ANALYSIS OF FINANCIAL DISTRESS OF CZECH COMPANIES USING REPEATED MEASUREMENTS

Mária Stachová – Lukáš Sobíšek

Abstract

In our contribution, we focus on relationship between financial distress of company and its financial indicators. Our previous studies of Slovak companies show strong correlation. Hence, we verify our findings by fitting a financial distress (bankruptcy) prediction model investigating a comparable set of financial predictors. Used data are collected over four consecutive years and thus their longitudinal character allows us to use RE-EM tree model that is able to incorporate time dynamic of selected predictors. Moreover, we try to investigate if the longer period of the data collection improves the prediction of bankruptcy and in order to verify this hypothesis, we compare the predictive power of selected financial indicators in three overlapping periods of different lengths, namely four, three and two years. Obtained results are compared with result from one-year classification tree model that has been built with CART algorithm. All our computations were done by using statistical system R.

Key words: RE-EM model, financial distress model, prediction, financial indicators

JEL Code: C38, G33

Introduction

There is a plenty of financial distress definitions. In general, we can say that it is a situation in which a company has a problem to pay off or has an inability to meet its financial obligations. The effort to determine whether a company is in the risk of financial distress has become an object of many studies. Since the well-known Altman's Z-score published in 1968 (Altman, 1968) and its revision (Altman, 1983), through the more recent studies that combine accounting data, market-based and macro-economic data (Tinoco & Wilson, 2013). Many of approaches use the static prediction models based on various statistical methods, such as discriminant analysis, logistic regression, decision trees. "In majority of cases these models are based on historical accounting data and corresponding financial ratios of a carefully selected sample of companies representing an economy of interest. The underlying idea is that the past values of appropriately selected financial and economic indicators are able to determine the financial health in the future. Unfortunately, such a microeconomic approach has well known shortcomings" (Kráľ et al., 2014).

We believe that knowledge about past trends in financial indicators can improve the prediction of future changes in financial health of company and moreover, we expect that the longer period (4 consecutive years) of data collection is used the better prediction accuracy can be achieved. As in (Stachová et al., 2015) is said: "this broader look enables us to observe the trend of the indicators and/or changes in proportion of the indicators. Besides, the longer period lowers the risk that the company is not genuine in their economic, i.e. the accounting of the company isn't influenced by any extraordinary transactions realized in one business year."

The main goal of this contribution is to investigate whether it is meaningful to extend the classical bankruptcy models based on supervised statistical methods (Brezigar-Masten, 2012) by adding information about the dynamics of financial ratios to a training data set.

In (Kráľ et al., 2014) the authors computed the changes in financial ratios between two consecutive years. In the next step, these changes were used as inputs to fit a prediction model. In this work the logistic regression and random forests were estimated to classify the companies into distress group and non-distress group. The study was illustrated on two different sets of Slovak industrial companies representing the period 2002-2004 and 2009-2010, respectively. The results of this study suggest that incorporating time dynamic in financial indicators could improve the prediction accuracy of models.

In order to verify our hypotheses and to follow the study (Stachová et al., 2015), we compare the predictive power of selected financial indicators in assessing the financial health of Czech companies in three overlapping periods of different lengths, namely four, three and two years. Financial indicators are collected during the years 2012 to 2015 and the financial health status of company comes from years 2013 to 2016. It means, that we always paired the financial indicators (predictor) from year t and financial status (objective) from year t+1. The second goal of this study is to compare the accuracy of two classification methods RE-EM and CART algorithm. RE-EM is designed for the analysis of repeated measurement data and its advantage is that it takes into account the within and between companies' causation. Moreover, RE-EM typically provides more realistic error rate in comparison to the unstable CART that is quite sensitive to small changes in data and can be prone to overfitting (Gatti, 2014).

The paper is organized as follows. In Section 2 we present our data and methodology procedure. A very short information of not so well-known RE-EM algorithm is provided. Section 3 describes results of our prediction of Czech companies' financial health. Finally, in the last Section, we discuss classification ability of fitted CART and RE-EM models and positives and negatives of the proposed methodology.

1 Data and Methodology

Our data set consists of 15 financial indicators (see Table 1) of small and medium enterprises. These indicators were chosen according to work (Boďa & Úradníček, 2016) and were scaled to avoid inconsistency in data over time. Data covers the period 2012-2015 and financial status comes from years 2013 to 2016. The dependent, financial status "default" describe the situation in which a company has negative equity, or its earnings after taxes are negative or went bankrupt. Our data set was extracted from Czech data repository Albertina, covering processing industry area denoted according to NACE classification as CZ-NACE C category. These data were used also in bachelor thesis (Dušek, 2017).

Tab. 1: Financial indicators

Financial status	Return on assets	Debt ratio
CF solvency	Total assets turnover ratio	Inventory turnover
Trade avrg. collection period	Trade payables turnover ratio	Average collection period
Return on sales	Leverage	Current ratio
Quick ratio	Cash position ratio	Return on equity

Source: Author's work.

At the beginning, we split the data into three different training sets and testing data sets to estimate the RE-EM models. The division into training and testing data set was made in 80:20 ratio. The first data set is covering the period 2012-2015, the second one the period 2013-2015 and the third one the period 2014-2015. The predictive ability of our models we test using the training data (i.e. 20% remaining data from each period). We observe whether the longer period used to estimate the model leads to higher predictive accuracy of the model.

The RE-EM model is a regression tree based model with random effects for panel data created via *REEMtree()* function implemented in R (R Core Team, 2013) package "REEMtree" (Sela et al., 2011b). The RE-EM tree model is combination of well-known CART algorithm and structure of mixed effect models for panel data (Sela et al., 2011a). It allows us to take advantages of flexibility of a regression tree „rpart" model without the restrictive parametric assumptions of other well-established models for panel data, such as mixed effects ones (see Baltagi, 2012, Pinheiro et al.. 2009).

The accuracy of RE-EM model is evaluated and compared with the predictive accuracy of regression tree model that is implemented in „rpart" (Therneau et al., 2014) R package and created with the function rpart(). The function "rpart" works with CART (Classification and Regression Tree) algorithm that is binary recursive partitioning method where each group of objects (in our case the object is an enterprise) can be split only into two groups. The splitting is based on a splitting rule that create the most homogeneous subgroups (see Breiman et al., 1984).

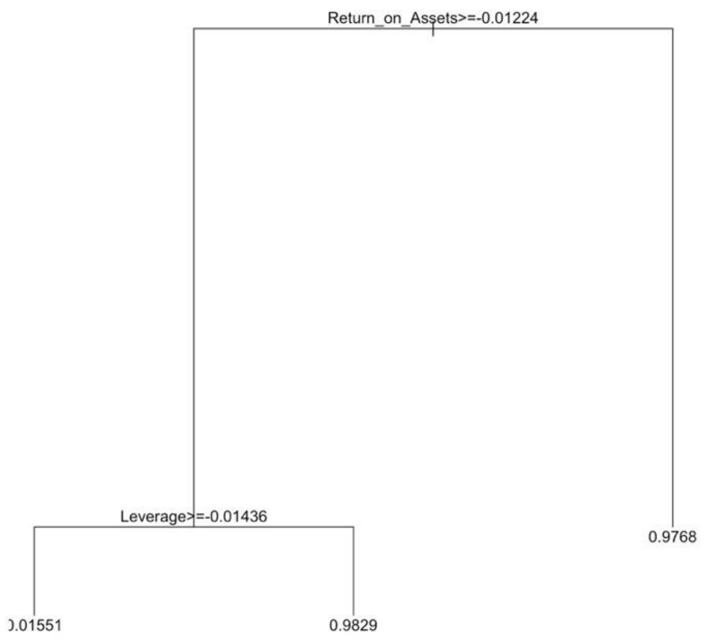
2 Results

We estimated the RE-EM tree model on three different overlapping data sets. First one takes the financial indicators from year 2012 to 2015, the second one from year 2013 to 2015 and the last one from year 2014 to 2015. The predicted variable is the classification of the enterprises as being in financial distress in the next year. The financial indicators of companies are taken as predictors. We thus obtained three RE-EM tree models and they can be seen in Figures 1-3.

Figure 1 shows the RE-EM tree combining all financial indicators from the years 2012-201. The root node is split according to the numeric variable namely the Return on Assets. If the condition of the root node is satisfied the respective company is placed in the left branches, all others being on the other side. The companies on the lower node are divided according to their Leverage. The terminal node (leaf) is ended with mean of our response for all observations assigned to this node.

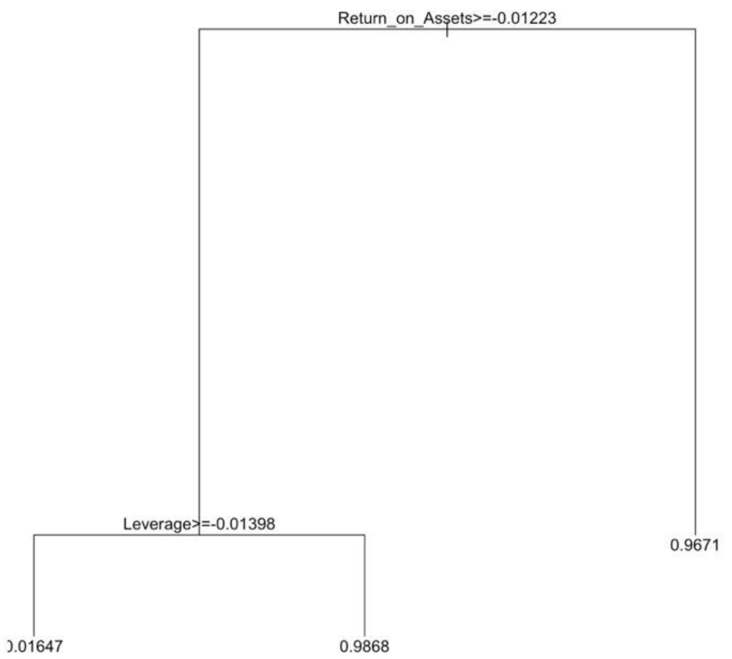
Figure 2 displays the RE-EM tree estimated on data from the period 2013 - 2015. The splitting variables were the same as in the first tree, the Return on Assets and the Leverage.

Fig. 1: RE-EM tree estimated on training set from the period 2012-2015.



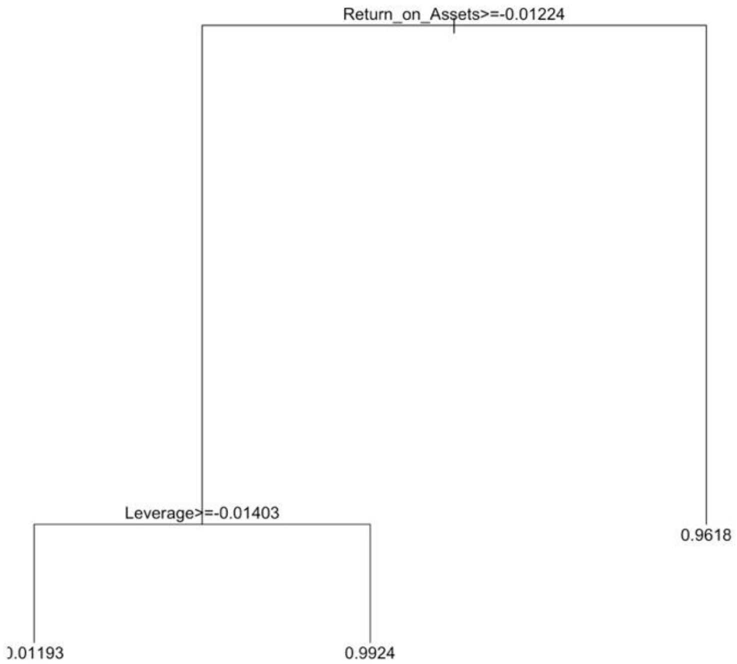
Source: Author's work.

Fig. 2: RE-EM tree estimated on training set from the period 2013-2015.



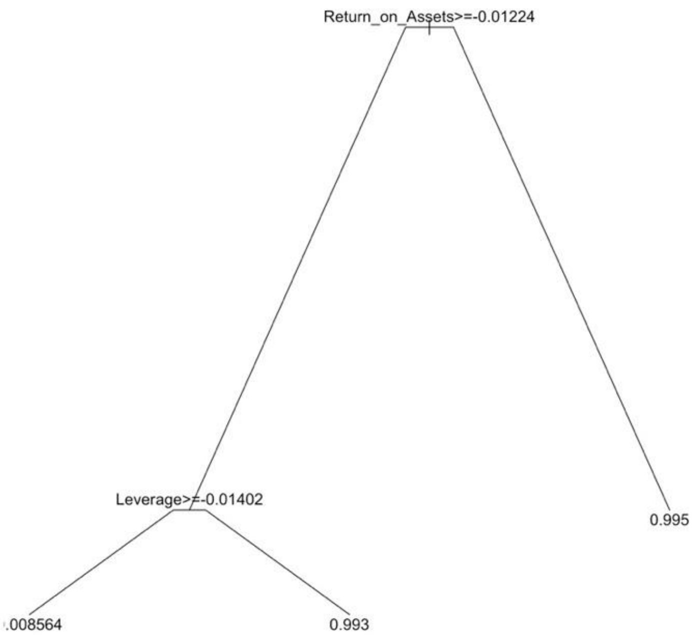
Source: Author's work.

Fig. 3: RE-EM tree estimated on training set from the period 2014-2015



Source: Author's work.

Fig. 4: CART tree estimated on training set from the year 2015.



Source: Author’s work.

To compare the predictive ability of RE-EM tree with statistical prediction model based on CART algorithm, we build the regression tree as is shown in Figure 4. This model is estimated only on data from the year 2015. The division of the root node is based on the same predictors as in case of RE-EM tree.

The predictive abilities of RE-EM tree models and CART tree model were evaluated on training data sets and expressed by confusion matrices and error rates in Tables 2 and 3. The threshold for labeling the company as being in financial distress was set at 0.5 in terminal node of tree.

Tab. 2: Confusion matrix of RE-EM model estimated on three different training sets.

	Predicted class					
	Training period 2012-2015		Training period 2010-2012		Training period 2011-2012	
Actual class	in distress	not in distress	in distress	not in distress	in distress	not in distress
in distress	1324	50	1160	54	516	28
not in distress	27	4472	35	4403	20	2295
	Accuracy rate:	98.7%	Accuracy rate:	98.4%	Accuracy rate:	98.3%

Source: Author’s work.

Tab. 3: Confusion matrix of Regression tree model estimated on data from the year 2015

	Predicted class	
	Regression tree	
Actual class	in distress	not in distress
in distress	1189	44
not in distress	6	5051
	Accuracy rate:	99.2%

Source: Author’s work.

Conclusion

In our study we focus on financial indicators of small and medium Czech enterprises to determine whether or not it is desirable to build the prediction model of financial health of these companies on longer period than one year and whether if we take 4 years data, it can improve the predictive accuracy of the model in comparison with model built on shorter time period. There are many reasons why it is important, for example for lender of the company to know how far to the past he has to look to get the most complete information that can indicate some ongoing financial problems of company and so on.

Although the predictive accuracy of RE-EM is similar to CART we prefer it over CART. It is because the RE-EM tree method is longitudinal model and thus combine the pros of mixed effect models with relative simple and uncomplicate regression tree. The RE-EM takes into account the within and between companies' causation as well. Moreover, RE-EM provides a more realistic error rate, even if it is a little bit higher, in comparison to the CART that tends to overfit. Our results indicate that the longer period (namely four consecutive years) of the data collection could lead to classifiers with lower error rates. Even the differences are very small, we can assume, that if we have longer time period of data available, the accuracy of the model can increase.

In a near future, we plan to work with different definitions of financial distress and different time periods to see whether the chosen financial indicators and algorithm are stable or whether they are strong depended on data set we used.

Acknowledgment

Mária Stachová have been supported by the project VEGA 1/0093/17 Identification of risk factors and their impact on products of the insurance and saving schemes.

References

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 583-609.

Altman, E. I. (1983). *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy.* New York: Wiley.

Baltagi, B.H. (2012). Econometric Analysis of Panel Data, Wiley, Chichester. 2012, ISBN 978-1-118-67232-7.

Boďa, M., & Úradníček, V. (2016). The portability of Altman's Z-score model to predicting corporate financial distress of Slovak companies. *In Technological and Economic Development of Economy*, 22(4), 532-553.

Brezigar - Masten, A., & Masten, I. (2012). CART-based selection of bankruptcy predictors for the logit model. *Expert Systems with Applications*, 39(11), 10153–10159.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). Classification and Regression Trees, Chapman and Hall (Wadsworth, Inc.), New York, ISBN 13: 978-0412048418.

Dušek, O. (2017). Analýza finančního zdraví podniku, Bachelor thesis, Faculty of Informatics and Statistics, University of Economics in Prague, 2017.

Gatti, Ch. (2014). Design of Experiments for Reinforcement Learning, Springer, Switzerland, 2014, ISBN 9783319121963.

Kráľ, P., Stachová, M. & Sobíšek, L. (2014). Utilization of repeatedly measured financial ratios in corporate financial distress prediction in Slovakia*:* In the 17th AMSE, international scientific conference, conference proceedings, Poland, 156-163.

Pinheiro, J. & Bates, D. (2009). Mixed-Effects Models in S and S-PLUS, Springer, New York. 2009, ISBN-13: 978-1475781441.

R Core Team (2013): R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2012, http://www.r-project.org/.

Sela, R. J. & Simonoff, J. S. (2011). RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data. *Machine Learning*, 86, pp.169-207.

Sela, R. J. & Simonoff, J. S. (2011). RE-EM Trees: Regression Trees with Random Effects, R package, version 0.90.3.

Stachová, M., Kráľ, P., Sobíšek, L., & Kakaščík, M. (2015). Analysis of financial distress of Slovak companies using repeated measurement. In 18th AMSE, International Scientific Conference Proceedings, Czech Republic.

Therneau, T., Atkinson, B. & Ripley, B. (2014). Rpart: Recursive Partitioning and Regression Trees. R package version 4.1-8.

Tinoco, M. H. & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables, *In International Review of Financial Analysis*, 30, pp. 394–419.

Contact

Mária Stachová
Faculty of Economics, Matej Bel University,
Tajovského 10
975 90 Banska Bystrica, Slovakia
maria.stachova@umb.sk

Lukáš Sobíšek
Faculty of Informatics and Statistics, University of Economics Prague
W. Churchill Sq. 1938/4
130 67 Prague 3 – Žižkov, Czech Republic
lukas.sobisek@vse.cz