Expert Systems with Applications 40 (2013) 5932–5944

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements

Nebojsa Nikolic *, Nevenka Zarkic-Joksimovic, Djordje Stojanovski, Iva Joksimovic

University of Belgrade, Faculty of Organizational Sciences, Jove Ilica 154, Belgrade, Serbia

ARTICLE INFO

Keywords:
Credit scoring
Logistic regression
Probability of default
Weight of evidence approach
Corporate entities

ABSTRACT

In this paper a brute force logistic regression (LR) modeling approach is proposed and used to develop predictive credit scoring model for corporate entities. The modeling is based on 5 years of data from end-of-year financial statements of Serbian corporate entities, as well as, default event data. To the best of our knowledge, so far no relevant research about predictive power of financial ratios derived from Serbian financial statements has been published. This is also the first paper that generated 350 financial ratios to represent independent variables for 7590 corporate entities default predictions'. Many of derived financial ratios are new and were not discussed in literature before. Weight of evidence (WOE) method has been applied to transform and prepare financial ratios for brute force LR fitting simulations. Clustering method has been utilized to reduce long list of variables and to remove highly correlated financial ratios from partitioned training and validation datasets. The clustering results have revealed that number of variables can be reduced to short list of 24 financial ratios which are then analyzed in terms of default event predictive power. In this paper we propose the most predictive financial ratios from financial statements of Serbian corporate entities. The obtained short list of financial ratios has been used as a main input for brute force LR model simulations. According to literature, common practice to select variables in final model is to run stepwise, forward or backward LR. However, this research has been conducted in a way that the brute force LR simulations have to obtain all possible combinations of models that comprise of 5–14 independent variables from the short list of 24 financial ratios. The total number of simulated resulting LR models is around 14 million. Each model has been fitted through extensive and time consuming brute force LR simulations using SAS® code written by the authors. The total number of 342,016 simulated models ("well-founded" models) has satisfied the established credit scoring model validity conditions. The well-founded models have been ranked according to GINI performance on validation dataset. After all well-founded models have been ranked, the model with highest predictive power and consisting of 8 financial ratios has been selected and analyzed in terms of receiver-operating characteristic curve (ROC), GINI, AIC, SC, LR fitting statistics and correlation coefficients. The financial ratio constituents of that model have been discussed and benchmarked with several models from relevant literature.

1. Introduction

Credit scoring models play an important role in contemporary banking risk management practice. They contribute to the key requirement in loan approval process, which is to accurately and efficiently quantify the level of credit risk associated with a customer. The credit scoring models objective is to predict future behavior in terms of credit risk by relying on past experience of customers with similar characteristics. The level of credit risk of a borrower is associated with probability that it will default on approved loan over given time horizon, usually 1 year. The main task of credit scoring model is to provide discrimination between the ones who do default and the ones who do not, i.e. between good and bad corporate entities in terms of their creditworthiness. Discrimination ability is the key indicator of model successfulness. The higher the discrimination power the more precise the credit scoring model will be.

The models can be established on judgmental basis or with support of statistical tools. Judgmental or expert-based models are established through set of formal 'rule-of-thumb' quantitative criteria. It is an easiest way to incorporate the best practices and the knowledge of credit managers into formal automated decision rules. On the contrast, statistical scoring models are built upon optimization algorithm which is applied on historical data of credit

* Corresponding author. Address: Svetozara Markovica 71/13, Belgrade, Serbia. Tel.: +381 603584674.

E-mail address: nebojsa.nikolic.bg@gmail.com (N. Nikolic).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.05.022

performance of both good and bad customers. For an extensive review of statistical methods and their credit scoring application we refer to study of Crook, Edelman, and Thomas (2007).

Contemporary risk management practice emphasizes and promotes the use of credit scoring models for various asset classes of bank's credit portfolio (BCBS., 2006). 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; Sustersic, Mramor, & Zupan, 2009). By employing process automation, the bank's staff costs are reduced, loan approval process is simplified, speeded up and more control on approval decision making process is attained (Blochlinger & Leippold, n.d.). 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.

Latest Serbian credit bureau data report states that about 68% of banking loan exposures belongs to corporate entities (ASB., 2013). In process of financial statement analysis, in order to evaluate the financial health of a corporate entity the financial ratios are commonly used as a part of fundamental analysis. Granting loans to corporate entities based solely on credit scoring models is generally performed only for smaller loan amounts and for particular standardized loan products. More often, credit scoring models are used as an additional tool or decision making criteria. Credit scoring models based on logistic regression (LR) modeling technique provide the results in form of probability of default (PD) level for particular corporate entity. The PD level represents a quantitative estimate of credit risk inherited into corporate entity and it plays a valuable role in credit risk assessment within bank. 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 business utilization of estimated PD serves for calculating: expected and unexpected losses, statistical rating of corporate entities, loan loss provisions and cost of risk component of interest rate (Altman & Sabato, 2007; Ruthenberg & Landskroner, 2008). Applying a credit scoring model with a higher discrimination power could result in lower capital requirements and more accurate PD estimates (Altman & Sabato, 2007).

The Basel II standards recommend portfolio segmentation[1] of corporate entities based on sales (BCBS, 2006). Many banks already follow this segmentation practice when modeling credit risk, but in academic literature according to Altman and Sabato (2007) a study that reveals significant benefit of such choice is lacking. The segmentation study of Bijak and Thomas (2012) has shown that segmentation does not always improve credit scoring model performance. Governed with this idea, the underlying principle of our study was to treat all corporate entities as one segment. The goal was to try to build "one-size-fits-all" credit scoring model for whole population of corporate entities.

According to relevant literature, two mostly used statistical methods for building credit scoring models for corporate entities are discriminant analysis (DA) and logistic regression (LR). The first attempt to link financial ratios of corporate entities to risk in terms of probability of bankruptcy was done by Beaver (1967). The well-known *Z*-score model, developed on small corporate entities sample by Altman (1968) and afterwards enhanced by Altman, Haldeman, and Narayanan (1977), was the first credit risk model to predict default probabilities of corporate entities using DA technique. For the first time LR was applied in default prediction study of (Ohlson (1980). The main benefits of LR over DA were emphasized in terms of less restrictive modeling assumptions. The 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. The first reported LR prediction results were of less predictive power than the ones reported in DA studies. Later on, studies have shown that LR is a sound and powerful statistical approach for modeling credit risk. Further researches of models for predicting business failures using LR are discussed and implemented by Johnsen and Melicher (1994), Dimitras, Zanakis, and Zopounidis (1996), Laitinen and Laitinen (2000), Becchetti and Sierra (2003) Westgaard and Wijst (2001), Altman and Sabato (2007), Kumar and Ravi (2007) and Chen (2011).

In the last decade the extensive development of credit scoring models has been done. Default prediction as a classification problem entails forecast of corporate entity failure likelihood given a number of independent variables in terms of financial ratios (Altman & Sabato, 2007; Fantazzini & Figini, 2009; Westgaard & Wijst, 2001). 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 study of Zekic-Susac, Sarlija, and Bensic (2004) compared LR results with other different estimation methodologies on Croatian bank dataset. The paper of Hermanto and 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, and 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, and 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 mentioned 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 behavioral scoring results have been revealed and compared to its application credit scoring counterpart. Finally, in the most recent study of Blanco, Pino-Mejias, Lara, and Rayo (2013) compared LR results with other non-parametric techniques, based on a sample of almost 5500 microfinance borrowers from Peru.

Recent studies for corporate entities show that beside financial ratios there is potential value added, in prediction power terms, when economic, environmental and non-financial information are included in the model as a default predictors (Blanco et al., 2013; Moon & Sohn, 2010).

Even with the existence of more sophisticated classification models for credit scoring, such as neural networks (Derelioglu & Gurgen, 2011; Lee, Han, & Kwon, 1996; Leshno & Spector, 1996), support vector machines (Kim & Ahn, 2012) and case based reasoning (Vukovic, Delibasic, Uzelac, & Suknovic, 2012) the popularity and usage of LR has continued mostly due to its practicality and theoretical soundness.

To the best of our knowledge, this is the first study that has examined all possible combinations of the models given the short list of financial ratios as input variables. In comparison to other studies we have generated long list of 350 financial variables and then tailored the principal component clustering technique in order to reduce this long list to a short list of 24 variables. We examined in details the predictive power of financial ratios as standalone variables, as well as, the all possible combinations of models that include 5–14 financial ratio variables.[2] The total num-

[1] Small medium enterprises (SME) segment is for sales less than €50 million, corporate (CO) over €50 million and large corporate (LC) entities over €500 million.

[2] Due to computational reasons the maximum number of variables considered in model has been set to 14. Analysis that follows shows this to be more than sufficient for identifying the most suitable model.

ber of estimated credit scoring models in this study, using logistic regression is around 14 million. As a final result of our research we proposed eight variable LR model with highest predictive power among all developed models. Predictive power of our proposed LR model, based on weight of evidence transformation of financial ratios, has been in line with results of Altman and Sabato (2007) and better when contrasted to results of (Sohn and Kim (2012).

The rest of this paper is organized as follows. A description of the Serbian dataset is given in Section 2. The details about research methodology are revealed in Section 3. Firstly, we propose approach for financial ratio construction and transformation based on official balance sheet and income statement dataset of corporate entities. Secondly, the attribute constellation of variables is discussed and performance measures together with LR estimates are presented. This section also emphasizes and describes the variable reduction technique using cluster analysis. In Section 4, we present and discuss LR simulation results and compare the prediction performance of the best models. Finally, we bring our conclusions in Section 5.

2. Data and variables

In the research we used end-of-year financial statements and historical performance data for 7590 corporate entities in Serbia. Only the corporate entities which have had material financial liabilities from credit-like products were taken into analysis. Dataset contains annual 5 year series of end-of-year balance sheets and income statements of Serbian corporate entities, covering a period from 2006 to 2010. This length of data series satisfies Basel II compliance condition of minimum existence of 5 year of data history (BCBS, 2006).

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). This assessment is based on corporate entity repayment behavior over a fixed performance window. According to Kennedy, Namee, Delany, O'Sullivan, and Watson (2013) an occurring default status can be defined either at the end of the outcome period (current status) or during the outcome period (worst status). The authors have shown that the worst status approach performed better on an outcome window of 12 months. This particular approach has also been used in our study, since Basel II definition of default (BCBS, 2006) does not explicitly prescribe which of those two definitions to use.

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 31.12.2005 through 31.12.2008, 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 31.12.2006 up to 31.12.2009 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 5 year economic cycle and to develop thought-the-cycle credit scoring model (Siddiqi, 2005).

To conclude previous elaboration we present example of how performance of the corporate entity has been scrutinized. For particular corporate entity, based on balance sheet and income statement data on 31st December of year $T-1$, we monitored its days past due performance during year T and conclude about firm's default status at 31st December of year T.

In order to perform a proper comparison of developed models the dataset is split into two disjoint sub-samples, train and validation (Banasik, Crook, & Thomas, 2003). The idea behind this is to set aside a part of the data as validation sample just 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 70:30 has been done using random stratified sampling on target variable by years shown in Table 1. The data partition also avoids possibility of overfitting. Namely, the danger comes from the fact that credit scoring model might be doing well on train sample, but performing poorly in practice. Overfitting in credit scoring models is typically characterized by large coefficients, high standard errors, or changes in coefficient for some variables that do not make logical sense (Dwyer, 2005). In data sample partitioning it is important to preserve the same default rate by years as shown in Table 1, on both sub-segments according to Crone and Finlay (2012) findings.

The ratio of total and total defaulted (bads) corporate entities in a non-partitioned dataset, over a 5 year period, implies thought-the-cycle default rate of 9.01%, as it is shown in Table 1. Train sub-sample used for model coefficients tuning purpose comprises of 5313 'firm-years' or 70% of observations. Validation sample data consists of 2277 'firm-years' or 30% of observations. In further research we used train and validation samples to evaluate model performance and to distinguish between discriminatory power differences between developed models.

3. Research methodology and experimental design

Although many corporate credit scoring studies have been successful in their particular aspects, there is no unified standard of financial ratios that should be used to evaluate the financial performance of a corporate entity (Sohn & Kim, 2012). After more than four decades of research, there is still no consensus on what financial ratios are good predictors and why (Westgaard & Wijst, 2001).

The raw financial position data from balance sheets and income statements are rarely used for modeling purposes. The best practice is to combine financial positions and to construct a wide range of financial ratio variables. There are significant advantages of transforming raw financial positions into relevant financial ratio form. The first is that use of ratios facilitates comparison between firms of different sizes. The second is that financial ratios quantify information from two positions in a single number (CFA Institute., 2012).

In many studies financial ratios are taken into the model as short list variables. In his novel paper (Altman, 1968) examined 22 potentially helpful financial ratios and ended up selecting five. Through his more recent study (Altman & Sabato, 2007), 17 variables have been analyzed and again five were selected as the best for the final model. The study of Fantazzini and Figini (2009) used only 16 predefined financial ratios, similarly (Kim & Ahn, 2012) in model comparison study used 14 financial ratios from different accounting categories. The longer list of financial ratios evaluated through the modeling research can be found in Lee et al. (1996) (Leshno and Spector (1996) and Lin and McClean (2001).

3.1. Financial statements data – variables construction

In literature, there is no theory that would clearly define which financial ratios should be always found in the models.

Table 1
Overview of stratified sampling of data set.

Performance year	Train		Validation		Total	
	Firms	% Default	Firms	% Default	Firms	% Default
2007	559	13.06	228	13.16	787	13.09
2008	931	7.63	406	6.16	1337	7.18
2009	1189	10.18	506	9.49	1695	9.97
2010	1302	8.76	530	8.68	1832	8.73
2011	1332	7.51	607	9.23	1939	8.05
TOTAL	5313	9.02	2277	9.00	7590	9.01

Across countries different accounting systems, economic conditions, funding structures and tax codes may affect the predictive power of the same financial ratio. According to Chen (2011) it is common practice that prior researches of financial variables as in Gentry, Newbold, and Whitford (1985) and Plattner (2002) have been used to define preliminary list of financial ratios for modeling purposes. In our study we did not use prior research recommendations. The goal was to apply bottom-up logic of constructing the long-list of financial ratios variables for default modeling purposes. Finally, we ended up with long-list of 350 ratios variables.

To the best of our knowledge, until now no relevant research has been conducted about predictive power of financial ratios on financial data of Serbian corporate entities. We decided to govern research according to CFA Institute (2012) recommendations and try to construct wide envelope of financial ratios. The primary goal was to construct the financial variables that can explain earnings quality, financial structure, debt repaying capacity, liquidity, activity and cash flow potential. Three main groups and corresponding subgroups of financial positions[3] were formed:

(1) *Balance sheet positions*:

(a) *Asset side*: assets (AST), adjusted assets (AAST = assets – intangible assets), cash and cash equivalents (CSH), marketable securities (MSEC), accounts receivable (ACCREC), inventories (INV), current assets (CURAST), fixed assets (FIXA), deferred taxes (DEFT), long term financial assets (LTFA).
(b) *Liabilities and Equity side*: liabilities (LIAB), adjusted liabilities (ALIAB = liabilities – cash and cash equivalents), accounts payable (ACCPAY), current liabilities (CURLIAB), long term debt (LTDBT), long term bank debt (LTBDBT), total debt (TDBT), total equity (EQT), core equity (CEQT), retained earnings (RTERNG).

By combining the asset, liabilities and equity side positions we also calculated synthetic positions as: working capital (WC), operating assets (OA), operating liabilities (OL), net operating assets (NOA), cash and marketable securities (CSH_MSEC).

(2) *Income statement positions*: COGS (costs of goods sold), personal costs (PSCST), depreciation/amortization (DEP), net expenses (NETEXP), adjusted net expenses (NETEXPADJ), financial expenses (FINEXP), financial income (FININC), sales (SLS), gross profit (GROSPRF), gross profit adjusted (GROSPRFADJ), net sales (NETSLS), EBIT, EBITDA, EBT, adjusted EBT (EBTADJ it includes extraordinary items[4]) and net profit (NETPRF).

(3) *Approximation of cash flow statement positions*: Income statement items such as net income can be manipulated because of accrual accounting.[5] Cash flow indicators, derived from income statement, are unaffected by potential managerial manipulation of net income items. According to CFA Institute (2012) the most relevant approximation of cash flow positions is done by constructing: cash flow from operations (CFO), free cash flow to firm (FCFF) and free cash flow to equity (FCFE). Firms which are unable to generate sufficient operating cash flow are more susceptible to bankruptcy. In study of Gupta, Wilson, Gregoriou, and Healy (2012) it is confirmed that the presence of operating cash flow information can improve the prediction power of credit scoring model. The cash flow usefulness has been also discussed by Aziz, Emanuel, and Lawson (2007).

The financial ratios have been constructed bearing in mind a set of all possible combinations of discussed raw financial positions. In doing so, it is important to stay in line with business logic and not to pair positions without economic sense. We applied this approach and the result indicated many new ratio combinations that were not discussed in literature.

Firstly, we constructed ratios by combining positions in-between balance sheet items, bearing the economic sense of ratio as a goal. We provide here an example of ratios that emerged using *cash* item in numerator and other raw financial positions in denominator. Produced cash related ratios are: CSH/ACCREC, CSH/ACCPAY, CSH/CURAST, CSH/CURLIAB, CSH/LTDBT, CSH/CEQT, CSH/LIAB etc. This procedure has been repeated many times but instead of cash in numerator, we substituted it by other balance sheet positions. The same logic applies to income statement positions, where for example we have: COGS/SLS, COGS/NETSLS, COGS/GROSPRFT, COGS/EBIT etc. Combination between balance sheet and income statement raw positions has also been considered. For example, net sales related ratios are: NETSLS/AST, NETSLS/LIAB, NETSLS/ALIAB, NETSLS/CEQT, NETSLS/EQT etc.

The second step was to construct ratios that use more raw data items in numerator and denominator, as well as, cash flow related ratios. At this stage the following ratios were recorded: defense interval, accruals ratio, total bank debt to capital ratio, payables processing period, receivables collection period, cash conversion cycle; FCFF/AST, FCFF/LIAB, FCFF/EQT, FCFF/TDBT etc.

The last step of variable construction process was to establish growth variables of some relevant ratios and financial positions. For instance, return on assets (ROA) growth. The growth is calculated as: growth = V_T/V_{T-1} where V stands for ratio value in year T and year $T-1$. The rationale is that growth ratio value greater than one stands for growth while value less than one stands for decline.

[3] Financial positions are the main constitutes of financial ratios. They should be placed in ratio numerator and denominator.

[4] The earnings before tax (EBT) are positioned before extraordinary items in Serbian income statements. Thus, EBT adjusted (EBTADJ) is calculated after including effect ("+" or "−" sign) of extraordinary items into EBT.

[5] Accrual basis of accounting specifies revenues are recognized when earned, and expenses are recognized when incurred, regardless of the timing of cash flow. Accrual accounting necessitates the use of discretion because of the many estimates and judgments involved. It can be "strategically" manipulated by management to achieve a desired result.

Peculiarities emerged during variable construction process were solved using the principle of special values. The first peculiarity we had to address with constructed ratios was the division by zero. If denominator is zero, which for instance can happen when financial expenses are equal to zero, the resulting ratio is not numerically defined. To account for this in the modeling process we imposed a special numeric value of −999,999,999 to all such observations. The second peculiarity involved addressing the cases when both numerator and denominator in a financial ratio can be a positive or a negative number (for example 'EBITDA/Gross profit'). To distinguish observations where both numerator and denominator are negative we imputed a special ratio value of −888,888,888.

3.2. Financial ratios transformation

In our study we decided to transform standardized financial ratio values since the study of Altman and Sabato (2007) has shown that transformed financial ratios can improve the credit scoring model. Variable transformation technique that we applied in this research is based on weight of evidence (WoE) approach (Siddiqi, 2005). We created two or more corresponding attributes[6] for each of 350 financial ratio variables and then calculated its WoE value. The grouping of financial ratio attributes requires an algorithm that can divide each variable into standard number of adequately sized attributes (groups).

Attributes creation within financial ratio variable facilitates understanding of its relationship with credit risk. For continuous variables division into attributes enables easier insight into behavior and economic sense of the variable (Hand & Niall, 2000). Attributes are formed by taking into account the target variable, i.e. the default rate. The aim of the grouping process is to maximize inter and minimize intra differences between created attributes. The better the differentiation is done, the more precious information will be extracted from the variable in terms of default rate prediction power. If the default rate of each attribute is close to average default rate of the sample[7] variable will probably be totally unusable. Attributes within variable with similar relative credit risks are usually merged. In attribute grouping it is a paramount that each attribute must have a sufficient number of default events. If this is not the case, the model obtained from such a variable might be wrong. The minimum of 10 default events and 10 non-default events in each financial ratio attribute is necessary according to Peduzzi, Concato, Kemper, Holford, and Feinstein (1996). Less than 10 defaults and 10 non-default events per attribute makes the model performance unsatisfactory. Some practitioners find this to be too strict, but obeying this, so called '10k rule', most probably keeps modeling process on the safe side.

For this study, financial ratios attributes were grouped in a way to preserve economic sense. Each attribute within financial ratio has been shaped to preserve consistent economic logic. For example, profitability ratios usually tend to have lower default rate for high ratio values and vice versa. Thus, the created attributes should preserve the same business logic. The number of outlined attributes should be as few as possible, since too many of them can cause greater complexity and increase potential for over fitting. On the other side, having too few attribute classes can distort valuable information (Siddiqi, 2005).

In this paper, creation of financial ratio attributes provided very effective way of dealing with outliers and special values. Imputed special values and financial ratio outliers have been treated as separate attributes within each variable. Taking into account imputed special values, if there were not enough default and non-default events to form stable attributes, the corresponding special values have been assigned to another attribute, usually the one with the largest number of observations. Financial ratio outliers have been treated in a similar manner, but they were merged with the first or last variable's attribute based on outlier direction. According to relevant literature (Anderson, 2007) and in line with our findings, no more than 7 attributes per financial ratio variable have been used in this study.

3.2.1. Algorithm for financial ratio attributes optimization

The process of transforming and dividing independent variables into corresponding attributes has been performed using bottom-up approach with an optimization algorithm (TrimTab Solutions., 2012). The goal of the optimization algorithm is to maximize prediction power of each standalone financial ratio by staying in line with economic logic. To perform the calculation, each variable is sorted into ascending order.

In *first stage*, the algorithm creates 64 fine classes of equal size. Three different approaches have been used to construct these 64 classes. Each of three grouping approaches produces a different grouping solution for every financial ratio. All results have been stored as different grouping arrays, so that at the end we have three possible groupings per each variable.

- The *first grouping procedure* is built by sorting financial ratio in ascending order and then creating 64 fine groups by preserving equal number of goods per group.
- The *second grouping procedure* is built by sorting financial ratio in ascending order and then creating 64 fine groups by preserving equal number of bads per group.
- The *third grouping procedure* is built by sorting financial ratio in ascending order and then creating 64 fine groups by preserving equal number of all observations.

During *second stage* each of three previously constructed fine grouping arrays are treated separately by the algorithm. The goal is to group 64 fine classes into larger groups until predefined number of attributes per variable is achieved. The initial grouping in second stage is done in order to form first two attributes. By choosing corresponding boundary of one of 64 fine classes as new separator, algorithm assembles 64 fine classes into two new groups. The separator cutoff has been set to maximize distinction between goods and bads between the two newly created groups. Separator cutoff between the new groups is recognized to maximize value $-\ln(p\text{-}value)$ measure calculated with chi-square statistics with. The chi-square statistics for k groups has $k-1$ degrees of freedom and is calculated as:

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{(\text{bads}_i - (\text{bads}_i + \text{goods}_i) \times default\_rate)^2}{(\text{bads}_i + \text{goods}_i) \times default\_rate} \right] \quad (1)$$

where bads and goods stands for number of bads and goods per group i, and *default_rate* equals percentage of bads in the whole population. After first separator cutoff has been established, each of the two new groups have been again separately treated with second stage process of algorithm. The underlying principle it to always leave separator cutoff which produces highest $-\ln(p\text{-}value)$ measure in comparison to separators from other groups. The second stage grouping process is repeated until we have been left with desired number of attributes.

Second stage algorithm has been worked through each of three resulting fine grouping arrays from the first stage. Finally we have been left with three different attributes constellations per each financial ratio. The one which produces highest standalone predictive power of particular financial ratio has been chosen as the final attribute constellation.

[6] In practice attributes can be referred to as categories or bins.

[7] No separation between goods and bads firms per financial ratio attribute.

The underlying algorithm is a slight modification of Chi-squared Automatic Interaction Detection (CHAID) decision tree rules (Kass, 1980) which automatically takes care of '10k rule' introduced by Peduzzi et al. (1996).

Where needed a manual adjustment of attribute boundaries has been done in order to preserve the economic logic of each variable and at the same time keep the predictive power of corresponding variable as high as possible. These adjustments are suggested by Bijak and Thomas (2012).

3.2.2. WoE approach

The Weight of Evidence (WoE) approach measures the information content of an attribute within variable and assigns to each attribute corresponding numerical value which quantifies its relation to default rate. The WoE value is calculated for each attribute within variable. The WoE value logic is to quantify difference between the proportion of good and bad observations in each financial ratio attribute (Thomas, 2009). WoE allows for the univariate assessment and comparison of the relative credit risk associated in financial ratio attributes (Bijak & Thomas, 2012). A ratio of goods to bads is referred to as the *odds* in credit scoring. The *population odds* represent a ratio of the proportion of goods to the proportion of bads in population. Suppose there are N attributes. The WoE value of attribute i is calculated according to Anderson (2007) as:

$$WoE = \ln\left(\text{goods}_i / \sum_{i=1}^{n} \text{goods}_i\right) - \ln\left(\text{bads}_i / \sum_{i=1}^{n} \text{bads}_i\right) \qquad (2)$$

The farther the WoE is from zero, the more informative an attribute is. Variable attributes with default rate higher than average sample default rate produce negative WoE values. The positive WoE values can be found in attributes where default rate is lower than average sample default rate.

Applying the WoE approach the problem of missing values, outliers and special values is eliminated as those values are now represented by calculated WoE value. In addition, applying the WoE approach facilitates the interpretation of credit scoring results (Anderson, 2007). After WoE transformation the number of distinct values per variable equals the number of attributes and these attributes are used instead of their corresponding financial ratio values.

3.2.3. Variable predictive power measure

The prediction power of a variable can be quantitatively estimated using the *information value* (IV). Information value is calculated according to formed N attributes, corresponding attributes WoE values, as well as, corresponding proportion of goods and bads within attributes.

$$\text{IV} = \sum_{i=1}^{n} \left(\text{goods}_i / \sum_{i=1}^{n} \text{goods}_i - \text{bads}_i / \sum_{i=1}^{n} \text{bads}_i\right) \times WoE_i \qquad (3)$$

According to Siddiqi (2005) estimated IV should be interpreted in the following way:

- IV < 0.05: almost no predictive power
- 0.05 ⩽ IV < 0.1: poor predictive power
- 0.1 ⩾ IV < 0.25: medium predictive power
- IV > 0.25: high predictive power

In this research we exploit IV to rank and compare financial ratios by their predictive power in order to come up with a short list of financial ratios.

3.3. Variables clustering – short list selection

In order to reduce the long-list of variables, the variable reduction process has been done using clustering technique. The SAS/STAT® software has been utilized for this purpose. Clustering routine PROC VARCLUS has been used as clustering tool, with the aim to find groups of variables that are as correlated as possible among themselves and as uncorrelated as possible with variables in other clusters. The underlying clustering algorithm we used is based on principal component analysis. According to SAS Institute Inc. (2011) it maximizes the sum of the variance across clusters explained by the cluster components. Since correlation matrix is used the algorithm treats all financial ratios as equally important[8] during the clustering.

Additional benefit is that by clustering we can also try to isolate highly correlated variables. This is done through variable reduction process which consists of choosing the most predictive variables from each available cluster. Just a few studies covering financial ratio clustering can be found in literature, the most notable of which are Gupta and Huefner (1972), and Wang and Lee (2008).

Correlation between created financial ratios[9] is the most likely to cause problems in modeling. Models built with too many correlated variables[10] can destabilize parameter estimates and distort variable interpretation within model (Hosmer & Lemeshow, 2000). Important property of cluster analysis that we exploit is to dissect financial ratios into "natural" groups suggested by the clustering algorithm, not defined a priori. In this way every financial ratio tends to be similar to all variables within its cluster, and dissimilar to variables outside of it. All variables that belong to a cluster should carry similar information and are linked to target variable in a similar manner. Finally, it helped us to reveal the level of resemblance of financial ratios.

The relocation of variables to corresponding clusters occurs in two phase algorithm. *The first* is a nearest component sorting phase, similar in principle to the nearest centroid sorting algorithms described by (SAS Institute Inc, 2011). By iterations, the cluster components are formed and each variable is assigned to the component with which it has the highest squared correlation. *The second phase* involves a search algorithm where each variable is tested to realize if it should be assigned to a different cluster. Variable is moved to another cluster only if doing so increases the amount of variance explained. If a variable is reassigned during the second phase, the components of the two clusters involved are recomputed before the next variable is tested. The algorithm stops splitting when the number of user defined clusters is reached (SAS Institute Inc, 2011). For our study, clustering procedure has been parameterized to create formation of 24 cluster groups.

Parameters we used for cluster quality assessment are presented in Table 2. The variable that represents each cluster is selected in terms of maximum IV within its cluster and it is shown in first column. Cluster ID and number of variables per cluster are also presented. Cluster association measure of each variable with parent cluster, named R-squared, represents the squared correlation of the variable with its parent cluster component. Parent cluster R-squared values should be, by logic, higher than the R-squared with any other cluster. The average IV of a cluster and maximum IV of its representative variable are given in last two columns of Table 2. After the clustering process ended the resulting clustering statistics denoted that 63.61% of total variation is explained by constructed 24 clusters.

The governing variable reduction principle is to come to 24 short-list variables from the constructed clusters. The selection

[8] If covariance matrix is used, variables with larger variances have more importance in the analysis.

[9] Because of the space, the correlation of the variables is not posted here, but is available upon request from the authors.

[10] In our study, large number of financial ratios that we generated can be highly correlated between each other since we used similar base of variables building blocks. For example, it is evident that, cash/assets will be highly correlated with accounts receivable/assets, since balance sheet item cash is the constituent of accounts receivable.

Table 2
Financial ratios clustering results.

Variables	Cluster ID	Variables in cluster	R-squared with parent cluster	IV cluster	IV variable
CURLIAB/NETSLS	CLUS19	4	0.883	**0.809**	0.906
EBTADJ/LIAB	CLUS04	30	0.897	**0.692**	0.923
EBITDA/FINEXP	CLUS01	31	0.610	**0.634**	0.821
CSH/CURLIAB	CLUS05	15	0.867	**0.550**	0.746
EBITDA/LIAB	CLUS13	28	0.881	**0.531**	0.768
GRSPRF/ALIAB	CLUS17	5	0.891	**0.502**	0.612
EBTADJ/EBT	CLUS23	11	0.673	**0.435**	0.547
ALIAB/AST	CLUS02	22	0.818	**0.363**	0.557
CSH/EBITDA	CLUS09	24	0.394	**0.344**	0.503
TBDBT/NETSLS	CLUS15	8	0.707	**0.322**	0.564
NETPRF_GRTH	CLUS20	6	0.867	**0.320**	0.419
CURAST/ALIAB	CLUS06	20	0.667	**0.265**	0.469
LTBDBT/EQT	CLUS21	9	0.494	**0.257**	0.380
CURAST/NETSLS	CLUS18	6	0.659	**0.234**	0.407
PAYPROCPRD	CLUS12	12	0.509	**0.198**	0.414
CSH/LTDBT	CLUS16	13	0.244	**0.195**	0.512
NETSLS/FINEXP	CLUS07	9	0.597	**0.183**	0.506
ACCREC/ALIAB	CLUS24	4	0.836	**0.144**	0.185
FCFE/FINEXP	CLUS10	13	0.462	**0.132**	0.198
NETSLS_GRTH	CLUS11	10	0.290	**0.103**	0.442
CSH/CEQTY	CLUS22	9	0.220	**0.085**	0.300
SHRTBDBT/CURAST	CLUS08	15	0.482	**0.066**	0.155
CURLIAB/ALIAB	CLUS03	13	0.418	**0.060**	0.158
TBDBT/CEQTY	CLUS14	9	0.449	**0.043**	0.062

has been done in Table 2, and the most predictive variable per cluster has been selected. For example, cluster 19 (CLUS19) has the highest average prediction power measure with IV = 0.809. The cluster CLUS14 has the poorest average predictive power with IV = 0.043. The most predictive variable according to the last column of Table 2 is EBTADJ/LIAB with IV = 0.923. Its economic sense is actually to show capacity for repaying liabilities. The higher the ratio, the greater the capacity is and less credit risk is inherited from corporate entity. The financial ratio TBDBT/CEQTY has the lowest discriminatory power in variable short-list.

3.4. Logistic regression

Probably the most common used technique for default prediction is logistic regression. It is employed in solving problems of assigning probability to an event where there is binary dependent target variable to predict. LR remains the most popular method applied by practitioners in financial services industry (Crook et al., 2007). The primary difference between linear and logistic regression is the use of a binary variable as modeling target.[11] Several LR modifications are considered in Cramer (2004) but the conclusions are that their performance on independent validation dataset is substantially the same as plain LR. The same finding has been asserted by Louzada et al. (2012). 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 (Crone & Finlay 2012). In modeling corporate credit risk, according to Hosmer and Lemeshow (2000) LR model identifies the probability of default event to occur as:

$$\pi(x_1,\ldots,x_n) = \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\ldots+\beta_kx_k)}} \quad (4)$$

where n is number of independent observations 'firm-year' in our research, β_i represents the particular coefficient in linear combination of k independent variables and $\pi(x_i)$ is the posterior conditional probability of default event for corporate entity in corresponding year. The LR coefficients are estimated on training dataset using the maximum likelihood (ML) method. The ML method is applied by constructing a likelihood function which expresses the probability of the observed data as a function of the unknown model coefficients. The solution for the unknown coefficients is recognized when ML function is maximized (Hosmer & Lemeshow, 2000). The resulting beta coefficients will agree most closely with the observed training dataset. The SAS/STAT® software procedure PROC LOGISTIC has been used to fit LR and to compute the ML estimators of the k variables (SAS Institute Inc, 2011).

3.5. Model performance measure

Once LR model is fitted on training data the process of model assessment begins. Modeling is done using performance measures that we used to evaluate whether the predicted values are an accurate representation of the observed values.

The first performance measure that we have used in our study is linked to the statistical performance of variables in fitted LR models. The logic of its use is that we want to take into account only models were each variable is statistically significant. The Wald test and its p-value have been used to examine the statistical significance of each β_i coefficients in the model. A Wald test calculates a Z statistics as:

$$W_i = \hat{\beta}_i/\sigma(\hat{\beta}_i) \quad (5)$$

The null hypothesis that an estimated coefficient is zero is tested using Eq. (4). If the p-value that the coefficient could be zero is greater than 0.05 the variable is considered as weak and it could be left out. The statistic follows standard normal distribution and literature suggests different significance thresholds that could be used for LR, for example refer to Min and Jeong (2009). We apply this statistic to distinguish between models that contain insignificant effect of particular variables. Thus, if Wald test indicates that any of variables is insignificant, we neglect this model from the further analysis. The rationale behind is to keep only financial ratios that have significant statistical influence on default event prediction.

The quality of the LR models has been evaluated with the following measures: GINI coefficient, Receiver Operating Characteris-

[11] Linear regression for binary type of dependent variable would not be appropriate. It violates several of the linear regression assumptions.

tic (ROC) curve, as well as, AUC referred to as area under ROC curve (Tasche, 2006).

ROC is an alternative way of expressing the relationship between the cumulative score distribution functions of the goods and the bads. It describes the differentiating property of the credit scoring model at various levels of cutoff points. It is represented graphically by plotting the fraction of true positives out of the total actual positives (Sensitivity) on the *y*-axis at various cutoff levels. In other words it is the true positive rate (ratio of predicted defaulters over total actual defaulters). The *x*-axis is represented by the fraction of false positives out of the total actual negatives (1-Specificity), in other words it is a false positive rate (ratio of "wrongly" predicted defaulters over the total actual non-defaulters) at various cutoff levels. ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones (BCBS., 2005).

AUC (the area under the ROC curve) provides a single value performance measure [0; 1] which quantifies the predictive power of the model (Thomas, 2009). For perfect discrimination AUC value is one. For a credit scoring which has purely random discrimination AUC value is 0.5. For any other scorecard, the AUC curve will be between 0.5 and 1 and the larger the area the better the discrimination. Any value less than 0.5 implies that the model is getting it wrong with some consistency and zero means the predictions are perfectly wrong.

GINI coefficient corresponds to twice the area under the ROC curve as $\text{GINI} = 2 \times \text{AUC} - 1$. Useless and perfect discriminations are quantified by 0 and 1, respectively. Another possibility to calculate the model performance using the related GINI coefficient is done by Trapezium rule (Crone & Finlay, 2012). GINI also enables comparing discrimination power results directly with other studies. By this reason we considered it to be main criterion for choosing the final prediction model.

Kolmogorov–Smirnov (KS) statistic is calculated using the cumulative distribution functions (CDFs) of financial ratio values. Firstly, CDFs are computed separately for goods and bads and the distance between these CDFs is captured using the formula $D = \max|\text{CDS}_{\text{goods}} - \text{CDS}_{\text{bads}}|$. KS takes values between 0 and 1 with higher values indicating stronger discriminatory power (Bijak & Thomas, 2012). In our study we use it to show train and validation sampling quality, as well as, to measure discriminatory power of short-list variables.

Akaike (AIC) (Akaike, 1974) and Schwarz (SC) (Schwarz, 1978) model fitting criteria[12] are also used as a model performance measures. Its goal is to avoid likelihood of overfitting by adding variables into model. Both of those two measures try to resolve overfitting problem by introducing a penalty term for the increase in number of variables in the model.

3.6. Model building technique

In most other studies, the variables selection process rests on literature recommended variables. After applying cluster analysis on long-list of variables we created short-list of 24 financial ratios which have been used for final model building. The modeling process in our study consists of finding the most predictive model as a combination of financial ratio variables.

The most commonly used techniques that discover the final model inputs are different LR algorithms that checks for the importance of variables and either includes or excludes variables on the basis of a predefined significance decision rule (Hosmer & Lemeshow, 2000). The significance of a standalone variable is defined in terms of a measured Wald test *p*-value of its coefficient. The most widely used LR algorithms that eliminate or include variables based on their importance are forward, backward and stepwise selection.

In a *forward selection*, beginning from no variables in the model, incrementally the variables are added. The variables that are firstly included are those with the most significant coefficient defined by Wald test. In a *backward selection*, beginning from the all variables model, variables with the least significant *p*-values are excluded. The *stepwise selection* is a procedure of alternate dynamic inclusion and exclusion of variables from a model based on the statistical significance of its coefficients. It combines the logic of forward and backward selection methods. Described LR variable selection routines can be found in studies of Min and Jeong (2009), Hermanto and Gunawidjaja (2010), Yazdanfar (2011) and Sohn and Kim (2012).

Described LR algorithms may suffer from the fact that several powerful variables with low *p*-values can dominate an algorithm. This may prevent other variables with higher *p*-value to enter the model. On other hand, the high *p*-value performance of a particular variable may turn out into significant value added, measured in predictive power terms. It may turn out that a variable with high *p*-value can significantly increases model predictive power in combination with other variables. Thus, high *p*-value standalone variables may have significant impact on predictive power when they are combined with other variables.

In order to avoid abovementioned peculiarities related to high *p*-value variables we finalize our research by running all possible model combinations containing 5–14 variables, out of our total short-list of 24 variables. The total number of LR that should be evaluated is given in Table 3.

According to Table 3 the total number of combinations can be calculated by function $n!/[k!(n-k)!]$. This approach can be referred to as "all possible combination" or "brute-force" LR approach. As it can be inferred from Table 3 it is a very computationally intensive process. In total, the number of 14,185,135 models have been built using brute force LR simulations.

The simulation process is based on authors SAS macro code that runs PROC LOGISTIC procedure with Fisher's coefficient optimization technique. The main inputs in SAS macro code are train and validation datasets each with WoE transformations of 24 financial ratios variables and corresponding target variable.

The huge number of simulations required to develop and adopt SAS macro code to run multiple independent *.exe* process on corresponding server with 16 processors. *The first macro step* was to generate all possible combinations of variable names that enter in 5 variables model. The resulting dataset with 5 variable combinations has then been divided into smaller sub-datasets. *In second macro step*, each sub-dataset has been assigned to an independent SAS macro process which runs necessary LR simulations through separate *.exe* processes. The model parameters have been estimated on train data and then tested on validation data. The results have been saved into corresponding 5 variables models sub-datasets which are finally merged after finishing each of separate *.exe* process. From Table 3 it can be observed that total number of 42,504 models have been estimated for 5 variables in the model. The same process is governed for the remaining 6–14 variable models.

The SAS macro simulation process has been also linked to a predefined set of fixed rules. The goal was to capture only statistically and economically correct and stable models, referred to as *well-founded* models. The following, *credit scoring model validity conditions* have been imposed to recognize well-founded and to reject the unstable models:

[12] The penalty term is larger for BIC than AIC. Given any two estimated models, the model with the lower value of BIC is the one to be preferred.

Table 3
Total number of LR models according to short-list variables.

LR simulations	Available variables to enter model (n)				
	20	21	22	23	24
Number of variables in model (k)					
5	15,504	20,349	26,334	33,649	42,504
6	38,760	54,264	74,613	100,947	134,596
7	77,520	116,280	170,544	245,157	346,104
8	125,970	203,490	319,770	490,314	735,471
9	167,960	293,930	497,420	817,190	1,307,504
10	184,756	352,716	646,646	1,144,066	1,961,256
11	167,960	352,716	705,432	1,352,078	2,496,144
12	125,970	293,930	646,646	1,352,078	2,704,156
13	77,520	203,490	497,420	1,144,066	2,496,144
14	38,760	116,280	319,770	817,190	1,961,256

Table 4
Well-founded models performance – train and validation sample.

Variables in model (v)	Well-founded models	Train sample		Validation sample	
		Average GINI (%)	St. Dev GINI (%)	Average GINI (%)	St. Dev GINI (%)
5	25,986	59.01	3.15	50.99	3.86
6	55,871	60.85	2.65	52.65	3.29
7	82,980	62.28	2.30	53.91	2.87
8	84,480	63.44	2.03	54.91	2.53
9	58,171	64.39	1.80	55.70	2.24
10	26,084	65.22	1.60	56.35	1.98
11	7248	65.95	1.42	56.94	1.70
12	1113	66.60	1.24	57.48	1.43
13	79	67.18	0.98	58.14	1.11
14	4	62.28	2.30	53.91	2.87

- The model might prove to be unstable if it produces high variance of its parameters which affects its p-value estimates.
- Some variables might be too correlated with other variables in the model. In extreme case when WoE approach is applied, it results in positive coefficient estimates and inverted logic of the variable attributes. For example, although the best corporate entities are those with highest EBIT related ratios, a positive coefficient parameter will make this variable behavior in the particular model of inverted logic.
- The model cannot be estimated if the MLE algorithm could not converge.

4. Results and discussion

After taking into account model validity conditions, a total of 342,016 (2.41%) well-founded models have been discovered out of about 14 million possible model combinations. The Table 4 summarizes the related statistics. The average GINI and GINI standard deviation for corresponding models on both train and validation sample has been presented. The highest standard deviation in GINI has been shown for 5 variable (5v)[13] models. From the results, it can be inferred that number of well-founded models[14] declines while prediction power on both train and validation sample increases as we go from 5 to 13 variables. The GINI standard deviation declines on both train and validation sample with increase in number of variables.

For 14 variables the sharp decline in prediction power is observed and only 4 well-founded models have been found out of about 1.96 million. The rejection of 14 variable models was probably due to multicollinearity effect.

Table 5 summarizes the quality assessment of 70:30 data partitioning on short-list variables before WoE transformation has been applied. The results in Table 5 show the distance (D) between goods and bads *cdf* (D stat.), KS statistics and its corresponding p-value. According to KS statistics it can be concluded that no statistical difference exists per variable between train and validation samples. The lowest observed p-value 0.093 (>0.05) is not statistically significant for 95% confidence level. From the results we draw to conclusion that no separation bias exists between train and validation sample within short-list variables when data partitioning of 70:30 has been applied.

Assessment of standalone variable's differentiation ability between goods and bads on whole population sample has been also done with KS statistics. From Table 5 just one variable TBDBT/CEQTY (p-value = 0.444) did not distinguish well between cumulative distribution function (cdf) of goods and bads. Referring to Table 2 it can also be inferred that this variable suffers from lack of predictive power. With GINI of 0.062 this is the only short-list variable with GINI less than 0.1 and poor predictive power. Thus, the evidence shown by KS statistics corresponds to that implied by GINI.

The last column of Table 5 shows in how many of well-founded models particular variable can be found. For example, regardless of poor standalone predictive power variable TBDBT/CEQTY shows high level of persistence and it can be found in 45.62% of well-founded credit scoring models.

Table 6 presents an overview of the best well-founded. The first column in Table 6 is used to mark the name of the each ranked model. The mark contains information about the number of variables enclosed in a model and model simulation ID.[15] The alloca-

[13] In following text, we denoted number of variables in the model as: 5 variables (5v) up to 14 variables (14v).

[14] Table 4 indicates that there are only 4 well-founded models with 14 variables, so we cannot use 14v models statistics for comparison purposes.

[15] For example, 5v_25117 corresponds to 5 variables model obtained in 25117th LR fitting out of 42504 possible models combinations fitting. (42504 is the total number of model combinations for 5 variables to be contained in the model out of 24 available variables)

Table 5
Short-list variables statistics.

Variables	Train vs. validation			Goods vs. bads on whole population			Goods in population		Bads in population		Variable persistence in well-founded models (%)
	D stat.	KS	*p*-value	D stat.	KSa	p-value	Mean	St. Dev	Mean	St. Dev	
SHRTBDBT/CURAST	0.015	0.567	0.905	0.163	4.018	<0.0001	0.273	0.297	0.377	0.366	30.08
CSH/CURLIAB	0.033	1.240	0.093	0.352	8.737	<0.0001	0.070	0.125	0.019	0.047	21.63
CSH/LTDBT	0.020	0.662	0.774	0.316	7.046	<0.0001	0.316	0.590	0.129	0.410	28.39
ALIAB/AST	0.023	0.857	0.456	0.254	6.256	<0.0001	0.614	0.240	0.758	0.207	26.02
ACCREC/ALIAB	0.018	0.685	0.737	0.143	3.543	<0.0001	0.538	0.448	0.395	0.282	35.87
CURAST/ALIAB	0.014	0.529	0.943	0.240	5.951	<0.0001	1.137	0.702	0.809	0.378	22.22
CURLIAB/ALIAB	0.021	0.796	0.551	0.150	3.719	<0.0001	0.840	0.256	0.788	0.221	15.58
CSH/CEQTY	0.027	0.872	0.432	0.220	4.965	<0.0001	0.290	0.441	0.198	0.396	38.21
TBDBT/CEQTY	0.018	0.634	0.817	0.038	0.864	0.444	27.515	53.986	28.284	58.149	45.62
LTBDBT/EQT	0.021	0.771	0.593	0.165	3.702	<0.0001	0.385	0.652	0.652	0.860	30.31
NETSLS/FINEXP	0.017	0.610	0.851	0.277	6.750	<0.0001	64.130	66.633	33.694	43.619	26.51
EBITDA/FINEXP	0.015	0.508	0.959	0.324	6.710	<0.0001	5.433	7.106	2.429	4.494	20.09
EBTADJ/EBT	0.017	0.535	0.937	0.119	2.023	0.001	0.983	0.630	0.928	0.728	28.67
PAYPROCPRD	0.013	0.499	0.965	0.256	6.110	<0.0001	151.333	149.859	233.667	185.738	43.10
CSH/EBITDA	0.029	0.999	0.271	0.240	4.923	<0.0001	0.187	0.219	0.114	0.194	36.77
CURAST/NETSLS	0.025	0.926	0.357	0.246	5.873	<0.0001	0.482	0.312	0.684	0.399	44.56
CURLIAB/NETSLS	0.014	0.509	0.958	0.365	8.684	<0.0001	0.398	0.301	0.690	0.402	15.71
TBDBT/NETSLS	0.014	0.526	0.945	0.300	7.193	<0.0001	0.222	0.269	0.433	0.391	20.03
EBITDA/LIAB	0.018	0.654	0.786	0.325	6.772	<0.0001	0.243	0.263	0.114	0.123	34.89
EBTADJ/LIAB	0.023	0.849	0.468	0.343	7.348	<0.0001	0.151	0.249	0.041	0.084	29.66
GRSPRF/ALIAB	0.018	0.674	0.753	0.303	7.252	<0.0001	0.691	0.589	0.378	0.400	37.75
FCFE/FINEXP	0.031	0.968	0.306	0.190	3.751	<0.0001	3.815	4.026	2.562	3.208	38.84
NETSLS_GRTH	0.026	0.918	0.368	0.298	6.739	<0.0001	1.091	0.371	0.884	0.388	46.72
NETPRF_GRTH	0.023	0.712	0.692	0.170	3.173	<0.0001	1.129	0.965	0.823	0.802	43.09

tion of the short-list variables along presented models is market by "*x*" in corresponding variable column. The ranking of the best models has been based on independent validation sample GINI, since the validation sample has not been used in LR coefficient fitting. For the ranking purposes the train sample GINI has been also provided in Table 6. It is evident that train sample GINI for each model is higher than its GINI estimated on validation sample. This also holds for each of 342,016 well-founded LR models in this study.

Ideally the most desirable model will have the highest possible predictive power. The same highest GINI on validation sample has been estimated for models 8v_159569 and 7v_260497. In credit scoring practice, another important decision making rule is the predictive power of the model on both train and validation sample. The model 8v_159569, has a little bit higher train GINI, so based on that difference we can regard it as the final and most desirable credit scoring model in this study. If the results from Tables 4 and 6 are compared it is evident that the best models comprising of 5–14 variables from Table 6 show much higher predictive power than average predictive power of its corresponding variables groups from Table 4.

On the other side, when deciding about final model in practice, it is preferred to have fewer variables as long as this does not compromise the best fitting results. This decision making rule would make some practitioners prefer model 8v_159569 (AIC = 2439.034 and BIC = 2497.385) since fewer variables in a model means less concern of overfitting.

It is interesting to observe the slight overfitting effect on the results presented in Table 6. Overfitting is manifested in the following way. Up to models with 8 variables both train and validation GINI increase gradually, but afterwards up to 13v models the train prediction power continues to grow while on validation it starts to decline. The main reason for overfitting is that by increasing number of variables a highly flexible mathematical function is attained on the train sample. If we compare model 8v_159569 with model 14v_912530 it can be inferred that model 14v_912530 produces lower GINI=0.593 on validation and higher GINI = 0.678 on train sample. Model 14v_912530 has calculated AIC = 2422.85 and BIC = 2520.102 on train sample. According to these findings, overfitting on train sample becomes evident; BIC has lower value for 8v_159569 in comparison to 14v_912530. Additionally, the continuing increase in maximum variable's Wald statistics *p*-value, presented in the last column of Table 6, indicates that more than 8 variables in a model may lead to suboptimal model.

The performance of the final model 8v_159569 has been benchmarked on train and validation sample using ROC curve (Fig. 1.). Corresponding calculated AUCs are 0.835 and 0.811. The comparison in Fig. 1 shows better ROC curve performance on train than on validation sample.

According to above mentioned predictive power performance using ROC curve, the model 8v_159569 could be chosen to represent best credit scoring model in this research. Due to the fact that it has been built on real-life financial statements and default events data it can be employed and implemented by the bank as a corporate credit scoring model. The LR model parameter estimates are presented in Table 7, while Table 8 indicates correlation coefficients between financial ratio variables.

Final model financial ratios are good representatives of main business factors of corporate entities. Different ratio groups are captured such as: debt ratio (SHRTBDBT/CURAST), leverage ratio (ALIAB/AST), liquidity ratio (CSH/CEQTY), activity ratio (CURAST/NETSLS), debt repaying capability (EBTADJ/LIAB and GRSPRF/ALIAB), cash generating capacity ratios (FCFE/FINEXP) and net sales growth ratio (NETSLS_GRTH).

These findings show that ratios from most important business related factor groups have entered in the final model through application of a purely statistical approach. The Table 8 shows that all correlation coefficients are less than 0.50 which indicates acceptable, modest correlation persistence within the model. Only one correlation coefficient is not significant on 95% confidence level and it is not bolded in Table 8. The highest correlation of 0.488 is captured between GRSPRF/ALIAB and ALIAB/AST.

5. Conclusion

In our research we propose corporate entity credit scoring model capable of predicting probability of bankruptcy in 1 year period. Dataset in this research comprised of 5 years of financial

Table 6
An overview of the most predictive LR models gathered from the list of 342,016 well-founded models.

Variables number & simulation ID	SHRTBDBT/CURAST	CSH/CURLIAB	CSH/LTDBT	ALIAB/AST	ACCREC/ALIAB	CURAST/ALIAB	CURLIAB/ALIAB	CSH/CEQTY	TBDBT/CEQTY	LTDBT/EQT	NETSLS/FINEXP	EBITDA/FINEXP	EBTADJ/EBT	PAYPROCPRD	CSH/EBITDA	CURAST/NETSLS	CURLIAB/NETSLS	TBDBT/NETSLS	EBITDA/LIAB	EBTADJ/LIAB	GRSPRF/ALIAB	FCFE/FINEXP	NETSLS_GRTH	NETPRF_GRTH	GINI - train	GINI - valid.	Rank - GINI train	Rank - GINI valid.	Max. coefficient p-value of Wald test
5v_25117				**x**				**x**								**x**				**x**			**x**		**0.658**	**0.608**	**68**	**1**	**0.000**
5v_25120				x				x								x					x		x		0.641	0.605	799	2	0.000
5v_25113				x				x								x			x				x		0.650	0.604	278	3	0.000
5v_38058								x								x				x	x		x		0.624	0.601	3276	4	0.000
5v_25153				x				x										x		x			x		0.661	0.601	44	5	0.000
6v_91390				**x**				**x**								**x**				**x**	**x**		**x**		**0.659**	**0.619**	**804**	**1**	**0.003**
6v_90910				x				x				x				x				x			x		0.668	0.613	140	2	0.000
6v_91371				x				x								x		x		x			x		0.668	0.613	141	3	0.000
6v_91381				x				x								x			x	x			x		0.660	0.613	698	4	0.018
6v_16262	x			x				x								x				x			x		0.664	0.612	325	5	0.003
7v_260497				**x**				**x**								**x**		**x**		**x**	**x**		**x**		**0.668**	**0.621**	**1060**	**1**	**0.022**
7v_57741	x			x				x								x				x	x		x		0.665	0.620	1867	2	0.006
7v_260516				x				x								x				x	x	x	x		0.667	0.619	1360	3	0.014
7v_259706				x				x				x				x				x	x		x		0.668	0.618	1087	4	0.045
7v_260488				x				x								x		x	x	x			x		0.670	0.617	662	5	0.010
8v_159569	**x**			**x**				**x**								**x**				**x**	**x**	**x**	**x**		**0.671**	**0.621**	**1964**	**1**	**0.021**
8v_598025				x				x								x		x		x	x		x	x	0.668	0.620	3328	2	0.011
8v_159562	x			x				x								x			x	x		x	x		0.672	0.619	1525	3	0.042
8v_598016				x				x								x		x	x	x		x	x		0.675	0.619	820	4	0.034
8v_159571	x			x				x								x				x	x		x	x	0.666	0.618	4758	5	0.006
9v_1126665				**x**				**x**								**x**		**x**		**x**	**x**	**x**	**x**	**x**	**0.673**	**0.620**	**2365**	**1**	**0.030**
9v_352876	x			x				x								x				x	x	x	x	x	0.672	0.620	3199	2	0.014
9v_351922	x			x				x				x				x				x	x		x	x	0.671	0.619	3610	3	0.040
9v_1126516				x				x						x		x		x		x	x	x	x		0.677	0.618	1088	4	0.028
9v_1126663				x				x								x		x	x	x		x	x	x	0.676	0.617	1492	5	0.038
10v_635300	x			x				x				x		x		x				x	x		x	x	0.677	**0.617**	1337	1	0.018
10v_1765006				x				x						x		x		x		x	x	x	x	x	0.678	0.617	926	2	0.015
10v_1764409				x				x				x		x		x		x		x	x		x	x	0.678	0.617	1015	3	0.031
10v_636212	x			x				x						x		x				x	x	x	x	x	0.677	0.616	1223	4	0.005
10v_593762	x			x		x		x						x		x				x	x	x	x		0.677	0.616	1274	5	0.034
11v_947227	**x**			**x**				**x**				**x**		**x**		**x**				**x**	**x**	**x**	**x**	**x**	**0.680**	**0.616**	**416**	**1**	**0.034**
11v_2319902				x				x				x		x		x		x		x	x	x	x	x	0.681	0.616	304	2	0.043
11v_2272932				x		x		x						x		x		x		x	x	x	x	x	0.678	0.614	637	3	0.043
11v_946434	x			x				x			x			x		x			x		x	x	x	x	0.676	0.614	967	4	0.027
11v_897672	x			x		x		x						x		x				x	x	x	x	x	0.678	0.614	663	5	0.036
12v_2570541				**x**				**x**	**x**			**x**		**x**		**x**		**x**		**x**	**x**	**x**	**x**	**x**	**0.686**	**0.610**	**18**	**1**	**0.047**
12v_1128811	x			x		x		x						x	x	x				x	x	x	x	x	0.680	0.609	121	2	0.043
12v_796583	x		x	x				x	x					x		x				x	x	x	x	x	0.687	0.606	12	3	0.026
12v_1346380	x							x		x		x		x	x	x				x	x	x	x	x	0.669	0.606	532	4	0.033
12v_440306	x	x		x				x	x	x				x		x			x	x		x	x		0.687	0.606	14	5	0.048
13v_149935	**x**	**x**	**x**	**x**				**x**	**x**					**x**		**x**			**x**	**x**		**x**	**x**	**x**	**0.689**	**0.601**	**1**	**1**	**0.046**
13v_1327816	x					x		x		x		x		x	x	x				x	x	x	x	x	0.672	0.600	51	2	0.038
13v_1048722	x		x					x	x	x				x		x			x	x	x	x	x	x	0.677	0.598	23	3	0.049
13v_657442	x	x				x		x	x	x				x		x			x	x		x	x	x	0.684	0.598	7	4	0.034
13v_779787	x		x	x	x			x	x					x		x				x	x	x	x	x	0.688	0.598	3	5	0.049
14v_912530	**x**		**x**			**x**		**x**	**x**	**x**				**x**	**x**	**x**				**x**	**x**	**x**	**x**	**x**	**0.678**	**0.593**	**2**	**1**	**0.045**
14v_1112419	x				x			x	x	x		x		x	x	x				x	x	x	x	x	0.676	0.588	4	2	0.044
14v_880172	x		x		x			x	X	x		x		x		x				x	x	x	x	x	0.679	0.586	1	3	0.046
14v_880706	x		x		x			x	X	x				x	x	x				x	x	x	x	x	0.678	0.585	3	4	0.044

statements and performance data (default event data) dating from 2006 to 2011. The long list of 350 financial ratios has been initially constructed based on corporate financial statements and default event data. The research has taken into account 7590 corporate entities. The study used brute force LR to come up with the most predictive credit scoring model. The WoE data transformation technique has been applied in order to divide financial ratios into corresponding attributes and to eliminate problems connected to special values and outliers in financial ratios. A specialized grouping algorithm has been utilized in order to achieve the highest predictive power per financial ratio measured in information value (IV) terms.

Data partitioning of the whole population on train and validation sample has been performed using stratification by default rate over the years. We found that partitioning algorithm achieved good partitioning quality over variables and that default rate per year between train and validation sample was balanced.

The variable clustering results have been used to select the final short list of 24 variables. The association of each cluster with its containing variables has been measured with R squared ratio. We found that the clustering results were satisfying since 62.31% of total variability has been explained by 24 clusters. We have proposed the main representatives of each cluster by finding the financial ratio variable with highest prediction power within each

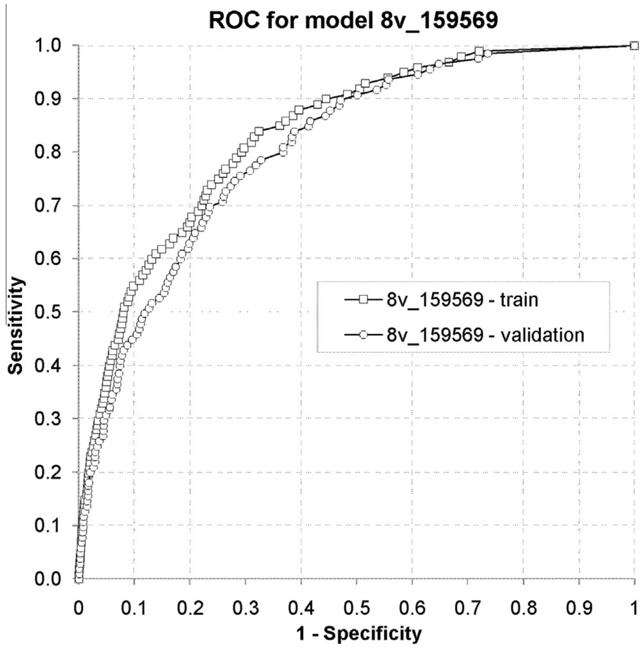

Fig. 1. Receiver operating characteristic curve – the comparison of 8v_159569 model on train and validation sample. *Source*: Authors' calculations.

Table 7
LR estimates from train sample: model 8v_159569.

Parameter	LR estimate	Standard error	Wald stat. *p*-value
Intercept	−2.1969	0.0582	<0.0001
SHRTBDBT/CURAST	−0.3407	0.1379	0.014
ALIAB/AST	−0.6152	0.0808	<0.0001
CSH/CEQTY	−0.7742	0.0972	<0.0001
CURAST/NETSLS	−0.5601	0.0861	<0.0001
EBTADJ/LIAB	−0.4714	0.0713	<0.0001
GRSPRF/ALIAB	−0.1997	0.0868	0.021
FCFE/FINEXP	−0.3964	0.1326	0.003
NETSLS_GRTH	−0.6452	0.0781	<0.0001

cluster. The most predictive variable in our research turned out to be EBTADJ/LIAB which achieved information value of 0.923. It comes from the group of ratios explaining corporate entity's repayment potential.

In order to check partitioning quality of train and validation sample in statistical terms we applied Kolmogorov–Smirnov (KS) test. No statistically significant differences were found in cumulative distribution function (*cdf*) between train and validation samples over 24 analyzed financial ratios. According to such result we concluded that the partitioning results were satisfying. Using the same KS statistics we have tested the short list variables ability to discriminate between good and bad corporate entities. The results have shown statistically significant ability to discriminate. Only one variable (TBDBT/CEQTY) has shown low and non-significant discrimination power, but we found that this variable comes from a low prediction power cluster.

In this paper a brute force logistic regression (LR) modeling approach has been utilized to find the best model. The transformed WoE values of financial ratios have been used as inputs for LR simulation fitting. The brute force LR simulation goal has been to obtain all possible combinations of models that comprise of 5–14 financial ratio variables from the short list of 24 financial ratios. The total number of resulting models that have been fitted through time consuming brute force LR simulations process is around 14 million. By predefined validity rules we rejected models that have shown significant multicollinearity and other unstable signs. Finally, we ended up with 342,016 correct and stable models, referred to as well-founded models. The validation sample GINI has been used to benchmark and find the best models comprising of 5–14 variables. By analyzing the "well-founded" models performance we concluded that validation GINI is lower than GINI on train dataset for each model.

Finally, we proposed the credit scoring model that comprises of eight variables that has shown the best prediction power performance among all well-founded models. In the last step of research we presented the LR estimates and found a modest correlation persistence structure within the final model variables. We have compared the ROC performance analysis of the proposed model on both train (AUC = 0.835) and validation (AUC = 0.811) sample. Our LR results based on WoE transformation of financial ratios according to AUC results has been in line with results of Altman and Sabato (2007) which compared original values LR results and LR logarithm transformed predictors results. Our final model AUC values are better when contrasted to results of Sohn and Kim (2012).

According to elaborated findings, quality of the obtained results and the fact that the credit scoring model is based on actual financial statements and default event data we conclude that the final model can be implemented as a deployable credit scoring model within a bank that operates in Serbia or in the region of South Eastern Europe.

However, the model developed in this study has its limitations. The default probability prediction of corporate entities in other economies using this model might be questionable because of different accounting treatments, corporate default rates and economic drivers of other countries. The model may be additionally improved in future research by supplementing it with macroeconomic indicators and non-financial information in order to higher its predictive power.

Table 8
Correlation coefficients structure in model 8v_159569.

Correlation coefficients*	SHRTBDBT/CURAST	ALIAB/AST	CSH/CEQTY	CURAST/NETSLS	EBTADJ/LIAB	GRSPRF/ALIAB	FCFE/FINEXP	NETSLS_GRTH
SHRTBDBT/CURAST	1							
ALIAB/AST	**0.257**	1						
CSH/CEQTY	**0.101**	**−0.069**	1					
CURAST/NETSLS	−0.028	**0.028**	**0.140**	1				
EBTADJ/LIAB	**0.208**	**0.397**	**0.273**	**0.190**	1			
GRSPRF/ALIAB	**0.222**	**0.488**	**0.134**	**0.248**	**0.470**	1		
FCFE/FINEXP	**0.183**	**0.287**	**0.045**	**0.067**	**0.235**	**0.323**	1	
NETSLS_GRTH	**0.054**	**0.054**	**0.091**	**0.243**	**0.184**	**0.108**	**0.065**	1

* Bolded value have significant correlation coefficients on a 5% significance level.

References

Akaike, H. (1974). Information theory and an extension of the maximum likelihood principle. In *Second interantional symposium on information theory* (pp. 267–281).

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.

Altman, E., Haldeman, R., & Narayanan, P. (1977). ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance, 1*(1), 29–54.

Anderson, R. (2007). *The credit scoring toolkit*. New York: Oxford University Press.

ASB. (2013). *Association of Serbian Banks*. Retrieved from <http://www.ubs-asb.com/>.

Aziz, A., Emanuel, D. C., & Lawson, G. H. (2007). An investigation of cash flow based models. *Journal of Management Studies, 25*(5), 419–437.

Banasik, J., Crook, J., & Thomas, L. (2003). Sample selection bias in credit scoring models. *Journal of the Operational Research Society, 54*(8), 822–832.

BCBS. (2005, May). Studies on the validation of internal rating systems. Basel Committee on Banking Supervision.

BCBS. (2006). International convergence of capital measurement and capital standards – A revised framework. Basel Committee on Banking Supervision.

Beaver, W. (1967). Financial ratios predictors of failure. *Journal of Accounting Research, 4*, 3.

Becchetti, L., & Sierra, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking and Finance, 27*(11), 2099–2120.

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

Blanco, A., Pino-Mejias, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications, 40*, 356–364.

Blochlinger, A., Leippold, M. (n.d.). Economic benefit of powerful credit scoring. *Journal of Banking and Finance* 30(3), 851–873.

CFA Institute. (2012). Financial Reporting and Analysis (CFA® Program Curriculum) (Vol. 2). Pearson.

Chen, M.-Y. (2011). Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications, 38*, 11261–11272.

Cramer, J. S. (2004). Scoring bank loans that may go wrong: A case study. *Statistica Neerlandica, 58*(3), 365–380.

Crone, S. F., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting, 28*, 224–238.

Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*(3), 1447–1465.

Derelioglu, G., & Gurgen, F. (2011). Knowledge discovery using neural approach for SME's credit risk analysis problem in Turkey. *Expert Systems with Applications, 38*, 9313–9318.

Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failure with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research, 90*(3), 487–513.

Dwyer, D. (2005). Examples of overfitting encountered when building private firm default prediction models. Moody's KMV, New York, www.moody's. com/whitepapers.

Fantazzini, D., & Figini, S. (2009). Random survival forests model for SME credit risk measurement. In *Methodology and computing in applied probability*, Springer.

Fidrmuc, J., & Hainz, C. (2010). Default rates in the loan market for SMEs: Evidence from Slovakia. *Economic Systems, 34*(2), 133–147.

Gentry, J., Newbold, P., & Whitford, D. (1985). Classifying bankrupt firms with funds flow components. *Journal of Accounting Research, 23*(1), 146–160.

Gupta, J., Wilson, N., Gregoriou, A., Healy, J. (2012). The Value of Operating Cash Flow in Modelling Credit Risk for SMEs. Available at SSRN 2180313.

Gupta, M. C., & Huefner, R. J. (1972). A cluster analysis study of financial ratios and industry characteristics. *Journal of Accounting Research*, 77–95.

Hand, D., & Niall, A. (2000). Defining attributes for scorecard construction in credit scoring. *Journal of Applied Statistics, 27*(5), 527–540.

Hermanto, B., Gunawidjaja, S. (2010). Default prediction model for SME's: Evidence from indonesian market using financial ratios. SSRN 1666703.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.

Jain, K., Gupta, P., & Sanjiv, M. (2011). Logistic predictive model for SMEs financing in India. *The Journal of Business Perspective, 15*(4), 331–346.

Johnsen, T., & Melicher, R. (1994). Predicting corporate bankruptcy and financial distress: Information value added by multinominal logit models. *Journal of Economics and Business, 46*(4), 269–286.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics, 29*(2), 119–127.

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.

Kim, K.-j., & Ahn, H. (2012). A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach. *Computers and Operations Research, 39*, 1800–1811.

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

Laitinen, E. K., & Laitinen, T. (2000). Bankruptcy prediction application of the Taylor's expansion in logistic regression. *International Review of Financial Analysis, 9*, 327–349.

Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid neural network models for bankruptcy predictions. *Decision Support Systems, 18*(1), 63–72.

Leshno, M., & Spector, Y. (1996). Neural network prediction analysis: The bankruptcy case. *Neurocomputing, 10*, 125–147.

Lin, F., & McClean, S. (2001). A datamining approach to the prediction of corporate failure. *Knowledge Based Systems, 14*, 189–195.

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.

Min, J. H., & Jeong, C. (2009). A binary classification method for bankruptcy prediction. *Expert Systems with Applications, 36*, 5256–5263.

Moon, T., & Sohn, S. (2010). Technology credit scoring model considering both SME characteristics and economic conditions: The Korean case. *Journal of the Operational Research Society*, 666–675.

Ohlson, J. (1980). Financial ratios and the probabilistic of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131.

Peduzzi, P., Concato, J., Kemper, E., Holford, T., & Feinstein, A. (1996). A simulation of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373–1379.

Plattner, D. (2002). Why firms go bankrupt. The in factors on the insolvency probability of small and medium sized enterprises. *KfW Research, 28*, 37–51.

Pluto, K., & Tasche, D. (2010). Internal-ratings-based approach. *Encyclopedia of Quantitative Finance*.

Ruthenberg, D., & Landskroner, Y. (2008). Loan pricing under Basel II in an imperfectly competitive banking market. *Journal of Banking and Finance, 32*(12), 2725–2733.

SAS Institute Inc. (2011). SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.

Siddiqi, N. (2005). *Credit risk scorecards: Developing and implementing intelligent credit scoring*. New York: John Wiley & Sons, Inc.

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 (pp. 1–13).

Sustersic, M., Mramor, D., & Zupan, J. (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications, 36*, 4736–4744.

Tasche, D. (2006). Validation of internal rating systems and PD estimates. *The Analytics of Risk Model Validation*, 169–196.

Thomas, L. C. (2009). *Consumer credit models: Pricing, profit, and portfolios*. New York: Oxford University Press Inc.

TrimTab Solutions. (2012). ScoreStar® – Credit Scoring Software. <http://www.trimtab.rs>.

Vukovic, S., Delibasic, B., Uzelac, A., & Suknovic, M. (2012). A case-based reasoning model that uses preference theory functions for credit scoring. *Expert Systems with Applications, 39*, 8389–8395.

Wang, Y.-J., & Lee, H.-S. (2008). A clustering method to identify representative financial ratios. *Information Sciences, 178*(4), 1087–1097.

Westgaard, S., & Wijst, N. v. (2001). Default probabilities in a corporate bank portfolio: A logistic model approach. *European Journal of Operational Research, 135*, 338–349.

Yazdanfar, D. (2011). Predicting bankruptcy among SMEs: Evidence from Swedish firm-level data. *International Journal of Entrepreneurship and Small Business, 14*(4), 551–565.

Zekic-Susac, M., Sarlija, N., & Bensic, M. (2004). Small business credit scoring: A comparison of logistic regression, neural networks, and decision tree models. In *26th International conference on information technology interfaces*. Croatia.