**UDK: 658.14/.17| DOI:10.5937/etp2301001K**
Datum prijema rada: 13.12.2022.
Datum korekcije rada: 23.12.2022.
Datum prihvatanja rada: 09.01.2023.

ORIGINALNI NAUČNI RAD

EKONOMIJA
TEORIJA I PRAKSA
Godina XVI • broj 1
str. 1–22

# CONSTRUCTION OF BANKRUPTCY PREDICTION MODEL USING DISCRIMINANT ANALYSIS AND FINANCIAL RATIOS

Kušter Denis[1]

***Abstract:*** *The main aim of this research is to develop a statistical model that can reliably predict bankruptcy of Serbian companies one year before bankruptcy proceedings start. The main motive for the research is the fact that there are not many scientific papers focusing on this important issue in Serbia. Bankruptcy prediction model may be useful for future researchers, but also for business owners and other stakeholkders. Research was conducted using financial ratio indicators and discriminant analysis in IBM's SPSS v.26 program. Initially 100 companies from the territory of Serbia were included in the research, but after data screening and meeting all the assumptions for discriminant analysis, 74 of them were included in the final modelling process. It was confirmed that the commonly used financial ratios and discriminant analysis can be useful in creating a bankruptcy prediction model, since the classification power of the developed model is 71.6% for original grouped cases, and 70.3% for cross-validated cases.*

***Keywords:*** *bankruptcy prediction / business failure / discriminant analysis / statistical analysis / financial analysis / financial ratios.*

## INTRODUCTION

The issue of bankruptcy is a topic that has intrigued researchers for a long time. The first research papers date from the early 1930s. This topic is always current and important, since there is no national economy that is

[1] University of Novi Sad - Faculty of Economics, Segedinski put 9-11, 24000 Subotica, e-mail: kuschter@yahoo.com

not affected by the bankruptcy of companies. Bankruptcy prediction models have evolved over the years, from discriminant analysis, through logistic regression, to modern machine learning methods such as decision trees and neural networks. „Some experts point out the complexity of artificial intelligence methods. They claim that traditional mathematical and statistical methods are comparable to artificial intelligence methods in terms of the accuracy of companies' classifications. As a result, many prediction models based on traditional prediction techniques have still been developed around the world. Given the different opinions of experts on various prediction methods, it can be argued that every method has its advantages and disadvantages, and also limitations of its use" (Svabova, et. al, 2020). The main role of the business failure prediction models is to identify financial problems at a certain time before starting of bankruptcy proceedings. In a large number of studies so far, it is a prediction period of one year before the sign of financial difficulties, although researchers are of the opinion that it is necessary to predict problems at least two to three years before their occurance, in order to be able to react in a timely manner.

The main aim of this research is to generate a bankruptcy prediction model that could predict financial problems one year before bankruptcy proceedings start. That could be useful both for business owners who need to understand if their company is on the right track, but also to external stakeholders who need to make decisions about cooperation with various companies. There are not many insolvency prediction models related to developing countries, like Serbia, which emphasizes the importance of the research. The research is organized as follows: after the introductory part, literature overview is presented. The next chapter focuses on the sample, variables and methodology. The following chapter presents data screening and research results. The final part includes discussion and concluding remarks.

## LITERATURE OVERVIEW

In the scientific community, there are various methods of bankruptcy analysis and prediction, and one of them is the so-called multivariate discriminant analysis (MDA) that will be the focus of this research paper.

As presented in the Table 1, from early 60s to 00s, discriminant analysis was the dominant method in bankruptcy prediction modelling, with descending trend starting from 90s.

**Table 1.** *Statistical methods overview*

| Timeline | MDA | LOGIT Analysis | PROBIT Analysis | Neural Networks | Other |
|---|---|---|---|---|---|
| 1960s | 2 | 0 | 0 | 0 | 1 |
| 1970s | 22 | 1 | 1 | 0 | 4 |
| 1980s | 28 | 16 | 3 | 1 | 7 |
| 1990s | 9 | 16 | 3 | 35 | 11 |
| 2000s | 2 | 3 | 0 | 4 | 3 |
| Total | 63 | 36 | 7 | 40 | 26 |

Source: Bellovary, Giacomino, & Akers, 2007, p. 6

Altman (1968) is one of the pioneers of MDA analysis in terms of bankruptcy prediction. He combined MDA statistical technique with ratio analysis on a sample of 66 companies. The result of the research is a well-known Z-score with 6 variables and overall prediction power of 79%. Deakin (1972) developed MDA model with 14 variables. The prediction power of the model is 77% for bankrupted and 82% for healthy companies one year before bankruptcy proceedings. Taffler (1984) made prediction model for UK companies using MDA technique. His model with 5 factors has the accuracy of 60% for bankrupted companies. Luoma and Laitinen (1991) developed MDA model with 7 factors with predictive power of 64.7% for bankrupted companies and 76.5% for non-bankrupted companies. Research was based on 36 Finnish small and medium limited companies. Odom and Sharda (1993) made general business failure prediction model based on MDA analysis. Sample consisted of 129 companies. Their model with 5 factors has prediction power of 59.26% for bankrupted companies and 89.29% for non-bankrupted companies. Alici (1996) focused on UK manufacturing companies and developed the MDA model with 4 factors and predictive power of 60.12% for bankrupted, and 71.07% for non-bankrupted companies. Dimitras and co-authors (1999) generated MDA model with 12 factors based on a sample of 40 Greek firms. Its accuracy is 63.2% for bankrupted and 68.4% for non-bankrupted companies. Du Jardin (2010) developed an MDA model that has an overall prediction power of 87.2%. It was based on a balanced sample of 1020 companies. Yoon and

Kwon (2010) applied MDA analysis on a balanced sample of 10.000 companies and developed a model with overall classification accuracy of 70.1%. Kim (2011) used stepwise method of variables selection in MDA on a balanced sample of 56 companies, and developed a model with predictive power of 72.6%. Zhou, Lai and Yen (2012) developed an MDA model with the classification accuracy of 64.4%. Lee and Choi (2013) used a non-balanced sample of 1775 companies and developed an MDA model with the overall predictive power of 82%. Zhou et al. (2014) developed a discriminant model based on a sample of 2010 companies. The overall classification accuracy of the model is 71.7%. Slefendorfas (2016) focused on developing bankruptcy prediction model for Lithuanian limited corporations. He included 156 ratios and did stepwise multivariate discriminant analysis. The sample included 145 SMEs, and a model with the overall classification power of 89% (for original grouped cases) was created. Nyitrai (2019) conducted discriminant analysis on a sample of 3370 companies and developed MDA model with 81.55% overall accuracy of classification for training set, and 81.52% for the testing sample.

## RESEARCH SAMPLE, VARIABLES AND METHODS

The research was conducted using instruments of financial and statistical analysis. The financial analysis has been based on financial ratio indicators that are presented in the Table 2. There are 12 commonly used variables in most traditional (Altman, 1968; Beaver, 1966; Deakin, 1972; Chesser, 1974; Ohlson, 1980; Zavgren, 1983; Taffler, 1983; Zmijewski, 1984; McKee, 1995, etc.), but also modern (Pervan & Vukoja, 2011; Cultera and Brédart, 2016; Slefendorfas, 2016; Obradovic, et. al, 2018; Korol, 2019; Papana & Spyridou, 2020; Svabova, et. al, 2020; Vukovic, et. al, 2020; Sfakianakis, 2021, etc.) bankruptcy prediction models, regardless of the statistical method applied. All the financial ratios (Table 2) are independent (explanatory) variables, while the status of being bankrupt or solvent is a dependant variable. Therefore, the dependant variable is dichotomous; it can take only two values. For the purpose of statistical analysis, the dependant variable was coded as follows: 0 = bankrupt companies and 1 = solvent companies. The financial ratios (independent variables) have been calculated for the period of one year before bankruptcy proceedings, since the aim was to develop a model that can predict bankruptcy one year before it occurs.

**Table 2.** *Independent variables overview*

| Variable Symbol | Calculation method | |
|---|---|---|
| | **Numerator** | **Denumerator** |
| V1 | Working Capital | Total Assets |
| V2 | Retained Earnings | Total Assets |
| V3 | Gross Result | Total Assets |
| V4 | Net Result | Total Assets |
| V5 | Equity | Total Assets |
| V6 | Net Cash Flow | Total Assets |
| V7 | Net Cash Flow | Total Liabilities |
| V8 | Net Result | Total Liabilities |
| V9 | Total Liabilities | Total Assets |
| V10 | Current Assets | Current Liabilities |
| V11 | Current Liabilities | Total Assets |
| V12 | Current Assets | Total Assets |

Source: Author

The research sample includes 100 companies from Serbia (large, medium and small). A company is considered bankrupted if it initiated bankruptcy proceedings, while those companies that did not initiate bankruptcy proceedings are considered as solvent. The sample is balanced, meaning that the number of bankrupted companies is equal to the number of solvent companies (50:50), which is the case in most research papers related to bankruptcy prediction. „Starting with Altman (1968), bankruptcy prediction models generally have been based on balanced samples, in which the proportion of failed and non-failed firms is equal, which offers two major advantages. Firstly, it allows the models to concentrate equally on both types of firms to design the classification rules. Secondly, the accuracy rate, one of the simplest and most popular evaluation metrics, can be properly applied only with this procedure" (Veganzones & Severin, 2020, p. 209). Financial statements used in the analysis cover timeline from 2016 to 2021. All the financial data was gathered from "Serbian Business Registers Agency". The selection of bankrupted companies was based on data available at the "Agency for Licensing of Bankruptcy Trustees" webpage, while selection of solvent companies was done on random basis. When it comes to statistical method, multivariate discriminant analysis will be applied. The main aim of multivariate discriminant analysis (MDA) is to classify observations (companies) using a set of independent variables $X = (x_1, x_2, x_n)$ into one of two or more categories (in this case bankruptcy and non-bankruptcy). If each observation's discriminant score $Z_i$ is a linear function of $X_i$, it is possible to write a discriminant function that linearly separates the observations as (Bogdan, Sikic & Baresa, 2021, p. 655):

$$Zi = \beta_0 + \beta_i X_{i1} + \beta_{i2} X_{i2} + \dots + \beta_n X_{im}$$

While discrimination boundary Z* is defined by the set of points where (Bogdan, Sikic & Baresa, 2021, p. 656):

$$\beta_0 + \beta_i X_{i1} + \beta_{i2} X_{i2} + \dots + \beta_n X_{im} = Z'$$

Statistical analysis will be conducted using IBM's SPSS v.26 program. Research hypothesis is defined as follows: *financial ratios and discriminant analysis can be useful in predicting bankruptcy of Serbian companies one year before bankruptcy proceedings start*.

## DATA SCREENING AND RESEARCH RESULTS

For proper application of discriminant analysis, it is necessary to meet several requirements. Discriminant analysis is highly sensitive to *outliers* in data set (Tabachnick & Fidell, 1996). Identification of outliers was performed using SPPS. The analysis was done for each variable individually. After the entire process of outliers' removal, 75 of the initial 100 observations remained. After exclusion of the outliers, the sample was rebalanced once again, in order to maintain a 50:50 ratio of bankrupt and solvent companies. The final sample included in model creation has 74 entities (37 bankrupt and 37 solvent).

Since discriminant analysis is fairly robust to failures of *normality* (Patterson, 1996, p. 84), testing of distribution normality was performed using Kolmogorov-Smirnov and Shapiro-Wilk tests. The results for the above-mentioned tests are presented in the Table 3.

**Table 3.** *Kolmogorov-Smirnov and Shapiro-Wilk normality tests*

| Tests of Normality | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Kolmogorov-Smirnov | | | Shapiro-Wilk | | |
| | Statistic | df | Sig. | Statistic | df | Sig. |
| V1 | 0.137 | 74 | 0.001 | 0.945 | 74 | 0.003 |
| V2 | 0.147 | 74 | 0.000 | 0.875 | 74 | 0.000 |
| V3 | 0.216 | 74 | 0.000 | 0.760 | 74 | 0.000 |
| V4 | 0.226 | 74 | 0.000 | 0.736 | 74 | 0.000 |
| V5 | 0.254 | 74 | 0.000 | 0.692 | 74 | 0.000 |
| V6 | 0.192 | 74 | 0.000 | 0.821 | 74 | 0.000 |
| V7 | 0.234 | 74 | 0.000 | 0.796 | 74 | 0.000 |
| V8 | 0.219 | 74 | 0.000 | 0.794 | 74 | 0.000 |
| V9 | 0.138 | 74 | 0.001 | 0.910 | 74 | 0.000 |
| V10 | 0.197 | 74 | 0.000 | 0.839 | 74 | 0.000 |
| V11 | 0.109 | 74 | 0.031 | 0.927 | 74 | 0.000 |
| V12 | 0.090 | 74 | .200* | 0.948 | 74 | 0.004 |

Source: Authors calculation, SPSS output

The Kolmogorov-Smirnov test (K-S) indicates that the distribution of all the variables except V12 significantly differs from a normal distribution, i.e. V1-V11 are not normal ($p < 0.05$). On the other hand, the Shapiro-Wilk (S-W) test indicates that no variable has normal distribution ($p < 0.05$). According to Field (2009), Sharpo-Wilk test is more powerful in detecting differences from normality. Therefore, S-W test results will be focused in further analysis. Since it is not recommended to generate a discriminant model with variables that do not have a normal distribution, it is necessary to perform a transformation of variables. Logarithm (LOG) transformation of all the variables was done (Tabachnick & Fidell, 1996, p. 98, Pallant, 2007, p. 88; Field, 2009, p. 155, etc.), considering the fact that no variable met the requirement of normality according to the S-W test. After normalization of the variables, another normality testing with LOG transformed variables was performed. The results are presented in the Table 4.

**Table 4.** *Kolmogorov-Smirnov and Shapiro-Wilk normality tests (performed on LOG normalized variables)*

| Tests of Normality | | | | | | |
|---|---|---|---|---|---|---|
| Variable | Kolmogorov-Smirnov | | | Shapiro-Wilk | | |
| | Statistic | df | Sig. | Statistic | df | Sig. |
| LOG_V1 | 0.092 | 74 | 0.199 | 0.974 | 74 | 0.128 |
| LOG_V2 | 0.191 | 74 | 0.000 | 0.809 | 74 | 0.000 |
| LOG_V3 | 0.046 | 74 | .200* | 0.988 | 74 | 0.692 |
| LOG_V4 | 0.061 | 74 | .200* | 0.989 | 74 | 0.768 |
| LOG_V5 | 0.193 | 74 | 0.000 | 0.837 | 74 | 0.000 |
| LOG_V6 | 0.062 | 74 | .200* | 0.993 | 74 | 0.958 |
| LOG_V7 | 0.053 | 74 | .200* | 0.990 | 74 | 0.849 |
| LOG_V8 | 0.078 | 74 | .200* | 0.981 | 74 | 0.310 |
| LOG_V9 | 0.110 | 74 | 0.028 | 0.978 | 74 | 0.236 |
| LOG_V10 | 0.074 | 74 | .200* | 0.989 | 74 | 0.778 |
| LOG_V11 | 0.122 | 74 | 0.008 | 0.971 | 74 | 0.090 |
| LOG_V12 | 0.144 | 74 | 0.000 | 0.891 | 74 | 0.000 |

Source: Authors calculation, SPSS output

According to K-S test, the following variables are normal (p >0.05): LOG_V1, LOG_V3, LOG_V4, LOG_V6, LOG_V7, LOG_V8 and LOG_V10. On the other hand, according to S-W test, the following variables met the criteria of normality (p > 0.05): LOG_V1, LOG_V3, LOG_V4, LOG_V6, LOG_V7, LOG_V8, LOG_V9, LOG_V10 and LOG_V11. Considering the fact that S-W test has more power in detecting deviations from normality, all the variables that passed S-W normality test are qualified for further model development.

Another aspect to be mentioned is *multicollinearity* Multicollinearity occurs when there are predictor variables that are redundant (Hair, Anderson, Tatham & Black, 1998). This aspect was tested using VIF (Variance Inflation Factor) analysis. Some of the variables caused huge VIF values, so they were removed from the further modelling. The variables that qualified for further model development after the initial VIF analysis are presented in Table 5.

**Table 5.** *VIF Analysis*

| Variables | Collinearity Statistics | |
|---|---|---|
| | Tolerance | VIF |
| LOG_V1 | 0.849 | 1.178 |
| LOG_V6 | 0.841 | 1.190 |
| LOG_V8 | 0.714 | 1.401 |
| LOG_V10 | 0.336 | 2.977 |
| LOG_V11 | 0.324 | 3.089 |
| LOG_V9 | 0.301 | 3.326 |

Source: Authors calculation, SPSS output

The results of VIF analysis indicate that there are no problems of multicollinearity with variables LOG_V1, LOG_V6, LOG_V8, LOG_V9, LOG V_10 and LOG_V11. The problem exists when the VIF value is greater than 10 (Choen et. al., 2003; O'brien, 2007).

After data analysis, preparation and screening procedures were completed, the discriminant analysis algorithm was started in IBM's SPSS v.26 program. Table 6 shows the number of entities included in the modelling: 74. Also, another assumption of the discriminant analysis is confirmed - the absence of missing data.

**Table 6.** *Multiple Discriminant Analysis – Case Processing*

| Analysis Case Processing Summary | | | |
|---|---|---|---|
| Unweighted Cases | Category | N | Percent |
| Valid | | 74 | 100.0 |
| Excluded | Missing or out-of-range group codes | 0 | 0.0 |
| | At least one missing discriminating variable | 0 | 0.0 |
| | Both missing or out-of-range group codes and at least one missing discriminating variable | 0 | 0.0 |
| | Total | 0 | 0.0 |
| Total | | 74 | 100.0 |

Source: Authors calculation, SPSS output

The testing of group means equality is presented in the Table 7. The p-values for LOG_V9, LOG_V10 and LOG_V11 are less than 0.05, meaning that these variables are potentially important predictors for discriminant model.

**Table 7.** *Group means equality analysis*

| Tests of Equality of Group Means | | | | | |
|---|---|---|---|---|---|
| Variable | Wilks' Lambda | F | df1 | df2 | Sig. |
| LOG_V1 | 0.992 | 0.596 | 1 | 72 | 0.443 |
| LOG_V6 | 0.998 | 0.176 | 1 | 72 | 0.676 |
| LOG_V8 | 0.986 | 1.024 | 1 | 72 | 0.315 |
| LOG_V9 | 0.711 | 29.274 | 1 | 72 | 0.000 |
| LOG_V10 | 0.704 | 30.254 | 1 | 72 | 0.000 |
| LOG_V11 | 0.732 | 26.389 | 1 | 72 | 0.000 |

Source: Authors calculation, SPSS output

In order to avoid robustness of the model and choose only relevant variables, the *stepwise* method of discriminant analysis was selected. In this type of analysis, model is being built step-by-step by selecting only variables that can best contribute to discrimination between groups. After two steps, LOG_V10 and LOG_V9 were included in discriminant model, as shown in the Table 8.

**Table 8.** *Stepwise discriminant analysis – selected variables*

| Variables Entered/Removed[a,b,c,d] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Step | Variable Entered | Wilks' Lambda | | | | | | | |
| | | Statistic | df1 | df2 | df3 | Exact F | | | |
| | | | | | | Statistic | df1 | df2 | Sig. |
| 1 | LOG_V10 | 0.704 | 1 | 1 | 72.000 | 30.254 | 1 | 72.000 | 0.000 |
| 2 | LOG_V9 | 0.668 | 2 | 1 | 72.000 | 17.651 | 2 | 71.000 | 0.000 |
| At each step, the variable that minimizes the overall Wilks' Lambda is entered. a. Maximum number of steps is 12; b. Minimum partial F to enter is 3.84; c. Maximum partial F to remove is 2.71. d. F level, tolerance, or VIN insufficient for further computation. | | | | | | | | | |

Source: Authors calculation, SPSS output

In Table 9, the whole process of variable selection is presented from the beginning to the last step. All the variables that qualified for model creation are presented in the initial moment – step 0. Selection of variables is based on „F to enter" values (Kinnear & Gray, 2004). „The F-to-enter is a partial multivariate F statistic which tests the additional discrimination introduced by the variable being considered after taking into account the discrimination achieved by the other variables already entered. If this F is small, we do not want to select that variable, because it is not adding enough to the overall discrimination" (Klecka, 1980, p. 57). In the step 1, variable LOG_V10 was selected, considering its highest F value of 30.254. In the step 2, variable LOG_V9 was selected due to its highest F value of 3.851. All the other predictor variables are removed from the analysis in the step 2, because their F values are lower than 3.84.

**Table 9.** *Stepwise discriminant analysis – the process of variables selection*

| Variables Not in the Analysis | | | | | |
|---|---|---|---|---|---|
| **Step** | | **Tolerance** | **Min. Tolerance** | **F to Enter** | **Wilks' Lambda** |
| 0 | LOG_V1 | 1.000 | 1.000 | 0.596 | 0.992 |
| | LOG_V6 | 1.000 | 1.000 | 0.176 | 0.998 |
| | LOG_V8 | 1.000 | 1.000 | 1.024 | 0.986 |
| | LOG_V9 | 1.000 | 1.000 | 29.274 | 0.711 |
| | LOG_V10 | 1.000 | 1.000 | 30.254 | 0.704 |
| | LOG_V11 | 1.000 | 1.000 | 26.389 | 0.732 |
| 1 | LOG_V1 | 0.999 | 0.999 | 0.628 | 0.698 |
| | LOG_V6 | 0.996 | 0.996 | 0.002 | 0.704 |
| | LOG_V8 | 0.983 | 0.983 | 2.095 | 0.684 |
| | LOG_V9 | 0.560 | 0.560 | 3.851 | 0.668 |
| | LOG_V11 | 0.555 | 0.555 | 2.692 | 0.678 |
| 2 | LOG_V1 | 0.988 | 0.554 | 0.330 | 0.665 |
| | LOG_V6 | 0.991 | 0.558 | 0.007 | 0.668 |
| | LOG_V8 | 0.924 | 0.527 | 3.697 | 0.634 |
| | LOG_V11 | 0.449 | 0.449 | 0.710 | 0.661 |

Source: Authors calculation, SPSS output

Log determinants are presented in Table 10. They range from -3.856 to -3.976. As per rule, the log determinants should be equal (Shanthi, 2019, p. 265) or nearly equal. The log determinants of the research are relatively equal, indicating homogeneity of covariance matrices between the groups (Meyerrs, Gamst & Guarino, 2006, p. 272).

**Table 10.** *Log Determinants*

| Log Determinants | | |
|---|---|---|
| Status | Rank | Log Determinant |
| 0 | 2 | -3.976 |
| 1 | 2 | -3.813 |
| Pooled within-groups | 2 | -3.856 |
| The ranks and natural logarithms of determinants printed are those of the group covariance matrices. | | |

Source: Authors calculation, SPSS output

Box's M is 2.742 with F value of 8.887 which is not significant at $p > 0.05$ (Table 11). This indicates equal covariance matrices of the predictors; thus, the assumption of equal covariance is met (Meyerrs, Gamst & Guarino, 2006, p. 272).

**Table 11.** *Box's M test*

| Test Results | | |
|---|---|---|
| Box's M | | 2.742 |
| F | Approx. | 0.887 |
| | df1 | 3 |
| | df2 | 933120.000 |
| | Sig. | 0.447 |
| Tests null hypothesis of equal population covariance matrices. | | |

Source: Authors calculation, SPSS output

Table 12 presents the Eigenvalues. The larger the eigenvalue is, the better is the variance of the dependent variable explained by the developed discriminant function. The canonical correlation of 0.576 exceeds the criterion of 0.5 for a strong relationship (Meyerrs, Gamst & Guarino, 2006, p. 271).

**Table 12.** *Eigenvalues*

| Eigenvalues | | | | |
|---|---|---|---|---|
| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
| 1 | .497[a] | 100.0 | 100.0 | 0.576 |
| a. First 1 canonical discriminant functions were used in the analysis. | | | | |

Source: Authors calculation, SPSS output

Below the Eigenvalues (Table 13), there is Wilks' Lambda indicator. It tests the significance of the eigenvalues. Wilks' Lambda value is 0.668, with p-value < 0.05, meaning further that developed discriminant function explains the variation well.

**Table 13.** *Wilks' Lambda*

| Wilks' Lambda | | | | |
|---|---|---|---|---|
| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
| 1 | 0.668 | 28.656 | 2 | 0.000 |

Source: Authors calculation, SPSS output

After the analysis and screening of the data was done, and after it was confirmed (tested) that the discriminant function meets all the assumptions and criteria, Table 14 gives an overview of the coefficients. Final discriminant function model includes two variables:

- LOG_V9: Log(Total Liabilities / Total Assets)
- LOG_V10: Log(Current Assets / Current Liabilities)

**Table 14.** *Discriminant Function Coefficients*

| Canonical Discriminant Function Coefficients | |
|---|---|
| Variable | Function<br>1 |
| LOG_V9 | -1.359 |
| LOG_V10 | 1.137 |
| (Constant) | -0.446 |
| Unstandardized coefficients | |

Source: Authors calculation, SPSS output

The developed discriminant function model looks more familiar if it is written in the following form:

$$D = -0.446 - 1.359(Log\_V9) + 1.137(Log\_V10)$$

$$\mathbf{D = -0.446 - 1.359\left[Log\left(\frac{Total\ Liabilities}{Total\ Assets}\right)\right] + 1.137\left[Log\left(\frac{Current\ Assets}{Current\ Liabilities}\right)\right]}$$

Also, it is important to mention what are the cut-off values for classification. They are presented in the Table 15.

**Table 15.** *Group Centroids*

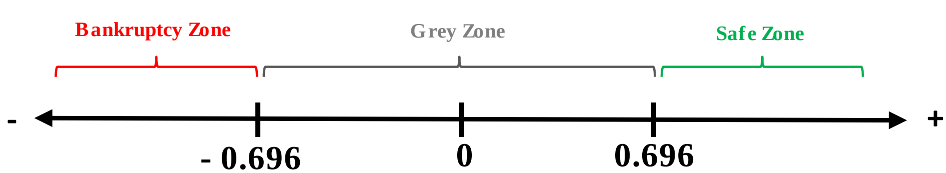
| Functions at Group Centroids | |
|---|---|
| Status | Function<br>1 |
| 0 | -0.696 |
| 1 | 0.696 |
| Unstandardized canonical discriminant functions evaluated at group means | |

Source: Authors calculation, SPSS output

There are three zones of discrimination:

1. D ≤ -0.696 can be interpreted as "*Distress / Risk / Bankruptcy*" Zone,
2. -0.696 < D < 0.696 can be interpreted as "*Grey*" Zone, where it is hard to say whether company is experiencing financial difficulties or not,
3. D ≥ 0.696 can be interpreted as "*Safe / Solvent*" Zone.

**Figure 1.** *Visual representation of discrimination zones*




Source: Author

The results of classification of the observations, according to developed model, are presented in Table 16. The total of 71.6% of original grouped cases are correctly classified, while 70.3% of cross-validated grouped cases are correctly classified.

**Table 16.** *Confusion matrix*

| Classification Results[a,c] | | | | | |
|---|---|---|---|---|---|
| **Status** | **Category** | **Status** | **Predicted Group Membership** | | **Total** |
| | | | **0** | **1** | |
| Original | Count | 0 | **28** | **9** | 37 |
| | | 1 | **12** | **25** | 37 |
| | % | 0 | **75.7** | **24.3** | 100.0 |
| | | 1 | **32.4** | **67.6** | 100.0 |
| Cross-validated[b] | Count | 0 | **27** | **10** | 37 |
| | | 1 | **12** | **25** | 37 |
| | % | 0 | **73.0** | **27.0** | 100.0 |
| | | 1 | **32.4** | **67.6** | 100.0 |

a. **71.6%** of original grouped cases correctly classified.
b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
c. **70.3%** of cross-validated grouped cases correctly classified.

Source: Authors calculation, SPSS output

Considering previous research and models, this classification power can be considered as good. It is important to mention that there are two types of relevant errors that can occur in classification and confirmation of bankruptcy prediction model: Type I and Type II errors. Type I error is present when an entity that has initiated bankruptcy proceedings is classified as “safe/solvent” by a model, while Type II error is present when a solvent entity is classified as “bankrupted”. The developed model has better performances when it comes to the classification of companies that have bankrupted:

- 75.7% (28/37 correct) for original sample and
- 73.0% (27/37 correct) in case of cross-validation.

This further means that Type I errors are lower than Type II errors, which is always the better scenario. Classifying solvent entity as bankrupted is considered as lost opportunity, while classifying risky (bankrupted to be) entity as solvent can lead to wrong investment or other business decisions, reputation damage, loss or even court costs.

## DISCUSSION

There is no national economy for which bankruptcy issues are not important. This topic is especially important for developing countries like Serbia. The main aim of the research was to develop a model that can reliably classify companies into two groups: bankrupted and solvent. A model that has a classification power of 71.6% on the original sample, and 70.3% on the cross-validated sample has been developed. Considering previous research, this accuracy can be considered as good. The overall classification power of the developed model (cross-validated) is in the range of 70-75%, which is the same as the models developed by the following authors: Luoma and Laitinen (1991), Odom and Sharda (1993), Yoon and Kwon (2010), Kim (2011), Zhou et. al (2014). Classification accuracy of the developed model is higher than the following models: Taffler (1984), Alici (1996), Dimitras et. al (1999) and Zhou, Lai and Yen (2012). The following models performed better compared to the developed model, with accuracy of 77% or higher: Altman (1968), Deakin (1972), Du Jardin (2010), Lee and Choi (2013), Sledendorfas (2016) and Nyitrai (2019). The important aspect to mention is that the developed model performs better in predicting bankruptcy than solvency, which is desirable. (Type I errors are lower than Type II errors). The model correctly classified ≈76% of bankrupted companies on the original sample, and 73% of bankrupted companies on the cross-validated sample. That being said, it can be concluded that research hypothesis can be accepted: *financial ratios and discriminant analysis can be useful in predicting bankruptcy of Serbian companies one year before bankruptcy proceedings start.*

## CONCLUSION

The results of this research are important for future researchers, business owners, but also external stakeholders. Based on the detailed data screening and analysis, future researches can identify variables that potentially have importance in creation of bankruptcy prediction models, but also fully understand how discriminant analysis has to be conducted. On the one hand, business owners can use this discriminant function to calculate their *D score* in order to see whether there is a risk of future bankruptcy present for their companies. On the other hand, this model can be useful for external stakeholders since they can use it to calculate risk of starting business relationship with a specific company. However, this research has several limitations that must be pointed out. To begin with, research sample is relatively small, which is the case in most research papers that are exploring bankruptcy predictions. This limitation is reduced by the precise analysis and preparation of the data and the use of advanced software. Furthermore, the sample is dominated by trading and manufacturing companies, meaning that the model may not be best option to predict bankruptcy in some specific economic activities. Also, Serbian economy is dominated by small companies, thus the research sample is also dominated by those companies. This means that the developed model may not show best precision if used for big entities. To conclude with, the model involves only financial data in form of financial ratios, and in order to get a better performing model, it is necessary to include some statistical, non-financial and external (macroeconomic) variables. That can serve as the idea for future research. With respect to the above-stated, it is important to point out that the basic assumption for this prediction model to work properly are correct and non-manipulated positions of financial statements.

## REFERENCES

1. Alici, Y. (1996). Neural networks in corporate failure prediction: The UK experience. Article in *Neural Networks in Financial Engineering*, A. Refenes, Y. Abu-Mostafa, J. Moddy and A. Weigend (eds.). Singapore: World Scientific, 393-406
2. Altman, E. I. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *Journal of Finance 23*(4), 589-609. DOI: 10.2307/2978933

3. Beaver, W. H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 71-111. DOI:10.2307/2490171
4. Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial education*, *Vol.33*, 1-42. Available at: https://www.jstor.org/stable/41948574
5. Bogdan, S., Sikic, L., & Baresa, S. (2021). Predicting Bankruptcy Based on the Full Population of Croatian Companies. *Ekonomski pregled*, *72*(5), 643-669
6. Chesser, D. (1974). Predicting Loan Noncompliance. *The Journal of Commercial Bank Lending*, *56*(12), 28-38.
7. Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2003*). Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates
8. Cultrera, L., & Bredart, X. (2016). Bankruptcy prediction: the case of Belgian SMEs. *Review of Accounting and Finance, 15*(1), 101–119. DOI: 10.1108/raf-06-2014-0059
9. Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research, 10*(1), 167–179. DOI: 10.2307/2490225
10. Dimitras, A. I., Slowinski, R., Susmaga, R., & Zopounidis, C. (1999). Business Failure prediction using rough sets. *European Journal of Operational Research, 263-280.*
11. Du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. *Neurocomputing, 73*(10-12), 2047–2060. DOI:10.1016/j.neucom.2009.11.034
12. Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998). *Multivariate Data Analysis, Fifth Edition*. Upper Saddle River, New Jerseey: Prentice Hall.
13. Kim, S. Y. (2011). Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis*. The Service Industries Journal*, *31*(3), 441–468. DOI:10.1080/02642060802712848
14. Kinnear, P. R., & Gray, D. C. (2004). *SPSS 12 Made Simple.* New York: Psychology Press – Taylor & Frands Group

15. Klecka, W. R. (1980). *Discriminant Analysis*. London: SAGE Publications, Inc.
16. Korol, T. (2019). Dynamic Bankruptcy Prediction Models for European Enterprises. *Journal of Risk and Financial Management, 12*(4), 185–. DOI:10.3390/jrfm12040185
17. Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, *40*(8), 2941–2946. DOI:10.1016/j.eswa.2012.12.009
18. Luoma, M. (1991). Survival analysis as a tool for company failure prediction. *Omega, 19*(6), 673–678. DOI:10.1016/0305-0483(91)90015-l
19. McKee, T. E. (1995). Predicting bankruptcy via induction. *Journal of Information Technology, 10*(1), 26-36. DOI:10.1177/026839629501000104
20. Meyers, L., S., Gamst, G., & Guarino, A., J. (2006). *Applied Multivariate Research – design and interpretation.* London: SAGE Publications
21. Nyitrai, T. (2019). Dynamization of bankruptcy models via indicator variables. *Benchmarking: An International Journal*, *26*(1), 317-332. DOI:10.1108/bij-03-2017-0052
22. O'brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors*. Quality & Quantity, 41*(5), 688
23. Obradovic, B. D., Jaksic, D., Rupic, B. I., & Andric, M. (2018). Insolvency prediction model of the company: the case of the Republic of Serbia. *Economic Research, 31*(1), 139-157, DOI: 10.1080/1331677X.2017.1421990
24. Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *IJCNN International Joint Conference on Neural Networks*. DOI:10.1109/ijcnn.1990.137710
25. Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109-131. DOI: 10.2307/2490395
26. Pallant, J. (2009). *SPSS: Guide for survival. Belgrade*: Mikro knjiga
27. Papana, A., & Spyridou, A. (2020). Bankruptcy Prediction: The Case of the Greek Market. *Forecasting*, *2*(4), 505–525. DOI:10.3390/ forecast2040027
28. Patterson, D. W. (2000). *Bankruptcy prediction: A model for the casino industry*. UNLV, Retrospective Theses & Dissertations, 2457.

29. Pervan, I., Pervan, M., & Vukoja, B. (2011). Prediction of company bankruptcy using statistical techniques – case of Croatia. *Croatian Operational Research Review, 2*(1), 158-167. Available at: https://hrcak.srce.hr/96660
30. Sfakianakis, E. (2021). Bankruptcy prediction model for listed companies in Greece. *Investment Management and Financial Innovations, 18*(2), 166-180. DOI: 10.21511/imfi.18(2).2021.14
31. Shanthi, R. (2019). *Multivariate Data Analysis Using SPSS and AMOS.* New Delhi: MJP Publishers
32. Slefendorfas, G. (2016). Bankruptcy Prediction Model for Private Limited Companies of Lithuania. *Ekonomika*, *95*(1), 134-152. DOI: 10.15388/Ekon.2016.1.9910.
33. Svabova, L., Michalkova, L., Durica, M., & Nica, E. (2020). Business Failure Prediction for Slovak Small and Medium-Sized Companies. *Sustainability, 12*(11), 4572. DOI: 10.3390/su12114572
34. Tabachnick, B. G. & Fidell, L. S. (1996). *Using Multivariate Statistics*. New York: Harper Collins College Publishers.
35. Taffler, R. (1984). Empirical models for the monitoring of UK corporations. *Journal of Banking and Finance, 8*(2), 199-227.
36. Taffler, R. J. (1983). The Assessment of Company Solvency and Performance Using a Statistical Model. *Accounting and Business Research, 13*(52), 295–308. DOI:10.1080/00014788.1983.9729767
37. Veganzones, D., & Severin, E. (2020). Corporate failure prediction models in the twenty-first century: a review. *European Business Review*, *33*(2), 204–226. DOI:10.1108/ebr-12-2018-0209
38. Vukovic, B., Milutinovic, S., Milicevic, N. & Jaksic, D. (2020). Corporate Bankruptcy Prediction: Evidence from Wholesale Companies in the Western European Countries. *Ekonomický časopis, 68*(5), 477-498. Available at: https://www.researchgate.net/publication/340917901_Corporate_Bankruptcy_Prediction_Evidence_from_Wholesale_Companies_in_the_Western_European_Countries
39. Yoon, J. S., & Kwon, Y. S. (2010). A practical approach to bankruptcy prediction for small businesses: Substituting the unavailable financial data for credit card sales information. *Expert Systems with Applications*, *37*(5), 3624–3629. DOI:10.1016/j.eswa.2009.10.029

40. Zavgren, C. (1983). The prediction of corporate failure: the state of the art. *Journal of Accounting Literature*, *2( )*, 1-33.
41. Zhou, L., Lai, K. K., & Yen, J. (2012). Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers & Mathematics with Applications, 64*(8), 2484–2496. DOI:10.1016/j.camwa.2012.06.003
42. Zhou, L., Lai, K. K., & Yen, J. (2014). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science*, *45*(3), 241–253. DOI:10.1080/00207721.2012.720293
43. Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, *22( )*, 59-82. DOI: 10.2307/2490859
44. www.alsu.gov.rs (Available at: *https://alsu.gov.rs/stecaj/statistika-stecajnih-postupaka/*; Access date: 01.04.2022)
45. www.apr.gov.rs (Available at: *https://pretraga2.apr.gov.rs /unifiedentitysearch*; Access date: 01.04.2022)

# KONSTRUKCIJA MODELA PREDIKCIJE BANKROTSTVA UPOTREBOM DISKRIMINANTNE ANALIZE I FINANSIJSKIH RACIJA

Denis Kušter

**Sažetak:** *Osnovni cilj ovog istraživanja je da se razvije statistički model koji može pouzdano predvideti bankrot srpskih preduzeća godinu dana pre početka stečajnog postupka. Osnovni motiv istraživanja je činjenica da u Srbiji nema mnogo naučnih radova koji se bave ovom važnom temom. Model predviđanja bankrota može biti koristan za buduće istraživače, ali i za vlasnike preduzeća i druge zainteresovane strane. Istraživanje je sprovedeno korišćenjem finansijskih racio pokazatelja i diskriminantne analize u IBM-ovom SPSS v.26 programu. U početku je u istraživanje bilo uključeno 100 preduzeća sa teritorije Srbije, ali je nakon skrininga podataka i ispunjavanja svih pretpostavki za diskriminantnu analizu, njih 74 uključeno u proces finalnog modelovanja. Potvrđeno je da uobičajeno korišćeni finansijski racio* *pokazatelji i diskriminantna analiza mogu biti korisni u kreiranju modela predviđanja bankrota, budući da je klasifikaciona moć razvijenog modela 71,6% za originalno grupisane jedinice posmatranja iz uzorka, a 70,3% za unakrsnu validaciju.*

***Ključne reči:*** *predikcija bankrotstva, poslovni neuspeh, diskriminantna analiza, statistička analiza, finansijska analiza, finansijska racija.*