# COMPARISON OF SELECTED STATISTICAL METHODS FOR THE PREDICTION OF BANKRUPTCY

**Maria Misankova – Viera Bartosova**

**Abstract**

The aim of the article is to compare selected statistical methods for the prediction of bankruptcy. Prediction of bankruptcy has been a subject of research´s interest since the beginning of 20th century. Large failures and corporate scandals in recent years have pointed to the need for extended and deepened research on financial distress. During this period various statistical methods have been used starting with the most popular univariate, linear and multivariate discriminant analysis, logistic regression, probit regression, decision trees, neural networks, rough sets, linear programming, principal component analysis, data envelopment analysis, survival analysis and so on. In the presented article are detected the most used statistical methods for the prediction of bankruptcy based on the carefully conducted literature review of more than 500 different models of bankruptcy prediction following with the theoretical specification of selected methods. These are multivariate discriminant analysis and logistic regression while these methods are not only the most used but also they provide the highest prediction ability. At the end of the article are these methods applied for the construction of unique models of bankruptcy prediction of companies in the Slovak Republic.

**Key words:** Bankruptcy, Multivariate Discriminant analysis, Logistic Regression, Financial health of the company.

**JEL Code:** C53, C38

## Introduction

Bankruptcy of the company is a critical event that could cause a great loss to management, shareholders, employees, customers and nation. The development of bankruptcy prediction models has been one of important issues in accounting and finance research fields since beginning of the 20th century, while large corporate´s failures in recent years have pointed out the significance of prediction of financial health of the company.

Fitzpatrick (1932) is considered as a pioneer with the initial study concerning the use of ratio analysis to predict future bankruptcy. His work had become the inspiration for many of applied studies that began to emerge in the middle of 60´s of the 20th century. The most widely recognized univariate study is that of Beaver (1966) followed by Altman´s (1968) multivariate study, which still remains very popular. The most important studies include also: Deakin (1972), Tamari (1976), Taffler (1976), Springate (1978), Ohlson (1980), Beerman (1982), Fulmer (1984), Zmijewski (1984), Zavgren (1985), Kralicek (1990), Laitinen (1991), Shirat (2002) and others. (Ravi, Ravi, 2007, Calabrese, et. al., 2016)

Nowadays there are several hundreds of prediction models that have been developed at a particular time and in terms of specific national economies. So there exists a great variety in bankruptcy prediction models from how many and which factors are considered to what methods are employed to develop the model. Based on the above mentioned can be detected the main subject of the discussion about the suitability of foreign prediction models of financial health of the company. So the research question is whether the model based on the data characterizing companies from one country can be successfully used to predict the financial situation of companies in other countries. (Kollar, et. al., 2015, Svabova, Durica, 2016)

The main goal of the presented article is to review various method used for the construction of prediction models of bankruptcy and based on the obtained results to compare and apply selected statistical methods for the construction of unique models of bankruptcy prediction of companies in Slovak Republic.

# 1 Literature review

As was mentioned in recent years a large number of researchers and practitioners have worked on the prediction of bankruptcy of companies. After an extensive but not necessarily exhaustive review of journal articles and works related to the issue of corporate´s failure, we identified a list of 528 different prediction models constructed worldwide with the focus on the method used. A variety of methodologies have appeared in the literature for modeling business failures. Each method has its own assumptions and different contributions in the field of business failure.

The primary methods that have been used for model development are univariate discriminant analysis (UDA), linear discriminant analysis (LDA), multivariate discriminant analysis (MDA), quadratic discriminant analysis (QDA), principal component analysis

(PCA), logistic regression (LOGIT), probit regression (PROBIT), neural network (NN), rough sets (RS), linear programming (LP), data envelopment analysis (DEA), survival analysis (SA), fuzzy techniques (FA), decision trees (DT) and others. (Kliestik, et. al., 2015, Dimitras, et. al., 1996, Liang, et. al., 2016, Bellovary, et. al., 2007, Delina, Packova, 2013)

We can state that the most important dimension of the review is the type of methodology applied. In the table 1 are shown results of our review of 528 prediction models constructed worldwide and their comparison with 30 prediction models constructed in countries of Visegrad based on the used methodology and this table is completed by the brief description of the most used statistical methodologies. This comparison of the most used methods between these groups of models is important for the further construction of prediction models in Slovak Republic, because we assume that the prediction models constructed in Visegrad countries have better explanatory power and methods used in these models are preferable considering the fact that these countries have similar type of economies, common history based on the transition from centrally planned to market economy and other relevant factors.

**Tab. 1: Results of the provided analysis of bankruptcy´s prediction models**

| *Statistical method* | *Number of foreign models* | *Number of models constructed in countries of Visegrad four* |
|---|---|---|
| UDA | 8 | - |
| LDA | 12 | 4 |
| **MDA** | **151** | **12** |
| QDA | 1 | - |
| PCA | 7 | - |
| **LOGIT** | **166** | **8** |
| PROBIT | 17 | - |
| NN | 112 | 2 |
| RS | 5 | - |
| LP | 21 | - |
| DEA | 4 | 2 |
| SA | 2 | - |
| FA | 2 | - |
| DT | 20 | 2 |
| ∑ | **528** | **30** |

Source: self-processed

The table 1 shows that there is not a significant difference between the most often used methods in prediction models constructed abroad and prediction models constructed in countries of Visegrad four. These methods are in both groups – Multivariate Discriminant Analysis and Logistic Regression and they will be also applied for the construction of unique models of bankruptcy prediction of companies in the Slovak Republic.

## 2 Sample and methods used

The sample used for the construction of unique models of bankruptcy prediction was provided by 4250 manufacturing companies (*SK NACE Rev.2, main section C: manufacturing*) based in the Slovak Republic. These companies include 2254 non-bankrupt companies and 1996 companies that went bankrupt in 2014. The data have been obtained from the Register of financial statements provided by Ministry of Finance of the Slovak Republic and processed by the use of the program IBM SPSS Statistics 20. In constructing models we used 13 explanatory variables, based on the study of models constructed in Visegrad four provided by Kliestik and Majerova (2015). They detected the list of variables (table 2), which have been used in more than 3 of these models.

**Tab. 2: The most significant explanatory variables used in prediction models constructed in Visegrad countries.**

| *Explanatory variable* | *Number of models that include* |
|---|---|
| Net income / Total assets | 9 |
| Current assets / Current liabilities | 8 |
| Retained earnings / Total assets | 6 |
| EBIT / Total assets | 5 |
| EBIT / Interest | 5 |
| (Current assets – Inventory) / Current liabilities | 4 |
| Total assets / Total debt | 4 |
| Total debt / Total assets | 4 |
| Net income / Sales | 3 |
| Net income / Equity | 3 |
| EBITDA / Sales | 3 |
| Resources / Daily sales | 3 |
| Liabilities against suppliers / Daily sales | 3 |

Source: Kliestik, Majerova (2015)

Methods used for the construction of models are Multivariate Discriminant Analysis and Logistic Regression. These methods were detected as the most often used based on the provided research of 528 prediction models constructed abroad and also 30 prediction models constructed in countries of Visegrad four. The basic ideas of these methods are described in following subsections.

### 2.1 Multivariate Discriminant Analysis

Discriminant analysis is a multivariate analytical method, which was used in a large number of studies, as was confirmed by the research provided in the table 1, for the development of models able to predict the business failure.

Firstly, we consider that any company *i* is characterized by vector X elements of which are measurement of *n* independent variables *x* (predictors). For two groups, bankrupt and non-bankrupt companies, it is assumed that the independent variables are distributed within each group according to a multivariate normal distribution with different means but equal dispersion matrices. The main goal of this method is to obtain the linear combination of the independent variables that maximizes the variance between the groups relative to within group variance. The method estimates a discriminant function that is a coefficient vector *A*, ($\alpha_1$, $\alpha_2$,…, $\alpha_n$ – *standardized classification coefficients*) and a constant term $\alpha_0$. The linear combination of the variables provides for each company a Z-score, according to:

$$Z_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \ldots + \alpha_n x_{in} \qquad (1)$$

where $Z_i$ is the Z-score for firm *i* and $x_{i1}$, $x_{i2}$, …, $x_{in}$ are the *n* independent variables for firm *i*. A cut-off score is calculated according to the a-priori probabilities of group membership and the costs of misclassification. Based on the calculated value of Z-score and the cut-off score, a company is classified into the group of bankrupt or non-bankrupt companies. (Altman, 1968)

### 2.2 Logistic regression

The main goal of the logistic regression is to find a model describing the relation between response (dependent) variable and a group of explanatory (independent) variables. At the center of the logistic regression analysis is the task estimating the log odds of an event. Mathematically logistic regression estimates a multiple linear regression function:

$$\log it(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} \qquad (2)$$

Odds and the probability are then gained through the downward transformation:

$$\frac{p}{1-p} = e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_n x_{in}} \qquad (3)$$

where logit is (p * 100) – percent quantile of logistic distribution.

The logit method provides the probability for a company to go bankrupt and then given that probability the company is classified as bankrupt or non-bankrupt, using a cut-off probability, attempting to minimize the type I and type II errors. The coefficients of a model are obtained by maximizing the log-likelihood function. The type I error arises when the bankrupt company is classified as non-bankrupt and the type II error arises when the non-bankrupt company is classified as bankrupt. The model coefficients are obtained by maximizing the log-likelihood function. (Ravi, Ravi, 2007)

## 3 Results and discussion

The application of selected mathematical-statistical methods: *Multivariate Discriminant Analysis* and *Logistic regression* have leaded to the construction of two unique models for the prediction of financial health of manufacturing companies in conditions of the Slovak Republic. During the construction of these models were selected explanatory variables entering into final models from the list of variables listed in the table 2. Results of the application of MDA for the construction of bankruptcy prediction model are in tables 3 and 4.

**Tab. 3: Summary of canonical discriminant functions**

| **Eigenvalues** | | | | |
|---|---|---|---|---|
| *Function* | *Eigenvalue* | *% of Variance* | *Cumulative %* | *Canonical Correlation* |
| 1 | ,007[a] | 100,0 | 100,0 | *,079* |
| a. First 1 canonical discriminant functions were used in the analysis | | | | |
| **Wilks´ Lambda** | | | | |
| *Test of Function(s)* | *Wilks´ Lambda* | *Chi-square* | *df* | *Sig.* |
| 1 | ,994 | 397,003 | 5 | *,000* |

Source: self-processed

Despite the relatively low value of the canonical correlation coefficient, the constructed prediction model is statistically significant (p-value=0,000). The final bankruptcy prediction model constructed by the application of MDA includes 5 explanatory variables and

constant with the overall prediction accuracy of the model 82,3%. The calculated value of the cut-off point is -0,021 which means that higher values classify company as non-bankrupt and lower values classify company as bankrupt.

**Tab. 4: Results of the application of MDA for construction of the bankruptcy prediction model**

| *Explanatory variable* | | *α* |
|---|---|---|
| Current assets / Current liabilities | $x_1$ | ,0029 |
| Total debt / Total assets | $x_2$ | ,000439 |
| EBIT / Total assets | $x_3$ | ,0039 |
| Liabilities against suppliers / Daily sales | $x_4$ | ,000298 |
| Retained earnings / Total assets | $x_5$ | ,000312 |
| Constant | | -,026 |
| **Cut-off point** | | **-,021** |

Source: self-processed

For the purposes of the presented paper is important the comparison with results obtained by the application of Logistic regression for the construction of bankruptcy prediction models. These are shown in table 5 and 6.

**Tab. 5: Model summary and Omnibus Tests of Model Coefficients**

| **Model Summary** | | | | |
|---|---|---|---|---|
| *Step* | | *-2 Log likelihood* | *Cox & Snell R Square* | *Nagelkerke R Square* |
| 4 | | 47066,865 | ,243 | ,364 |
| **Omnibus Tests of Model Coefficients** | | | | |
| | | *Chi-square* | *df* | *Sig.* |
| Step 4 | Step | ,165 | 1 | ,046 |
| | Block | 1763,543 | 8 | 0,000 |
| | Model | 1763,543 | 8 | 0,000 |

Source: self-processed

The value of Nagelkerke R Square shows that model explains 36,4% variability of dependent variable. Based on the Omnibus Tests can be concluded that overall model is statistically significant (level of significance $\alpha$=0,05) and model was constructed in 4 steps containing 8 explanatory variables and constant.

**Tab. 6: Results of the application of LOGIT for construction of the bankruptcy prediction model**

| *Explanatory variable* | | *β* | *S.E.* | *Wald* | *df* | *Sig.* | *Exp(β)* |
|---|---|---|---|---|---|---|---|
| Current assets / Current liabilities | $X_1$ | -,0018 | ,000 | 53,342 | 1 | ,000 | ,9982 |
| Total debt / Total assets | $X_2$ | ,087 | ,004 | 458,043 | 1 | ,000 | 1,087 |
| Total assets / Total debt | $X_3$ | -,0012 | ,000 | 103,219 | 1 | ,000 | 0,9988 |
| EBIT / Total assets | $X_4$ | -,0134 | ,003 | 15,654 | 1 | ,000 | 0,9866 |
| Net income / Equity | $X_5$ | ,0021 | ,000 | 86,234 | 1 | ,000 | 1,0021 |
| Liabilities against suppliers / Daily sales | $X_6$ | ,0001 | ,000 | 5,432 | 1 | ,013 | 1,0001 |
| Resources / Daily sales | $X_7$ | ,0001 | ,000 | 5,634 | 1 | ,048 | 1,0001 |
| Net income / Sales | $X_8$ | ,00012 | ,000 | 3,453 | 1 | ,016 | 1,00012 |
| Constant | | -1,1254 | ,010 | 13345,449 | 1 | 0,000 | 0,3436 |

Source: self-processed

The overall accuracy of bankruptcy prediction models constructed by the application of MDA and LOGIT are in the table 7.

**Tab. 7: Classification table**

| *MDA* | *Non-bankrupt* | *Bankrupt* | *Percentage Correct* | *LOGIT* | *Non-bankrupt* | *Bankrupt* | *Percentage Correct* |
|---|---|---|---|---|---|---|---|
| *Non – bankrupt* | 2142 | 112 | 95,0 | *Non – bankrupt* | 2198 | 56 | 97,5 |
| *Bankrupt* | 640 | 1356 | 67,9 | *Bankrupt* | 382 | 1614 | 80,9 |
| ***Overall Percentage*** | | | **82,3** | ***Overall Percentage*** | | | **89,7** |

Source: self-processed

It is also important to note that both constructed models have 4 common explanatory variables and are statistically significant at the 5% level of significance. Because of the limited extend of the paper we present only final calculations of constructed models not the whole process of their creation. Also the verification and test of constructed models is not presented in the paper.

## Conclusion

Bankruptcy prediction has been an issue of interest for many years. A better understanding of it and mainly the possibility to predict the financial health of the company have tremendous consequences for management of different subjects, from companies, suppliers, customers to

banks, government and so on. Based on previously conducted researches can be pointed out the need for the use of an appropriate validation method when developing and testing bankruptcy prediction models. Given the provided analysis and comparison of methodologies used in bankruptcy prediction models constructed abroad followed by analysis of methodologies used in bankruptcy prediction models constructed in countries of Visegrad four have been detected two most often used methods, which are Multivariate Discriminant Analysis and Logistic regression. These methods were subsequently applied for the construction of unique models of bankruptcy prediction of manufacturing companies in Slovak Republic. Results of these models have shown that better prediction accuracy has the LOGIT model with 8 explanatory variables. Results of the paper can be subsequently applied for further researches on selected issue.

## Acknowledgment

This research was financially supported by the Slovak Research and Development Agency – Grant NO. APVV-14-0841: Comprehensive Prediction Model of the Financial Health of Slovak Companies.

## References

Altman, EI. (1968). Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance*, 23(4), pp. 589-609.

Bellovary, J., Giacomino, D., Akers, M. (2007). A Review of Bankruptcy Prediction Studies: 1930 to Present. *Journal of Financial Education*, 33, pp. 1-43.

Calabrese, R., Marra, G., Osmetti, SA. (2016). Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model. *Journal of the Operational Research Society*, 67(4), pp. 604-615.

Delina, R., Packova, M. (2013). Prediction bankruptcy models validation in Slovak business environment. *E & M Ekonomie a Management*, 16(3), pp. 101-112.

Dimitras, A.I., Zanakis, S.H., Zopounidis, C. (1996). A survey of business failure with an emphasis on prediction method and industrial applications. *European Journal of Operational Research*, 90, pp. 487–513.

Kliestik, T., Majerova, J. (2015). Selected issues of selection of significant variables in the prediction models. *Proceedings of 10th International Scientific Conference Financial management of Firms and Financial Institutions*, Ostrava, Czech Republic, pp. 537-543.

Kliestik, T., Majerova, J., Lyakin, A. N. (2015). Metamorphoses and Semantics of Corporate Failures as a Basal Assumption of a Well-founded Prediction of a Corporate Financial Health. *3rd International Conference on Economics and Social Science (ICESS 2015), Advances in Education Research*, 86, pp. 150–154.

Kollar, B., Kramarova, K., Valaskova, K. (2015). The Financial Distress Analysis of the Stated-Own Company by Applying Chosen Default Prediction Models in Condition of Slovakia. *3rd International Conference on Economics and Social Science (ICESS 2015), Advances in Education Research*, 86, pp. 160–167.

Liang, D., Lu, CC., Tsai, CF, Shih, GA. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(1), pp. 561-572.

Ravi Kumar, P., Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180(1), pp. 1-28.

Reznakova, M., Karas, M. (2014). Bankruptcy Prediction Models: Can the prediction power of the models be improved by using dynamic indicators? *Procedia Economics and Finance*, 12, pp. 565-574.

Svabova, L., Durica, M. (2016). Correlation Analysis of Predictors Used in Bankruptcy Prediction Models in Slovakia. *Ekonomicko-manazerske spektrum*, 10(1), pp. 2-11.

**Contact**

Ing. Maria Misankova, PhD.
University of Zilina, Faculty of operation and economics of transport and communications, Department of economic
Univerzitna 1, 010 26 Zilina, Slovak Republic
maria.misankova@fpedas.uniza.sk

doc. Ing. Viera Bartosova, PhD.
University of Zilina, Faculty of operation and economics of transport and communications, Department of economic
Univerzitna 1, 010 26 Zilina, Slovak Republic
viera.bartosova@fpedas.uniza.sk