

Predicting the Bankruptcy of Construction Companies: A CART-Based Model

Michal Karas, Maria Reznakova

Brno University of Technology
Kolejni 4, 612 00 Brno, Czech Republic
E-mail. karas@fbm.vutb.cz, reznakova@fbm.vutbr.cz

crossref <http://dx.doi.org/10.5755/j01.ee.28.2.16353>

Company bankruptcy is a frequent research topic, and with regard to current economic developments, its importance and relevance is undisputable. Models created in different conditions do not achieve the accuracy claimed by their authors. The aim of this article is to present the results of research associated with the development of a bankruptcy model designed for construction companies. Due to the properties of financial data, the quality of the model, i.e. its discrimination ability, is strongly influenced by the choice of the method used to derive the model. The model was developed for the years 2011–2014 based on 29 financial indicators of companies operating in the construction industry, which were calculated on the basis of accounting data of the companies contained in the AMADEUS database. A non-parametric method of Classification and Regression Trees (CART) was used to derive this model. The discrimination ability of the model was evaluated based on the percentage of correctly differentiated companies and the percentage of Type I and Type II errors. To test the discrimination accuracy of models, the Receiver Operating Characteristic curve (ROC curve) and the Area Under Curve (AUC) were also used. Based on these tests, it is possible to graphically and numerically measure the discrimination ability of the models. The discrimination accuracy of the model created on the basis of the data of construction companies was compared with other selected models, which were not created using the data of construction companies. The comparison clearly showed that the model created especially for the construction companies achieves the highest discrimination ability.

Keywords: *Financial Ratios, Bankruptcy Prediction, Classification and Regression Trees, ROC Curve, Construction.*

Introduction

Although, in specialist literature, we can find a number of studies on the construction of bankruptcy models; most of the models, however, were derived either from the data of companies from the manufacturing industry, or from companies from different industries, i.e. not taking into account the field of business. As demonstrated by our previous research, the accuracy of models significantly decreases if they are used in a different environment (or a different industry). Other authors also reached similar conclusions, e.g. Platt and Platt (1990), Grice and Dugan (2001), Niemann *et al.*, (2008) and Wu, Gaunt and Gray (2010). Heo, Yang (2014), Thomas Ng, Wong and Zhang (2011) point out that the existing models are unsuitable for predicting bankruptcy in the construction industry, and new models have to be created for this sector. According to the overview provided by Marzal-Martinez, Barrachina-Martinez and De la Poza-Plaza (2014), the construction of bankruptcy models for the construction industry is a very rare phenomenon.

Lee, Choi (2013) compared the accuracy of their model based exclusively on the data from construction companies with a similar model, but based on the data of companies from different industries. The model especially designed for construction companies reached a 6–12 % higher classification accuracy compared to the model created on the data of companies from different industries. The authors believe that the accuracy of the model would be even higher if predictors specific for the construction industry were used.

They believe that new specific models should be formulated for the construction industry.

The construction industry is a sector very sensitive to the economic development of the country or - in the conditions of globalization - to the global economic development (see e.g. Carling *et al.*, 2007; Gertler, 2015). It was clearly shown in the period after 2008, when, e.g., construction output in the Czech Republic decreased by 8.3 % by 2013 and the number of jobs fell by 6.1 %. In the same period, return on invested capital fell by 7.8 percentage points as measured by ROE, and by 4 percentage points as measured by ROA. The following years were associated with a slight increase in this sector. One of the main reasons for this development was the decline in the number and volume of public projects due to the high indebtedness of the public sector and restrictions of bank financing, which was - among other things - the result of the deteriorating financial situation of construction companies. In the monitored period, the value of capital tied up in net working capital (NWC) in relation to sales increased about 2.5 times (measured by the 'net working capital-to-sales' ratio) or by more than one third, if measured by the NWC-to-total assets ratio. This development was caused by the drop in sales, but also by an increase in overdue receivables. The consequence was an increase in the risk of secondary insolvency, i.e. the risk of inability to pay one's liabilities, which can result even in the bankruptcy of companies. In the monitored period (2008–2013), the number of bankruptcies in the Czech Republic increased by about 50 %; and a further increase came in 2014, while the year-on-year growth was the highest: the year-on-year increase in the number of company bankruptcies was 58 %, with construction companies accounting for almost 20 % of the increase. One

can assume that the long-term stagnation of demand or its decline is fatal for some businesses. These facts clearly show that the question of bankruptcy prediction for the construction industry is highly topical.

According to Heo, Yang (2014), the typical characteristic for the construction industry is an uneven cash flow from projects that is usually concentrated in the projects' later stages. This leads to the situation that this sector is characterized by high levels of liquidity and high indebtedness. With regard to the specifics of this industry, Sun, Liao, Li (2013) say: "*The construction industry is a capital-intensive industry that requires long-term project periods and huge investments, and takes a long time to receive returns from the investment. Therefore, it has a different capital structure from other industries, and the same criteria used for other industries cannot be applied to effectively evaluate its financial risk*" (see Sun, Liao, Li, 2013 in: Heo, Yang, 2014).

This conclusion is confirmed by another study by Barrie, Paulson (1992): "*Due to the distinctive operational behaviours of the construction industry, its financial characteristics also differ from other industries*" (see Barrie, Paulson 1992 in: H. P. Tseng *et al.*, 2012).

Bankruptcy prediction in construction companies in the Czech Republic was discussed by Kubenka, Kralova (2013) and Spicka (2013). Spicka (2013) concluded that typical bankruptcy manifestations in construction companies in the Czech Republic included high indebtedness due in particular to current liabilities, low labour productivity and a negative return on assets.

The above reasons have led us to create a bankruptcy prediction model based on the data from construction companies in the Czech Republic. The fundamental step in creating a model is the formulation of a decision-making rule on the basis of which it is possible - with the highest possible accuracy - to distinguish between companies threatened by bankruptcy and those that are financially stable, i.e. companies able to continue to pay their debts (referred to as "active companies" in the paper). The discrimination ability of the model will be tested on a sample of data of the companies that were not used in the creation of the model, and at the same time will be compared with the discrimination ability of a number of selected models.

The aim of this article is to present the results of our research and the construction of the new bankruptcy model designed for construction companies in the Czech Republic. We proceeded from the following assumptions:

H1: The models not designed specifically for construction companies are not efficient in predicting bankruptcy of construction companies.

H2: A new model designed especially for construction companies outperforms the other models not developed specifically for construction companies.

H3: Creating a more comprehensive version of the new model will increase its efficiency.

Methods Used to Create Bankruptcy Models

When creating a bankruptcy model, it is necessary to also pay attention to the choice of a method, because that predetermines to a large extent its discrimination ability.

Methods used to create bankruptcy models can be divided into statistical methods and methods of artificial intelligence.

Statistical methods include, e.g., a discriminant analysis (Altman, 1968), a logit model (Martin, 1977; Ohlson, 1980), Cox's model (Henerby, 1996; Schumway, 2001), etc. The discriminant analysis followed by the logit model are the two most frequently used method for the creation of bankruptcy models. These methods, however, are limited by the assumptions on which they are based. Chuang, Lin (2009) state that the methods "... are designed for the relationships between variables that are linear, which caused them less accurate in credit scoring." Also, the discriminant analysis method is sensitive to deviations from the multivariate normality assumption (West, 2000).

The methods of artificial intelligence include, e.g., artificial neural networks (Back, Laitinen, Sere, 1996; Kim, Kang, 2010), Data Envelopment Analysis (DEA) methods (Cielen, Peeters, Vanhoof, 2004; Ding, Song, Zen, 2008), Classification and Regression Trees (CART), see Li, Sun, Wu (2010), etc. According to Brezigar-Masten, Masten (2012), the common feature of these methods is "*a fully non-parametric specification of both the distributional form of variables and functional relations among them.*" The importance of non-parametric methods for predicting bankruptcy is also highlighted in the study by De Andres, Lorca, de Cos Juez, Sanchez-Lasheras (2011) due to the specific properties of financial data, such as non-normality and heteroscedasticity.

Sample and Method Used

The data were obtained from AMADEUS (Analysis Major Database for European Sources). The bankrupt companies in our sample declared bankruptcy in 2011 and 2014. The examined sector is the construction industry. Only small and medium-sized construction companies were included in the sample. The selection criterion was the size of their asset, i.e. companies studied had in at least one of the monitored periods assets between 2 and 50 million EUR. This criterion was met by 1,257 active companies and 283 bankrupt companies, and they made up the original sample. For comparison, Eurostat data indicate that there were 170,806 companies operating in the Czech Republic in 2014, but only 4,795 of them with more than 9 employees. To analyze the data, we applied 29 financial ratios used to evaluate the financial situation of the companies. Because the database included companies that did not publish complete financial statements, the selection of observations was further limited, and we focused on companies whose financial statements contained at least the data necessary for the calculation of selected ratios. The model was thus based on data from 630 active companies and 24 bankrupt companies.

Investigated Ratios

The following ratios are often used in studies on bankruptcy prediction (see Altman, 1968, Tian *et al.*, 2015,

Gordini, 2014, Laitinen *et al.*, 2014, Kwak *et al.*, 2014, Banyiova *et al.*, 2014, Faltus, 2014, Carling *et al.*, 2007, Karas, Reznakova, 2013, Cut, 2014).

The List of Investigated Ratios

No.	Ratio	Abbreviation	Type	No.	Ratio	Abbreviation	Type
1.	Current ratio	CR	L	16.	Sales/Debtors	S/Deb.	T
2.	Working capital/total assets	WC/TA	L	17.	Quick assets/sales	QA/S	T
3.	Working capital/sales	WC/S	L	18.	Current liabilities/total assets	CL/TA	I
4.	EBIT/total assets	EBIT/TA	P	19.	Long-term liabilities/total assets	LTL/TA	I
5.	EBITDA/total assets	EBITDA/TA	P	20.	Debt-equity ratio	DER	I
6.	EAT/equity	ROE	P	21.	EBIT/Interest	EBIT/Int.	I
7.	Cash flow/total assets	CF/TA	P	22.	EBITDA/Interest	EBITDA/Int.	I
8.	Cash flow/sales	CF/S	P	23.	EBIT 5y volatility	EBIT (5-vol)	SF
9.	Cash flow/total liabilities	CF/TL	P	24.	logarithm of total assets	LogTA	SF
10.	EAT/total assets	EAT/TA	P	25.	logarithm of sales	LogS	SF
11.	EBIT/Sales	EBIT/S	P	26.	Fixed assets/total assets	FA/TA	SR
12.	EBITDA/Sales	EBITDA/S	P	27.	Sales/Operating revenue	S/OR	SR
13.	Retained Earnings/total assets	RE/TA	P	28.	Added Value/Sales	AD/S	SR
14.	Sales/total assets	S/TA	T	29.	Cost of employees	CE/S	SR
15.	Sales/Stocks	S/St.	T				

Note to type of ratios: P – profitability, L – liquidity, T- turnover, I – indebtedness, SF -size factors, SR-structural ratios.

Source: Altman (1968), Tian et al. (2015), Gordini (2014), Laitinen et al. (2014), Kwak et al. (2014), Banyiova et al. (2014), Fatuus (2014), Carling et al. (2007), Karas, Reznakova (2013), Cut (2014).

The research methodology is influenced by the properties of the data that are the subject of examination. The following table provides an example of a descriptive characteristic of the examined sample (of all companies).

Since the number of variables tested was very high, three ratios were selected to demonstrate the sample properties. Values of ratios for active companies are designated with (A), for bankrupt companies with (B).

Table 2

The Descriptive Statistics of the Sample – All Companies

	Mean	Median	Grubbs' Test-stat.	p-val.	Min.	Max.	Std. Dev.	Skew.	Kurt.
EBIT/TA (A)	0.0358	0.0236	13.31	0.000000	-1.40	0.70	0.11	-2.62	44.86
EBIT/TA (B)	-0.3670	-0.0014	6.58	0.000000	-8.50	0.50	1.24	-4.86	26.39
S/TA (A)	1.0325	1.0059	8.06	0.000000	0.00	8.40	0.91	1.27	4.34
S/TA (B)	1.6560	1.2936	4.04	0.002126	-0.60	9.24	1.88	1.79	3.77
WC/TA (A)	0.2039	0.2097	9.37	0.000000	-3.40	1.00	0.38	-1.33	7.17
WC/TA (B)	-107.5750	-0.1199	9.53	0.000000	-9420.00	0.48	977.59	-9.60	92.44

According to the Grubbs' test, almost every ratio includes an outlier. The sample exhibits kurtosis, which clearly indicates a non-normal distribution of data.

Grubbs' Test

Grubbs' test (see Grubbs, 1969) tests the null hypothesis, i.e. that the examined sample does not contain any outliers. The Grubbs' test statistic is the largest absolute deviation from the sample mean in units of the sample standard deviation (see Grubbs, 1969):

$$G = \frac{\max_{i=1,2,\dots,n} |Y_i - \bar{Y}|}{s}$$

where

Y_i –value of individual ratios tested, \bar{Y} - mean value of ratios tested, s – standard deviation.

Classification and Regression Trees (CART)

For the creation of the model, we chose a non-parametric method of Classification and Regression Trees (CART). This method was first used for the prediction of bankruptcy by Frydman et al. (1985). It is suitable for the selection of predictors (see, e.g., Brezinger-Masten, Masten, 2012), as well as for the formulation of a decision-making

rule (see, e.g., Gepp, Kumar, 2015, Liang, Lu, Tsai, Shih, 2016). The main advantages of CARTs include:

- 1) The resulting classification rule is easy to interpret (Brezinger-Masten, Masten, 2012);
- 2) This is a non-parametric method, which is also able to capture complex relationships between the variables (Brezinger-Masten, Masten, 2012);
- 3) The method is very robust with regard to the existence of outliers in the sample (Di Marco, Nieddu, 2014).

When creating a model, the objective is to find a decision-making rule that will be able to split - with the highest possible accuracy - the studied group of companies into those that are threatened by bankruptcy and those that are financially sound. Another requirement is that the final rule should also be generalizable to companies outside the studied sample. The consequence of an incorrect splitting of the group is either a Type I error (i.e. a bankrupt company is included among active companies), or a Type II error (an active company is included among bankrupt companies). According to Zhou, Elhag (2007), the Type I error is 2 to 20 times more serious (i.e. costly) than the Type II error. In using the CART method, it is possible to attribute different weights to the Type I and II errors. This characteristic was

also used in the creation of the model for the construction industry. The weight, i.e. the importance of the Type I error, was set twice as high than that of the Type II error.

When applying the CART method, a frequently discussed topic is a stability of the created tree, i.e. the dependence of its structure on the studied sample. It is common that the created model is able to very correctly define companies in the sample of the companies that were used to create the model; but outside the learning sample, the accuracy is significantly lower. We solved this problem by using k-fold cross validation with $k = 10$, similarly as Liang, Lu, Tsai, Shih (2016). In this approach, the studied sample is split into ten learning and test sets, which are used to train and test the model, and the whole process is repeated ten times. The final rule is derived as an average of all ten results obtained.

Principle of the CART Method

CART analysis is applied, *inter alia*, to classification problems where a dependent variable Y is to be described using two independent variables X_1 and X_2 . The basic idea behind the Trees is the repeated splitting of a complex problem of feature space into two smaller sub-spaces known as regions (R), which can then be described through simpler models (for example constants). For a two-dimensional classification problem, it is possible to describe the approach using the following diagrams (see Figs 1 and 2). The diagrams document the subdivision of two-dimensional feature space into regions using the constant t .

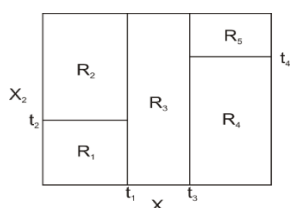


Figure 1. The division of two-dimensional space into regions
Source: Authors' own elaboration based on (Hastie et al., 2009, p. 306)

Alternatively, the same division can be shown using trees, as in the following diagram.

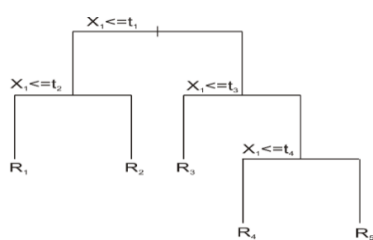


Figure 2. Division of two-dimensional space into regions using trees
Source: Authors' own elaboration based on (Hastie et al, 2009., p. 306)

The central problem of the method of using trees is establishing the optimal divisional boundaries t between regions R . The boundaries are established in such a way that the demarcated regions, or the trees, satisfy specific defined criterion.

This property of the regions, or the trees, is defined as a *node impurity* and the aim of the method is its minimization. For classification purposes, where the output can attain the value $1, 2, \dots, K$, it is possible to describe *node impurity* in the following way, see (Hastie et al, 2009, p. 306):

In the m -th node, representing the m -th region R_m with N_m , the number observed is a proportion of the group k in the node m :

$$\hat{p}_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k)$$

where

y_i - is the i -th observation of dependent variable Y .

It is then necessary to define the majority of observed elements of the k -th group in the node m as:

$$k(m) = \arg \max_k \hat{p}_{mk}$$

Node impurity of the tree T or $Q_m(T)$ can be defined using several standards, e.g., according to the misclassification error, Gini index or cross-entropy (or deviance). Cross-entropy as a level of node impurity was used here as part of the presented research. Cross-entropy can be expressed as:

$$-\sum_{k=1}^K \hat{p}_{mk} \log \hat{p}_{mk}$$

Selected Models Used for Comparison

To compare the differentiating accuracy of our model, four previously created models were used. Historically, the first multi-dimensional model was the Altman model from 1968 (see Altman, 1968), which was later revised to suit companies not listed on the stock exchange. The model was based on a dataset from manufacturing companies. The revised model has the following form (see Altman, 2000):

$$Zscore = 0.717 \cdot WC/TA + 0.847 \cdot RE/TA + 3.107 \cdot EBIT/TA + 0.42 \cdot E/D + 0.998 \cdot S/TA$$

where

WC/TA is the net working capital/total assets ratio, RE/TA is the retained earnings/total assets ratio, $EBIT/TA$ is the EBIT (Earnings before Interest and Taxes)/total assets ratio, E/D is the book value of equity/total debt ratio, and S/TA is the sales/total assets ratio.

The grey zone of the model is in interval $<1.23; 2.9>$. A company with the Z-score greater than 2.9 is classified as active, and those with the Z-score lower than 1.23 are classified as threatened by bankruptcy.

According to Mandru et al (2010), Li and Ragozar (2012) or Satish and Janakiram (2011), Altman model is still robust and gives good results, and it was therefore used for the comparison with the model created by us.

The above model has been revised by many authors in an effort to better match it to the environment of their respective countries. The model by Sorins and Voronova of 1997 (see Camska, 2016) can serve as an example. The model uses the same variables as its original version, with the exception of the variable EBIT/TA, in which EBIT (earnings before interest and taxes) is replaced by EBT (earnings before taxes). The coefficients of the model and the gray zone boundaries were calculated on the basis of Latvian retail food companies:

$$Z(modif) = 2.5 \cdot WC/TA + 3.5 \cdot RE/TA + 4.4 \cdot EBT/TA + 0.4 \cdot E/D + 0.7 \cdot S/TA - 2.4$$

Compared to the original version, the modified model does not include the gray zone. If $Z(modif) < 0$, the company is under high risk of bankruptcy, if $Z(modif) = 0$, the risk of bankruptcy is moderate, and for $Z(modif) > 0$, the probability of bankruptcy is low.

Another tested model is the Fulmer H-score. The authors developed the model for the evaluation of small and

medium-sized companies in response to the fact that those companies go bankrupt more often (e.g. Ding et al, 2008, Niemann et al, 2008, Psillaki, Tsolas, Margaritis, 2009). Fulmer H-score can be written as follows (see Fulmer et al, 1984, Kalupa, 2001):

$$H = 5.528 \cdot RE/TA + 0.212 \cdot S/TA + 0.073 \cdot EBT/RE + 1.270 \cdot (EAT+dep.)/D - 0.120 \cdot D/A + 2.335 \cdot CL/A + 0.575 \cdot \log(TFA) + 1.083 \cdot WC/D + 0.894 \cdot \log(EBIT/Int) - 6.075$$

where

EBT/RE - earnings before taxes/retained earnings ratio, *(EAT+dep.)/D* - sum of earnings after taxes plus depreciation and amortization/total debt ratio, *D/A* - total debt/assets ratio, *CL/A* - current liabilities/assets ratio, *log(TFA)* - the decimal logarithm of the value of tangible fixed assets, *WC/D* - net working capital/total debt ratio, *log(EBIT/Int)* - the decimal logarithm of the earnings before taxes and interest expense/interest expense ratio.

If $H < 0$ the company is threatened by bankruptcy, if $H > 0$ the company is classified as active (financially sound).

Another model compared with the CART analysis is the bankruptcy index developed for manufacturing companies in the Czech Republic on the basis of data from the years 2008 - 2010. Its structure is a combination of a linear discriminant analysis and Box-Cox transformation of variables (see Box, Cox, 1964). The tested form of the model is as follows (see Karas, Reznakova, 2013):

$$IB = -11.8356 \cdot (S/TA + 0.9306)^{-0.4949} + 9.9934 \cdot (QA/S + 1.1965)^{-1.4560} + 10.9205 \cdot TA^{0.0765}$$

where

S/TA - a turnover of total assets (sales/total assets ratio), *QA/S* - the so-called quick assets (current assets minus inventories)/sales ratio, *TA* - total assets [in CZK]

The model classifies a company as bankrupt if $IB < 23.826$; otherwise it is classified as active. It was used for comparison because it was created in the Czech Republic on the basis of recent data. It should therefore correctly discriminate between active and bankrupt companies. If not, our assumption that models created in other industries have lower discrimination ability when used in the construction industry will be confirmed.

Results and Discussion

In our research, we developed a model that was based on only three variables. It will be referred to as the basic version or Model 1. Since the number of variables was low compared to previously published models of other authors, we developed a more comprehensive version of Model 1 where one node is further branched. That version will be referred to as Model 2.

Model 1

Model 1 is a tree with three non-terminal nodes and four terminal nodes. The model contains only three variables, namely the current liabilities/total assets (CL/TA) ratio, the quick assets/sales (QA/S) ratio, and the factor of the company size in the form of the log of total assets (LogTA). Model 1 can be described by the following diagram (Figure 3).

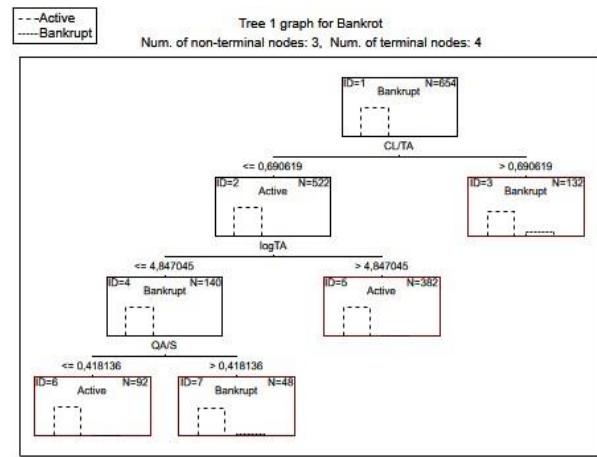


Figure 3. CART Model 1

The diagram clearly illustrates the decision-making rule: the first node (ID=1) evaluates current liabilities. Any value under 69.06% (see node 2) is considered safe. Higher values of current liabilities are classified by the model as risky, i.e. the company is classified as being threatened by bankruptcy. If the values are lower, the company is classified as active only subject to the fulfilment of two additional conditions:

- a) The decimal logarithm of total assets is less than 4.847045 (the logTA indicator), which corresponds to an approximate value of assets less than EUR 2.5 mil;
- b) The quick assets-to-sales ratio is greater than 41.81% (the QA/S ratio).

The CART analysis makes it possible to split created nodes and thus to create more comprehensive models. Only nodes that contain a sufficient number of observations are split. In our case, it was node 3. Model 2 was created by splitting node 3.

Model 2

The right branch of the tree is limited to the boundary value of a single ratio ($CL/TA > 0.6906$), and 132 companies were assessed by that rule, i.e. almost all bankrupt companies (23 of 24) and a considerable number of active observations (111 of 630). In order to enhance model accuracy, node 3 was further split and another branch was formed. Successive divisions of node 3 allowed for an inclusion of another two variables, and the number of terminal nodes increased by 3. Model 2 is shown in the following diagram (Figure 4).

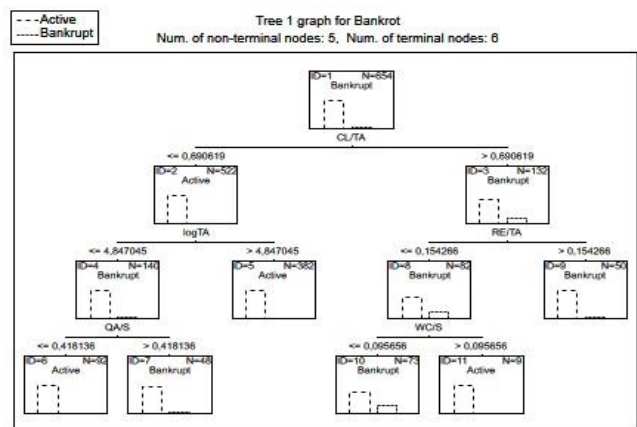


Figure 4. CART Model 2

Companies are evaluated as bankrupt if they meet the same conditions as those in Model 1, but if their current liabilities are greater than 69.06 %, at least one of the two conditions below need to be met if a company is to be evaluated as bankrupt:

a) the retained earnings/assets (RE/TA) ratio is greater than 15.42 %;

b) the retained earnings/assets (RE/TA) ratio is less than 15.42 %, and, at the same time, the net working capital/sales (WC/S) ratio is less or equal to 9.56 %.

This means that even in the case of high current liabilities ($CL/TA > 0.6906$), the model evaluates the company as active if it has a sufficiently high net working capital ($WC/S > 0.095656$), i.e. it is a company capable of generating sales and has potential revenues tied up in outstanding receivables.

According to Spicka (2013), one of the typical characteristics of bankrupt construction companies is their extreme debt ratio (sometimes exceeding 100%), the problem being mainly in current liabilities. Our study has corroborated his conclusions.

Model 2 evaluates higher RE/TA as a high risk factor, which is rather surprising. A possible explanation is that overindebted companies that are at the same time profitable retain their earnings and do not pay dividends. After some time, the amount of the generated profit is no longer sufficient for further operation of the company and the company is threatened by bankruptcy or goes bankrupt. This hypothesis is supported also by the fact that none of the tested profitability ratios is contained in the model, which implies that profitability indicators are assessed as insignificant.

Both models agree on the importance of company size, where bigger companies are evaluated as not threatened by bankruptcy. From the information point of view, this introduces the market position aspect into the models (see Ding et al, 2008, Niemann et al, 2008, Psillaki, Tsolas,

Margaritis, 2009). Shumway (2001) considers the company size factor as a significant predictor of bankruptcy, but he derives that indicator from market data. Wu, Gaunt, Gray (2010) add that bigger firms are considered more capable of surviving tough economic times and less prone to bankruptcy. This assumption is generally associated with the larger companies, which are perceived by their surroundings as a more stable business partner.

Testing of the Model and Comparing its Accuracy with Previously Created Models

Both versions of our model were tested both in the sample used in creating the model, and outside it. Results of the testing of the CART models (Model 1 and Model 2) are given in Table 3. The models were based on a sample of companies whose data made it possible to quantify all the 29 tested variables (“learning sample”). Since the models contain only 3 and 5 variables, respectively, it was possible to test them also on those companies in the sample from which we did not have complete accounting data, but where it was possible to calculate the ratios used in the models (“test sample”).

The accuracy with which the model is able to correctly identify a company as active or bankrupt was calculated by the following formulas:

$$\frac{A(OK)_T}{A(Valid)_T} \text{ (for Type I errors)} \quad \frac{B(OK)_T}{B(Valid)_T} \text{ (for Type II errors)}$$

Error rate of models (Type I error or Type II error) was calculated as follows:

$$\frac{A(E)_T}{A(Valid)_T} \text{ (for Type I errors)} \quad \frac{B(E)_T}{B(Valid)_T} \text{ (for Type II errors)}$$

Where A or B - number of observations of active or bankrupt companies; OK - correctly identified observations; E - wrongly identified observations; $Valid$ - valid observation; T - number of years before bankruptcy.

Table 3

Accuracy of the CART Models

CART	Model 1				Model 2			
	Learn		Test		Learn		Test	
	Active	Bankrupt	Active	Bankrupt	Active	Bankrupt	Active	Bankrupt
Valid	630	24	511	59	630	24	511	59
Correctly (OK)	474	24	312	56	490	23	320	54
Accuracy	75.24 %	100.00 %	61.06 %	94.92 %	77.78 %	95.83 %	62.62 %	91.53 %

Both versions of the model achieve relatively high accuracy on a sample of bankrupt companies, but less on a sample of active companies. There can be two reasons for it: It can be either a consequence of the fact that the number of non-bankrupt companies in the set (but also in the economy) was much higher. However, it could be also affected by setting up a higher weight to Type I error when creating the model.

The accuracy with which Model 1 can recognize a bankrupt company in the sample intended for deriving the model and for its testing is 75.24 and 61.06 %, respectively. A similar accuracy in identifying bankrupt companies is 100 % in companies from the learning sample and 94.92 % in those from the test sample.

The accuracy of Model 2 on the learning sample and the test sample was 77.78 % and 62.62 %, respectively.

A similar accuracy regarding the bankrupt companies was 95.83 % (learning sample) and 91.53 % (test sample).

Despite the small number of variables, Model 1 achieves slightly better results than Model 2.

Other models were tested for comparison purposes. To achieve maximum credibility for the comparison, a sub-sample of companies was chosen that met the following criteria:

The data of the companies were not used to derive the CART models.

The data contain at least those variables that are included in the tested models.

These criteria were met by 156 companies, of which 143 were active and 13 bankrupt.

Table 4

Accuracy of all Tested Models

Model	Active	Bankrupt	Grey zone (A)	Grey zone (B)	Type I error	Type II error
CART model 1	57.34 %	92.31 %	0 %	0 %	7.7 %	42.7 %
CART model 2	58.74 %	92.31 %	0 %	0 %	7.7 %	41.3 %
Z score	16.50 %	28.60 %	62.00 %	64.30 %	29 %	21 %
Z (modif)	62.30 %	67.90 %	0 %	0 %	32.10 %	37.70 %
Fulmer H	69.10 %	57.10 %	0 %	0 %	42.90 %	30.90 %
Bankruptcy index	43.10 %	82.10 %	0 %	0 %	17.90 %	56.90 %

Only Altman’s model operates with a gray zone; other models divide companies into two groups only (active x bankrupt). Altman’s model, which was created with the dataset from US manufacturing companies, was found to be least accurate. In contrast, its modified version created on the data from Latvian enterprises had a much better accuracy.

Interestingly, the highest accuracy in the evaluation of active companies was achieved by Fulmer H-index, which - according to the authors - was designed for SMEs. Their model, however, also reached the highest error rate in the assessment of bankrupt companies (Type I error) - up to 42.9 % of bankrupt companies were designated as active. Given that Type I error is much most costly than Type II errors, this incorrect classification is a more serious mistake than the inclusion of an active company among bankrupt companies (Type II error), which is high in all models created by the authors of the article (i.e. the CART models, the Bankruptcy Index).

All bankruptcy models operate on the principle of searching for similarities, i.e. the company is included into an appropriate group on the basis of similarities with other companies constituting the relevant group. The comparison is based on the suitable financial ratios constituting a functional model. In the case of the presented models, decisions were made on the basis of a cut-off (discrimination) score (one or more). When assigning the company to the respective group (bankrupt or active companies), probability of an erroneous assessment should be considered against the backdrop of the costs of the errors, i.e. Type I or Type II errors.

The discrimination ability of the models was tested also by the ROC curve, which is suitable to assess the quality of the discrimination rule for including an entity into one of two classes. The criterion used was the size of the area under the ROC curve (AUC). The advantage of ROC curves as a tool for testing bankruptcy models compared to the above approach is that they assesses the accuracy of the model for all possible settings of the cut-off score or the gray zone boundaries. ROC curves were plotted for all tested versions of the model (Figure 5).

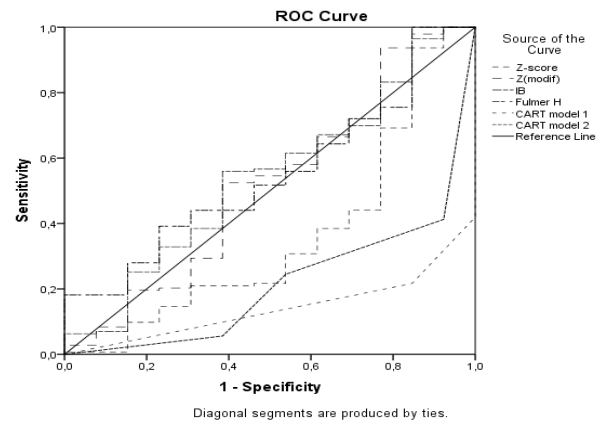


Figure 5. ROC Curve of the Tested Models

The more the ROC curve approaches the upper left corner, the more accurate the model is. The information illustrated by the ROC curve can be supplemented with information on the area under the curve (AUC).

Table 5

Values of Under Curve Area in Individual Models

Model	AUC
Z score – ZETA	0.358
Z (modif)	0.520
Bankruptcy index – IB	0.539
Fulmer H score -Fulmer H 1	0.547
CART model 1	0.859
CART model 2	0.785

AUC of 0.5 is considered a critical value as it represents the boundary of random choice (coin flip). The model is useful when it provides more accurate information than the random choice and is considered effective when AUC is greater than 0.8 (see, e.g., Camska, 2016). While AUC in compared models ranges from 0.358 (ZETA model) to 0.547 (Fulmer H-score), the CART Model 2 had AUC of 0.785, and Model 1 even 0.859. The AUC limit value of 0.8 was only reached by CART Model 1, i.e. it is the model with the highest discrimination ability.

Comparison of the accuracy of the models that were created based on the data of the companies operating in the construction industry brings a surprising conclusion that a model with a lower number of variables is more accurate than a model with a larger number of variables, i.e. the more comprehensive model. This means that additional information does not increase the accuracy of the model; on the contrary, it may lead to a more limited applicability due to the potential lack of data.

Conclusions

Based on the conclusions of previously published studies and our previous research, a hypothesis (H1) was formulated that the models created in the past and not specifically designed for construction companies are not efficient in predicting bankruptcy of construction companies. This hypothesis was verified by testing the differentiating accuracy of the selected models on the current data sample of companies operating in the construction industry. We proved that their accuracy was significantly lower than that originally reported by their authors. ROC curves and AUC data were used to compare model accuracy regardless of the setting of their gray zone or cut-off score. The AUC value in historical models ranges from 0.358 (ZETA model) to 0.547 (Fulmer H-score). AUC of 0.8 is considered a critical value, which the models do not achieve. We can therefore consider this hypothesis as confirmed.

Our research output were two new models. The starting point for the creation of the models was the selection of a suitable classification algorithm for typical properties of the data, and the creation of a rule that is easy to interpret and apply. Descriptive statistics of the studied sample showed that the data exhibit non-normal distribution and also contain outliers.

When creating the previous model (Bankruptcy Index), we decided to solve the problem of non-normal data distribution by transforming variables (see Karas, Reznakova, 2013). In the research presented here, we opted for a non-parametric algorithm, namely the classification

regression trees method (CART). The reasons for our selection were stated above and can be summarized in three points: it is a non-parametric method capable of capturing even a complex relationship between the variables; the method is immune to the influence of outliers; and it produces an easily interpretable rule. The frequent problem of the trees (the model created by the CART method) is their stability. We solved this problem by using the k-fold cross validation method, at $k=10$.

Two models were created by the CART method. Model 1 includes only three variables, namely current liabilities to total assets ratio (CL/TA), the logarithm of total assets (logTA), and the quick assets (current assets minus inventories)-to-sales (QA/S) ratio. According to the results of the model, the critical value for current liabilities is 69.09 %: at higher values, Model 1 will straightaway classify the company as bankrupt. Model 2 contains two additional variables, i.e. the retained earnings/total assets (RE/TA) ratio and the net working capital/sales (WC/S) ratio. Our models achieved significantly higher AUC values, i.e. 0.785 (Model 2) and even 0.859 (Model 1). Since the accuracy of both our models developed specifically for construction companies is higher than that of the compared models, we consider the second hypothesis (H2) also confirmed.

The last assumption (H3) was not confirmed: Model 2, which contains five variables and is more comprehensive, had lower accuracy measured by the AUC value than Model 1, i.e. the model that contained only three variables. Besides, a model with a fewer variables can be successfully used even in data-scarce environments.

References

- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I. (2000). Predicting Financial Distress of Companies: Revisiting the Z-score and Zeta® Models. Available from internet: <<http://pages.stern.nyu.edu/~ealtman/PredFncldistr.pdf>>.
- Aziz, M., & Dar, H. (2006). Predicting Corporate Bankruptcy: Where We Stand? *Corporate Governance*, 6(1), 18–33. <https://doi.org/10.1108/14720700610649436>
- Banyiova, T., Bielikova, T., & Piterkova, A. (2014). Prediction of Agricultural Enterprises Distress using Data Envelopment Analysis. *European Financial Systems 2014*, Lednice, 12-13.6.2014
- Barrie, D. S., & Paulson, B. C. (1992). Professional construction management: Including CM, design-construct, and general contracting, Hightstown: McGraw-Hill.
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, 26(2), 211–252.
- Brezingar-Masten, A., & Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications*, 39(11), 10153–10159. <https://doi.org/10.1016/j.eswa.2012.02.125>
- Carling, K., Jacobson, T., Linde, J., & Roszbach, K. (2007). Corporate credit risk modeling and the macroeconomy. *Journal of Banking & Finance*, 31(3), 845–868. <https://doi.org/10.1016/j.jbankfin.2006.06.012>
- Cut, S. (2014). Prediction of Company Financial Distress Using Neural Network Based on the Radial Basis Function, 2nd International conference in humanities, social sciences and global business management, Singapore, 21-22.6.2014.
- De Andres, J., Lorca, P. De Cos Juez, F. J., & Sanchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). *Expert Systems with Applications*, 38(3), 1866–1875. <https://doi.org/10.1016/j.eswa.2010.07.117>
- Di Marco, L., & Nieddu, L. (2014). Trigger factors that influence bankruptcy: a comparative and exploratory study. *Rivista Italiana di Economia Demografia e Statistica*, 68,(3/4), 191–198.
- Ding, Y., Song, X., & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support

- vector machine. *Expert systems with application*, 34(4), 3081–3089. <https://doi.org/10.1016/j.eswa.2007.06.037>
- Frydman, H., Altman, E. I., & Kao, D. L. (1985). Introduction Recursive Partitioning for Financial Classification: *The Case of Financial Distress*. *Journal of Finance*, 40(1), 269–291. <https://doi.org/10.1111/j.1540-6261.1985.tb04949.x>
- Faltus, S. (2014). Firm Default Prediction Model for Slovak Companies, *European Financial Systems 2014*, Lednice, 12–13.6.2014.
- Fulmer J. G, Moon J. E., Gavin T. A., & Erwin M. J. (1984). A bankruptcy classification model for small firms. *Journal of Commercial Bank Lending*, 66(11), 25–37.
- Gepp, A., & Kumar, K. (2015). Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques, *Procedia Computer Science* 54(1), 396–404. <https://doi.org/10.1016/j.procs.2015.06.046>
- Gertler, L. (2015). Interactions of Unconventional Monetary Policy Measures with the Euro Area Yield Curve. *Czech Journal of Economics and Finance*, 65(2), 106–126.
- Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. *Expert Systems with Applications*, 41(14), 6433–6445. <https://doi.org/10.1016/j.eswa.2014.04.026>
- Grice, J. S., & Dugan, M. T. (2001). The Limitations of Bankruptcy Prediction Models: Some Cautions for the Researchers, *Review of Quantitative Finance and Accounting*, 17(2), 151–166. <https://doi.org/10.1023/A:1017973604789>
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1), 1–21. <https://doi.org/10.1080/00401706.1969.10490657>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer-Verlag. <https://doi.org/10.1007/978-0-387-84858-7>
- Henerby, K. L. (1996). Do Cash Flows Variables Improve the Prediction Accuracy of a Cox Proportional Hazards Model for Bank Failure? *The Quarterly Review of Economics and Finance*, 36(3), 395–409. [https://doi.org/10.1016/S1062-9769\(96\)90023-X](https://doi.org/10.1016/S1062-9769(96)90023-X)
- Heo, J., & Yang, J. Y. (2014). AdaBoost based bankruptcy forecasting of Korean construction companies. *Applied Soft Computing*, 24(1), 494–499. <https://doi.org/10.1016/j.asoc.2014.08.009>
- Chuang, C., & Lin, R. (2009). Constructing a reassigning credit scoring model. *Expert Systems with Applications*, 36(2), 1685–1694. <https://doi.org/10.1016/j.eswa.2007.11.067>
- Kalupa, L. (2001). Model H-score w badaniu kondycji małych i srednich przedsiębiorstw. *Ruch prawniczy, ekonomiczny i socjologiczny*, 63, (4), 207–220.
- Karas, M., & Reznakova, M. (2013). Bankruptcy Prediction Model of Industrial Enterprises in the Czech Republic. *International Journal of Mathematical Models and Methods in Applied Sciences*, 7(5), 519–531.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119–127. <https://doi.org/10.2307/2986296>
- Kubenka, M., & Kralova, V. (2013). Vyuziti Z" score pri hodnoceni financniho zdravi odvetvi stavebnictvi. *E + M Ekonomie a Management*, 16(1), 101–112.
- Kwak, W., Shi, Y., Gong, G., & Yan, N. (2014), Bankruptcy Prediction for Chinese Firms: Comparing Data Mining Tools With Logit Analysis, *Journal of Modern Accounting and Auditing*, 10 (10), 1030–1037.
- Laitinen, E., Lukason, O., & Suvas, A. (2014). Behaviour of Financial Ratios in Firm Failure Process: An International Comparison. *International Journal of Finance and Accounting*, 3(2), 122–131.
- Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, 40, (8), 2941–2946. <https://doi.org/10.1016/j.eswa.2012.12.009>
- Liang, D., Lu, C., Tsai, C., & Shih, G. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561–572. <https://doi.org/10.1016/j.ejor.2016.01.012>
- Mandru, L., Khasman, A., Carstea, C., David, N., & Patrascu, L. (2010), The Diagnosis of Bankruptcy Risk Using Score Function, 9th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, Cambridge, 20-22.2.2010.
- Martin, D. (1977). Early Warning of Bank Failure: A Logit Regression Approach. *Journal of Banking & Finance*, 1(3), 249–276. [https://doi.org/10.1016/0378-4266\(77\)90022-X](https://doi.org/10.1016/0378-4266(77)90022-X)
- Min, J. H., & Jeong, C. (2009). A Binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36(3), 5256–5263. <https://doi.org/10.1016/j.eswa.2008.06.073>

- Niemann, M., Schmidt, J. H., & Neukirchen, M. (2008). Improving performance of corporate rating prediction models by reducing financial ratio heterogeneity. *Journal of Banking & Finance*, 32(3), 434–446. <https://doi.org/10.1016/j.jbankfin.2007.05.015>
- Marzal-Martinez, E. Barrachina-Martinez, I., & De la Poza-Plaza, E. (2014). Forecasting the Bankruptcy of Spanish Construction Companies throughout Multivariate Analysis, *Vision 2020: Sustainable Growth, Economic Development, and Global Competitiveness*, Valencia, 13-14.5.2014
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109–131. <https://doi.org/10.2307/2490395>
- Platt, D. H., & Platt, M. B. (1990). Development of a Class of Stable Predictive Variables: The Case of Bankruptcy Prediction. *Journal of Business Finance & Accounting*, 17(1), 31–51. <https://doi.org/10.1111/j.1468-5957.1990.tb00548.x>
- Psillaki, M., Tsolas, I. T., & Margaritis, M. (2010). Evaluation of credit risk based on firm performance. *European Journal of Operational Research*, 201(3), 873–881. <https://doi.org/10.1016/j.ejor.2009.03.032>
- Satish, Y. M., & Janakiram, B. (2011). Turnaround Strategy Using Altman Model as a Tool in Solar Water Heater Industry in Karnataka. *International Journal of Business and Management*, 6(1), 199–206.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, 74(1), 101–24. <https://doi.org/10.1086/209665>
- Sun, J. Liao, B., & Li, H. AdaBoost and bagging ensemble approaches with neuralnetwork as base learner for financial distress prediction of chinese con-struction and real estate companies. *Recent Patents on Computer Science*, 6(1), 47–59.
- Spicka, J. (2013). The financial condition of the construction companies before bankruptcy. *European Journal of Business and Management*, 5(23), 160–166.
- Taffler, R. J. (1982). Forecasting Company Failure in the UK Using Discriminant Analysis and Financial Ratio Data. *Journal of the Royal Statistical Society*, 145(3), 342–358. <https://doi.org/10.2307/2981867>
- Thomas Ng, S. T., Wong, J. M. W., & Zhang, J. (2011). Applying Z-score model to distinguish insolvent construction companies in China. *Habitat International*, 35(1), 599–607. <https://doi.org/10.1016/j.habitatint.2011.03.008>
- Tian, S., Yu Y., & Guo, H. (2015). Variable Selection and Corporate Bankruptcy Forecasts. *Journal of Banking & Finance*, 52(1), 89–100. <https://doi.org/10.1016/j.jbankfin.2014.12.003>
- Tserng, H. P., Chen, P. Huang, W. Lei, M. C., & Tran, Q. H. (2014) Prediction of default probability for construction firms using The logit model. *Journal of Civil Engineering and Management*, 20(2), 247–255. <https://doi.org/10.3846/13923730.2013.801886>
- West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27(11/12), 1131–1152. [https://doi.org/10.1016/S0305-0548\(99\)00149-5](https://doi.org/10.1016/S0305-0548(99)00149-5)
- Wu, Y., Gaunt, C., & Gray, S. (2010). A Comparison of Alternative Bankruptcy Prediction Models. *Journal of Contemporary Accounting & Economics*, 6(1), 34–45. <https://doi.org/10.1016/j.jcae.2010.04.002>

The article has been reviewed.

Received in October, 2016; accepted in April, 2017.