ASSESSING CREDIT DEFAULT USING LOGISTIC REGRESSION AND MULTIPLE DISCRIMINANT ANALYSIS: EMPIRICAL EVIDENCE FROM BOSNIA AND HERZEGOVINA

Deni Memić*

Department of Economics – Sarajevo School of Science and Technology
Sarajevo, Bosnia and Herzegovina

DOI: 10.7906/indecs.13.1.13
Regular article

Received: 29 September 2014.
Accepted: 12 January 2015.

ABSTRACT

This article has an aim to assess credit default prediction on the banking market in Bosnia and Herzegovina nationwide as well as on its constitutional entities (Federation of Bosnia and Herzegovina and Republika Srpska). Ability to classify companies info different predefined groups or finding an appropriate tool which would replace human assessment in classifying companies into good and bad buckets has been one of the main interests on risk management researchers for a long time. We investigated the possibility and accuracy of default prediction using traditional statistical methods logistic regression (logit) and multiple discriminant analysis (MDA) and compared their predictive abilities. The results show that the created models have high predictive ability. For logit models, some variables are more influential on the default prediction than the others. Return on assets (ROA) is statistically significant in all four periods prior to default, having very high regression coefficients, or high impact on the model's ability to predict default. Similar results are obtained for MDA models. It is also found that predictive ability differs between logistic regression and multiple discriminant analysis.

KEY WORDS

Bosnia and Herzegovina, default prediction, logistic regression, multiple discriminant analysis, banking

CLASSIFICATION

JEL: G17, G33, G53

*Corresponding author, η: deni.memic@ssst.edu.ba; + 387 62 345 565;
University Sarajevo School of Science and Technology, 71 000, Sarajevo, Bosnia and Hezegovina

INTRODUCTION

Ability to classify companies info different predefined groups is an important business research issue [1], which can be utilized a strong risk management tool. Default prediction has been an important area of business interest for many researchers, from the theoretical and practical aspect [1-29], as it is an intergral part of the credit risk, which is considered to be one of the most important banking risks [2]. Expert default prediction interests a wide range of stockholders such as banks, microcredit organizations, insurance companies, other creditors, auditors and more. The increase of default occurrence can be linked to the latest global financial crisis and appropriate credit risk management.

Many economists consider the latest global financial crisis to be the worst crisis since the Great Depression. Many European countries like Greece, Portugal, Italy and Ireland are facing severe financial and liquidity crisis, which are likely to lead to further issues s like mass demonstrations, European currency crisis, further jobs reduction. Financial crisis is defined in the relevant literature as a "disturbance to financial markets that disrupts the markets capacity to allocate capital – financial intermediation and hence investment come to a halt" [30]. It is believed that one of the main causes of the crisis lays in the collapse of large financial institutions, generally banks around the world.

International Convergence of capital measurement and capital standards segments three main parts of the minimal capital requirements for banks. The main three parts used for calculation of minimal capital requirements are credit, operational and market risks [31].

Assessing the probability of occurrence of credit default is the main interest of this research. According to the definition given by Basel Committee on Banking Supervision [31] credit default occurs when one or more of the following takes place:

- "It is determined that the obligor is unlikely to pay its debt obligations (principal, interest, or fees) in full;
- A credit loss event associated with any obligation of the obligor, such as charge-off, specific provision, or distressed restructuring involving the forgiveness or postponement of principal, interest, or fees;
- The obligor is past due more than 90 days on any credit obligation or
- The obligor has filed for bankruptcy or similar protection from creditors."

In predicting credit default existing literature uses several classification techniques such as multiple discriminant analysis, linear probability, logit analysis, probit analysis, multinomial logit, decision trees, and artificial neural networks.

The main purpose of this study is to assess the probability of default occurrence on the banking market from Bosnia and Herzegovina. In other words the main purpose of the study is predict credit default, or to create a prediction model that distinguishes defaulted and non-defaulted companies, based on the financial data obtained from their financial statements. Since no such research with data including both Bosnia and Herzegovina entities, has been performed, the purpose of this study to construct the first credit default prediction model on the country, as well as on the entities levels, using traditional statistical such as logistic regression (logit) and multiple discriminant analysis (MDA). Since Bosnia and Herzegovina is constructed of two separate entities (Federation of Bosnia and Herzegovina and Republic of Srpska) the purpose of the research is focused on the entities as separate banking markets and on the country level as a whole. The only similar published study in the past constructed a default prediction model using data one-year prior to default and was focused on data from Federation of Bosnia and Herzegovina [3].

Additionally, the data used in this study includes financial figures for a large sample of companies, up to four years prior to default, enabling creation of several default prediction models one, two, three and four years prior to default. Comparison of predictive ability between different techniques is assessed on the basis of time to default, in order to define which technique has the best default predictive ability, when applied to the banking dataset from Bosnia and Herzegovina. Corporate financial data in different research fields is also used for other purposes such as tax evasion detection [32], which have an aim at optimizing the business processes, improve financial stability and public revenues.

The results of this research should have a practical application in an everyday risk assessment performed by creditors, especially banks operating on the market whole of Bosnia and Herzegovina as well as in the banks registered in only one of the Bosnia and Herzegovina entities, since it will be the first default prediction model. The results of the research are expected to contribute to better risk management, of existing risk as well as in the process of assessing the risk of new potential clients in financial institutions. The costs of wrong credit risk assessment in financial institutions, especially banks, can be enormous. Approving loans to companies with high risk (represented through probability of default) consecutively lead to high loan loss provisions.

Creating an effective default prediction model using an original data from Bosnia and Herzegovina banking market will help different stakeholders manage risks more efficiently. Banking market of Bosnia and Herzegovina is divided into two parts. Each entity is regarded as a separate banking market having its own regulating authority. Some large foreign banking groups have two separate banks in Bosnia and Herzegovina, one in Federation and one in Republic of Srpska.

The main question behind the study is determining a way to detect of discriminate bad from good credit applicants, without having to make long, detailed and thorough human-based financial analysis. It is usually a case, that credit officers are engaged and responsible for conducting such financial analysis, using different combinations of financial ratios. The calculated financial ratios are than compared to historically define optimal ranges, as comments are given to each one of them. The main issue emerging in conducting such an analysis appears in the form of constructing an overall assessment whether the analyzed company is creditworthy or not. In other words it is a very difficult task to assess whether an analyzed company is healthy or has a potential to default in the coming period. If no other data is available, it is also not an easy task to differ already defaulted and financially healthy companies.

The main research questions, aimed at solving such a problem are defined as follows:

1) Is it possible to predict credit default in Bosnia and Herzegovina using traditional logistic regression?
2) Is it possible to predict credit default in Bosnia and Herzegovina using multiple discriminant analysis?
3) Is credit default prediction identical for all groups of companies in Bosnia and Herzegovina, with regards to their size and geographical location?

Once the research questions are answered, it should be possible to discriminate between defaulted and healthy companies, on the basis of information included in companies' financial statements. The research questions include two different groups of model creation methods, giving the study a multidimensional approach to credit default prediction issue.

The organization of the paper is as follows. Section 2 presents the reviewed literature on the default prediction. Section 3 presents the used methodology while Section 4 shows the data used and results obtained. Section 5 summarizes the paper with the concluding remarks.

LITERATURE REVIEW

Unlike in many modern European countries, banking market in Bosnia and Herzegovina has very complex legal and inefficient and ineffective structure. Even though the banking market in Bosnia and Herzegovina is considered as one of the most developed, its complex and inefficient structure leaves a large space for further improvements. Undeveloped structure of the financial market in Bosnia and Herzegovina leaves the banking sector as the main source of financing the economy.

The Constitution of Bosnia and Herzegovina defines that the country structure consists of two separate entities: Federation of Bosnia and Herzegovina and Republic of Srpska and one district. Likewise, each entity has its own banking market. Another difference between Bosnian banking market and other developed European markets is in the regulator authority. Central Bank of Bosnia and Herzegovina is the authority in charge of the monetary policy, however the authority of regulating the banking sector is granted to each entity. Each entity has its own law on banks as well as its own banking agency (Banking Agency of Federation of Bosnia and Herzegovina and Banking Agency of Republic of Srpska).

Credit risk is defined as “the degree of value fluctuations in debt instruments due to changes in the underlying credit quality of borrowers and counterparties” [4]. According to one study [5] three main variables affecting the credit risk of a financial asset are Probability of default (PD), Loss-given-default (LGD) and Exposure at default (EAD).

Many risk management models have been constructed previously, such as Algorithmics, CreditMetrics, CreditRisk+, KMV’s Portfolio Mager, Loan Pricing Corporation, McKinsey’s Credit Portfolio View and others. Most of them were constructed by large global banks such as J.P. Morgan and Credit Suisse [6].

Another study [7] divided all bankruptcy prediction models are divided into three main groups: statistical model group (MDA, logit), artificial intelligence models (RP and ANN) and theoretical models (entropy theory).

One of the earliest studies [8] reports that the ratio analysis was born with the construction of current ratio as the first indicator of credit worthiness. It has developed ever since, as it includes multiple different ratios. It is used very often for decision-making processes by lenders, rating agencies, investors, regulators, management and others. The study uses data from financial statements that are created according to internationally accepted reporting standards. The final sample chosen from the Moody’s Industrial Manual database included 79 failed companies form the period 1954-1964. These 79 companies belonged to 38 different industries. Non-failed companies were added to the sample matching the industry and size of failed ones. Five years before failure financial statements were collected. Six different groups of financial ratios were calculated including cash flow ratios, net-income ratio, debt to total asset ratios, liquid asset to total asset ratios, liquid asset to current debt ratios and turnover ratios. Mean values of each ratio were compared between failed and non-failed companies.

Altman [9] created the fundaments of all bankruptcy prediction models. The purpose of his study was to assess the quality of ratio analysis. The sample used in the study included corporate manufacturing companies and it was divided into two main groups. The first group included 33 corporate manufacturers that filed a bankruptcy petition under Chapter X of the National Bankruptcy Act in the period of 1946-1965, while the other group included the same number of non-bankrupt companies. The average size of the sample companies was $ 6,4 million. Set of financial ratios proposed in the literature was used in the study to as potential predictors of bankruptcy. Financial data up to five years prior to bankruptcy were taken into consideration. Used financial ratios were divided into five standardized groups: liquidity, profitability, leverage, solvency and activity.

Altman [10] revised two models for assessing the distress of industrial corporations. The models assessed are called Z-Score model and ZETA® credit risk models. These models study the unique characteristics of business failures by assessing the influential financial variables that can predict the occurrence of default. Ratio analysis of financial variables is the fundament in this study. This study is an extension to Altman's previous studies, as it includes companies that are not traded publicly, non-manufacturing companies. The model also includes a new bond-rating equivalent model for emerging markets corporate bonds. The list of 22 variables (ratios) was used in the final analysis and they represented five groups of financial ratios: liquidity, profitability, leverage, solvency, and activity. Finally five ratios were used out of the starting 22 since they showed the best bankruptcy prediction ability.

The main purpose of another study was to construct a bankruptcy model, which would classify companies up to five years prior to failure of bankruptcy [11]. Used data set covered the period between 1969 and 1975. Linear and quadratic analysis was used in the study. The variables used in the study were return on assets, stability of earnings, debt service, cumulative profitability, liquidity, capitalization and size. Their results show that 7-variable ZETA bankruptcy model successfully predicts 90 % of the sample one year prior to bankruptcy and about 70 % of the sample five years prior to bankruptcy. The used sample in the study consisted of 53 bankrupt companies and 58 non-bankrupt ones. Linear and quadratic methods were used as the linear model showed 96,2 % for the bankrupt and 89,7 % accuracy for the non-bankrupt companies. The total linear model accuracy was 92,8 %. Quadratic model showed 94,3 % for the bankrupt, 91,4 % accuracy for the non-bankrupt companies and total accuracy of 92,8 %.

The first and only study on default prediction using financial data from Bosnia and Herzegovina used data from one of the entities in Bosnia and Herzegovina to create the first default prediction model in Bosnia and Herzegovina [3]. The study included 599 companies registered in Federation of Bosnia and Herzegovina and were divided into an original and holdout sample. Each company in the sample was assigned 15 financial ratios from the following ratio groups: profitability, liquidity, activity, leverage and coverage. Six variables that were extracted from the PCA were used in the logit analysis, which produced the default prediction model. The total classification accuracy of the model was 81,3 % with 95,6 % of correct classification of non-defaulted companies and 36,1 % of defaulted companies.

Deakin [12] analyzed possibility of business failure prediction by using multiple discriminant analysis, which finds a linear combination of ratios that discriminated between the desired groups, which should be classified. The author's goal was to replicate previous studies conducted by [8, 9] using the same financial ratios. The sample consisted of 32 failed companies that experienced failure between 1964 and 1970, where by failure any form of bankruptcy, insolvency or liquidation was included. The failed companies were matched with non-failed ones on the basis of industry classification, year and the asset size. Financial data (financial ratios) for sampled companies were collected for the period of up to five years prior to bankruptcy. The author found that the relative importance of the variables changes over the five years prior to bankruptcy and that almost all of the variables contribute significantly to the discriminant ability of the function. The results showed that the prediction ability in the first three years was larger than in the fourth and fifth year prior to failure. The author concludes that the statistical technique of discriminant analysis can be used accurately for failure or bankruptcy prediction up to three years in advance. Although the study brings an interesting and a significant contribution to the bankruptcy prediction theory, it must be stressed out that the sample used in the study is relatively small.

Lau [13] presented a model that differs from the previous models in a way that instead of the failing and non-failing dichotomy it represents a five financial states that approximate the financial state of a company, and instead of classifying the firm into one of two financial states, this study provides a probability that a company will enter each of the five predefined financial states. The five presented financial states are: financial stability state, omitting or reducing dividend payments state, technical default and default on loan payments state, protection under X or XI of Bankruptcy Act and bankruptcy and liquidation. The results showed that each of the used explanatory variables assumed different values for companies in different predefined states. The paired t-tests showed that for each variable the five state means are significantly different. In order to construct the model multinominal logit analysis (MLA) was used. The authors also used multiple discriminant analysis (MDA) to compare the study results. The results showed that MLA outperformed MDA. The author reported overall 96 % prediction accuracy for the model that used one year before default data, 92 % prediction accuracy for the model that used two years before default data and 90 % prediction accuracy for the model that used three years before default data.

One study utilizes three different methods for bankruptcy prediction and classification: discriminant analysis, logistic regression analysis and genetic algorithms, whereas each of them has a different set of assumptions. Data consisted of financial statements of 74 Finish companies three years before default, equally divided between defaulted and non-defaulted. Failures used in the study occurred between 1986 and 1989. Total of 31 financial ratios were classified into three main groups of liquidity, solidity and profitability. Factor analysis with Varimax rotation was conducted to whether variables in different models are describing different financial dimensions. For the data one year before bankruptcy neural networks had the lowest Type I (5,26 %), Type II (0,00 %) and total error (2,70 %) between the three methods. Discriminant analysis had the lowest Type I (24,32 %), Type II (18,92 %) and total error (21,62 %) between the three methods for the data two years before bankruptcy. Three years before bankruptcy neural network had the best predictive ability or the lowest error rates as well [14].

Ohlson [15] studied bankruptcy prediction using the logit model and it was chosen instead of MDA to overcome the problems associated with MDA such as statistical requirements imposed on the distributional properties of the predictor variables. The data used in the study are from the period between 1970 and 1976 from the Compustat database, and the sample excluded small or privately held corporations, utilities, transportation companies, and financial services companies. Financial data for the failed companies were collected for the period of three years before bankruptcy. The final sample included 105 failed companies. Total number of 2058 healthy companies was added to the sample. Three different models including an intercept and the nine independent variables were made, see Appendix A. Their predictive abilities were 96,12 %, 95,55 % and 92,84 %.

Another study used a sample of Croatian companies to classify them into distressed of non-distressed group, based on 15 financial ratios [16]. The methods of logit analysis, MDA and multidimensional scale (MDS) were used. Financial statements of Croatian companies from 2002 were obtained. Seven different industries were represented in the sample. Unlike many of the previous bankruptcy prediction studies, the companies in the sample were classified into three different groups: good, bad and medium, representing different levels of credit risk. MDA results showed that three variables were statistically significant as they represented liquidity and profitability ratios. MDA model correctly classified all good companies, while the classification of bad companies was very low with a hit ratio of 57 %. The overall predictive accuracy of the MDA model on the Croatian data set was 89 %. MDS model correctly classified 97 % companies (100 % of good and 88 % of the bad companies) and it included six financial ratios. Logit analysis included three variables in the study and had an overall predictive ability of 93 % (95 % of good and 88 % of the bad companies). These results showed that MDA had the best predictive ability among the used techniques.

Šarlija [17] used logistic regression, decision trees and neural networks to predict bankruptcy on a sample of 200 loan applications. Out of the 200 applications 67 were defaulted afterwards, while the remaining 133 were good loan decisions. The data was obtained from one Croatian credit-deposit institution. Seven groups of predictor characteristics were used (business idea, growth plan, marketing plan, entrepreneur characteristics, business characteristics, credit program characteristics, lending institution-entrepreneur relationship). The best predictor of credit default on this study was logistic regression with 83 % of correct classification. Different models of ANNs (from 44 % to 69 %) and DTs (from 44 % to 79 %) had lower predictive ability.

Croatian sample was also used in the study that aimed at predicting short term illiquidity. The results of the study have shown that financial ratios from all five groups are statistically significant in short term liquidiy prediction. The study utilized additional variables indicating the company's geographical location and indistry type, which show that liquidity differs among companies located in different geographical locations as well as for companies from different industries [7].

METHODOLOGY

Based on the previous default prediction research, list of most frequently used financial ratios was assessed, and calculated for each defaulted and healthy company in the sample. The data patterns were analyzed for the total data set and for each of the groups of companies separately (defaulted and healthy group).

Two main groups of methods were used to test the posed hypothesis and answer the research questions. The collected data was analyzed by a group of traditional statistical methods represented by logistic regression and multiple discriminant analysis. The outcome of the study is a set of default prediction models created using two different techniques:

- on a country level,
- on constitutional entities level,
- for SMEs,
- for large corporate companies.

Hypothesis testing has been performed after the prediction models were created. The predictive ability of created prediction models using logistic regression and MDA (see Appendix B) was compared in order to detect optimal combination of the applied technique and used subset of data.

The research aims at creation of the first default prediction models in Bosnia and Herzegovina. The first and only prediction model for Federation of Bosnia and Herzegovina was created [3] as one-year, prediction model created using data from one of the two entities in Bosnia and Herzegovina, using a single statistical technique. This research aims at creating a multi-period prediction models with a utilization of a wider set of statistical and artificial intelligence methods.

Logistic regression analysis is a statistical technique often used for in different fields of research such as medical and social sciences, marketing, movie industry, finance and many more. It was introduced in late 1960s as an alternative to ordinary least squares (OLS) regression. It found a wide application in statistical software programs during 1980s [33].

Its main goal is to find the best fitting model that best describes the relationship between an outcome and the set of independent variables [34, 35]. The main mathematical concept under the logistic regression is the logit or the natural logarithm of an odds ratio.

Logistic regression as a statistical method is suited and usually used for testing hypothesis about relationships between a categorical dependent or an outcome variable and one or more categorical or continuous predictor or independent variables. The dependent variable in logistic regression is binary or dichotomous. The maximum likelihood method, which yields values for the unknown parameters, is used for estimating the least squares function. Logistic regression solves such problems applying the logit transformation. Logistic regression predicts the logit of Y to X. Since the logit is the natural logarithm (ln) of odds of Y, and the odds are the ratios of probabilities (π) of Y happening to probabilities ($1-\pi$) of Y not happening. The dependent variable in logistic regression can be presented as follows:

$$Y = \begin{cases} 1 \\ 0 \end{cases}. \tag{1}$$

Let us denote the p independent variables by the vector X' = (x_1, x_2, … , x_p). If the conditional probability of the outcome is $\pi(Y = 2 \mid x)$ $\pi(x)$, the logistic regression model is given by the following equation

$$\pi(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p . \tag{2}$$

The logistic regression has the following form

$$\text{Logit}\,(Y) = \text{natural\_log}(odds) = \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x\,, \tag{3}$$

$$\pi = \text{probability}(Y = \text{outcome of interest} \mid X = x\text{, a specific value of } X) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}, \tag{4}$$

where π is probability of the outcome of interest, α intercept of Y and β regression coefficient. The logistic regression can also be stated in the form of odds:

$$\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \beta_1 X_i)\,, \tag{5}$$

It can also be presented in terms of probability:

$$\pi_i = \frac{\exp(\alpha + \beta_1 X_i)}{1+\exp(\alpha + \beta_1 X_i)}\,, \tag{6}$$

In logistic regression the value of the coefficient β determines the direction of the relationship between X and the logit of Y, while α and β are usually determined by maximum likelihood method (ML). The data are entered into the analysis as 0 and 1 coding for the dichotomous outcome. Usually, the null hypothesis of the logit model states that all β-s are equal to zero. If there exists at least one β different from zero the null hypothesis is rejected, and the logistic regression model predicts the probability of an outcome better than only the mean of the dependent variable marked as Y [36]. The probability of X or π(X) does not have a linear relation to coefficients in the logistic function, and the maximum likelihood is used. The maximization of the likelihood function expresses the probability of the data set as a function of the unknown parameters:

$$L(b) = \prod_{i=1}^{n} \pi_i(x_i)^{y_i}\left[1-\pi_i(x_i)\right]^{1-y_i} . \tag{7}$$

The use of logistic regression requires certain data pre-processing and model building methodology. In data pre-processing phase a question of variable selection is raised. The dependent variable is dichotomous representing a state of distress or no distress, while the dependent variables can be dichotomous or continuous usually representing financial ratios. Logistic regression is less sensitive to statistical assumptions than the other statistical techniques. Missing values are also an important topic in the use of logistic regression in bankruptcy forecasting, since financial statements differ between companies from different industries. Missing value problem can be solved either by deleting cases with any missing value, by filling the gap with the mean value of other cases or by computing the missing variable value from a regression equation estimated against the other variables. Many different model-building methodologies exist in logistic regression such as stepwise procedure (forward and backward development). Wald statistic with the 10 % significance level is probably is used for variable removal. The model classification power for the logistic regression is computed by the chi-square statistic [18, 36].

Discriminant analysis is a statistical technique used in many different fields and it includes a discriminant variety and represents a linear combination of two or more independent variables that discriminate between the objects in the a priori defined groups [37]. Discriminant analysis is mainly used for solving classification and prediction problems [19]. Similar to the logistic regression the dependent variable is dichotomous. The discriminant function has the following form:

$$Z_{jk} = \alpha + w_1 x_{1k} + w_2 x_{2k} + \ldots + w_n x_{nk}, \tag{8}$$

where Z_{jk} is discriminant Z score of the discriminant function, α the intercept, w_i the discriminant weight for independent variable i, while x_{ik} is the independent variable i for object k.

The probability that a case with a discriminant score of Z belongs to group i is estimated by the following equation:

$$\pi(G_i|D) = \frac{\pi(D|G_i)\pi(G_i)}{\sum_{j=1}^{n} \pi(D|G_j)\pi(G_j)}, \tag{9}$$

where the prior represented by $\pi(G_i)$ is an estimate of the likelihood that a case belongs to a certain group [1]. The objects are classified into one or the other group on the basis of the obtained Z score, whether it is higher or lower than the predefined cut off value. Multiple discriminant analysis computes the discriminant coefficients. The main advantage of MDA is in the ability to classify objects, by analyzing the total variable profile of one object simultaneously [9]. The discriminant analysis creates a vector of weights in a way that the sum of the products of the each element of the vector times the corresponding ratio produces a score that maximizes the distinction between the predefined groups. The distance between the centroids of the two groups is used to test the statistical significance of the discriminant model. If the overlap is small enough the two groups can be considered as different, and vice versa [12].

Multiple discriminant analysis assumes several statistical assumptions such as: normal distribution, homogeneity of variances/covariance, correlations between means and variances, multicollinearity.

Since the main aim of the study is to predict occurrence of default, the dependent variable shows the existence or non-existence of the state of default. Banking data including indicators of default was used for determining the dependent variable. Companies being in delay of bank debt service more than 90 days, according to Basel II principles [31], are regarded as defaulted or distressed. There are several reasons why state of default, as a separation limit was chosen instead of state of bankruptcy in this research:

- defaulted companies are the ones being in delay of debt servicing for more than 90 days. Bankruptcy in practice usually occurs long time after the state of default took place. Bankrupt companies can be in default for several years before the bankruptcy announcement. This is the main reason why default was used as a trigger instead of bankruptcy,

- banks are obliged to internally exclude booking of regular interest income for clients once they are defaulted,
- defaulted assets are regarded as non-performing assets.

DATA AND RESULTS

Once all of the companies being in the state of default are extracted, only the ones transferred from healthy to defaulted, during the year of 2011 were retained in the sample. For each of the company that defaulted during the year of 2011, financial data was collected. The financial data for this study will obtained from AFIP[1] database for Federation of Bosnia and Herzegovina and from APIF[2] database for Republic of Srpska data. These databases consist of financial statements of all of the companies registered in Federation of Bosnia and Herzegovina and Republic of Srpska. The total number of legal entities registered in AFIP database of Federation of Bosnia and Herzegovina in the year of 2010 exceeds 20 000 while in APIF database of Republic of Srpska the number of legal entities exceeds 9 000. Companies with inconsistent and incomplete financial statements were left out. For all of the sampled defaulted companies from both entities, financial statements for up to four years before default were collected. Healthy companies from both entities were added to the sample. Financial statements for the healthy companies were collected for at up to four consecutive years. The years of financial statements are matched for defaulted and healthy companies.

Sampling procedure begun with choosing defaulted companies. Available bank data bases were used to detect defaulted companies (Basel II default methodology). To ensure proper dispersion, several banks' default data bases were combined. Based on the total created default dataset, which included companies from Federation of Bosnia and Herzegovina, Republic of Srpska, large corporate companies and SMEs, sample was chosen using random method. Based on the number of randomly chosen defaulted companies, healthy companies were added, also using random sampling method. The ratio between defaulted and non-defaulted companies was created to ensure proper representation of default-healthy ratio expected on the market in the coming periods. The default-healthy ratio used was assessed based on historical data and trends, which are clearly indicating further deterioration on loan portfolios of banks operating in Bosnia and Herzegovina. Larger share of defaults was intentionally included in the final sample, since Type II errors, represented by misclassification of defaults as healthy companies, are several times more costly than Type I errors.

Table 1. Data structure on country level (D – defaulted, H – healthy). Source: author's calculations.

	t – 4			*t* – 3			*t* – 2			*t* – 1		
	D	**H**	**Total**	**D**	**H**	**Total**	**D**	**H**	**Total**	**D**	**H**	**Total**
FB&H	272	279	**755**	292	484	**776**	292	484	**776**	292	484	**776**
RS	113	280	**393**	113	280	**393**	113	280	**393**	113	280	**393**
B&H	**385**	**763**	**1,148**	**405**	**764**	**1,169**	**405**	**764**	**1,169**	**405**	**764**	**1,169**

Table 2. Data structure on company size level (D – defaulted, H – healthy). Source: author's calculations.

	t – 4			*t* – 3			*t* – 2			*t* – 1		
	D	**H**	**Total**	**D**	**H**	**Total**	**D**	**H**	**Total**	**D**	**H**	**Total**
COR	194	279	**473**	198	279	**477**	198	279	**477**	198	279	**477**
SME	191	484	**675**	207	485	**692**	207	485	**692**	207	485	**692**
B&H	**385**	**763**	**1,148**	**405**	**764**	**1,169**	**405**	**764**	**1,169**	**405**	**764**	**1,169**

Sampling was also done in order to represent Federation of Bosnia and Herzegovina and Republic of Srpska based companies as well as large corporate companies and SMEs.

The chosen dependent variable is represented by an indicator of existence or non-existence of default for a certain company included in the sample. Based on the indicator of days of delay, companies are identified as defaulted and non-defaulted (healthy). A dichotomous variable is used to represent the dependent variable, or to separate defaulted from the healthy companies. If the value of the dichotomous dependent variable is 1 it represents a defaulted company, while the value of the dependent variable of 0 is given to a healthy company.

$$Y = \begin{cases} 1, & \text{defaulted}, \\ 0, & \text{healthy}. \end{cases} \tag{10}$$

Financial ratios are believed to be the best representatives of the financial health of one company. Even though they are not a substitute for a crystal ball, they can however summarize large quantity of financial data in order to compare financial results of different companies [38]. They are widely used in for internal management purposes, external evaluation and very often in finance research, especially in default prediction issues, as they were found to be good predictors of default in previous research. Financial ratios represent relationship between two or more financial positions and their correct interpretation can lead to better understanding of the financial state of one company [39]. They provide us with necessary information regarding company's profitability and riskiness [40].

Companies with financial problems are showing statistically significant differences in the main financial ratios from the healthy companies. Based on the previous theoretical background [7, 41], list of 31 financial ratios was chosen for the study. All of the variables were chosen to represent five main groups of financial ratios: liquidity, profitability, leverage, activity and efficiency.

Two additional dichotomous variables representing the entity in which the company is registered (FB&H or RS) and company size (corporate or small and medium enterprise) were added on the list, totaling 33 variables. Table 3 shows the used 31 financial ratios and additional two dichotomous independent variables. The most frequently used financial ratios from each of the five groups were included in the study, and calculated for all of the companies in the sample. Return on investment (ROI) was consecutively omitted from the study, due to unavailability of data regarding investments in collected financial statements of companies in the sample.

To ensure the validity in this research, several steps were initiated. Since the default prediction model creation, was the main aim of the study, detailed research methodology was constructed. Defaulted companies were detected, with a corresponding year of default occurrence. Data was used from several banks operation in Bosnia and Herzegovina to ensure that the sample represents majority of the banks. It was also important that the chosen data sample, includes companies from both entities, and all major regions of the two entities. Once the defaulted companies from both entities and most regions were identified, total NPL share was assessed. To create a sample which would represent the average Bosnia and Herzegovina NPL share, an appropriate number of healthy companies was added to the defaulted list. The sample was chosen to have a needed share of both SME and large corporate companies, as well as to represent all major regions of the two entities. Original financial data was the basis for financial ratio calculation. Choice of financial ratios, as default predictors was done on the published literature basis. To create default prediction models, financial data was collected in the way to include data up to four consecutive years prior to default for defaulted companies, and up to four consecutive years, making sure that they correspond to the years of defaulted cases.

A large sample size and high confidence level was selected to be able to generalize the results to the population. To ensure research reliability, the total data set was divided twice: on two different entities and on the basis of company size. In other words, each result was validated four times for reliability several times by the created subsets. To ensure proper level of validity and reliability only the best performance models are presented. The basis for the decision on the presentation of the best performing models are: statistical, economic and logic.

Correlation matrices were created for all five datasets in order to detect multicollinearity presence for MDA analysis. Each variable having correlation with another variable higher than 0,50 or lower than –0,50 was removed from further analysis. Variance inflation factor (VIF) and tolerance values were also used to detect multicollinearity among the predictor variables. Table 4 shows the variables removed from the five data sets due to multicollinearity.

Logistic regression and multiple discriminant analysis are used to create default prediction models, on each of the five data sets. For each of the data sets, the techniques are applied on four observed years, for which the financial data are collected. The output of this research set-up is 32 prediction models. To ensure proper level of validity and reliability only the best models are presented. The basis for the decision on the presentation of the best performing models is threefold: statistical, economic and logic criteria.

Statistical criteria consist of several different tests which create the basis for the decision on the best prediction model. Predictive ability of the model is assessed by its hit ratio. Every model has three different hit indicators: (a) healthy companies correct classification, (b) defaulted companies correct classification and (c) total model hit ratio. Depending on the type of method used, different indicators of goodness-of-fit statistics are used to compare different models. Each of the predictor variables is analyzed with its relation to the dependent variable and compared to the expected relationship prior to the model creation in order to assure its economic justification. Models are also compared on the basis of the logic of the relationship between dependent and independent variables. Only the models created with the use of different methods, which have the optimal combination of statistical indicators (hit ratios and goodness-of-fit), economic justification and logical results are presented.

The final output of the research is 8 default prediction models (marked m1 to m8). Each group of methods used will have four representing models. Models m1, m2, m3 and m4 are the optimal logistic regression models for the periods four, three, two and one years before default. Models m5, m6, m7 and m8 are the optimal multiple discriminant analysis models.

LOGISTIC REGRESSION MODELS

The optimal logistic regression models are chosen for all four periods prior to default. The decision is made on the basis of the scored hit ratios represented by classification tables and goodness-of-fit statistics represented by Cox & Snell R^2 and Nagelkerke R^2. All chosen logit models are given in Appendix C.

The first chosen model m1 representing logistic regression in the period t-4 and is chosen from the large corporate companies (COR) data set. Enter was chosen over the stepwise method in this model since it gave better predictive ability and better model fit (stepwise logit model had the predictive ability of 74,2 % and Nagelkerke R^2 of 0,37). The model was also tested with a classification cut off point of 0,50. After the comparison between the models created with classification cut off points of 0,50 and 0,45 the later was retained, since it predicted defaulted companies significantly better than the model with the higher classification cut off point.

The first model shows that only 4 variables are statistically significant ($p < 0,05$), including 2 financial ratios, one dummy variable and a constant variable. The independent variables that are statistically significant are: return on equity, indebtedness factor, Entity Dummy and a constant.

Return on assets has a logistic coefficient B of –5,281 and exponentiated coefficient exp(B) of 0,049 (p-value < 0,05), showing that for one unit change of the return on assets variable will reduce the odds of a default by 99 %. Indebtedness factor has a logistic coefficient B of 0,022 and exponentiated coefficient exp(B) of 1,023 (p-value < 0,05), showing that for one unit change of the indebtedness factor variable will increase the odds of a default by 2 %. The model shows that companies registered in Federation of Bosnia and Herzegovina have 80 % lower odds of defaulting than companies registered in Republic of Srpska. The model has Wald statistic value of the Omnibus Test of Model Coefficients of 160,207 (p = 0,000), and it can be concluded that the proposed model fits the data well. Hosmer and Lemeshow test compares the fitted expected and actual values by groups. Hosmer and Lemeshow test has an insignificant Wald statistic of 3,384 (p = 0,908) also suggesting and confirming Omnibus Test of Model Coefficients that the created model fit at a desired level. It can be concluded that the model has satisfactory goodness-of-fit confirmed by both chosen statistics.

The –2 Log likelihood of the final model has a value of 437,234 and has decreased significantly from the base model in which the –2 Log likelihood amounted 533,827. It can be concluded that adding statistically significant variables (return on equity, indebtedness factor, Entity Dummy) improved the model χ^2 (3, N = 442) = 96,593, p < 0,001. Two additional descriptive measures of goodness-of-fit presented in Table 16 are Cox and Snell R^2 and Nagelkerke R^2, as they have values of 0,304 and 0,410, respectively indicating also fairly good goodness-of-fit. The two R^2 statistics show the amount of variability explained by the model. This can be considered as a high level of goodness-of-fit since the data used for the model is four years distant from default occurrence. Higher values of the two R^2 indicators suggest a better model goodness-of-fit.

The second chosen model m2 representing logistic regression in the period t – 3 is created from the large corporate companies (COR) data set. Enter was chosen over the stepwise method in this model since it gave better predictive ability and better model fit (stepwise logit model had the predictive ability of 77,4 % and Nagelkerke R^2 of 0,44). The model was also tested with a classification cut off point of 0,45. After the comparison between the models created with classification cut off points of 0,50 and 0,45 the model with classification cut off of 0,50 was retained, since it predicted healthy companies significantly better than the model with the higher classification cut off point. There is no significant difference of defaulted companies predictive ability between the two compared models.

The second model shows that 6 variables are statistically significant (p < 0,050), including 4 financial ratios, one dummy variable and a constant variable. The independent variables that are statistically significant are: short term assets to total assets, return on assets, shareholder equity ratio, account payables days, Entity Dummy and a constant.

Short term assets to total assets has a logistic coefficient B of –3,294 and exponentiated coefficient exp(B) of 0,211 (p-value = 0,00), showing that for one unit change of the short term assets to total assets variable will increase the odds of a default by 88 %. Return on assets has a logistic coefficient B of –11,090 and exponentiated coefficient exp(B) of 0,000 (p-value < 0,05), showing that for one unit change of the return on assets variable will decrease the odds of a default by 100 %. Shareholder equity ratio has a logistic coefficient B of –2,803 and exponentiated coefficient exp(B) of 0,061 (p-value = 0,00), showing that for one unit change of the shareholder equity ratio variable will decrease the odds of a default by 94 %. Account payables days has a logistic coefficient B of –0,002 and exponentiated coefficient exp(B) of 0,997 (p-value < 0,05), showing that for one unit change of the account payables days variable will decrease the odds of a default by less than 1 %. The model three years prior to default shows that companies registered in Federation of Bosnia and

Herzegovina have 79 % lower odds of defaulting than companies registered in Republic of Srpska. The model has Wald statistic value of the Omnibus Test of Model Coefficients of 220,275 (p = 0,000), and it can be concluded that the proposed model fits the data well. Hosmer and Lemeshow test however has a significant Wald statistic of 16,419 (p = 0,037) unlike Omnibus Test of Model Coefficients showing that the created model does not fit at a desired level. Since the Hosmer and Lemeshow test shows the p-value close to the insignificance level of 0,50, and that Omnibus Test of Model Coefficients has a significant p-value it can be concluded that the model has satisfactory goodness-of-fit confirmed by one of the chosen statistics. Since all of the other models tested on the $t – 3$ data had a significant Hosmer and Lemeshow test, the model created from the corporate data set is presented, having better predictive ability and goodness-of-fit.

The –2 Log likelihood of the final model has a value of 399,887 and has decreased significantly from the base model in which the –2 Log likelihood amounted 555,937. It can be concluded that adding statistically significant variables (short term assets to total assets, return on assets, shareholder equity ratio, account payables days, Entity Dummy) improved the model significantly $\chi^2(5, N = 458) = 156{,}050$, $p < 0{,}001$. Additional descriptive measures of goodness-of-fit are Cox and Snell R^2 and Nagelkerke R^2, as they have values of 0,385 and 0,515, respectively indicating also good goodness-of-fit. The two R2 statistics show the amount of variability explained by the model. It can be interpreted that shown by Cox and Snell R^2 the presented model explains 38,5 % of the variance, while the Nagelkerke R^2 shows that the model explains 51,5 % of the variance of default. This can be considered as a high level of goodness-of-fit since the data used for the model is three years distant from default occurrence.

The third chosen model m3 representing logistic regression in the period $t – 2$ is from the large corporate companies (COR) data set. Stepwise was chosen over the enter method in this model since it gave better predictive ability and better model fit (enter logit model had the predictive ability of 81,1 % and Nagelkerke R^2 of 0,44). The model was also tested with a classification cut off point of 0,45. After the comparison between the models created with classification cut off points of 0,50 and 0,45 the model with classification cut off of 0,50 was retained, since it predicted healthy companies significantly better than the model with the higher classification cut off point and there was no significant difference of defaulted companies predictive ability between the two compared models.

The model shows that 6 variables are statistically significant ($p < 0{,}050$), including 5 financial ratios and one dummy variable. Constant variable has an insignificant p-value. The independent variables that are statistically significant are: coefficient of financial stability, return on assets, liabilities to total asset ratio, total assets turnover ratio, account receivables days and Entity Dummy.

Coefficient of financial stability has a logistic coefficient B of,481 and exponentiated coefficient exp(B) of 1,618 (p-value < 0,05), showing that for one unit change of the coefficient of financial stability variable will increase the odds of a default by 62 %. Return on assets has a logistic coefficient B of –16,523 and exponentiated coefficient exp(B) of 0,000 (p-value = 0,00), showing that for one unit change of the return on assets variable will decrease the odds of a default by 100 %. Liabilities to total asset ratio has a logistic coefficient B of 0,009 and exponentiated coefficient exp(B) of 1,009 (p-value < 0,05), showing that for one unit change of the liabilities to total asset ratio variable will increase the odds of a default by 1 %. Total assets turnover ratio has a logistic coefficient B of –0,693 and exponentiated coefficient exp(B) of 0,500 (p-value < 0,05), showing that for one unit change of the total assets turnover ratio variable will decrease the odds of a default by 50 %. Account receivables days has a logistic coefficient B of 0,002 and exponentiated coefficient exp(B) of1,002 (p-value < 0,05), showing that for one unit change of the account receivables days variable will increase the odds of a default by less than 1 %.

The chosen model included the dummy variable coding companies coming from different entities. The model two years prior to default shows that companies registered in Federation of Bosnia and Herzegovina have 77 % (logistic coefficient B equals –1,482) lower odds of defaulting than companies registered in Republic of Srpska. The model has Wald statistic value of the Omnibus Test of Model Coefficients of 191,095 (p = 0,000), and it shows that the proposed model fits the data well. Hosmer and Lemeshow test has an insignificant Wald statistic of 12,348 (p = 0,132) and confirms the Omnibus Test of Model Coefficients, showing that the created model fits at a desired level. It can be concluded that the model has satisfactory goodness-of-fit confirmed by both chosen statistics. The model fit is estimated with the value of –2 times the log of the likelihood known as –2 Log likelihood. The –2 Log likelihood of the final model has a value of 424,112 and has decreased significantly from the base model in which the –2 Log likelihood amounted 511,711. The total difference of the two –2 Log likelihood statistics is 87,599. It can be concluded that adding statistically significant variables (coefficient of financial stability, return on assets, liabilities to total asset ratio, total assets turnover ratio, account receivables days and Entity Dummy) improved the model significantly $\chi^2(6, N = 454) = 87{,}599$, $p < 0{,}001$. Two additional descriptive measures of goodness-of-fit Cox and Snell R^2 and Nagelkerke R^2 have values of 0,344 and 0,463, respectively indicating also good goodness-of-fit. Cox and Snell R^2 and Nagelkerke R^2 improved significantly as they amounted 0,204 and 0,275, respectively, in the base model. The two R2 statistics show the amount of variability explained by the model. It can be interpreted that shown by Cox and Snell R^2 the presented model explains 34,4 % of the variance, while the Nagelkerke R^2 shows that the model explains 46,3 % of the variance of default. This can be considered as a relatively high level of goodness-of-fit since the data used for the model is two years distant from default occurrence.

The fourth chosen model m4 representing logistic regression in the period t – 1 is from the large corporate companies (COR) data set. Stepwise was chosen over the enter method in this model since it gave better predictive ability and better model fit (enter logit model had the predictive ability of 84,7 % and Nagelkerke R^2 of 0,53). The model was also tested with a classification cut off point of 0,45. After the comparison between the models created with classification cut off points of 0,50 and 0,45 the model with classification cut off of 0,50 was retained, since it predicted healthy companies significantly better than the model with the higher classification cut off point and there was no significant difference of defaulted companies predictive ability between the two compared models.

The model shows that 5 variables are statistically significant (p < 0,050), including 4 financial ratios and one dummy variable. Constant variable has an insignificant p-value. The independent variables that are statistically significant are: coefficient of financial stability, return on assets, debt asset ratio, total assets turnover ratio and Entity Dummy.

Coefficient of financial stability has a logistic coefficient B of 0,174 and exponentiated coefficient exp(B) of,187 (p-value = 0,00), showing that for one unit change of the coefficient of financial stability variable will increase the odds of a default by 19 %. Return on assets has a logistic coefficient B of –8,255 and exponentiated coefficient exp(B) of 0,000 (p-value = 0,00), showing that for one unit change of the return on assets variable will decrease the odds of a default by 100 %. Debt asset ratio has a logistic coefficient B of 2,495 and exponentiated coefficient Exp(B) of 12,126 (p-value = 0,00), showing that for one unit change of the debt asset ratio variable will increase the odds of a default by more than 1,000 %. Total assets turnover ratio has a logistic coefficient B of –1,974 and exponentiated coefficient exp(B) of

0,139 (p-value = 0,00), showing that for one unit change of the total assets turnover ratio variable will decrease the odds of a default by 86 %.

The chosen model included the dummy variable coding companies coming from different entities. The model one years prior to default shows that companies registered in Federation of Bosnia and Herzegovina have 81 % (logistic coefficient B of –1,676) lower odds of defaulting than companies registered in Republic of Srpska. The model has Wald statistic value of the Omnibus Test of Model Coefficients of 246,049 (p = 0,000), and it shows that the proposed model fits the data well. Hosmer and Lemeshow test has an insignificant Wald statistic of 15,245 (p = 0,055) and confirms the Omnibus Test of Model Coefficients, showing that the created model fits at a desired level. It can be concluded that the model has satisfactory goodness-of-fit confirmed by both chosen statistics. The model fit is estimated with the value of –2 times the log of the likelihood known as –2 Log likelihood. The –2 Log likelihood of the final model has a value of 363,314 and has decreased significantly from the base model in which the –2 Log likelihood amounted 469,258. The total difference of the two –2 Log likelihood statistics is 105,944. It can be concluded that adding statistically significant variables (coefficient of financial stability, return on assets, liabilities to total asset ratio, total assets turnover ratio, account receivables days and Entity Dummy) improved the model significantly $\chi^2(5, N = 452) = 105{,}944, p < 0{,}001$.

Two additional descriptive measures of goodness-of-fit Cox and Snell R^2 and Nagelkerke R^2 have values of 0,420 and 0,567, respectively indicating also good goodness-of-fit. Cox and Snell R^2 and Nagelkerke R^2 improved significantly as they amounted 0,267 and 0,360, respectively, in the base model. The two R^2 statistics show the amount of variability explained by the model. It can be interpreted that shown by Cox and Snell R^2 the presented model explains 42,0 % of the variance, while the Nagelkerke R^2 shows that the model explains 56,7 % of the variance of default.

The key goal of each of created prediction models is to predict the data behavior to the maximum possible extent. The next table compared the predictive abilities, through hit ratios, of the four chosen logistic regression models.

Table 3. Logit model hit ratio comparison. Source: author's calculations.

Logit hit ratio comparison				
	m1 (t – 4), %	**m2 (t – 3), %**	**m3 (t – 2), %**	**m4 (t – 1), %**
Healthy	77,90	84,07	83,15	86,30
Defaulted	70,00	76,06	78,61	82,97
Overall correct	**74,40**	**80,79**	**81,28**	**84,96**

MULTIPLE DISCRIMINANT ANALYSIS MODELS

All multiple discriminant analysis models were adjusted in order to fulfil all statistical assumptions including normal distribution, homogeneity of variances/covariance, correlations between means and variances and multicollinearity. All chosen logit models are given in Appendix D.

The first chosen model m5 representing MDA in the period t – 4 is created from the large corporate companies (COR) data set. The model was created using seven model creation steps including creation of the base discriminant model, checking the database for outliers, testing of variables for normality assumption, substitution of variables not fulfilling normality assumption with transformed variables, checking the variables for multicollinearity, checking the model for homogeneity of variances/covariance assumption.

The model has Wilks' lambda highly significant ($p < 0,000$) and has a value of 0,754 shows relatively low separation between defaulted and healthy companies, as 75,4 % of variability is not explained by the model. The χ^2 statistic tests whether the canonical correlation of the function is equal to zero, or whether the model has any discriminating power. The null hypothesis of the χ^2 statistic is that the model has no discriminating power. The χ^2 statistic's significance level is calculated by the p-value. The presented model has a χ^2 of 117,190 (p = 0,000). The Box's M statistic has a value of 519,355 with a significant p-value (0,000) which means that the null hypothesis is not rejected and the assumption of equal dispersion is violated. The large sample size however makes this violation not too important for the further model interpretation and classification ability. The model consists of seven predictor variables including: net working capital, return of assets, net profit margin, indebtedness factor, total assets turnover ratio, account receivables turnover ratio and a dummy variable indicating an entity where the company is registered. The model shows that variables return on assets and EntityDummy have the highest coefficients and strong impact on the default prediction.

The second chosen model m6 representing MDA in the period t – 3 is from the large corporate companies (COR) data set. The model has Wilks' lambda highly significant ($p < 0,000$) and has a value of 0,741 shows relatively low separation between defaulted and healthy companies, as 74,1 % of variability is not explained by the model. The χ^2 statistic tests whether the canonical correlation of the function is equal to zero, or whether the model has any discriminating power. The null hypothesis of the χ^2 statistic is that the model has no discriminating power. The χ^2 statistic's significance level is calculated by the p-value. The presented model has a χ^2 of 128,270 (p = 0,000). The Box's M statistic has a value of 883,232 with a significant p-value (0,000) which means that the null hypothesis is not rejected and the assumption of equal dispersion is violated. The large sample size however makes this violation not too important for the further model interpretation and classification ability.

The model consists of five predictor variables including: return of assets, total assets turnover ratio, liabilities to total asset ratio, return on equity, and financial efficiency ratio. The model shows that variables return on assets and total assets turnover ratio have the highest coefficients and strong impact on the default prediction.

The third chosen model m7 representing MDA in the period t – 2 is from the Federation of Bosnia and Herzegovina (FB&H) data set. This model initially included 756 cases, before the variables transformation. In the model creation procedure, 448 cases were excluded due to missing variables, giving the total of 308 valid cases used in the model creation procedure. The model has Wilks' lambda highly significant ($p < 0,000$) and has a value of 0,691 shows relatively low separation between defaulted and healthy companies, as 69,1 % of variability is not explained by the model. The χ^2 statistic tests whether the canonical correlation of the function is equal to zero, or whether the model has any discriminating power. The null hypothesis of the χ^2 statistic is that the model has no discriminating power. The χ^2 statistic's significance level is calculated by the p-value. The presented model has a χ^2 of 112,558 (p = 0,000). The Box's M statistic has a value of 5,607 with an insignificant p-value (0,478) which means that the null hypothesis is rejected and the homogeneity of variances/covariance assumption fulfilled.

The model consists of three predictor variables including: return of assets, net profit margin and liabilities to total asset ratio and all with high coefficients and strong impact on the default prediction. All of the included variables are transformed using their logarithmic value.

The fourth chosen model m8 representing MDA in the period t – 1 is from the Federation of Bosnia and Herzegovina (FB&H) data set. The model initially included 756 cases. This model has highly significant Wilks' lambda ($p < 0,000$) and has a value of 0,607 shows relatively low separation between defaulted and healthy companies, as 60,7 % of variability is not explained by the model. The χ^2 statistic tests whether the canonical correlation of the function is equal to zero, or whether the model has any discriminating power. The null hypothesis of the χ^2 statistic is that the model has no discriminating power. The χ^2 statistic's significance level is calculated by the p-value. The presented model has a χ^2 of 367,675 (p = 0,000). The Box's M statistic has a value of 3008,662 with a significant p-value (0,000) which means that the null hypothesis is not rejected and the assumption of equal dispersion is violated. The large sample size however makes this violation not too important for the further model interpretation and classification ability.

This model consists of nine predictor variables including: short term assets to total assets return of assets, return on assets, debt asset ratio, total assets turnover, account receivables days, account payables days, main activities efficiency ratio and financial efficiency ratio. The model shows that variables main activities efficiency ratio, debt asset ratio and return on assets have the highest coefficients and strong impact on the default prediction.

Table 4 compares the predictive abilities, through hit ratios, of the four chosen multiple discriminant analysis models. Table 5 gives the comparative analysis of obtained models' predictive abilities and used predictor variables.

Table 4. MDA model hit ratio comparison. Source: author's calculations.

MDA hit ratio comparison				
	m5 (t – 4), %	**m6 (t – 3), %**	**m7 (t – 2), %**	**m8 (t – 1), %**
Healthy	68,80	69,00	69,62	84,40
Defaulted	74,87	85,40	83,73	76,45
Overall correct	**71,40**	**75,98**	**73,80**	**81,45**

Table 5. Logit and MDA model hit ratio comparison. Source: author's calculations.

Logit hit ratios				
	m1 (t-4)	**m2 (t-3)**	**m3 (t-2)**	**m4 (t-1)**
Healthy	77,90%	84,07%	83,15%	86,30%
Defaulted	70,00%	76,06%	78,61%	82,97%
Overall % correct	**74,40%**	**80,79%**	**81,28%**	**84,96%**
Model variables	ROA, IF, EntityDummy	STATTA, ROA, SER, ALD, EntityDummy	CFS, ROA, LTTAR, TAT, ARD, EntityDummy	CFS, ROA, DAR, TAT, EntityDummy
MDA hit ratios				
	m5 (t-4)	**m6 (t-3)**	**m7 (t-2)**	**m8 (t-1)**
Healthy	68,80%	78,71%	69,62%	84,40%
Defaulted	74,87%	84,49%	83,73%	76,45%
Overall % correct	**71,40%**	**81,19%**	**73,80%**	**81,45%**
Model variables	ROA, TAT, MAEF, NPM, EntityDummy, NWC, ARTR, IF	ROA, TAT, LTTAR, FER, ROE	ROA, NPM, LTTAR	ROA, TAT, MAEF, FER, STATTA, ARD, DAR, ALD, ROE

CONCLUSIONS

This article investigated the bankruptcy prediction on banking market in Bosnia and Herzegovina and its constitutional entities using logistic regression and multiple discriminant analysis. Problem of finding an appropriate tool which would replace human assessment in classifying companies into good and bad buckets has been one of the main interests on risk management researchers for a long time [42]. The main reason for such interest in this topic lays in the increasing default rates which endanger profitability of financial institutions.

This research is based on the detailed review of the relevant literature. All of the reviewed literature treats risk modeling, mainly models predicting default occurrence among legal entities. Literature gives a wide overview of models created using different data sets and different model-creation techniques. Default was also assessed from the theoretical aspect, considering definitions given by Basel Committee, Fedral Banking Agency and Banking Agency of Republic of Srpska. Main criteria for a company to be considered defaulted are companies being in delay of debt servicing for more than 90 days.

Banking sector of Bosnia and Herzegovina was analized, as the sample for the research was created. Initial sample includes companies from both B&H entities, as they are presumed as almost separate banking markets. Data for the study was collected from several data bases, independently for default data and corresponding financial data, as no integral data base exists in Bosnia and Herzegovina. Defaulted companies were detected, with a corresponding year of default occurrence. Data was used from several banks operation in Bosnia and Herzegovina to ensure that the sample represents majority of the banks.Financial ratios, as main default predictors were chosen, based on the relevant literature. The research includes 31 financial and 2 dummy predictor variables, which were collected for all sampled companies, up to four periods prior to default. They were matched with default data, indicating whether a company is defaulted or healthy. The final sample included 1148 companies for the period $t-4$, and 1169 companies for the periods $t-3$, $t-2$ and $t-1$.

This study has a main purpose to assess the probability of default occurrence on the banking market from Bosnia and Herzegovina. In other words the main purpose of the study is predict credit default, or to create a prediction model that distinguishes defaulted and non-defaulted companies, based on the financial data obtained from their financial statements using multi-method approach. The methods used in this study are: logistic regression (logit) and multiple discriminant analysis (MDA).

The outcome of the study is a set of default prediction models created using four different techniques on five different groups of data sets: on a country level, on constitutional entities level, SMEs and large corporate companies.

The results show that the created models have high predictive ability. Among the created models, some variables seem to be more influential on the default prediction issue than the others. Observing logistic regression models, return on assets is statistically significant in all four periods prior to default, having very high regression coefficients, or high impact on the model's ability to predict default. Entity dummy representing the companies coming either from Federation of Bosnia and Herzegovina or Republic of Srpska, is also statistically significant in all four periods prior to default, with its negative regression coefficients showing that companies registered in Federation of Bosnia and Herzegovina in general have lower odds of defaulting than companies registered in Republic of Srpska. Variables coefficient of financial stability, total assets turnover ratio are included in two out of the four models.

Among four logistic regression models, some variables are more influential on the default prediction than the others. Observing logistic regression models, return on assets is statistically significant in all four periods prior to default, having very high regression coefficients, or high impact on the model's ability to predict default. Entity dummy representing the companies coming either from Federation of Bosnia and Herzegovina or Republic of Srpska, is also statistically significant in all four periods prior to default, with its negative regression coefficients showing that companies registered in Federation of Bosnia and Herzegovina in general have lower odds of defaulting than companies registered in

Republic of Srpska. Variables coefficient of financial stability, total assets turnover ratio are included in two out of the four models.

Best performing multiple discriminant analysis default prediction models were selected based on scored hit ratios represented by classification tables, eigenvalues, canonical correlations and Wilks' lambda values of created MDA models, as well as on the strength and direction of the impact of chosen predictors on default. Two models were created using large corporate data sets, while the other two were created on the basis of Federation of Bosnia and Herzegovina data sets.

In the four multiple discriminant analysis models return on assets (or its logarithmic transformation) is the only predictor variable, present in all four MDA models, having a high impact on the model's predictive ability and on default. Total assets turnover ratio is included in three out of four models. Main activities efficiency ratio, net profit margin, liabilities to total asset ratio and financial efficiency ratio are included in two models.

Credit default predictive ability differs between logistic regression and multiple discriminant analysis. Logistic regression exhibited better predictive ability than multiple discriminant analysis Based on the models created it can be confirmed that there is no significant difference between credit default prediction of large corporate companies and SMEs. The study also showed that I there is no significant difference between credit default prediction models for companies in Federation of Bosnia and Herzegovina and in Republic of Srpska, with a remark that some models showed that companies registered in Republic of Srpska are more likely to default, and carry higher credit risk accordingly, than companies registered in Federation of Bosnia and Herzegovina.

It is expected that this research motivates another researchers and practitioners to give more attention to a very complex issue of risk management and more particularly default prediction in Bosnia and Herzegovina. This area has been fairly undeveloped in practice and almost not considered in business research areas in Bosnia and Herzegovina so far.

REMARKS

[1]State agency for financial, information and intermediation services (Agencija za finansijske, informatičke i posredničke usluge d.d. Sarajevo).

[2]State agency for intermediation, information and financial services (Agencija za posredničke, informatičke i finansijske usluge a.d. Banja Luka).

APPENDICES

Appendix A

Table 6. Dependent and independent variables. Source: author's data.

Group	Ratio	Equation	Abbriviation	Expected default impact
LIQUIDITY	Current liquidity	Current assets/current liabilities	CL	+
	Acid test	(Current assets-inventories)/current liabilities	AT	+
	Quick liquidity ratio	Cash/current liabilities	QLR	+
	Net working capital	Current assets-current liabilities	NWC	+
	Short term assets to total assets	Short term assets/total assets	STATTA	+
	Coefficient of financial stability	Long term assets/(equity+long term liabilities)	CFS	+
PROFITABILITY	Return on assets	Net result/total assets	ROA	+
	Return on equity	Net result/total equity	ROE	+
	Return on investment	Net result/investments	ROI	+
	Net profit margin	Net result/total revenues	NPM	+
	Gross profit margin	Gross result/total revenues	GPM	+
LEVERAGE	Indebtedness factor	Total liabilities / (retained earnings+depreciacion)	IF	-
	Debt asset ratio	Total financial liabilities/total assets	DAR	-
	Interest coverage	EBIT/interest expences	IC	+
	Shareholder equity ratio	Equity/total assets	SER	+
	Liabilities to total asset ratio	Total liabilities/total assets	LTTAR	-
	Coverage ratio I	Equity/long term assets	CR I	+
	Gross interest coverage	Gross result/interest expences	GIC	+

Group	Ratio	Equation	Abbriviation	Expected default impact
ACTIVITY	Total assets turnover ratio	Total revenues/total assets	TAT	+
	Long-term assets turnover ratio	Total revenues/long term assets	LTAT	+
	Short-term assets turnover ratio	Total revenues/short term assets	STAT	+
	Inventories turnover ratio	COGS/inventories	ITR	+
	Account receivables turnover ratio	Total revenues/ account receiveables	ARTR	+
	Account payables turnover ratio	COGS/account payables	ALTR	-
	Inventory days	365/inventories turnover ratio	ID	-
	Account receivables days	365/account receivables turnover ratio	ARD	-
	Account payables days	365/account payables turnover ratio	ALD	+
EFFICIENCY	Total efficiency ratio	Total revenues/total expenditures	TEF	+
	Main activities efficiency ratio	Main activities revenues /Main activities expenditures	MAEF	+
	Extraordinary activities efficiency ratio	Extraordinary revenues /Extraordinary expenditures	EAEF	+
	Financing efficiency ratio	Financial revenues/financial expenditures	FER	+
CAT	Entity	FB&H/RS	ENTITY	neutral
	Size	COR/SME	SIZE	neutral

Appendix B

Table 7. MDA multicollinear variables removed. Source: author's calculations.

Variable removed	$t-4$	$t-3$	$t-2$	$t-1$
AT	×	×	×	×
GIC	×	×		×
TEF	×		×	×
SER	×			×
DAR		×	×	
GMP		×		
IC			×	
QLR				×

Appendix C

Table 8. Logistic regression models. Source: author's calculations.

Predictor	m1 (t-4)		m2 (t-3)		m3 (t-2)		m4 (t-1)	
	B	sig.	B	sig.	B	sig.	B	sig.
CL	-,029	,354	-,046	,554				
QLR	,149	,611	,273	,505				
NWC	,000	,180	,000	,396				
STATTA	-,969	,293	-2,125	,033**				
CFS	-,187	,491	-,394	,103	,481	,016**	,174	,009***
ROA	-5,281	,049**	-11,090	,001***	-16,523	,000***	-8,255	,000***
ROE	-,070	,549						
NPM	,322	,221	,051	,317				
IF	,022	,000***	-,001	,683				
DAR	,603	,370					2,495	,000***
IC	-,006	,310	-,017	,418				
SER			-2,803	,000***				
LTTAR	-,005	,360	,008	,163	,009	,027**		
CRI	,085	,244	,040	,211				
TAT	-,442	,082	-,483	,154	-,693	,001***	-1,974	,000***
LTAT	-,004	,556						
STAT	,014	,868	-,206	,092				
ITR	-,002	,673	,001	,624				
ARTR	,017	,055	-,001	,606				
ALTR	-,018	,149	-,002	,503				
ID	,002	,231	,002	,094				
ARD	,002	,330	,002	,067	,002	,036**		
ALD	-,003	,066	-,003	,017**				
MAEF	-,715	,213	,360	,606				
EAEF	-,014	,093	,014	,201				
FER	,029	,733	-,320	,348				
EntityDummy(1)	-1,630	,000***	-1,555	,000***	-1,482	,000***	-1,676	,000***
Constant	1,820	,019**	3,294	,002**	,203	,520	-,268	,530
Omnibus Tests of Model Coefficients	160,207	,000***	191,095	,000***	191,095	,000***	246,049	,000***
Hosmer and Lemeshow Test	3,384	,908	12,458	,132	12,458	,132	15,245	0,055
-2 Log likelihood	437,234		399,887		424,112		363,314	
Cox & Snell R Square	,304		,382		,344		,420	
Nagelkerke R Square	,410		0,515		0,463		,567	

**p*–value < 0,01
***p*–value < 0,05
****p*–value < 0,1

Appendix D

Table 9. Multiple discriminant analysis models. Source: author's calculations.

MDA model variable comparison (Canonical discriminant coefficients)				
Variable	**m5 (t-4)**	**m6 (t-3)**	**m7 (t-2)**	**m8 (t-1)**
NWC	0,000			
STATTA				-0,154
ROA	-5,166	12,499		-1,821
Log_ROA			-2,423	
ROE		-0,392		
NPM	0,178			
Log_NPM			1,898	
IF	0,012			
DAR				1,909
LTTAR		-0,008		
Log_LTTAR			1,147	
TAT	-0,359	0,632		-0,388
ARTR	0,120			
ARD				0,001
MAEF				-1,169
FER		0,242		-0,014
EntityDummy	1,463			
Segment Dummy				0,895
Constant	-0,368	-0,858	-1,337	-0,070

REFERENCES

[1] Jo, H. and Han, I.: *Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction*.
Expert Systems with Applications **11**(4), 415-422, 1996,
http://dx.doi.org/10.1016/S0957-4174(96)00056-5,

[2] Šarlija, N.; Pejić-Bach, M. and Jaković, B.: *Managing Credit Risk in Croatian Banking Sector*. In Wagner, U. et al., eds.: *28th Scientific Symposium "The Future of Monetary and Financial Stability in Europe"*. Fakultat fur Wirtschaft und Recht & Faculty of Economics, Osijek, pp.125-140, 2008,

[3] Memić, D. and Rovčanin, A.: *On the main financial predictors of credit default - Evidence from Federation of Bosnia and Herzegovina*.
Our Economy 1-2/2012,

[4] Lopez, J.A. and Saidenberg, M.R.: *Evaluating credit risk models*.
Journal of Banking & Finance **24**(1-2), 151-165, 2000,
http://dx.doi.org/10.1016/S0378-4266(99)00055-2,

[5] Altman, E.; Resti, A. and Sironi, A.: *Analyzing and Explaining Default Recovery Rates*.
https://www.isdadocs.org/c_and_a/pdf/Analyzing_Recovery_rates_010702.pdf,

[6] Fatemi, A. and Fooladi, I.: *Credit risk management: a survey of practices*.
Managerial Finance **32**(3), 227-233, 2006,
http://dx.doi.org/10.1108/03074350610646735,

[7] Šarlija, N.; Penavin, S. and Harc, M.: *Forecasting Enterprises Illiquidity in Croatia*. In Croatian.
Zbornik Ekonomskog fakulteta u Zagrebu **7**(2), 21-36, 2009,

[8] Beaver, W.H.: *Financial Ratios as Predictors of Failure*.
Journal of Accounting Research **4**, 71-111, 1966,
http://dx.doi.org/10.2307/2490171,

[9] Altman, E.I.: *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*.
The Journal of Finance **23**(4), 589-609, 1968,
http://dx.doi.org/10.1111/j.1540-6261.1968.tb00843.x,

[10] Altman, E.I.: *Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETATM Models*.
http://pages.stern.nyu.edu/~ealtman/PredFnclDistr.pdf,

[11] Altman, E.I.; Haldeman, R.G. and Narayanan, P.: *ZETATM analysis: A new model to identify bankruptcy risk of corporations*.
Journal of Banking & Finance **1**(1), 29-54, 1977,
http://dx.doi.org/10.1016/0378-4266(77)90017-6,

[12] Deakin, E.B.: *A Discriminant Analysis of Predictors of Business Failure*.
Journal of Accounting Research **10**(1), 167-179, 1972,
http://dx.doi.org/10.2307/2490225,

[13] Lau, A. H.-L.: *A Five-State Financial Distress Prediction Model*.
Journal of Accounting Research **25**(1), 127-138, 1987,
http://dx.doi.org/10.2307/2491262

[14] Back B.; Laitinen T. and Sere. K.: *Neural networks and bankruptcy prediction: funds flows, accrual ratios and accounting data*.
Advances in Accounting **14**, 23-37, 1996,

[15] Ohlson, J.A.: *Financial ratios and the probabilistic prediction of bankruptcy*.
Journal of Accounting Research **18**(1), 109-131, 1980,
http://dx.doi.org/10.2307/2490395,

[16] Novak, B. and Crnković, I.: *Classification of Bank Debtor Distress Based on Official Financial Statements*. In Croatian.
Ekonomski pregled 58(1-2), 41-71, 2007,

[17] Šarlija, N.: *Credit risk models in assessing the creditworthiness of small businesses*. Ph.D. Thesis. In Croatian.
University of Zagreb, Zagreb, 2002,

[18] Tucker, D.J.: *Neural networks versus logistic regression in financial modelling: a methodological comparison*.
In Proceedings of the 1996 world first online workshop on soft computing, 1996,

[19] Pejić-Bach, M.: *Application of Discriminant Analysis and Financial Ratios in Prediction of Business Bankrutpcy*. In Croatian.
Računovodsvo i financije **1**(11), 515-532, 1997,

[20] Ahn, B.S.; Cho, S.S. and Kim, C.Y.: *The integrated methodology of rough set theory and artificial neural network for business failure prediction*.
Expert Systems with Applications **18**(2), 65-74, 2000,
http://dx.doi.org/10.1016/S0957-4174(99)00053-6,

[21] Altman, E.I.: *The success of business failure prediction models: An international survey*.
Journal of Banking and Finance **8**(2), 171-198, 1984,
http://dx.doi.org/10.1016/0378-4266(84)90003-7,

[22] Altman, E.I.; Marco, G. and Varetto, F.: *Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)*.
Journal of Banking & Finance **18**(3), 505-529, 1994,
http://dx.doi.org/10.1016/0378-4266(94)90007-8,

[23] Charalambous, C.; Charitou, A. and Kaourou, F.: *Comparative Analysis of Artificial Neural Network Models: Application in Bankruptcy Prediction*.
Annals of Operations Research **99**(1-4), 403-425, 2000,
http://dx.doi.org/10.1023/A:1019292321322,

[24] Dimitras, A.I.; Zanakis, S.H. and Zopounidis, C.: *A survey of business failures with an emphasis on prediction methods and industrial applications*.
European Journal of Operational Research **90**(3), 487-513, 1996,
http://dx.doi.org/10.1016/0377-2217(95)00070-4,

[25] Efrim Boritz, J. and Kennedy, D.B.: *Effectiveness of neural network types for prediction of business failure*.
Expert Systems with Applications **9**(4), 503-512, 1995,
http://dx.doi.org/10.1016/0957-4174(95)00020-8,

[26] Kim, S.Y.: *Prediction of hotel bankruptcy using support vector machine, artificial neural network, logistic regression, and multivariate discriminant analysis*.
Service Industries Journal **31**(3), 441-468, 2010,
http://dx.doi.org/10.1080/02642060802712848,

[27] Min, J.H. and Lee, Y.-C.: *Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters*.
Expert Systems with Applications **28**(4), 603-614, 2005,
http://dx.doi.org/10.1016/j.eswa.2004.12.008,

[28] Turku, B.; Back, B.; Oosterom, G. and Sere, K.: *A Comparative Study of Neural Networks in Bankruptcy Prediction*.
In proceedings of the 10th conference on artificial intelligence in research, pp.140-148, 1994,

[29] Muller, G.H.; Steyn-Bruwer, B.W. and Hamman, W.D.: *Predicting financial distress of companies listed on the JSE – comparison of techniques*.
South African Journal of Business Management **40**(1), 21-32, 2009,

[30] Hunter, W.C. et al.: *The Asian financial crisis: origins, implications, and solutions*.
Springer, 1999,

[31] Basel Committee on Banking Supervision: *International convergence of capital measurement and capital standards: A revised framework comprehensive version*.
Bank of International Settlements, 2006,

[32] Pejić Bach, M.; Zoroja, J. and Mahović-Komljenović, M.: *Usage of information system for discovering tax evasion: Case study of Croatia*.
In Yasuda, T., ed.: Proceedings of the 8th International Conference of the Japan Economic Policy Association. Japan Economic Policy Association, Tokyo, pp.205-212, 2009,

[33] Peng, C.-Y.J.; Lee, K.L. and Ingersoll, G.M.: *An Introduction to Logistic Regression Analysis and Reporting*.
The Journal of Educational Research **96**(1), 3-14, 2002,
http://dx.doi.org/10.1080/00220670209598786,

[34] Hosmer, D.W. and Lemeshow, S.: *Applied logistic regression*. 2nd edition.
Wiley & Sons, 2000,
http://dx.doi.org/10.1002/0471722146,

[35] Žugaj, M.; Dumičić, K. and Dušak, V.: *The foundations of scientific research. Methodology and Teaching Methods*. 2nd revised edition. In Croatian.
Faculty of Organisation and Informatics – University of Zagreb, Varaždin, 2006,

[36] Dumičić, K. and Bahovec, V., eds.: *Business Statistics*. In Croatian.
Element, 2011,

[37] Hair, J.F. and Anderson, R.E.: *Multivariate data analysis*.
Prentice Hall, 2010,

[38] Breadley, R.A. and Myers, S.C.: *Principles of corporate finance*. 4th edition.
McGraw-Hill, Inc., 1991,

[39] Horne, J.C.V. and Wachowicz, J.M.: *Fundamentals of financial management*.
Pearson Education, 2008.

[40] Damodaran, A.: *Corporate finance, theory and practice*.
John Wiley & Sons, 1997,

[41] Samuelson, P.: *Economics*.
McGraw-Hill/Irwin, 2005,

[42] Zoroja, J.; Pejić Bach, M. and Ćurko, K.: *Data Mining Applications Framework for Business Organizations: Business Functions Approach*.
The Business Review Cambridge **22**(1), 119-126, 2014.

PROCJENA STATUSA NEISPUNJAVANJA KREDITNIH OBVEZA KORIŠTENJEM LOGISTIČKE REGRESIJE I VIŠESTRUKE ANALIZE DISKRIMINANTE: EMPIRIJSKI PODACI ZA BOSNU I HERCEGOVINU

D. Memić

Ekonomski fakultet – Sarajevo School of Science and Technology
Sarajevo, Bosna i Hercegovina

SAŽETAK

Cilj ovog rada je procijeniti status neispunjavanja obveza u kreditnom predviđanje na bankarskom tržištu Bosne i Hercegovine te bankarskim tržištima njenih konstitutivnih entiteta (Federacija Bosne i Hercegovine te Republika Srpska). Sposobnost klasificiranja tvrtki u različite unaprijed definirane grupe, odnosno nalaženje prikladnog alata za zamjenu ljudske procjene pri klasificiranju tvrtki, već je duže vrijeme jedan od glavnih interesa istraživača upravljanja rizikom. Istraživana je sposobnost za polazno predviđanje i pripadna točnost pomoću tradicionalnih statističkih metoda logistiičke regresije (logit) i višestrukom diskriminantnom analizom. Njihove su sposobnosti predviđanja uspoređene. Rezultati pokazuju visoku sposobnost predviđanja postavljenih modela. Za logit-modele neke varijable su utjecajnije u predviđanjima od drugih. Povrat na imovinu statistički je značajan u sva četiri razdoblja prije polaznog, karakteriziraju ga visoki koeficijenti regresije ili veliki utjecaj na sposobnost modela da predviđa status neispunjavanja obveza. Slični rezultati dobiveni su za modele višestruke diskriminantne analize. Također je ustanovljeno da se sposobnost predviđanja razlikuje između logističke regresije i višestruke diskriminantne analize.

KLJUČNE RIJEČI

Bosna i Hercegovina, predikcija statusa neispunjavanja obveza, logistička regresija, višestruka diskriminantna analiza, bankarstvo