18th AMSE Applications of Mathematics and Statistics in Economics
International Scientific Conference | Jindřichův Hradec, 2-6 September 2015

ANALYSIS OF FINANCIAL DISTRESS OF SLOVAK COMPANIES USING REPEATED MEASUREMENTS

MÁRIA STACHOVÁ, PAVOL KRÁĽ
Matej Bel University in BanskáBystrica, Faculty of Economics, Department of Quantitative Methods and Information Systems, Tajovského 10, 975 90 BanskáBystrica
email: maria.stachova@umb.sk, pavol.kral@umb.sk

LUKÁŠ SOBÍŠEK, MARTIN KAKAŠČÍK
University of Economics in Prague, Faculty of Informatics and Statistics, Department of Statistics and Probability, W. Churchill Sq. 4, 130 67 Prague 3
email: lukas.sobisek@vse.cz, xkakm01@vse.cz

Abstract
From the previous studies it is obvious that there is a strong relationship between financial distress and quantitative characteristics, e.g. financial ratios of companies. The main goal of our contribution is to investigate if the longer period (four consecutive years) of the data collection improves the ability of financial indicators to predict company bankruptcy using a data set consisting of financial indicators of Slovak companies collected over a short period of time, namely years 2009-2013. We apply two different approaches (RE-EM and CART) to this data set in order to predict a risk of financial distress of companies during the next period.

Key words: *models of financial distress, financial analysis ex-ante, repeated measurements.*

1. Introduction

Financial distress of companies is defined as an inability to pay off or a difficulty to meet company financial obligation. To determine whether a company is a potential candidate for financial distress has become a subject of many studies since the well-known Altman's Z-scorepublished in 1968 (Altman, 1968). Many of similar studies and approaches are based on static classification models constructed using various statistical methods, e.g. 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).The more recent studies combine accounting data, market-based and macro-economic data (Tinoco and Wilson, 2013).

In the paper (Kráľet al., 2014) we proposed a simple two-step methodology how to incorporate minimal knowledge about changes in financial ratios to a well-known static approach forcorporate financial distress modelling.In this study we focus on possibility to extend the classical bankruptcy models based on supervised statistical methods (Balcean andOoghe, 2006, Brezigar-Masten, 2012) by adding information about the dynamics of financial ratios to a training data set. In the first step, we compute changes in financial ratios between two consecutive years. In the second step, these changes are used as inputs for fitting a classification model. The methodology was illustrated on two different sets of Slovak industrial companies representing the period 2002-2004 and 2009-2010, respectively. Logistic regression and random forests were utilized for construction of classifiers. Results of the analysis presented in (Kráľ et al., 2014) indicated that incorporating dynamics in financial ratios could improve prediction ability of classifiers.

We hypothesize that knowledge about past trends in financial indicators can help us to predict future changes in financial health.Moreover, we expect that if the longer period of the data collection is used, we can get the better predictive accuracy. 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 economy, i.e. the accounting of the company isn't influenced by any extraordinary transactions realized in one business year.The main goal of our contribution is to investigate if the longer (four years) period of the data collection improves the prediction of bankruptcy.

In order to verify our hypotheses, we compare the predictive power ofselected financial indicatorsin assessing the financial health of Slovak companies in three overlapping periods of different lengths, namely four, three and two years. We use 6921companies' financial indicators collected over five consecutive years (2009-2013). The second goal of this study is to comparethe accuracy of two classification methods RE-EM andCART algorithm. RE-EM is designed for the analysis of repeated measurement data and its advantageis that it takes into account the within and between companies causation. Moreover, RE-EM typically provides more realistic error rate in comparison to the unstableCART that isquite sensitive to small changes in data and can be prone to overfitting (Gatti, 2014).

The paper is organized as follows. In Section 2 we explain our data analysis procedure. A very brief description of not so well-known RE-EM algorithm is provided. Section 3 describes analyzed data set of Slovak companies. Finally, in Section 4 we discuss classification ability of fitted CART and RE-EM models and pros and cons of the proposed methodology.

2. Data and Methodology

Our data set consists of 30 financial indicators (see Table 1) from 6921 companies, 20 % of them labeled as being in financial distress (unhealthy). Data covers the period 2009-2013. For the purpose of our analysis we define financial distress as a situation in which a company went bankrupt, has some ongoing liquidation or has some overdue obligations in the current accounting period. Our data set was extracted from the data set purchased from CRIF – Slovak Credit Bureau, s.r.o, covering economic activities 1110 – 96060 according to SK NACE classification.

Table 1. Financial indicators

Financial distress	Accrued assets	Accruals
Number of changes to the Board of directors	Equity	Income from ordinary operation
Number of registered employees (in intervals in actual period)	Prior period retained loss	Value added
Long-lived intangible assets	Net income after taxes	Personal costs
Long-lived tangible assets	Provisions	Revenue from disposal of

		long-lived assets and material
Long-term financial assets	Short-term liabilities	Income from financial operations
Inventory	Long-term liabilities	Interest expense
Long-term receivables	Bank loans and assistance	Ordinary income after taxes
Short-term receivables	Long-term bank loans	Extraordinary expenses
Short-term financial assets	Short-term bank loans	Net income after taxes

Source: Author's work.

In the first step of our analysis we cluster the companies in order to determine whether there are differences among size of companies. We use the *K*-means clustering method with the following variables: Number of employees, Total revenue and Total assets from year 2009. We assume one to ten clusters.Clustering shows that there is only one homogenous cluster and thus we can say that companies in our data set are similar from this point of view and values of their financialindicators are comparable.

In the second step of our analysiswe fit prediction models that should be able to predict financial health of the selected Slovak companies.

We split the data intothree different training sets to estimate the RE-EM models. The first one is covering the period 2009-2012, the second one the period 2010-2012 and the third one the period 2011-2012. The predictive ability of our models we testusing the last year data, i.e. the data from the year 2013. 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" (Selaet al., 2011b). This model combines the flexibility of tree-based predictive models with the structure of mixed effects models for longitudinal data (Sela et al., 2011a). It allowsus to take advantages of a regression tree „rpart“ model and avoids the drawbacks, e.g. restrictive parametric assumptions of other well-established models for panel data, such as mixed effects ones, see (Baltagi,2012, Pinheiro et al., 2009).

The predictive accuracy of RE-EM model is compared with the predictive accuracy of well-known regression tree model that is implemented in „rpart“ (Therneau et al., 2014) R package and created with the function of the same name.

CART (Classification and Regression Tree) algorithm is binary recursive partitioning method where each group of objects (in our case the object is a company) can be split only into two groups. Thus each node can be split into two child nodes and this process can be applied over and over again. The splitting is performed according to a splitting rule to create the most homogeneous subsamples, see (Breiman et al., 1984).

3. Results

We estimated the RE-EM tree model on three different data sets covering the periods 2009-2012, 2010-2012 and 2011-2012. The outcome variable is the classification of the companies as being in financial distress in the year 2013. The companies' financials absolute values were employed as the predictors. We obtained three RE-EM tree models as indicated in Figures 1-3.

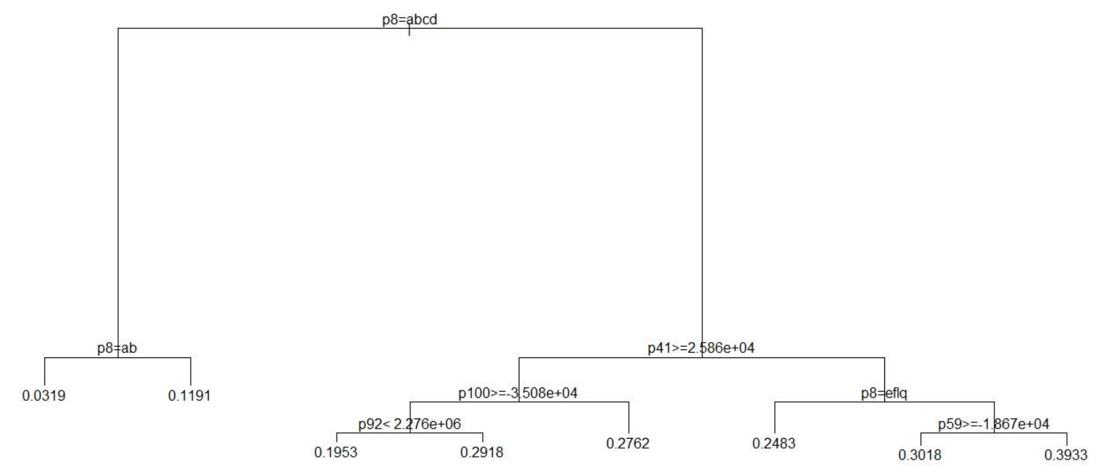

Figure 1. RE-EM tree estimated on training set from the period 2009-2012.
Source: Author's work.

Fig. 1 shows the RE-EM tree combining all financial indicators from the years 2009-2012. The root node is split according to the categorical variable namely the number of employees (p8). 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 nodes are divided according to their short-term financial assets (p41), income from financial operations (p100), personal costs (p92) and prior period retained loss (p59), etc. until the splitting stop condition is satisfied. The terminal node (leaf) is ended with mean of our response for all observations assigned to this node.

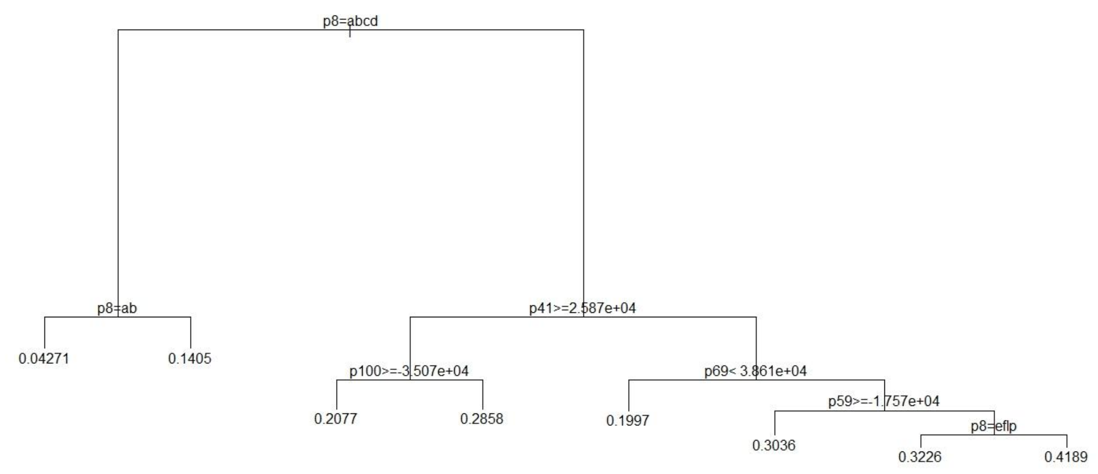

Figure 2. RE-EM tree estimated on training set from the period 2010-2012.
Source: Author's work.

Fig. 2 displays the RE-EM tree estimated on data from the period 2010-2012. The splitting variables were similar as in the first tree, the number of employees (p8), short-term financial assets (p41), income from financial operations (p100), liabilities (p69) and prior period retained loss (p59).

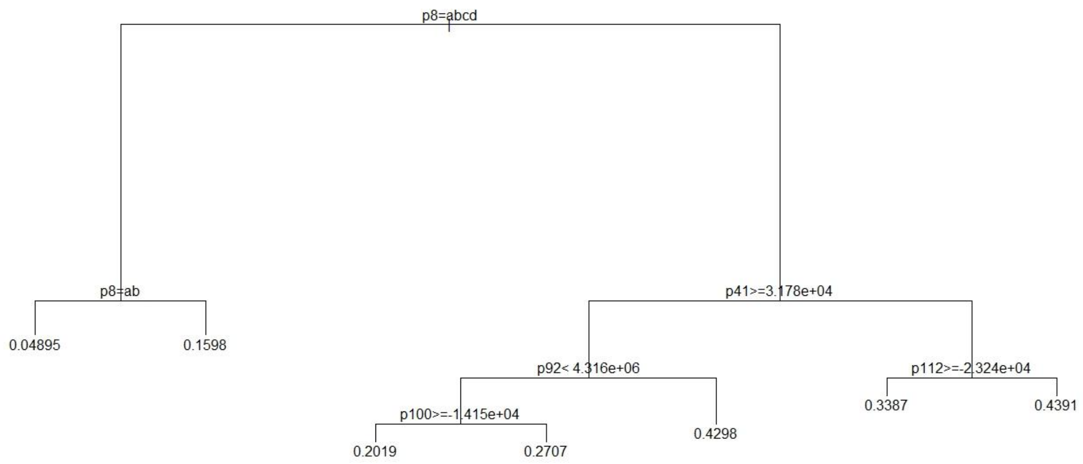

Figure 3. RE-EM tree estimated on training set from the period 2011-2012.
Source: Author's work.

Fig. 3 presents the RE-EM tree trained on data from years 2011-2012. The predictors in the root node and sub-nodes are the number of employees (p8), short-term financial assets (p41), taxes (p112), income from financial operations (p100) and personal costs (p92).

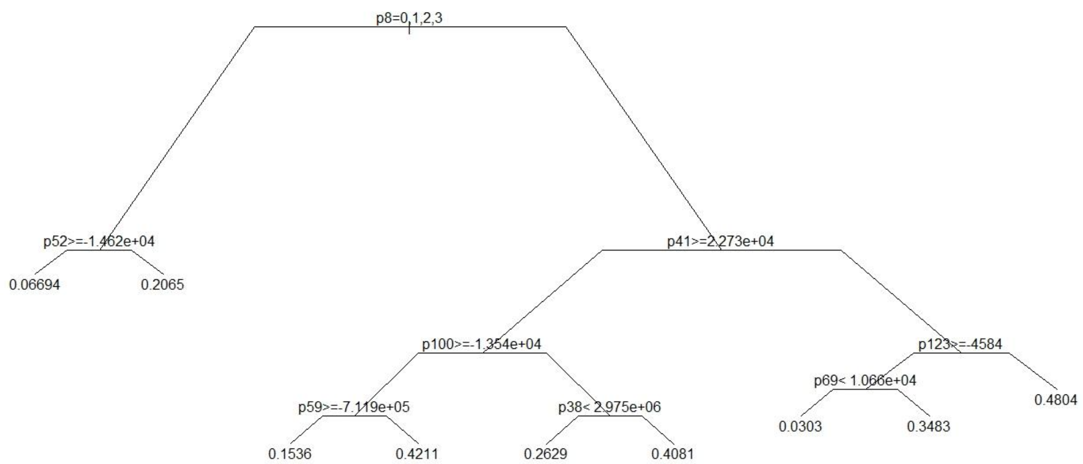

Figure 4. Regression tree estimated on training set from the year 2012.
Source: Author's work.

The regression tree in Fig. 4 illustrates the model estimated only on data from the year 2012. The division of the root node is based on the number of employees (p8), the other key nodes being split by short-term financial assets (p41), net income after taxes (p123), equity (p52), etc.

The predictive abilities of RE-EM tree models and regression tree model were evaluated on data from the year 2013 and expressed by confusion matrices and error rates in Tab.2 and Tab.3. The threshold for labeling the company as being in financial distress was set at 0.3 in terminal node of tree.

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

	Predicted class					
	Training period 2009-2012		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	516	1167	598	1115	829	884
not in distress	654	4554	809	4399	1223	3985
	Error rate:	26.7%	Error rate:	27.8%	Error rate:	30.4%

Table 3. Confusion matrix of Regression tree model estimated on data from the year 2012

	Predicted class	
	Regression tree	
Actual class	in distress	not in distress
in distress	136	1557
not in distress	112	5096
	Error rate:	24%

4. Conclusions

Since financial ratios and their relative changes were not suitable in this case due to incomplete (case-missing) data, we restricted ourselves to absolute values of financial indicators. It was justified by cluster analysis using baseline values (2009) where our companies formed only one cluster. In addition, we do not expect radical change in the company's size in consecutive five years. Although the predictive accuracy of RE-EM is similar to CART we prefer it over CART because RE-EM is designed for the analysis of repeated measurement data and it takes into account the within and between companies causation. Moreover, RE-EM provides a more realistic error rate, even if it is a little bit higher, in comparison to the CARTthat 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.

In ourfuture research we plan to determine an optimal period for collecting financial indicators and investigate whether resamplingmethods enhance the estimation accuracy. We intend to apply a classificationmethoddesignedforcross-sectionaldatatoadatasetthatwouldincludeboththevaluesfromagivenyearandauxiliaryvariablesrepresentingthedynamicsin the data.

Acknowledgements

MáriaStachová and Pavol Kráľ have been supported by the project VEGA 1/0647/14.

References

1. ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. In The Journal of Finance, 1968, vol. 23, iss. 4, pp. 589-609.
2. BALCAEN, S., OOGHE, H. 2006. 35 Years of Studies on Business Failure: an Overview of the Classic Statistical Methodologies and Their Related Problems, In.The British Accounting Review, 2006, vol.38, iss.1, pp. 63-93.
3. BALTAGI, B.H. 2012. Econometric Analysis of Panel Data, Wiley, Chichester. 2012, ISBN 978-1-118-67232-7.
4. BREZIGAR-MASTEN, A., MASTEN, I. 2012. CART-based selection of bankruptcy predictors for the logit model, In. Expert Systems with Applications, 2012, vol. 39, iss.11, pp. 10153-10159.
5. BREIMAN, L., et al. 1984. Classification and Regression Trees, Chapman and Hall (Wadsworth, Inc.), New York, 1984, ISBN 13: 978-0412048418.
6. GATTI, Ch., 2014. Design of Experiments for Reinforcement Learning, Springer, Switzerland, 2014, ISBN 9783319121963.
7. KRÁĽ, P., STACHOVÁ, M., SOBÍŠEK, L. 2014. Utilization of repeatedly measured financial ratios in corporate financial distress prediction in Slovakia: In 17th AMSE, international scientific conference, conference proceedings, Poland. 2014, pp.156-163, ISBN 978-83-7695-421-9.
8. PINHEIRO, J., BATES, D. 2009. Mixed-Effects Models in S and S-PLUS, Springer, New York. 2009, ISBN-13: 978-1475781441.
9. R CORE TEAM 2013: R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2012, http://www.r-project.org/.
10. SELA, R. J., SIMONOFF, J. S. 2011. RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data: Machine Learning, 2011, vol. 86, pp.169-207.
11. SELA, R. J., SIMONOFF, J. S. 2011. RE-EM Trees: Regression Trees with Random Effects, R package, version 0.90.3.
12. THERNEAU, T., ATKINSON, B., RIPLEY, B. 2014. Rpart: Recursive Partitioning and Regression Trees. R package version 4.pp. 1-8.
13. 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, vol. 30, pp. 394-419.