study, emphasis was placed on the liquidity of companies, but also on how efficient are used the raw materials, the basic elements in the current financial crisis. This should give us a clearer picture of companies that are ready to get over this difficult time.
Key-Words: - Discriminant analysis, Cluster analysis, Pattern recognition, Stock exchange, Portfolio analysis, Classifiers.

1. Introduction

In most human activities appears the need to surround, to make the difference, to group or classify certain objects in the form of categories whose determination must be very clear and very natural.

Research aimed at structuring and differentiation of many items on specific categories or classes, depending on the fundamental properties of objects, are known under various names, such as grading, clustering, group or discrimination.

In general, we can say that discrimination and clustering represent activities of arranging objects, individuals or observations, in the form of groups, classes or categories, depending on the degree of similarity or contrast between them.

The overall aim of the pattern recognition theory is identifying at the level of many complex and heterogeneous forms or objects of structures, groups, classes or clusters existing at the latent level in that crowd and those who shape in a natural way, depending on the similarities and differences between many of these items.

2. The Separation of Classes in Forms Space

The first and most difficult problem to be solved in the discriminant analysis is the separation of classes in the prediction of the Ω set. The most direct way of separating Ω set classes is represented by *defining the space of separation surfaces or decision surfaces.* These areas of separation are those which cause the offset of the classes of prediction $\omega_1, \omega_2, ..., \omega_k$ and it pass, necessarily, by the set of objects belonging to the intersection of the classes that separate them.

For reasons like simplification of the classification process, usually, there are used linear separation surfaces, like straight lines, planes or hyperplanes[1]. Separation surfaces are defined by functions known as discriminant functions. Finding an effective way to separate the set elements on disjunctive classes is a difficult problem, especially because of the existence in the set Ω of some objects that belong simultaneously to two different real classes. Affecting of this kind of objects to a class or another could be possible only through the probabilistic calculus

The main problem to be solved in the discriminant analysis is that of constructing criteria or rules of classification, and based on it, we can make predictions about affiliation of new forms, with initially unknown affiliation.

Criteria for classification are known as *classifiers*, and the deduction of these criteria is called *training* of the classifier. The classifier is actually an algorithm which determines the most likely affiliation of a form to a prediction class. The training of the classifier is based on the information contained in a sample form whose affiliation is known before and which is called *training set*.

2.1. Linear Classifier

The first way of approaching problems with classification of discriminant analysis techniques dates from 1933 and it was proposed by Fisher. Subsequently such approaches have developed steadily, and applications based on discriminant analysis were extended to even more areas of activity and have diversified increasingly more.

Most of them and the most useful application of discriminant analysis based on Fisher s criterion are met in the financial-banking field, area in which these kind of techniques are called *credit-scoring techniques* and they are the most important tools to substantiate decisions on granting loans.

Method of discriminant analysis proposed by Fisher is a parametric method, characterized by simplicity and robustness, and offers possibilities of interpretation very useful for analysis. The simplicity of this method stems from the fact that using it does require only the evaluation of estimations of population and its classes parameters, parameters represented by averages, variants or covariants. This is a very important advantage of Fisher's discriminant analysis, in comparison, for example, with Bayes's techniques, whose use involves knowing of the aprioric probabilities.

The theoretical basis of Fisher's discriminant analysis is represented by the variant analysis. Fisher s Criterion defines a way to deduct the discriminant functions on the basis of comparative analysis between intragroup variability and intergroup variability, at the level of classes or analyzed population groups. The discriminant functions deducted on the basis of Fisher s criteria are called also *score functions* and they are *linear* functions.

From a certain point of view, the discriminant analysis can be considered as similar to principal components analysis, which aims to identify general axes relative to the variability of objects to be maximum[2]. The main difference between discriminant analysis and principal components analysis is related to the fact that in principal components analysis the causal space is considered in its entirety, without making any differentiation between its elements in terms of a specific criteria.

In case of principal components analysis, the variability is viewed as a general characteristic of the population analyzed, without taking into account the existence of any structure on this population group or class. Consequently, the variability which is the object of principal components analysis is considered as a whole, without any possibility of decomposition in relation to a certain structure of causal space analyzed.

In contrast, in case of discriminant analysis it is considered that population is divided into groups or classes, and the variability of this population can be split in two components: intergroup variability and intragroup variability.

In addition to the difference mentioned before, in the discriminant analysis the new directions to be identified should not necessarily be orthogonal, unlike principal components analysis in which the directions of maximum variability should check the orthogonal property.

The most important issue of the Fisher's criterion of discrimination between classes of a population is related to the decomposition of variability of this population[3]. We will detail how to split the population variability in relation to the two meanings of it: *simple variability* expressed through the total amount of square deviations and *mixed or composed variability* measured through mixed products of deviations matrix. It is obvious that mixed variability can be defined only for multidimensional objects.

2.2. Defining Fisher's discriminant functions

A Fisher discriminant function is determined as a linear combination of discriminant variables, whose combination coefficients are components of the eigenvector of the matrix $\sum_w^{-1} \cdot \sum_b$. From this way of defining, result, by default, that it can be identified more discriminant functions. The maximum number of possible discriminant functions that can be identified on Fisher s is equal to the number of the distinct and strictly positive values of the matrix $\sum_w^{-1} \cdot \sum_b$.

Since this matrix has the size $n \times n$, when it is strictly positive defined and it has the maximum rank, the result is that the total number of discriminant functions that can be determined is equal to"n". We will present next the way it can be determined all possible discriminant functions.

For this we will note the "n" values of the matrix $\sum_w^{-1} \cdot \sum_b$ with $\lambda_1, \lambda_2, \ldots, \lambda_n$ and we will assume that they are ordered in terms of values that they have as follows:

$$\lambda_1 > \lambda_2 > \ldots > \lambda_n \quad 0 \quad (1)$$

We note with $\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(n)}$, its own"n" vectors of the matrix $\sum_w^{-1} \cdot \sum_b$, associates, in order, with their own values $\lambda_1, \lambda_2, \ldots, \lambda_n$. The first discriminant function is defined using the vector itself, which corresponds to the higher own value, and has the following form:

$$D_1(x) = \beta_0^{(1)} + \beta_1^{(1)} \cdot x_1 + \beta_2^{(1)} \cdot x_2 + \ldots + \beta_n^{(1)} \cdot x_n \quad (2)$$

Since this function corresponds to the highest possible value of the report between the intergroup variant and intragroup variant, it provides the best separability of the classes, in terms of mixed criterion mentioned above. This means that the object projections on the new axe determined by the vector of coefficients $\beta^{(1)}$ can be separated into classes that differentiate in the greatest degree possible and that has the highest possible degree of homogenity.

Similarly, the second discriminant function is defined using the eigenvector which corresponding to the second eigenvalues, namely:

$$D_2(x) = \beta_0^{(2)} + \beta_1^{(2)} \cdot x_1 + \beta_2^{(2)} \cdot x_2 + \ldots + \beta_n^{(2)} \cdot x_n \quad (3)$$

Being determined on the basis of the second eigenvalues of the matrix $\sum_w^{-1} \cdot \sum_b$, this discriminant function corresponds to a smaller value of the report between the intergroup variant and intragroup variant. Consequently, it provides a smaller resolution in terms of separability leave of the set. From this point of view, it is possible that projections of objects on the new axe which has the vector as support to match the classes that are less homogeneous and differentiate less between them.

Finally, with eigenvector associated with the lower eigenvalue, that is vector $\beta^{(n)}$, it determines the last discriminant function, namely:

$$D_n(x) = \beta_0^{(n)} + \beta_1^{(n)} \cdot x_1 + \beta_2^{(n)} \cdot x_2 + \ldots + \beta_n^{(n)} \cdot x_n \quad (4)$$

By comparison with other discriminant functions, this latter discriminant function ensures the poorest separability between classes of the Ω set. The effective number of discriminant functions that must be retained in the analysis depends directly on the number of classes and the number of discriminant variables.

3. The Bucharest Stock Exchange case.

An analysis of 45 companies listed permanently on Bucharest Stock Exchange (BSE) in 2002-2006 has already been done[4]. In an attempt to group them using cluster analysis techniques, it were used hierarchical clustering methods, like single linkage method or Ward s method and iterative methods of clustering. The results were satisfactory for k-means algorithm, an iterative refinement heuristic, invented in 1956 [5] and later developed by Loyd[6]. K means algorithm is an algorithm that tries to group n objects in k clusters, where k n. The objective is to minimize total intracluster variance or to maximize the intercluster variance.

For each case were used 8 variabiles, defined as: liquidity ratios (current ratio) ; solvability ratios (debt-to-equity ratio, total debt ratio) ; efficiency ratios (total assets turnover) ; profitability ratios (return on assets, return on equity, net profit Margin) ; indicators of market value (earnings per share).

3.1. What is new

In today s global crisis, we thought it would be interesting to add two new differentiation variables. The first would be cash ratio, which is the most conservative liquidity ratio, which relates only the cash items (cash and short-term financial assets) to current liabilities. With regard to current accounts, measures of liquidity alone are generally inadequate because differences in the structure of a firm s current assets and current liabilities can significantly affect its true liquidity so the second variable is Inventory turnover. This ratio indicates the efficiency with which the firm uses its inventory. At the same time, the implied ratio indicates the processing time of the company s products. Generally, the higher the value of the ratio, the higher the efficiency of using inventory, but, as in the case of the previous ratio, this ratio is strictly dependent on the industry. Also, it is important to note that a low value for this ratio is problematic, since it might be a sign of capital being tied up in inventory and problems with selling the final products.

Thus, having 10 variables, we standardized and we applied again k-means algorithm. Noting that splitting into 4 classes offers 2 classes almost similar, we tried a partitioning of the objects in only 3 categories for 38 companies listed permanently between 2002-2007.

3.2. Results obtained

Thus, in 2007 we got the first class that is composed of Financial Investment Companies (FICs) and several other companies. Members of this class are characterized by the following aspects: solvability rates have the lowest values which indicate a low level of indebtedness, normally for financial investment companies, total debt ratio has the lowest value, return on assets and return on equity values are the most closer to 0, namely those close to the general average, so natural for these companies that own shares in several companies. Also, net profit margin and earnings per share recorded the highest values. As you see, investments in FIC sites are the most profitable.

The 2nd Cluster is a clusters of average. Net profit margin, earnings per share and total assets turnover have comparable values in standardized variables.

As the difference between the first two classes would be that the first comprises less indebted companies, which have a noticeably high liquidity, profitability and a slightly higher. In other words, we can find in the second group companies with good levels of liquidity and profitability, but under the FICs. Obviously the second option would be investing in companies of the 2nd cluster.

The last category includes companies that have real problems

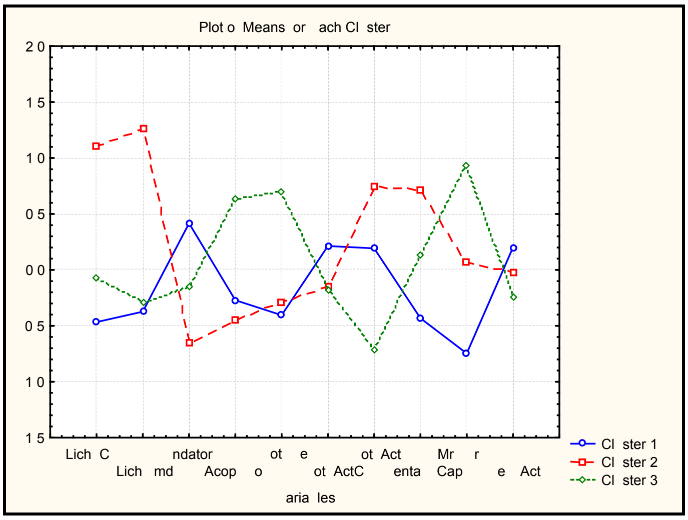

Figure 1

Companies in the latter cluster were the lowest solvability ratios, lowest liquidity ratios and lowest profitability. In figure 1 you can see the attributes of clusters determined with the k-means algorithm.

We have repeated all these operations for the last 5 years and tried to see how companies have evolved as they migrated from one cluster to another. We could see that FICs were part of the same class every year, even if previous years were joined by other companies.

3.3. Discriminant analysis

After finding a proper way to separate the elements of the Ω set (those 38 companies) on classes of prediction ω_1, ω_2, ω_3 (the 3 classes found), the main task of discriminant analysis is to decide on the membership of the 3 classes of new objects from the Ω set or to make predictions concerning the affiliation of these objects. This means that the problem of classification using discriminant analysis can be made as follows:

Giving an object that is known vector x of values of its characteristics, is required to determine the object belonging to one of the 3 classes possible, ω_1, ω_2, ω_3 of the set Ω.

We are trying to verify the data obtained by the K-means algorithm. Classification Matrix (generated by statistics 7.1) checks how many cases were good predictioned and how many were wrong. Thus, for each element A_{ij}, we can interpret: Item A was calculated as belonging to the class i and actually belong to the class j. Figure exposed below (table 1) noted that besides the main diagonal, all the matrix elements are equal to 0. As seen, prediction is considered 100% correct.

	Classification Matrix (Stefan IND 11) Rows: Observed classifications Clolumns: Predicted classifications			
Group	Percent Correct	G 1:1 p=.44737	G 2:2 p=.21053	G 3:3 p=.34211
G 1:1	100,000	17	0	0
G 2:2	100,000	0	8	0
G 3:3	100,000	0	0	13
Total	100,000	17	8	13

Table 1

3.4. Mahalanobis distances:

In table 2 appear the distances between centroids of the 3 classes. Obviously, the main diagonal is zero.

	Squared Mahalanobis Distances (Stefan IND 11)		
Var Clas	G 1:1	G 2:2	G 3:3
G 1:1	0.00000	23.46597	30.84500
G 2:2	23.46597	0.00000	27.93896
G 3:3	30.84500	27.93896	0.00000

Table 2

We could also generate a table in which in the first two columns are showed the cases and belonging to the groups determined by the K-means algorithm. Mahalanobis distances of the following 3 columns are the distances from each firm to centroid of each class. We can notice that for all those 38 companies, the smallest distance corresponds the cluster centroid of which belongs the company. (for lack of space we can not reproduce than a few lines of this table)

	Observed Classif.	1 Mah. Distances	2 Mah. Distances	3 Mah. Distances
OtelInox	2	67,24908	23,58100	68,67287
Azomures	1	10,08497	27,23722	47,59443
Turbomecanica	3	38,10748	42,20765	5,17626

Table 3

Interesting is that for any case, a new company, can predict belonging to a class calculating the for Mahalanobis distances. Obviously, the minimum will show which class belongs to the newcomer.

3.5. Discriminant Functions

The main problem to be solved within discriminant analysis is the construction of criteria or rules of classification, from which it can make predictions concerning the affiliation of new forms, with initial affiliation unknown. Criteria for classification are known as the classifiers, and the deduction of these criteria is called training of the classifiers.

The classifier is actually using an algorithm which determines the most likely belong to a form to a certain class of prediction. The training of the classifier is based on the information contained in a sample form whose affiliation is known aprioric which is called training set.

Determination of discriminant function is equivalent to finding some directions, or vectors, in relation with whom the intragroup variability would be minimal, and intergroup variability to be high.

These directions will define the axes of discriminant space against and can be identified in the form of linear combinations of descriptions variables selected in the analysis.

In conclusion, we can say that Fisher's discriminant functions are linear functions with the following form:

$$D(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \quad (5)$$

Where $\beta_0 = -(\beta_1 \cdot \mu_1 + \beta_2 \cdot \mu_2 + \ldots + \beta_n \cdot \mu_n)$ is the free element, and coefficients $\beta_1, \beta_2, \ldots, \beta_n$ are components of an eigenvector of the matrix $\sum_w^{-1} \cdot \sum_b$.

Classification Function parameters, calculated using Statistics 7.1 appear in table 4:

Variable	Classification Functions; Grouping Var Clas(Stefan IND 11) G 1:1 p=.44737	G 2:2 p=.21053	G 3:3 p=.34211
Lich CRT	-2.42295	0.72274	2.72371
Lich Imd	0.68960	3.00506	-2.75105
Indator	1.79983	-2.46766	-0.83506
Acop Dob	-1.63140	-0.84026	2.65045
Vrot Deb	-1.83965	-1.87618	3.56026
Vrot ActC	1.12591	-0.94167	-0.89285
Vrot ActT	1.14229	2.86493	-3.25680
Rentab Cap	-1.98692	1.05747	1.94753
Mrj Br V	-2.07831	1.25443	1.94583
Rez Act	1.74793	-0.62868	-1.89887
Constant	-3.82054	-6.67595	-6.01773

Table 4

The discriminant functions will be:

$D_1(X) = -3,82 – 2,42X_1 \quad 0,69X_2 \quad \ldots \quad 1,75X_{10}$ (6)
$D_2(X) = -6,68 \quad 0,72X_1 \quad 3,01X_2 \quad \ldots - 0,63X_{10}$ (7)
$D_3(X) = -6,02 \quad 2,72X_1 – 2,75X_2 \quad \ldots - 1,90X_{10}$ (8)

The coefficients for the first discriminant function are derived so as to maximize the differences between the group means. The coefficients for the second function are also derived to maximize the difference between the group means, but the values of the functions are not correlated. The second function is orthogonal to the first and the third is orthogonal to the second. Variables will be even values of the 10 indicators normalized.

What is most interesting about the functions of classification is that it may set belonging to a set of classes for any new company whose indicators are known but unknown membership.

3.6. The a prioric and posterior probabilities:

To estimate the aprioric probabilities P_1, P_2, P_3, is calculated the number of cases or the number of firms in each class using the information on the K-means algorithm. Then we determine T_i which represents the number of firms in class i and then calculate the relative frequencies $P_i = T_i \quad T$.

Class	Prior Pro a ilities o Classi ications Ste an 11 1 17 000	2 8 000	3 13 000
Pro a ility	0 447368	0 210526	0 342105

Table 5

We will assume that the probability density of the classes are normal, in other words are like the following:

$$f_{\omega_i}(x) = \frac{1}{(\sqrt{2\cdot\pi})^n \cdot |\Sigma_i|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\cdot(x-\mu^{(i)})^t \cdot \Sigma_i^{-1} \cdot (x-\mu^{(i)})} \quad i=1,2 \ (9)$$

a posterior probability will be:

$$P(\omega_i / x) \approx \frac{P(\omega_i) * f_{\omega_i}}{f_\Omega} \quad i = 1, 2, 3 \quad (10)$$

In the following tables we have in the first two columns firms and classes predicted using the K-means algorithm. Thus, the next 3 columns are presented as a posterior probability object to belong to the class 1, 2 or 3 and in the second table, the last three columns can be found standings probability, for example classes which belongs to the highest probability.

Note that the algorithm K has been very precisely, in most cases with a probability of over 99% respectively as subject to belong to the predicted class of K-means algorithm. (for lack of space we can not reproduce than a few lines of this tables)

	ser ed Classi	1 pro	2 pro	3 pro
tel no	2	0 0000	1 0000	0 0000
A om res	1	0 9999	0 0000	0 0000
r omecanica	3	0 0000	0 0000	1 0000

Table 6

	ser ed Classi	Highest Pro	Second Highest	hird Highest
tel no	2	2	1	3
A om res	1	1	2	3
r omecanica	3	3	1	2

Table 7

4 Conclusion

With discriminant analysis model assumptions we have checked K-means algorithm and we have succeeded in calculating the classification of features that may help in future predictions.

The most useful application of discriminant analysis are seen in the banking area in which techniques are called and credit-scoring wich are the most important tools for the decision on the granting of loans. Such firms can be divided into classes of trust and credit decision to make depending on membership class.

Another area would be marketing, clients can be divided into different classes of interest for those who do studies. Last but not least the establishment of development areas can be based on algorithms presented in this paper.

This is an important step towards the following researches and also represents an efficient instrument in the context of the global financial crisis. given that the last period were recorded depreciation on all markets, BSE has not been an exception, remains to be seen if the clusters found will remain approximately similar to or will change radically.

Remains a bad opinion of the authors for the Romanian economy is not fully reflected in BSE, many companies are not listed.

5. What to do

As a continuation of this work, with application on the capital market, we can see the following points:

- an analysis of evolution in the dynamic; observation of clusters changing from one year to another and explanations of migration cases from one cluster to another;
- combination of companies with indicators on the evolution of capital market. to view the indicators that most influence the decision to invest;
- determining the structural movement of the Romanian economy depending on size and performance of firms;
- deduction of certain classes of risk and that the risk classes of companies listed on BSE;
- construction of models for character developments phenomenon Romanian scholar; building portfolios based on clusters, instead of the classical.

References:

[1] Ruxanda Gh., *Analiza datelor*, Bucuresti, Editura ASE, 2001

[2] Chris Ding and Xiaofeng He. K-means Clustering via Principal Component Analysis . *Proc. of Int'l Conf. Machine Learning* (ICML 2004), pp 225-232.

[3] Fisher R.A . The use of multiple measurements in taxonomic problems, *Ann. of Eugenics*,7, 1936, p179-188.

[4] Ionescu, S.A.; Gheorghe, C.M. Zaheu, I. Using Pattern Recognition Techniques To Determine Profitability on Financial Markets. 0641-0642, *Annals of DAAAM for 2008 & Proceedings of the 19th International DAAAM Symposium*, pp 321, 2008

[5] H. Steinhaus. *Sur la division des corp materiels en parties*. Bull. Acad. Polon. Sci., C1. III vol IV, 1956. :801– 804,

[6] S. Lloyd, Last square quantization in PCM's. Bell Telephone Laboratories Paper (1957). *Published in journal much later: S. P. Lloyd. Least squares quantization in PCM. Special issue on quantization, IEEE Trans. Inform.* Theory, 28, 1982.:129–137.