CONSIDERATIONS ON THE USE OF THE DISCRIMINANT ANALYSIS IN EVALUATING BANKRUPTCY RISK

Introduction

The analysis of the bankruptcy risk has been a necessity and a great challenge for banks and not only, and the solutions suggested have been varied and particular to national economies. This paper brings forward the idea of the necessity of using discriminant analysis to create a multidimensional model for the evaluation of the bankruptcy risk. This model should be specific to the economical environment of Romania.

Initially developed for biological research, the discriminant analysis represents an instrument that can identify subspecies having relatively similar features. At present, the method is widely applied in various fields, such as: marketing, social activities, electoral analyses, engineering, etc.

The working hypothesis in applying the discriminant analysis is structured on the following aspects:

- The population observed is made up of at least two distinct subpopulations;
- A linear function can be determined on the basis of certain figures or attributes of the population that will allow the analyst to distinguish between the two subpopulations.

Economic units form the population studied by us. Les us suppose one tries to identify which of the companies applying for a credit are in the bankruptcy risk zone and which ones have a safe future. The subjects – the researched population – are differentiated according to various aspects: fields of activity, environment, capital, turnover, number of employees, markets, etc.

Most of these aspects can be quantified in which case we can suppose there exists a linear combination of these figures that can allow the specialist to discriminate between companies on the brink of bankruptcy and those the bank can rely on. The difficulty lies in achieving the linear function in such a way to minimize the number of errors.

An. Inst. cerc. ec. „G. Zane", t. 10–11, Iași, 2001–2002, p. 103–111

We would like to emphasize the fact that the discriminant analysis should not be mistaken for the cluster analysis, which focuses on population groupings. The population observed is already divided into two or several subpopulations. The discriminant analysis assists only in determining membership to one group or another.

Consequently, one must determine a linear function of the type:

$$z=\sum_{i=1}^{n} p_i R_i + k$$

where p_i = the quotient of the factor i,
R_i= the factor (predictor) i taken into account,
k = constant.

The discriminant analysis will be briefly presented here. In Figure 1 it is evident that, given the difficulties in delimiting exactly the two subpopulations (safe companies (F_s) and risk companies (F_r)), one can not be able to adequately classify the subjects in the zone $[z_1, z_2]$.

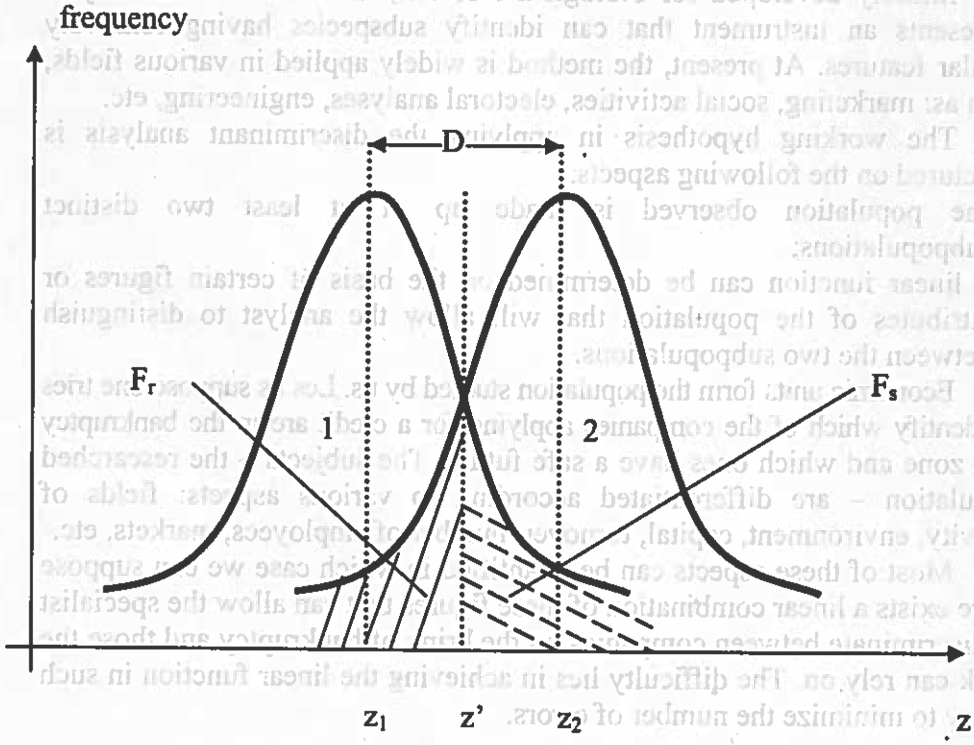

Figure 1. Normal distribtion of safe companies and risk companies

One of the solutions is that for every $z_1<z<z'$ the subject would belong to companies posing high risk while for $z'<z<z_2$, the subject should be listed among safe firms. In this case, all the subjects with bankruptcy risk and that meet the $z'<z<z_2$ condition shall be classified as safe. Similarly, all safe companies meeting the condition $z_1<z<z'$ shall be considered unsafe. Bankruptcy is an irreversible process that has a negative impact not only on the company involved but also on those that have established economic relations with it. Minimizing these classification errors is therefore a prerequisite.

Two methods of tackling the issue have developed in time – one developed by Newman and Pearson, the other by Fisher.

Newman and Pearson ([1], p. 265) suggest the discriminant analysis should be treated as an attempt to minimize the total number of classification errors, for both safe and unsafe companies.

This aim must be attained using a mathematical model, expressed through the function z.

Fisher ([1], p. 265) regards the discriminant analysis as a function that creates the highest number of separations between the two subpopulations. This is equivalent to a minimization of errors, but has a different mathematical model as its basis.

Models used in evaluating bankruptcy risk

Creating a mathematical model for evaluating company bankruptcy risk has been of major interest ever since the 1960's. Some of the most widely used models abroad, are the Altman model, the Conan and Holder model, the model of the French Central Bank, etc. All these models propose a single linear function by which firms are classified as follows: safe or low risk companies (when z is higher than a safety level), uncertain companies and high bankruptcy risk companies (z being lower than a minimum threshold) (see [1], p. 428).

The models mentioned above have limited applicability and are used in the countries where they were created. They usually tackle the big-company section in the industrial production field. Also, they cannot always predict bankruptcy cases at a particular time. For instance, the Altman model managed to predict 75 % of bankruptcy cases (see [3], p. 56).

Due to the specific moment of their emergence and the singularity of every national economy these models cannot be applied to Romanian companies. In Romania, the most widely used models are the one used by the Romanian Commercial Bank (see [2], p. 418) and the Cămăşoiu – Negoescu model (see [3], p. 58).

The Cămăşoiu – Negoescu model is based on a function of the type below:

$$z = \frac{\sum_{i=1}^{10} k_i R_i}{100 \cdot \sum_{i=1}^{10} k_i} \tag{1}$$

which, by the substitution care, $p_i = \frac{k_i}{100 \cdot \sum_{i=1}^{10} k_i}$ will have the linear form:

$$\Leftrightarrow z = \sum_{i=1}^{10} \frac{k_i}{100 \cdot \sum_{i=1}^{10} k_i} \cdot R_i \Leftrightarrow z = \sum_{i=1}^{10} p_i R_i \tag{2}$$

where: i = the index of the factor $i = \overline{1,10}$;

k_i = the weights of the factor R_i;

R_i = the predictor i.

It is worth noting that in the case of this model, apart from data that characterise economic output, elements of analysis of the employed personnel are taken into consideration (age, the rate of commuters).

The model of the Romanian Commercial Bank is rather special and contains several indicators that characterise the financial situation of the company([2], p. 498 and [3], p. 62), each of them being rated by a number of points: liquidity (x_1) – 4 points; solvency (x_2) –6 points; financial profitability (x_3) - 4 points; asset rotation (x_4)– 4 points; dependence upon supply and sale markets (x_5) – 4 points; collaterals (x_6) – 4 points.

There result 5 classes from the evaluation made by adding the respective indicator rates:

A	20 points	safe
B	16-20 points	
C	11-15 points	uncertainty
D	6-10 points	high risk
E	0-5 points	

If the indicators were normalized and, starting from the rating given, we determined their specific weights, one would obtain a function of the type:

$$z = \sum_{i=1}^{6} p_i R_i \tag{3}$$

where: i = is the index of the factor, $i = \overline{1,6}$;
k_i = represents the weights of the factor R_i;
R_i = factor i.

Both techniques provide a discrimination of firms into high risk, uncertain and financially sound companies. We have shown that starting from models already existing in our country, we can determine a linear function that can allow us to discriminate between the two subpopulations. The application hypotheses of the discriminant analysis are secured.

The advantages of the discriminant analysis

A very important stage in applying this analysis method is the selection of the predictors. The selection requires the knowledge and understanding of the studied phenomenon in order to identify the most consequential values through which we can optimally characterize the studied phenomenon and exactly capture its particularities.

The models presented previously are constructed on the basis of a single discriminant function and usually lead to a classification of companies in two large groups. The discriminant analysis allows the construction of several functions, in accordance with the number of groups we intend to obtain and the selected predictors. By using several discriminant functions we construct multidimensional space, the number of dimensions being provided by the number of discriminant functions. By its method of construction, a discriminant function integrates the maximum level of information. The rest of the information remains to be explained by constructing the second discriminant function.

Determining the dimensions by which groups differ maximally and predicting membership of one group or another both depend on the number of subpopulations and the nature of the predictors used. Their selection is made for theoretical reasons as well as for practical motivations (the costs of collecting and processing data).

The maximum number of functions that can be constructed is the following:

$$\min(k-1, r) \tag{4}$$

where: k is the number of groups into which we have to classify the subjects, and r is the number of predictors taken into account.

The model with a single function (unidimensional) provides both advantages and disadvantages. The favorable points include:

– it includes the most significant predictors;

– easily applied model;

– requires a limited quantity of data;

– lower costs, etc.

The major disadvantages of the multidimensional model are the following:

- the unidimensional model incorporates 'most' of the required data – which is provided especially by the most significant predictor on which the function is constructed, a fact which proves to be ever more insufficient. Consequently, a more effective coverage of the data is required (over 90%).
- it discriminates between 3 subpopulations at the most. The increase in the number of populations will lead to increased homogeneity. The main problem is classification in cases where heterogeneity is high. Thus, the subjects tend to be distributed in widely dispersed groups.
- it uses a limited number of predictors (5 or less) because of the occurrence of multicolinearity when several predictors are used. On the other hand, specialised software packages (SPSS, STATISTCICA, STATGRAPHICS) have implemented a minimum tolerance level in tackling multicolinearity.

Consequently, we propose the construction of a model with two or more discriminant functions, and these functions can be seen as being spatial dimensions. We propose an apriori grouping of the potential factors (predictors) to be used in specific classes (personnel, results, etc). The formation of these classes is focused on theoretical economic reasons (efficiency, classification, etc) as well as on practical bases (for example the method of collecting data).

Supposing we construct P classes of predictors to which correspond P functions, determining such a discriminant function is similar to unidimensional models and involves some problems related to selecting and creating classes of predictors and / or making the discrimination.

The selection and grouping into classes of the best factor has to start from the final objectives of an analysis. In our case, in order to classify firms into safe companies and high-risk ones both the internal efficiency elements and external ones that can lead to bankruptcy must be taken into consideration.

The selection of the best predictors can be made on the basis of previous research or even by employing the discriminant analysis.

The discriminant analysis solves two issues:

1. Choosing most significant factors (predictors) from the point of view of information;
2. Making a classification according to these predictors.

By using the discriminant analysis, the predictors can be classified hierarchically in relation to the quantity of data they incorporate.

Classifying predictors into groups remains an activity specific to the analysis of the efficiency and ratability of a company and not only.

As previously shown, primarily the number of subpopulations forming the classification limits the number of discriminant function analysis.

The idea of a model with several functions is to incorporate a larger mass of data by using by using several predictors more effectively. At the same time, one should note that a high number of functions and predictors will not ensure the achievement of the proposed objectives, due to both the complex calculus and the need to periodically update the model. ([6], p. 144).

Supposing that for the proposed model we obtain n functions for the m classes of predictors. (n ≤ m). Each function manages to explain p_i%, $i = \overline{1, n}$ of the data needed to classify the companies in subpopulations, independently from one another. For example, the first function captures 70% of the data needed for classification, regardless of the second function that explains 60%. Adding up the result will exceed 100%.

$$\text{So}: \sum_{i=1}^{n} p_i > 100\% \tag{5}$$

An optimization criterion in selecting the most significant functions is the quantity of information each of them provides independently.

In practice, if the population examined is formed by K subpopulations the maximum number of discriminant functions will be K-1. Supplementing the number of functions is not optimal above this value. It is more effective to determine a function for each subpopulation by which a subject can or cannot be viewed as belonging to that community.

At the same time, it is noticeable that after the first four functions, the supply of information is increasingly insignificant.

In the present case, we can consider the population observed as being grouped into:

– very safe companies P1
– safe companies P2
– relatively stable companies P3
– uncertain companies P4
– high bankruptcy risk companies P5

Suppose the model has n functions.

A subject (company) is classified through each discriminant function z_i in a particular subpopulation (P1 ,..., P5). Establishing membership to the population P_k is achieved through the relation:

$$k = \min_{i \in M} i$$

where M is the multitude of subpopulation indices into which the analyzed subject was classified. This criterion corresponds to maximum aversion to risk on the part of the lender.

For example, we can say that if the company "X" is defined by the function Z_1 as being very safe, and by function Z_2 as being safe, then it will qualify for the subpopulation of safe companies. The selected class will consequently be the least favorable subpopulation into which the subject was listed according to one of the functions of the model.

The probability that a subject belong to a certain subpopulation becomes a composed probability in relation to each function of the model.

This methodology supposes a reorientation of the strategies of the banks that used to group companies in 2-3 populations only. Banks will be able to provide credit lines with differentiated interest and supplementary collaterals in proportion to the risk they are facing.

The model presented here supposes a hierarchical classification of subpopulations and aims at reducing the banks' risk of collecting detrimental credits, as the emphasis will be on determining all the companies in a difficult situation (other than those in the group of safe and very safe companies).

Conclusions

The discriminant analysis is a relatively new method that is applicable in various fields. The greatest difficultty is the selection of the particular factors that best characterize the phenomenon studied– bankruptcy – and, mainly, the construction of a model that provides the largest mass of data.

These predictors cannot only be static images of the aspects of a company's activity, but they should capture its dynamics. Obtaining an effective function requires the testing of a large number of specific factors in a homogenous set of companies or one made up of companies that are as homogenous as possible.

In order to cover a complex phenomenon, such as risk, its analysis should be made not exclusively by a single discriminant function, but by taking into consideration several dimensions of the characterization space.

A discrimination model is always liable to emendation by taking into account other predictors or dimensions necessitated by the dynamics of the economic environment. In the case of Romanian companies, its creation requires taking account all elements specific for the national environment. The creation of this model involves the co-ordination and combination of the financial participation by banks and of economic research institutes.

References

1. Peter, R.S., Arnim H.M., *Survey Sampling and Multivariate Analysis for Social Scientist and Engineers*, Editura D.C. Heath and Company Lexingthon, Toronto, 1979;
2. Niculescu, Maria, *Diagnostic global strategic*, Editura Economică, Bucureşti, 1997;
3. Isaic-Maniu, A., Mitruţ, C., Voineagu, V., *Statistica pentru managementul afacerilor*, Editura Economică, Bucureşti, 1999;
4. Negoescu, Gh., *Risc şi incertitudine în economia contemporană*, Editura Alter-Ego Cristian, Galaţi, 1995;
5. Gil Lafuente, Ana Maria, *Analiza financiară în condiţii de incertitudine*, Editura AIT Laboratories S.R.L., Bucureşti, 1994;
6. Stancu, I., *Finanţe*, Editura Economică, Bucureşti, 1997;
7. Tacu, Al.P., Jaba, Elisabeta (coordonatori), *Statistica în cercetarea economico-socială*, Editura Junimea, Iaşi, 2001.