Miloš Vujnović[1]
Vesna Bogojević Arsić[2]
Nebojša Nikolić[3]

Estimation of Default Probability for Corporate Entities in Republic of Serbia

Abstract*: In this paper a quantitative PD model development has been excercised according to the Basel Capital Accord standards. The modeling dataset is based on the financial statements information from the Republic of Serbia. The goal of the paper is to develop a credit scoring model capable of producing PD estimate with high predictive power on the sample of corporate entities. The modeling is based on 5 years of end-of-year financial statements data of available Serbian corporate entities. Weight of evidence (WOE) approach has been applied to quantitatively transform and prepare financial ratios. Correlation analysis has been utilized to reduce long list of variables and to remove highly interdependent variables from training and validation datasets. According to the best banking practice and academic literature, the final model is provided by using adjusted stepwise Logistic regression. The finally proposed model and its financial ratio constituents have been discussed and benchmarked against examples from relevant academic literature.*

Keywords*: Credit risk, Probability of default, Rating, Scoring model, Rating calibration*

Procena verovatnoće neizvršenja za privredna društva u Republici Srbiji

Apstrakt*: U ovom radu prikazan je razvoj kvantitativnog PD modela u skladu sa standardima Bazelskog sporazma o kapitalu. Set podataka za modeliranje*

---

[1] JUBMES banka a.d. Belgrade, milos.vujnovic@jubmes.rs
[2] University of Belgrade, Faculty of Organizational Sciences
[3] UniCredit bank Srbija a.d.

zasniva se na informacijama iz finansijskih izveštaja iz Republike Srbije. Cilj rada je da se razvije model kreditnog skoringa koji je sposoban da utvrdi procene PD sa visokom prediktivnom sposobnošću na osnovu uzorka privrednih društava. Modeliranje se zasniva na podacima iz godišnjih finansijskih izveštaja privrednih društava u Srbiji iz perioda od 5 godina. Pristup pondera izvesnosti (WOE) je primenjen kako bi se kvantitativno transformisala i pripremila finansijska racija. Korelaciona analiza je iskorišćena za skraćivanje dugačke liste promenjivih i za isključivanje visoko međuzavisnih promenjivih iz razvojnog i validacioniog seta podataka. U skladu sa najboljom bankarskom praksom i akademskom literaturom konačni model je dobijen korišćenjem prilagođene stepenaste logističke regresije. Konačno predložen model i njegovi konstitutivni elementi u vidu finansijskih racija obrazloženi su i upoređeni sa primerima iz relevantne akademske literature.

***Ključne reči**: kreditni rizik, verovatnoća neizvršenja, rejting, scoring model, kalibracija rejting modela.*

1. Introduction

Probability of default (PD) represents the credit risk parameter that plays an important role in contemporary banking risk management practice. It contributes as the key risk parameter in loan approval process and it is also used as the basis for rating class determination of the client. The aim of PD estimate is to accurately and efficiently quantify the level of credit risk inherited within a customer. The PD estimates is calculated by usage of credit scoring models, its objective is to predict future behavior of customer or transacion in terms of probability occurance of 90 days past due event in next 12 months[4] after the loan disbursment. The PD is estimated solely by relying on past experience of customers with similar characteristics. Thus, this parameter of credit risk of a borrower is associated with probability that it will enter in status of default on approved loan one-year horizon. The main tool for PD calculation is credit scoring model which has a goal to provide discrimination between the clients who do default and the ones who do not, i.e. between good and bad clients in terms of their creditworthiness. Discrimination ability is the key indicator of predictive model successfulness. The higher the discrimination power of credit scoring model the more precise the credit scoring model will be.

---

[4] In the contemporary IFRS 9 standard multiyear PD is suggested to be used as the basis for calculation of expected loss provisions for clients with significant credit deterioration. For further details see IASB, 2014.

Contemporary risk management practice and regulation emphasizes and promotes the use of credit scoring models for various asset classes of bank's credit portfolio (BCBS, 2006). Basel II framework emphasizes three approaches to quantifying of group PD: approach based on historical default, approach based on the statistical model and external mapping approach. The visibility and attractivness of PD models has also been recognized in new IFRS 9 standard (IASB, 2014). The new IFRS 9 standard extends the usage of PD models not only for calculation of risk weighted assets, as currently under Basel Capital Accord IRB approach, but for calculation of loan loss provision and allowances.

Retail banking practice uses application and behavioral credit scoring models for automation of loan approval process for individuals (Kennedy, Namee, Delany, O'Sullivan, & Watson, 2013). In retail banking decision to grant a loan based on fundamental analysis and credit analyst assessment is left to be applied only for high amount or non-standard loans. The PD estimate has been referred to as one of the main and most widely used risk factor in Basel II era (Pluto & Tasche, 2010).

The main benefits of logistic regression (LR) usage is that linearity, normality conditions, as well as, independence among independent variables is not assumed in LR approach which leaves more flexibility in working with real-life data. Later on, studies have shown that LR is a sound and powerful statistical approach for modeling credit risk.

In the recent years the extensive development of credit scoring models has been done. Credit scoring models were first built on data from developed world economies and only later they started to utilize data from different emerging markets. The paper of (Hermanto & Gunawidjaja, 2010) tested the performance of LR model on Indonesian SME data over the period of 2005-2007. The LR study performed on 700 SME loans in Slovakia between 2000 and 2005 pointed out that liquidity and profitability factors are important determinants of SME defaults (Fidrmuc & Hainz, 2010). The recent research of (Louzada, Ferreira-Silva, & Diniz, 2012) tried to reveal the LR models performance on state-dependent sample extracted from a portfolio of a Brazilian bank. Furthermore, the research of (Jain, Gupta, & Sanjiv, 2011) examined the behavior of default risk measures and explored the most significant financial variables for SMEs using LR technique. For the purpose of this research, the Indian database of about 3000 SMEs has been used, covering years from 2007 to 2009. Another research, based on Korean dataset (Sohn & Kim, 2012) tried to reveal the best behavioral credit scoring model for technology-based SMEs. The study (Muminović, Pavlović & Cvijanović, 2011) uses data from Serbian publicly listed companies of non-banking sector which were part of Belex15 index and non-banking sector stocks that entered into the composition of Belexline index, to test the

accuracy of Altman's Z-score model in predicting failure of Serbian companies. The same authors (Pavlović, Muminović & Cvijanović, 2011) applied the models developed by Sandin and Porporato to a sample of Serbian companies in order to test the usefulness of ratio analysis to predict bankruptcy in a period of stability of an emerging economy. The behavioral scoring results have been revealed and compared to its application credit scoring counterpart. Finally, in the most recent study of (Khemais, Nesrine & Mohamed, 2016) compared LR and discriminatory analysis results, based on a sample of small and medium enterprises for one Tunisian commercial bank.

2. Probability od default estimation methodology and experimental design

PD estimation for individual borrowers is the first step in estimation of credit exposure and potential losses that financial institutions face. When PD is known, it is simple to estimate the loss distribution which represents the basic element of risk estimation present in the economy and the financial systems (Avesani, Liu, Mirenstean i Salvati, 2005). However, PD estimation may be a challenge mainly due to limited data availability. Basel II emphasizes the need of banks to develop and apply only internal credit risk models and therefore quantitative models of PD estimation represent the basis for application of IRB approach for corporate entities. The banks are required to enable PD estimation based on the group of borrowers with similar characteristics. However, PD estimation may be a challenge mainly due to limitated data availability. Scoring models of PD estimation used in this research are based on the financial data related to a concrete company. The model is based on logistic regression.

In view of the fact that PD is the basic input of the credit risk model, the primary problem may occur in the case of small number of defaults. Also, there is the problem of instability of the number of clients in the status of default as the result of small number of borrowers in individual rating classes.

Basel II framework published in relation to the validation principles (2005) is considered to be sufficiently flexible so that even portfolios with a small number of default incidents are acceptable for application under the IRB systems. As in the case of other portfolios, they must fulfill minimum criteria established by the Basel framework that include requirements for a sensible risk differentiation and sensibly precise and consistent quantitative risk estimations. The choice of tools and techniques will considerably depend on the situation of the individual bank and the portfolio itself.

Rating model development process included the following phases:
1. definition of basic input dataset;

2. definition of sample for model development and validation;
3. definition of independent variables;
4. correlation analysis of variables;
5. regression analysis;
6. testing of scoring model and
7. rating definition.

For necessary calculations in the development and validation process, certain concepts of validation in the form of measurement of the discrimination strength of the model and the adequacy of its calibration have been applied. Approaches that were found to best measure the model quality were chosen, on the basis of the validation process of the model according to empirical results.

All necessary methods and models were applied on the relevant data obtained from a small bank's operations exposed to credit risk (measured by the balance amount and its share in the total balance amount of the banking sector) which predominantly performs its business activities with corporate clients (corporate entities) on the domestic market i.e. on the market of the Republic of Serbia.

2.1. Data collection and structure of dataset

In order to define the database which will serve as a source of data for the model development, it was necessary to define the model objectives. In this sense, the model developed here is intended for use in estimation of new borrowers (application model). The model included only quantitative financial data. The quantitative data are mostly standardized, which enables reliable overview of the borrower's financial position.

Data collection is the most demanding segment of the model development process. It was very importent for the model integrity to make sure that the empirical data fulfill the requirements such are the representativity of the segment for model application, the quantity (sufficient quantity to enable statistically important results), the quality (in order to avoid distortion as a result of unreliable data). In this context, it was necessary to perform preliminary analysis of the available database in order to enable insight into available data, deletion of double entries and identification of the nature of missing data.

Dataset representing the basis for creation of the possibility to perform adequate modelling had to fulfill the following conditions:

- obvious mistakes had to be removed;
- it had to include only homogeneous data, where relation between e.g. financial ratios and incidents of default could be comparable;

- data on default are available and reliable for all borrowers.

Basic aspect of data acquisition is defining of exact time framework in which data are acquired. The prrediction period of one year was used, which coincides with the financial reporting period. The use of longer observation periods may have favourable impact on the predictive strength of the model and, although it is not obligatory under the Basel II standards, banks are encouraged to use longer prediction periods when assigning internal ratings.

In the research, we used the possibility of acquisition of observations of different borrowers for the puropose of their inclusion into the database using different starting points in time. This approach of time stratification is desirable as it decreases dependence of data on a particular calendar year i.e. the economic cycle present at the moment to which the data refer. This is especially important from the viewpoint of work with small databases, such is the available database, and with the aim of generating a quality, statistically significant sample that may be used for the model creation. Available database contains both clients regular in repayment and those who are not.

The database was formed out of financial statements of clients of a small bank operating on the market of the Republic of Serbia, obtained from the database of the Serbian Business Registers Agency. The Agency represents a relevant source of finacial statements and other information about the borrowers on the domestic market. Legal regulations oblige legal entities to annually submit complete financial statements to the Serbian Business Registers Agency.

Financial statement data were matched with each firm's credit repayment performance over the 12-month period in order to construct the default status as a dependent (target) variable. The target variable is represented as a binary variable: 1- default status, 0- non-default status. The default status emerges if the firm in the subsequent year enters into material delinquency (more than 1% of exposure) on their obligation of more than 90 consecutive days past due. Such definition of target variable is compliant with (BCBS, 2006). Defaults are internal information of the bank and are recorded only for the bank's clients.

In view of the fact that the bank is small with a relatively limited database, the whole population was included in the analysis i.e. census research was performed in which all data were acquired for all elements of the population. In model development, the population is usually too large for the analysis based on the entire population, therefore a carefully chosen sample is used for the analysis.

In the research we used end-of-year financial statements and historical performance data for 718 corporate entities in Serbia. Basic input dataset used for development of scoring model and validation i.e. its testing is made

of information about borrowers – corporate entities in the reference period of 5 years: 2009-2013. Dataset contains annual five-year series of end-of-year balance sheets and income statements of Serbian corporate entities. Only the corporate entities which have had material financial liabilities from credit-like products were taken into analysis. This length of data series satisfies Basel II compliance condition of minimum existence of 5 year of data history (BCBS, 2006).

The structure of the dataset regarding the years from which the data used in research derive from is shown in the following table.

Table1: Structure of the dataset by years

Year	Number of borrower-years	%
2009	105	14.62
2010	125	17.41
2011	154	21.45
2012	156	21.73
2013	178	24.79
Total	718	100

Data acquired and cleared represent the entire dataset. However, it was necessary to make a division on the sample for creation (development) of the model and the validation sample. In this sence, a real division was applied that enabled development of the scoring model based on development sample of unknown dataset in the validation sample. The idea behind this is to set aside a part of the data as validation sample to test how well the model obtained from the train sample performs on the data that were not involved in coefficient estimation. The data partition of xx:yy has been done using random stratified sampling on target variable by years. The data partition contributes to lowering the possibility of model overfitting. Namely, the danger comes from the fact that credit scoring model might be doing well on train sample, but performing poorly in practice (Harrell, 2015).

Many of the performing corporate entities have been found in bank's data portfolio throughout the years. Each year an entity has different end-of-year financial statements, so we have settled the basic modeling observation to be 'firm-year'. It should be emphasized that one firm may be present in data set several times as different 'firm-year' row in dataset. For instance, if a firm has its financial statements from 31.12.2009 through 31.12.2012, it has been shown four times in the dataset, each time with different financial variable values. Its corresponding default event performance i.e. target variable is captured for each consecutive at year the end from of 31.12.2010 up to

31.12.2013 for each consecutive year. This reasoning is uniquely applied in this study and it makes each 'firm-year' appearance unique and suitable for modeling purposes. The rationale behind such dataset construction was to try to grasp a five-year economic cycle and to develop thought-the-cycle credit scoring model (Carlehed & Petrov 2012)

Population from which samples for development and validation of the model were created included corporate entities for which complete information are availabe including their financial standing, regularity of repayment, account blockages and other relevant data categories.

Bearing in mind that basic information (dependant variable) for the model development is borrower's status of default, it was necessary to exclude borrowers to which the bank had only short period of exposure after the staring date of observing. This was done with regard to the fact that PD, as one of the basic parameters that will be the result of the rating model, in accordance with Basel II standards on which the regulations of the National Bank of Serbia are based, is estimated for the period of one year. One-year period is the most suitable for estimation of PD as it coincides with the financial reporting and the auditing period for corporate entities.

Division on development and validation samples was performed in the following manner:

Table 2. Division on development and validation samples

Sample type	Development	%	Validation	%	Total	%
Number of non-defaulters	279	77.7159%	278	77.4373%	557	77.5766%
Number of defaulters	80	22.2841%	81	22.2567%	161	22.4234%
Total	359		359		718	
%	50%		50%			

It is not easy to determine the minimally acceptable sample size. In some approaches, it is recommended that, after division of samples according to ratings, there should be at least ten observations (Siddiqi, 2012) in the status of default per one rating. This also applies to groups that emerge upon transformation of data by application of weight of evidence (WoE) approach, i.e. groups of risk levels for analysis of individual attributes, indicators that are candidates for inclusion in the final scoring function. Minimum sample size in the reasearch is determined by application of 10k rule (Siddiqi, 2012) starting from the basic premise of the developing model and it is that it will be defined

with minimum number of parameters (e.g. financial indicators) in the model recommended by the literature, which is 6. In application of this approach, with determined percent of clients in default in the entire population, it was necessary to create the sample of 312 input data. On the basis of this, although recommended proportion of development and validation samples is 70:30, the proportion of 50:50 share of these two samples was applied here. Such proportion satisfied the applied rule for determination of the minimum sample size and offered a better basis for model validation as it provided sufficient number of default events for both creation and testing of the model.

2.2. Financial statements data – variables construction

After the quality of basic financial data was determined, the choice of potential variables that may describe the outcomes was made. The first step is to define and calculate all possible indicators from the availbale dataset. A number of indicators will be excluded in the first phase as the consequence of insufficient or inadequate availability of data. In order to enable the multivariate analysis, it was necessary to solve the problem of missing values. When it comes to models that are applied on coroprate entities, financial ratios are usually used for standardization of available information. This means financial ratios that indicate the structure and the financial position of the borrower as well as the trends of development of certain aspects of the borrower's business operations. Chosen input ratio indicators should represent the most important credit risk factors: leverage, liquidity, productivity, turnover, profitability, size, growth, etc. For the purpose of this research, 142 ratios were formed.

After the calculation of input financial indicator ratios, it was necessary to identify and eliminate potential elements that significantly deviate (outliers) as they can seriously distort evaluated model parameters. Outliers in ratios may exist even when their financial data are basically clear (e.g. when allowed determinator values are near zero). Financial ratio indicators are more desirable for use than the financial statement data in view of the fact that raw financial data depend on the size of the company.

First two tasks that were executed in the course of financial ratio processing were the overview of their economic meaning and establishment of working hypotheses of their relationship with the default status and review of the monotony structure of default rate at different levels of observed financial ratio.

An indicator is considered to be the reliable default risk prediction factor if it behaves in accordance with the economic theory. In this case, it can be presumed that it is not only accidentally correlated with default risk but it also indicates the facts that have significant economic relation to PD. Only when it

was possible to establih the working hypothesis and confirm it empirically, the financial ratio was used in futrther analysis.

If the empirical data have contradictory trends in relation to the financial theory, based on the model development sample, such financial ratio was excluded from the list of possible variables to be included in the multivariate analysis.

For relative indicators such as financial ratios, another executed analysis is the verification of sustainability of the hypothesis of monotonous relation to the default rate at various levels of observed financial ratio. This is the crucial requirement in modern multivariate statistical approaches to rating models based on logistic regression.

Graph 1. Trends of ratio Financial obligations/EBITDA and default rates
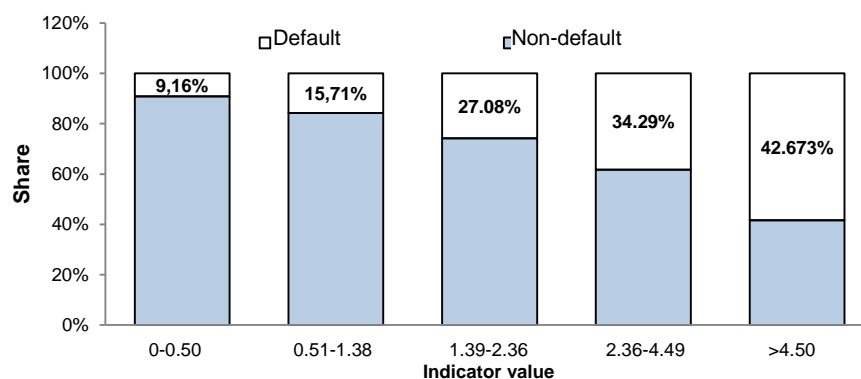
As it can be seen from the ratio example shown above, the growing rate of defaults is recorded with the growth of indicator's value. This is in accordance with economic logic that with growth of indebtedness in relation to business operation results, the ability of regular repayment of the corporate entity declines.

So, four aspects are considered for each financial indicator in the dataset:

- economic importance;
- working hypothesis related to their expected relation to PD;
- estimation of their structural monotony;
- possible solution in the case of structural non-monotony.

If non-monotonous behaviour is recorded, it is most desirable to exclude such indicators from further analysis. However, by expert estimation it can be determined that certain financial ratios may contain significant information for the model quality and the probability of default regardless of their monotony. If

the problem of non-monotonous relationship of the indicator with PD occurs, it is necessary to consider the possibility of financial ratio transformation. The transformation is performed by division of value indicators to value groups and calculation of WoE for appropriate bucket of indicator values.

Purpose of univariate analysis was to identify the characteristics of credit worthiness that have economic importance i.e. the dicrimination strength and the data appropriate for multivariate analysis with the aim of scoring function development. Univariate analysis considers one by one indicator making division within the sample to observations that fulfilled and did not fulfill the obligation and comparing their measures of central tendencies and distribution. The result of this activity is the list of independent variables eligible for inclusion in the multivariate analysis.

Objective of the univariate analysis, in view of the fact that there is a large number of possible financial ratios, is to make a selection of those that have the highest predictive strength i.e. represent quality indicators of customer's credit worthiness.

Linearity test represents the first checkup whether univariate dependence between the observed variable and PD is expected and has the character of importance with regard to explanation of interdependence.

A large number of indicators is usually available, but from the statistical viewpoint it is not recommended to include all indicators in the regression. Some indicators will be highly correlated so that evaluated coefficients will be significantly and sistematicaly biased. Therefore, it is recommended to previously make the choice of indicators based on univariate analysis and the correlation among the same.'

Process of pre-selection consists of the following: first, it is necessary to perform univariate logistic regression for all potential indicators as input data whose value for probability of default is estimated by calculation of measures that describe the importance of this relationship. The example of one such meausre is information value (IV). After that, the analysis of correlation between indicator pairs is performed in order to identify subgroups of highly correlated indicators, where such (e.g. with correlation above 50%) are grouped into one group. Finally, out of each correlation group (usually a group of indicators of the same risk type) an indicator is chosen to be included in the multivariate analysis, according to obtained values of used measure of importance of the relation.

Correlation between independent variables may lead to problems in model stability upon estimation of the scoring function coefficient. Highly correlated indicators may indicate similar information contained in the indicators and may lead to problematic conclusions. It is recommended to always use lowly correlated indicators, wherever possible. Between two correlated indicators, it

is necessary to exclude the one with lower discrimination strenght. Pearson parameter correlation and Spearman correlation may be used for the calculation.

The attribute strength is estimated by the use of WoE measure of predictive strength. WoE measures strength of the indicator or the group of indicators in discrimination of good and bad clients. This is the measure of difference between the proportion of the good and the bad ones in every attribute (i.e. the chance that the client with certan indicator is a good or a bad client). It is based on the calculation of likelihood logarithm:

$$Ln(distribution\ of\ good\ /\ distribution\ of\ bad)x100 \qquad (1)$$

that measures the chance of a good outcome. Negative values indicate that a certain indicator isolates a larger number of bad than good ones. At the same time, the application of this measure is possible on the group level of attributes within an observed indicator, when the analysis includes observation of the WoE range and trends for the chosen group of attributes.

Diskrimination strength of the model is measured by application of IV approach (Siddiqi, 2012) that indicates the level of discrimination of good and bad clients by the use of the individual indicator. IV is determined in the following manner:

$$\sum_{i=1}^{n}(distribution\ of\ good_i - distribution\ of\ bad_i) \times ln\left(\frac{distribution\ of\ good_i}{distribution\ of\ bad_i}\right) \qquad (2)$$

Indicator with IV above 0.5 is tested in view of extremely high strength of prediction, i.e. they can either stay outside the modelling process or they can be used in a controlled manner.

Except the missing data, other groups of indicators are realized so as to have linear relation to WoE i.e. they denote linear and logical relation between the value of the given indicator and the proportion of bad clients. Therefore, as one of the basic elements of the choice of the indicator process, it is necessary to perform operational considerations of WoE measurement trends.

The approach based on business logic in the indicator choice is better, primarily because logical relations ensure sensible final weighting in the model and also understanding and accepting by the final users. Business experience contributes to model improvement better than statistics. By grouping of indicators in a logical manner the possibility of model overfitting to available data is decreased.

In the case of existence of missing values, observations that have significant levels of missing values (above 50%) were excluded, especially if the same were considered to continue in the future. If there is a large number of missing

values in a particular observation (client-year), the same is excluded from further analysis, exept in the case of a client who has entered in the status of default, where addition of the missing data or application of the modelled value inputation system was desirable.

When it comes to extreme values of some attributes, which, although they exist, could be considered as a form of missing values, these data may have a negative impact on the regression results and are excluded if they are missing in the large number of observations.

2.3. Model building technique

Scoring model estimates client's credit worthiness i.e. represents the function of prediction of probability of client's transition to the status of default. Scoring model in this research is developed on the basis of historic data and by application of statistical methods. Historical data included information about the financial situation of the client in the previous period. Regression analysis is applied as the most accepted estimation procedure for scoring models. Results of the estimation are the scores of certain attributes. The idea of scoring is to determine the basic factors of default before it happens and to weight them into the quantitative score. This score can be directly interpreted as the probability of default or it can be used for development of internal rating system based on the probability of default.

Probably the most common used technique for default prediction is logistic regression (LR). It is employed in solving problems of assigning probability to an event where there is binary dependent target variable to predict. The primary difference between linear and logistic regression is the use of a binary variable as modeling target (Khemais, Nesrine & Mohamed, 2016) . Several LR modifications are considered in (Louzada, Ferreira-Silva, & Diniz, 2012) but the conclusions are that their performance on independent validation dataset is substantially the same as plain LR. The main reason for continuing usage of LR over other methods of estimation is that it provides suitable balance of: accuracy, efficiency and interpretability of the results (Core & Finlay 2012).

Bearing in mind that logistic models are considered to be the most popular approach, the same is applied in the reasearch for model establishing. Logistic regression models regress the probability function that certain elements will enter certain category of dependent variable Y, on the linear combination of X_i variables. General form of the model is:

$$Y = f(\beta_0 + \sum_{i=1}^{n} \beta_i x_i) \tag{3}$$

where β_0 is the constant and β_i are estimated weights of X_i, of transformed raw data. The expression on the right-hand side represents the input data in the distribution function, which is a form of logit distribution. The slope coefficient (β_0) offers data on the unit change in X on the probability function Y. In the logit function, the left-hand side is the logarithm of Y, e.g. the logarithm of likelyhood to enter certain category as opposed to the likelyhood not to enter the category.

In the case of bancruptcy prediction, the binary outcome, whose probability is estimated through the logit model is default, and for the same a large number of variables is used. The method adjusts models of linear logistic regression for data on binary or ordinal outcomes by means of the method of maximum certainty (Abdou & Pointon, 2011). The recent applications of logistic regression in the context of financial disturbances may be found also in Nie, G., (Rowe, Zhang, *et al* 2011). This technique weights independent variables and distributes Y score in the form of PD for each client in the sample.

Let y_i denote the outcome of the company i depending on the outcome of variable $x_1, \ldots, x_{ki}$. For example, let $Y = 1$ denote default and $Y = 0$ regular repayment (survival). By the use of logistic regression PD for the company is expressed as:

$$P(Y = 1 | x_1, \ldots, x_k) = f(x_1, \ldots, x_k) \tag{4}$$

Function f denotes the logistic distribution function such as to obtain:

$$P(Y = 1 | x_1, \ldots, x_k) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)} \tag{5}$$

Logistic regression function transforms the regression into interval (0,1). Logit (*x*) can be defined as

$$\log it(x) = \log\left(\frac{x}{1 - x}\right) \tag{6}$$

so that the model can be expressed in the following manner:

$$\log it(P(Y = 1 | x_1, \ldots, x_k)) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{7}$$

with real constants $\beta_0, \beta_1, \ldots, \beta_n$. Advantage of the model is that it does not represent the multivariate normality and equal covariance matrices as e.g. the discrimination analysis. Besides, logistic regression is well adjusted for the problems when the variable is binary or has multiple categorical values, or when there are multiple independent variables in the problem.

Function of logarithm certainty will be maximized (MLE):

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^{n} \{y_i \ln[P(x_i)] + (1 - y_i) \ln[1 - P(x_i)]\} \tag{8}$$

Wald Chi-square test (Allison, 2012) is used to test the statistical significance of individual coefficients. Under hypothesis that $\beta_i = 0$, the next test measure follows χ- square distribution with one level of freedom:

$$W_i = \frac{\hat{\beta}_i^2}{S\hat{E}(\hat{\beta}_i)^2} \tag{9}$$

Multivariate regression represents simultaneous observation for more than one variable. Multivariate regression uses regression for obtaining a small number of indicators which will be included in a final model and determination of weights of importance for the given indicators and functional forms of the model. Through estimation of statistical importance of each indicator it is tested whether the logistic model can be improved and made smaller by inclusion or exclusion of some indicators.

Multivariate regression is carried out through stepwise approach. Stepwise regression means inclusion and exclusion of indicators until the best combination of indicators in obtained on the basis of minimum p value or χ- square test for inclusion or retainment of indicators in the model.

Regression is repeated by the use of different combinations of indicators in different steps and with different levels of importance in the iterative process in order to obtain the highest model strength. Within each regression, the indicators are arranged from the weakest to the strongest. Characteristics included in the model in the previous step are included in the regression in each next step. Let X and Y be two variables tested for N statistical units that can take h values for X and k values for Y. Result of simultaneous classification of variables in the dependency tables may be summarized in pairs $\{(x_i^*, y_j^*), n_{xy}(x_i^*, y_j^*)\}$, where $n_{xy}(x_i^*, y_j^*)$ is the number of statistical units among N observed where the pair of values x_i^* i y_j^* is observed. It is called absolute joint frequency referring to the observed value pair. It is often denoted as n_{ij}. As $N = \sum_i \sum_j n_{xy}(x_i^* y_j^*)$ is equal to the total number of observations, relative joint frequencies may be calculated in the following manner:

$$p_{xy}(x_i y_j) = \frac{n_{xy}(x_i^* y_j^*)}{N}. \tag{10}$$

Variables chosen in the previous step are included in one model that has to be statistically important with little correlation between the variable and intuitive, i.e. the indicator must have economic sense.

Chosen model must have the highest value of the measure used for estimation of the model acceptability (e.g. Gini coefficient, AUROC) and satisfy statistical conditions. Dependant variable Y_{it} is the binary discrete variable that indicates whether or not the company has entered into default in the year t. General presentation of the model is:

$$Y_{it} = f(\beta_k, X_{it-1}^k) + \varepsilon_{it} \tag{11}$$

where X_{it-1}^k represents value k of the variable for the company i, one year before estimation of dependent variable. In order to analyze whether emprical data support the working hypothesis for the indicators considered, for each of them descriptive statistical measures can ve calculated for every group of clients, those who eneterd into the status of default and those who have not.

Based on the research of the relevant academic literature from the introduction, it was noticed that scoring models for corporate entities have at least 5 variables in the model (as Altman, 1968), in order to ensure a stable function. Models with small number of indicators often cannot stand the time test as they are sensitive to small changes in the profile of risk exposure.

On the basis of the final list of indicators, a large number of scoring functions was formed, which were then estimated on the basis of development sample. Decision on the final scoring function was made on the basis of the following criteria:

- consideration of coefficient signs;
- discrimination strength of the scoring function;
- stability of discrimination strength;
- statistical significance of certain coefficients;
- encompassness of relevant information categories.

After definition of all the functions that pass the first test criterion, it was necessary to calculate the dicrimination strength of the same and make dicision which is the most informative one. Stability of the scoring function was tested on the data that were not the element of development sample, i.e. on the basis of validation sample (out of sample testing). A moderate dicrease of discrimination strength was expected (up to 10%) in out of sample testing, so that beyond that level it was necessary to perform further optimization. Statistical importance was tested by application of previously mentioned statistical tests. Indicators that will be included in the model were determined on the basis of univariate and multivariate analysis.

Testing of discrimination strength was performed by application of relevant tests for the preliminary chosen model, in accordance with the level of Gini coefficient of development and validation sample. Complete testing of discrimination strength was performed for preliminary chosen model as an adequate candidate. Testing of discrimination strength was performed for

development and validation sample, for which comparative results are given in further text. The following approaches were used for the purpose of measurement of the discrimination strength of the scoring model:

- Cumulative accuracy profile (CAP);
- Receiver operating characteristic (ROC);
- Pietra index;
- Kolmogorov-Smirnov test;
- Information value;
- Proportional hazard test;
- Kulback-Leibler divergence;
- Mean value difference.

CAP represents visual presentation of the model discrimination strength (Siddiqi, 2012). Measure that comes out of the created curve for the model is called accuracy ratio (AR) and represents the ratio of results of the analyzed scoring model and the ideal model in which all empirical transitions into the status of default would be recorded in the class denoting the highest risk.

ROC curve (Chen & Li 2010) is obtained by mapping of cumulative probability densities of clients who have and those who have not defaulted on the horizontal and vertical axes, respectivelly. ROC curve analysis included the curve slope and measuring of the area below the curve (AUROC), as one of the basic measures on the basis of which the acceptable scoring model was chosen. For the needs of comparison of the candidate models obtained after the indicator correlation analysis and the multivariate regression, AUROC was determined empirically, i.e. the presumption of theoretical distribution was not established.

By application of Kolmogorov-Smirnov (KS) test (Bijak K., Thomas L. C., 2012), maximum absolute diference between cumulative distribution of good and bad clients was tested. Zero hypothesis that was tested assumed that the distributions were identical.

As an additional measure coming out of the ROC curve, the value of Pietra index (PI) was determined as a measure of distance between the diagonal (uninformative model) and obtained ROC curve (Allison, 2012).

Proportional hazard (PH) test offered information about the relationship between the cumulative distribution of clients that did not enter in the status of default and those that did (at the level of 50%).

Information value (IV) represents the sum of relative entropy of customers that have not entered the status of default under condition of distribution of customers that have entered the status of default and relative entropy of customers that have entered the status of default under condition of distribution of customers that have not entered the status of default.

Kullback-Leibler divergence (KL) is the measure od difference between two compeletely defined distributions of probability (BCBS, 2005). The divergence should be as large as possible because in such situation the benefit from the information obtained by the classification is highest.

Mean difference (MD) of probability of default analyzes the difference of clients who belong to distributions of those who have and those who have not entered the status of default.

Definition of ratings means determination of PD range per rating obtained by application of scoring model. This does not represent simple division of probabilities from 0% to 100% in seven equal ranges, but the definition is performed through fine adjustment by moving the range limits with the aim to realize their optimum structures i.e. the classification. Namely, the optimization means achievement of such structure of rating classes that will provide achievement of one or two (or both) objectives that are measured on the base of transition matrices. Objectives of the optimization were the following:

- maximization of rates on the diagonal of transition matrix, i.e. rates that show retainment of clients in the same rating class and monotony of decrease of the rate of transition to other classes the more it moves from the initial rating;
- monotony of growth of transition rate to the rating that represents default as it moves toward infavourable classification.

It is possible to perform optimization of the number of rating classes by application of various clustering techniques such is K-Means algorithm (Malik, & Thomas 2012) and other approaches as presented and done in (Fei, Fuertes & Kalotychou 2012).

Transition matrices are created on the basis of transition rates among rating classes in the prediction period of one year. This means recording of the rating class in which one client is, one year after the initial classification, under the condition that he is in the placement at the end of this period. So the cohort method is applied where each year within the period of observance is considered as a separate cohort within which any change in the period is included. Also, starting presumption that PD are independed from the business cycle, i.e. they are based on PIT philosophy.

After definition of the method of borrowers rating class determination, the model calibration was tested. Calibration was used in order to determine whether previous estimation of the probability of default (PD) significantly deviate from empirical results in view of risk measure. It was thought that the rating model and the classification that came from the rating model were well calibrated if estimated PD for every rating class corresponded to real (empirical) default rates, i.e. it devaited from default rates only marginally. In

practice, estimations of PD deviate from default rates, but basic question is whether deviations are accidenatal or they occur systematically. Calibration estimation was performed on development and validation samples that were used and upon testing of model discrimanation strength. Default rates were calculated on the basis of the number of exposures in the beginning of the observed period according to initial ratings and number of exposures that entered status of default in one-year period.

Calibration estimation was carried out by application of the binomial test (Allison, 2012) and the chi-square test. Besides, Brier score was calculated as the test of calibration and the model discrimination strength.

Binomial test was used for testing of the following hypotheses:

- Zero hypothesis (H_0): estimated PD is maximally equal to empirical DR
- Alternative hypothesis (H_1): estimated PD is underestimated in relation to empirical DR i.e. it is higher.

For hypothesis testing, a binomial distribution of defaults and rates per rating class risk was supposed, with confidence interval (CI) of 95%, on the basis of which critical value of deviation significance of empirical value from the estimated one was determined. Zero hypothesis is rejected with given CI if the number of observed empirical defaults, functions (of binomial distribution) distributed for certain risk class is higher or equal to the critical value of the number of defaults for the chosen risk class. If the number of observed defaults for certain risk class is higher than the critical value of the numer of defaults or default rate for certain class is higher than the critical value of default rate, it can be concluded with defined CI that estimated PD for certain risk class is underestimated i.e. that the real number of defaults is higher than estimated PD.

By chi-square test, statistical significance that chi-square value is better than random was measured. Starting hypothesis was that the sample came from the population characterized by normal distribution. As the accuracy measurement, p-value of chi-square test was used applied with 8 degrees of freedom (number of ratings+1). Chi-square test requires calculation of expected number of defaults per ratings and comparison with real number of defaults. Expected number of defaults is obtained on the basis of expected PD per rating class obtained from the model.

3. Empirical data results and discussion

First step in empirical analysis represent the assessment of the potential effect of independent variables. Defined financial indicators were analysed as

potential relevant independent variables. The total of 149 indicators were tested, out of which there are 142 financial indicators and the rest primarily refer to borrower's current account blockage. In the process of indicator analysis, it was established that indicators that refer to account blockage (both current and in the previous period) have very high significance in default prediction. These indicators were therefore excluded from further process of model development as they were too predictive and limited the significance of other indicators thus reducing the function to only two to three independent variables i.e. model parameters, which would endanger the stability of the same.

The main event that was analyzed, based on available data in defined sample, i.e. the data that will be predicted through the scoring function is the borrower's default i.e. entering into the status of default.

In the process of the scoring function creation, one of the basic steps was to group attributes of each analyzed variable into groups that have the same characteristics, taking into consideration the values of the attributes and the business logic limitations. In the case of financial indicators, grouping meant forming of 3-5 groups in which values of the observed indicator were in certain ranges (e.g. ranges of indebtedness rate 0-0.2, 0.2.-0.4., ...), so the arrangement of bad clients was such that increase or decline of share of such clients in the group (range of indicator value) corresponds to the business logic (e.g. groth of share with growth of indebtedness). Illogical grouping (and the graph that shows WoE) may lead to inadequate scoring function, and therefore to inadequate rating of clients. In this phase of creation of the scoring function, all questions regarding the attributes of analyzed variables were solved.

By grouping of attributes, it was ensured that for each analyzed indicator (variable), difference between the groups is maximized, according to the default rate, in order to achieve the best possible discrimination between the clients and as accurate as possible information that can be obtained from the indicator observation. Grouping was performed in the following manner: if the event (default) rate in the sample is 10%, grouping for a particular indicator was performed in 5 groups where the event rates per groups were 0.5%, 2.5%, 6.0%, 11.0% and 20.0%.

In this way, it was provided that such an indicator has certain information value in the estimation of probability of occurance of analyzed event. Otherwise, if the clients were arranged in groups in such a way that each group had the event rate of 10%, such indicator would have no information value.

In each group of attributes formed according to analyzed indicators, basic condition of grouping was to make sure that each group had sufficient
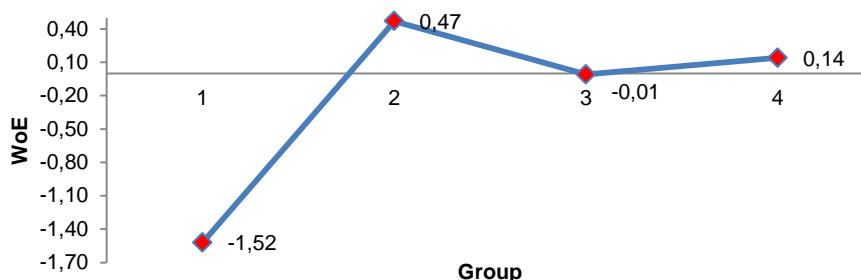
statistical data (of both good and bad clients). Sufficient statistical data for a group was obtained by making sure that each group had at least 10 clients (input data) belonging to good and bad clients (10k rule), taking care that there were at least 3 groups and obtaining as much as possible input data for each group.

Treatment of missing data (e.g. division in analyzed indicator by zero) or data with extereme values (e.g. EBITDA/Interest expenses, where they are very low and the indicator significantly deviate from averagely expected value of this indicator) was solved in this phase of model creation. Data with extreme values were grouped in several ways, in accordance with the business logic and maintaining the quality of grouping. This means that such data were:

- included in the group with the most similar characteristics;
- included in the group with the largest number of elements;
- included in the end groups (with minimum or maximum values).

Choice of independent variables appropriate for further analysis and creation of scoring model is made on the base of measurment of WoE, as the measure of information strength of every attribute of the variable and IV of each variable, as the measure of the information strength of this variable. WoE is used in the sense of analysis of the information strenght of each attribute of the indicator and the existence of rational financial logic for the use of a particular indicator for further activities in creation of the scoring function. In this context, indicators for which no adequate logic can be found through WoE analysis were excluded from futher analysis. For example, if indicator Net result/Sales revenues is observed, it can be found that IV for this indicator is 0.309, which in accordnace with the limits of estimation of predictive ability based on IV puts it in the category of indicators with high predictive ability. The problem is however with the shape of the WoE graph which shows zig-zag movement. Such pattern usually points out that no acceptable financial logic of the same can be determined and it is better to exclude such indicator from further analysis although IV is high. Graph 2 points out that after certain level of ratio between Net results and Sales revenues previously recorded trend of decline of share of bad clients changes its direction i.e. growth of the share of bad clients is recorded with a higher ratio value. This would mean that higher profitability results in higher expectancy of default, which has no economic logic so that the indicator is not appropriate for the scoring function.

Graph 2. WoE trend of indicator Net resault/Sales revenues

In opposition to this indicator, indicators chosen after the analysis of IV and WoE record growing or declining trend of movement of defaults share in the groups of attribute values (and therefore WoE) that can be explained by economic logic. Example of such logic movement of WoE in indicator Financial liabilities/Capital is shown in graph 3.

Graph 3. WoE trend of indicator Financial liabilities/Capital
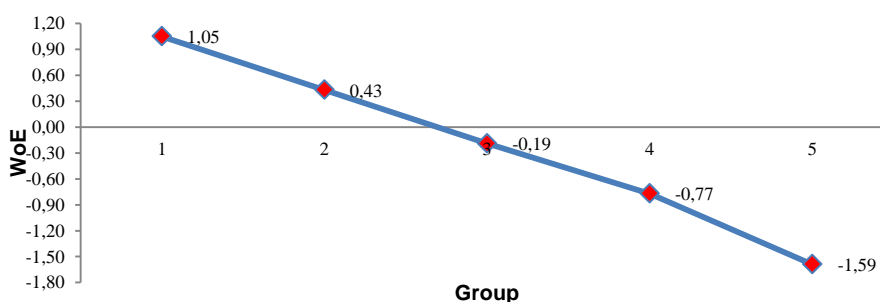
Variables with attributes that have high values of WoE have bigger prediction strength than variables that have WoE near to zero. WoE equal to zero corresponds to analyzed event equal to average rate of event occurrence in the sample. So, the higher rate of analyzed event occurrence, WoE is lower and vice versa. Segments (groups) with the share of bad clients higher than the average (clients in the status of default) will have low WoE, and groups with the share of good clients (active) higher than the average will have higher WoE.

Variables with low or without any ability of prediction are excluded from further analysis, as variables with low predictive strength. IV is through model development interpreted in the following manner (Siddiqi, 2012):

$IV < 0.05$: without predictive strength
$0.05 \leq IV < 0.1$: low predictive strength
$0.1 \leq IV < 0.25$: acceptable predictive strength
$IV > 0.25$: high predictive strength

On the basis of the above mentioned analyses of WoE graphs, business logic and measurement of IV, 93 indicators were selected for further process. For further analysis, indicators with highest IV were selected, for which it is possible to perform distribution of indicator attributes to adequate number of groups (at least 3) under previously mentioned conditions in view of number of elements in each group and for which WoE graph follows business logic for the given indicator. Certain indicators with significantly high IV, although they satisfy other conditions, have also been excluded form the process as this would endanger the possibility of creation of a model with a larger number of variables, having in mind that such indicators would have predominant significance in the model.

After the analyses of WoE and IV, the analysis of correlation between the indicators was performed in order to identify the groups (clusters) of indicators that are highly correlated, so that the model is not created out of indicators which basically show the same trend, that would practically reduce the use of a number of such indicators to the use of one indicator for rating.

Table 3. Segment of indicator correlation matrix

Indicators		66	67	70	72	73	75	76	78	79	80	82	83
	66	100.00%	10.96%	34.25%	41.40%	32.76%	39.14%	31.87%	32.86%	37.78%	39.59%	30.74%	0.23%
	67	10.96%	100.00%	19.32%	5.00%	6.09%	4.06%	5.04%	6.83%	1.07%	0.80%	1.09%	7.67%
	70	34.25%	19.32%	100.00%	25.25%	29.19%	25.54%	28.39%	31.18%	30.43%	27.83%	30.56%	9.49%
	72	41.40%	5.00%	25.25%	100.00%	22.31%	87.32%	8.59%	25.96%	47.55%	49.44%	15.32%	-26.50%
	73	32.76%	6.09%	29.19%	22.31%	100.00%	21.51%	88.62%	94.66%	56.11%	56.34%	57.25%	22.06%
	75	39.14%	4.06%	25.54%	87.32%	21.51%	100.00%	8.99%	23.87%	46.21%	45.02%	15.55%	-31.31%
	76	31.87%	5.04%	28.39%	8.59%	88.62%	8.99%	100.00%	84.57%	42.00%	41.87%	52.64%	20.32%
	78	32.86%	6.83%	31.18%	25.96%	94.66%	23.87%	84.57%	100.00%	57.37%	58.67%	56.66%	22.27%
	79	37.78%	1.07%	30.43%	47.55%	56.11%	46.21%	42.00%	57.37%	100.00%	94.79%	75.50%	22.82%
	80	39.59%	0.80%	27.83%	49.44%	56.34%	45.02%	41.87%	58.67%	94.79%	100.00%	74.70%	24.64%
	82	30.74%	1.09%	30.56%	15.32%	57.25%	15.55%	52.64%	56.66%	75.50%	74.70%	100.00%	49.74%
	83	0.23%	7.67%	9.49%	-26.50%	22.06%	-31.31%	20.32%	22.27%	22.82%	24.64%	49.74%	100.00%

Above table presents a segment of correlation matrix that shows correlations between individual indicators, on the basis of which clusters with high correlation have been identified.

After the identification of clusters of highly correlated indicators, the choice of indicators is primarily made on the basis of the highest value of IV within a cluster (the most informative indicator) and 33 indicators (shown in Appendix) have been chosen for further analysis.

Regression analysis has been carried out over the chosen 33 independent variables. Regression analysis is carried out by application of stepwise regression, which means that by analysis of combinations of variables by alternate inclusion of one additional variable in the model, a certain number of models is obtained with the regression results in the form of β and α factors for the variables included in given models.

Regression is carried out over the development sample and testing over the validation sample. The goal was to chose as the most acceptable model the one which through testing gives the most favourable results of testing that refer to the analysis of discrimination strength of the model, especially the Gini indicator, as well as the regression ability of the model.

By regression analysis, the following scoring models were obtained:

Table 4. Review of observed scoring models

Number of variables	Gini of development sample	Gini of validation sample	Max significance	Indicators													
				1	17	26	40	42	44	58	70	83	89	95	102	109	133
6	**75.94%**	**69.99%**	**0.029583**			x	X	x			x	X	x				
4	73.39%	68.03%	0.028661					x	x	X	x						
5	48.06%	39.57%	0.021752	x	x							X		x			x
5	74.11%	72.06%	0.029771				X	x			x		x		x		
5	74.65%	68.76%	0.036842			x	X	x			x		x				
5	73.71%	71.21%	0.021480				X	x			x	X	x				

After the analysis of discrimination strength and regression ability of the model, model marked in the previous table was chosen. Results of the performed regression analysis for the chosen model over development sample are given in the following table:

Table 5. Results of regression analysis of chosen scoring model

No.	Variable	Score range	Min score	Max score	β	Σ	T	P
0	α	-	-	-	-1.313181	0.172092	-7.63435	2.28964e-14
1	33	33.0056	65.927	98.9325	-0.97927	0.356854	-2.74418	0.00606639
2	47	56.002	53.7498	109.752	-0.737784	0.230044	-3.20714	0.00134075
3	49	36.9978	58.3545	95.3523	-0.610575	0.25618	-2.37754	0.0174285
4	77	82.6126	33.7186	116.331	-0.824538	0.134674	-6.12247	9.24954e-10
5	90	35.9832	65.1758	101.159	-1.1069	0.449372	-2.46322	0.0137696
6	96	28.9161	75.1327	104.049	-0.520198	0.239102	-2.17563	0.0129583

The following table gives the results of WoE analysis for variables that were included in the model, as well as the limits of variable values for which given WoE is applied. Values of WoE have the role in calculation of PD for each

individual client, i.e. for each element of the sample and thay depend on the grouping method for values of attributes of each variable.

Table 6. Number of points, range of values and WOE of variables of the chosen scoring model

No.	No. of points	Range of variable value	WOE
		Variable 26 (Net result/Capital)	
1	65.927	$x < 0.668398$	$WoE_{1,1} = -0.763677$
	78.0991	$0.668398 \leq x < 3.36431$	$WoE_{1,2} = -0.332894$
	89.1777	$3.36431 \leq x \leq 10.8966$	$WoE_{1,3} = 0.0570665$
	98.9325	$10.8966 \leq x$	$WoE_{1,4} = 0.404424$
		Variable 40 (Financial liabilities/Capital)	
2	109.752	$x < 0.501071$	$WoE_{1,1} = 1.04503$
	96.6688	$0.501071 \leq x < 1.3836$	$WoE_{1,2} = 0.430457$
	83.4482	$1.3836 \leq x < 3.3767$	$WoE_{1,3} = -0.190578$
	71.1218	$3.3767 \leq x < 2{,}571.4$	$WoE_{1,4} = -0.769612$
	53.7498	$2{,}571.4 \leq x$	$WoE_{1,5} = .1.58566$
		Variable 42 (Financial liabilities/EBITDA)	
3	95.3253	$x < 2.91623$	$WoE_{1,1} = 0.445411$
	86.2627	$2.91623 \leq x < 5.66949$	$WoE_{1,2} = -0.0705302$
	73.9466	$5.66949 \leq x < 10.5263$	$WoE_{1,3} = -0.769612$
	58.3545	$10.5263 \leq x$	$WoE_{1,4} = -1.65465$
		Variable 70 (Cash / Total assets)	
4	33.7186	$x < 0.00012408$	$WoE_{1,1} = -2.26079$
	65.3621	$0.00012408 <= x < 0.000884127$	$WoE_{1,2} = -0.930731$
	94.9293	$0.000884127 <= x < 0.00877981$	$WoE_{1,3} = 0.312051$
	116.331	$0.00877981 <= x$	$WoE_{1,4} = 1.21162$
		Variable 83 (Liabilities from operations/Total assets)	
5	65.1758	$x < 0.0537668$	$WoE_{1,1} = -0.699139$
	75.0817	$0.0537668 \leq x < 0.105112$	$WoE_{1,2} = -0.388984$
	86.2666	$0.105112 \leq x < 0.28837$	$WoE_{1,3} = -0.0387815$
	101.159	$0.28837 \leq x$	$WoE_{1,4} = 0.427505$
		Variable 89 (Net working capital/Total assets)	
6	75.1327	$x < -1.09678$	$WoE_{1,1} = -0.824302$
	81.6667	$-1.09678 \leq x < -0.949665$	$WoE_{1,2} = -0.388984$
	89.8727	$-0.949665 \leq x < -0.844019$	$WoE_{1,3} = 0.157729$
	104.049	$-0.844019 \leq x$	$WoE_{1,4} = 1.10219$

Probability of default based on logit model used here is calculated by the following formula:

$$PD_i = \frac{1}{1 - e^{-\left(\alpha + \beta_{1,(i)} \times WoE_{1,(i)} + \beta_{2,(i)} \times WoE_{2,(i)} + \beta_{3,(i)} \times WoE_{3,(i)} + \beta_{4,(i)} \times WoE_{4,(i)} + \beta_{5,(i)} \times WoE_{5,(i)} + \beta_{6,(i)} \times WoE_{6,(i)}\right)}} \qquad (12)$$

where: PD_i is probability of default of *i* borrower, α intercept of classification model, $\beta_{1,...,6,(i)}$ coefficients of significance of chosen classification indicators

of i borrower, and $WoE_{1,...,6,(i)}$ of transformed value of chosen quantitative indicator of i borrower.

Based on observance of all measures obtained by application of used approaches, the adequacy of discrimination strength of the model is determined, i.e. the satisfactory level of differentiation between good and bad clients by application of this model.

Table 7. Results of estimation of discrimination strength of the model

No.	Measure	Development sample	Validation sample
1	AR	0.7743	0.7045
2	AUROC	0.8872	0.8522
3	KS	0.6565	0.5283
4	PI	0.3536	0.3536
5	PH	0.9573	0.8938
6	IV	2.0712	1.6263
7	KL	1.1462	1.0417
8	MD	3.4376	3.0015

We have compared the results in the previous table the proposed model on both train (AUROC=0.8872) and validation (AUC=0.8522) sample. Our LR results based on WoE transformation of financial ratios according to AUROC results have been in line with results of (Altman & Sabato, 2007) which compared original values LR results and LR logarithm transformed predictors results. Moreover, proposed model has better performances when contrasted to results of (Sohn & Kim, 2012). Bearing in mind that, by the analysis of discrimination strength of the scoring model, the acceptability of the same was determined, by application of the same determination of PD was performed on the basis of data from both samples (development and validation).

Data thus obtained represented the basis for definition of rating classes. As Basel II standards define acceptable internal classification models of at least 7 classes for active clients and one class for clients in the status of default, and having in mind the size of available population (small population), this standpoint is accepted and the stated number of classes is defined.

The following table represents the matrix created on the basis of transition of clients after one-year period from the initial classification, in the observance period from 2009 to 2014.

Table 8. One-year transition matrix in period 2009 – 2014

		Final rating							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Initial rating	**1**	**61.29%**	17.74%	6.45%	9.68%	1.61%	1.61%	0.00%	1.61%
	2	20.00%	**33.33%**	16.00%	14.67%	5.33%	6.67%	1.33%	2.67%
	3	6.98%	13.95%	**30.23%**	13.95%	13.95%	13.95%	2.33%	4.65%
	4	1.64%	18.03%	8.20%	**22.95%**	19.67%	19.67%	1.64%	8.20%
	5	2.70%	4.05%	10.81%	13.51%	**28.38%**	27.03%	1.35%	12.16%
	6	1.14%	2.29%	2.29%	5.71%	10.86%	**42.29%**	6.86%	28.57%
	7	0.00%	0.00%	2.60%	2.60%	1.30%	7.79%	**12.99%**	72.73%
	8	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**100.00%**

Graph 4. Default rate curve pr rating classes
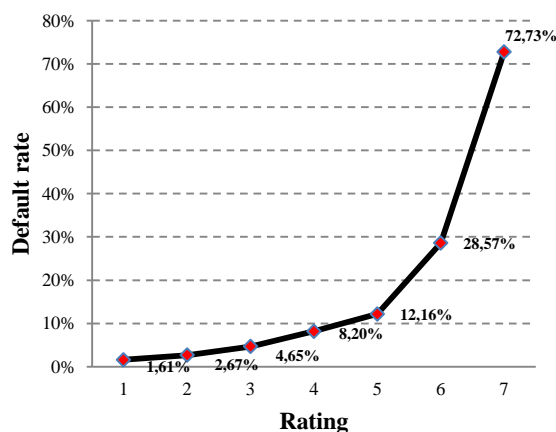
Rating classes with the following PD limits are determined:

Table 9. PD limits per rating

Rating	Lower limit of PD	Upper limit of PD
1	0.00%	1.66%
2	1.66%	2.91%
3	2.91%	4.12%
4	4.12%	6.18%
5	6.18%	12.22%
6	12.22%	49.46%
7	49.46%	100.00%

Model testing results in view of calibration through binomial test point out that for observed classification dataset model does not show divergence from adequate calibration in none of defined classes, in either development or validation sample.

Table 10. Results of binomial test

Initial classification	Development sample		Validation sample	
	Function of binomial distribution (N)	Significance of divergence ($N > 95\%$)	Function of binomial distribution (N)	Significance of divergence ($N > 95\%$)
1	92.70%	Insignificant	62.73%	Insignificant
2	70.25%	Insignificant	75.70%	Insignificant
3	37.66%	Insignificant	92.49%	Insignificant
4	58.03%	Insignificant	85.98%	Insignificant
5	80.25%	Insignificant	80.73%	Insignificant
6	65.94%	Insignificant	65.28%	Insignificant
7	60.83%	Insignificant	53.24%	Insignificant

On the basis of results of performed chi-square test, it can be concluded that zero hypothesis for the chosen model is rejected with 95% CI, as p-value in both development and validation sample is lower than 5%, i.e. the model calibration can be considered adequate.

Table 11. Chi-square test

Rating	Chi-square value	
	Development sample	Validation sample
1	0.7017937	0.4690058
2	0.0075279	0.0035290
3	0.9943689	1.0687647
4	0.1322188	0.5121140
5	0.3137860	0.3344067
6	0.0840068	0.0727917
7	0.0220097	0.0021376
Total	**2.255712**	**2.462750**
p-value	**2.784836%**	**3.655464%**

Brier value (Allison, 2012) represents the test of calibration and discrimination strength of the model. Brier value represents the method of estimation of quality of PD prediction obtained by model application, i.e. whether predictions of default rate deviate from sample default rate. It is also calculated as average deviation between predicted and empirical PD and depends on the default rate for the entire portfolio. Higher obtained Brier value

indicates a worse model and its advantage is that it does not depend on statistical presumptions. Brier value for data on default from development sample is 0.1142, and that from the validation sample is 0.1194. As values may vary from 0 to 1, these values indicate a quality prediction obtained through the model application.

4. Conclusion

In our research we have built corporate PD model capable of predicting probability of bankruptcy in one year period. The model has been developed on dataset which comprised of five years of financial statements data originated from corporate entities in Republic of Serbia.

The aim of the research was to design a consistent and complete framework of PD model development as well as an initial validation framework in order to confirm the soundness of the obtained results. During the model development process, various limitations and peculiarities have emerged, primarily with respect to data availability necessary for application of statistical analyses. All problems were successfully overcome and the final results produced the statistically profund model build on six financial ratios.

Quality of developed model was tested through established validation methodology in this paper and the proposed model passed all validation tests related to predictive power assessment widely accepted in the banking industry. The quality of obtained results and the fact that the developed PD model is based on actual financial statements financial ratios, as well as on default status data which have been proved to be statistically significant and sound. Accoring to results, we can conclude that the developed and proposed model can be implemented and employed within a bank that operates in Serbia or in the region of South Eastern Europe.

On the basis of the conducted research in this paper, with application of appropriate techniques widely accepted in academic and industry practice the empirically appropriate process of estimation of probability of default is proposed and implemented. The quality of the developed model is underlined by the fact that real financial statemets and default status data from the available database were used. Moreover, fully replicable through application of methods and described approaches were employed during the model deveopment. The resulting PD model's predictive power was compared with the similar ones from the studies in academic literature and it has been found that the result was in line or event better that one of the related model results in the benchmarked studies.

References

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management*, *18*(2-3), 59-88.

Allison, P. D. (2012). *Logistic regression using SAS: Theory and application*. SAS Institute.

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance, 23*(4)

Altman, E., & Sabato, G. (2007). Modeling Credit Risk for SMEs: Evidence from the US Market. *Abacus, 43*(3), 323-357.

Basel Committee on Banking Supervision Publications, Bank for International Settlement (2005), *Studies on validation of internal rating systems.*

Basel Committee on Banking Supervision Publications, Bank for International Settlement (2006). International convergence of capital measurement and capital standards – A revised framework.

Bijak, K., & Thomas, L. C. (2012). Does segmentation always improve model performance in credit scoring? *Expert Systems with Applications, 39*, 2433–2442.

Carlehed, M., & Petrov, A. (2012). A methodology for point-in-time-through-the-cycle probability of default decomposition in risk classification systems. *The Journal of Risk Model Validation*, *6*(3), 3.

Chen, F. L., & Li, F. C. (2010). Combination of feature selection approaches with SVM in credit scoring. *Expert Systems with Applications*, *37*(7), 4902-4909.

Fei, F., Fuertes, A. M., & Kalotychou, E. (2012). Credit rating migration risk and business cycles. *Journal of Business Finance & Accounting*, *39*(1-2), 229-263.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

International Accounting Standards Board (2014). *International Financial Reporing Standard 9 – Financial Instruments*.

Jain, K., Gupta, P., & Sanjiv, M. (2011). Logistic Predictive Model for SMEs Financing in India. *The Journal of Business Perspective, 15*(4), 331-346.

Khemais, Z., Nesrine, D., & Mohamed, M. (2016). Credit Scoring and Default Risk Prediction: A Comparative Study between Discriminant Analysis & Logistic Regression. *International Journal of Economics and Finance, 8*(4), 39.

Kennedy, K., Namee, B. M., Delany, S., O'Sullivan, M., & Watson, N. (2013). A window of opportunity: Assessing behavioural scoring. *Expert Systems with Applications, 40*, 1372–1380.

Louzada, F., Ferreira-Silva, P. H., & Diniz, C. A. (2012). On the impact of disproportional samples in credit scoring models: An application to a Brazilian bank data. *Expert Systems with Applications, 39*, 8071–8078.

Malik, M., & Thomas, L. C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, *28*(1), 261-272.

Muminović, S., Pavlović, V., Cvijanović, J.M. (2011) Predictive ability of various bankruptcy prediction z-score models for Serbian publicly listed companies. *Industrija*, vol. 39, br. 3, str. 1-12

Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*,

38(12), 15273-15285.

Pavlović, V., Muminović, S., Cvijanović, J.M. (2011) Application of Sandin & Porporato's bankruptcy prediction model on Serbian companies. *Industrija*, vol. 39, br. 2, str. 1-13

Pluto, K., & Tasche, D. (2010). Internal-Ratings-Based Approach. Encyclopedia of Quantitative Finance

Siddiqi, Naeem. *Credit risk scorecards: developing and implementing intelligent credit scoring.* Vol. 3. John Wiley & Sons, 2012.

Sohn, S. Y., & Kim, Y. S. (2012). Behavioral credit scoring model for technology-based firms that considers uncertain financial ratios obtained from relationship banking. *Small Business Economics*, 1-13.

Appendix

Table 12. Indicators chosen for multivariate analysis

No.	Indipendent variable calculation	IV	Number of buckets	WOE Trend
1	Operating expenses / Total expenses	0.177	4	Growing
4	Operating income / Total revenues	0.117	3	Growing
11	Capital / Fixed assets	0.206	4	Growing
15	Total liabilities / Capital	0.130	5	Declining
17	(Long-term provisions + Long-term liabilities) / Capital	0.105	3	Declining
19	(Net result + Interest expense + Tax expenses) / Interest expense	0.789	4	Growing
21	(Operating result +Costs of depreciation and provisions) / Interest expense	0.455	4	Growing
22	Operating result / Sales revenues	0.146	4	Growing
25	Net result / Total assets	0.763	4	Growing
26	Net result / Capital	0.215	4	Growing
40	Financial liabilities / Capital	0.683	5	Declining
42	Financial liabilities / Earnings before interest, taxes, depreciation and amortization	0.432	4	Declining
44	Net liquid assets / Operating income	1.018	3	Growing
52	Net operating funds needs / Sales revenues	0.375	3	Growing
53	Total net operating fund / Sales revenues	0.262	4	Growing
58	Earnings before interest, taxes, depreciation and amortization / (Capital + Financial liabilities)	0.551	5	Growing
59	Operating result / (Capital + Financial liabilities)	0.415	4	Growing
66	Earnings before interest, taxes, depreciation and amortization / Financial expenses (gross)	0.782	4	Growing
68	Inventories / Total assets	0.119	4	Growing
70	Cash / Total assets	1.473	4	Growing
78	Long-term liabilities / Total assets	0.280	4	Declining
79	Financial liabilities / Total assets	0.669	4	Declining
83	Liabilities from operations / Total assets	0.149	4	Growing
85	(Net profit + Amortization and depreciation expenses)/ Financial expenses	0.366	4	Growing
86	Operating result / Financial expenses	0.260	4	Growing
89	Net operating capital / Total assets	0.568	4	Growing
91	Operating assets / Total assets	0.170	3	Growing
95	(Cash and cash equivalents + Short-term fiinancial investments) / Total assets	0.136	4	Growing
97	Operating assets / Sales revenues	0.727	5	Declining
98	(Operating assets – Inventories) / Sales revenues	0.612	5	Declining
109	Financial expenses / Sales revenues	0.751	4	Declining
110	Receivables / Total liabilities	0.287	3	Growing
113	Liabilities from operations / Sales revenues	0.292	5	Declining
120	Days receivables outstanding (Receivables * 365 / Sales revenues)	0.388	4	Declining
121	Days liabilities outstanding (Liabilities from operations * 365 / (Cost of goods sold + Cost of material)	0.223	4	Declining
123	Earnings before interest and taxes / Total assets	0.506	4	Growing
124	(Earnings before interest and taxes + Interest income) / Total assets	0.469	4	Growing
126	Retained profit / Total assets	0.586	3	Growing
133	(Net result – growth of Receivables + growth of Liabilities from operations) / Total assets	0.351	4	Growing
135	(Net result – growth of Receivables) / Liabilities from operations	0.430	4	Growing
138	(Net reesult – growth of Receivables) / Sales revenues	0.662	3	Growing
141	Total liabilities / Sales revenues	1.160	4	Declining