STATISTICAL LEARNING METHODS IN CORPORATE FINANCIAL DISTRESS PREDICTION OF SLOVAK ENTERPRISES: COMPARISON OF ALTERNATIVE MODELS

Mária Stachová – Pavol Kráľ

Abstract

Historically, various financial distress prediction models based on standard statistical learning methods were constructed for Slovak enterprises. Usually, these models were created utilizing different training datasets, covering different time periods, considering different sets of predictors and various definitions of financial distress. Consequently, it is difficult to select the best model or group of models from such a heterogeneous family of models for adoption in decision-making processes in enterprises, having in mind convenience of their applicability in the Slovak economic environment and possibility to rely on them in the future. One way how to overcome this problem and support overall adoption of financial distress prediction models is to offer an easily applicable model, or set of models, with satisfactory prediction ability during a prespecified period and regular updates via refitting of the models using the most recent data. In our contribution, we present an update of one such model with preselected set of five predictors inspired by the original Altman's models and distress defined using financial indicators equity, earnings after tax and current ratio. The model was generated utilizing decision tree and random forest methods separately for enterprises covering economic activities manufacturing, construction and wholesale according to SK NACE classification.

Key words: financial distress prediction, alternative models, tree-based methods

JEL Code: C38, G33

Introduction

The models for classifying and predicting whether a company is a potential candidate for being financially distressed has become a subject of many studies since well-known Altman's Z-score (Altman, 1968) and its revision (Altman, 1983). All the models share two essential components needed for construction of financial health prediction model – set of financial indicators used as predictors and definition of financial health determining the response variable.

There are many different definitions of financial health or financial distress of enterprises. Mihoková et al. (2007) claim that it is caused by subjective selection of financial health indicators as to evaluate financial health of company and to ensure its sustainability, it is necessary to take into account many different financial measurements, such as a current ratio, payment ability, profitability and efficiency of enterprise.

The main starting point in financial health assessment is considering sustainability and sufficient profit of company. A company has to be able to keep its business at balance with the changing conditions and needs of an external environment but with regard to interests of groups that are involved in the decision-making and functioning of the business.

Managers and financial experts, as a critical component of the company's internal environment, are largely responsible for the current and future status of the business influenced by external and internal factors. They are expected to be able to make the right decisions at the right time, by using different methods and tools. In the other words, they need also to consider the right financial indicators that can signalize the incoming financial problems and they have to use proper and suitable methods to analyze these indicators.

According to Lesáková et al. (2015), we can divide the financial indicators of enterprises into several groups:

- current ratio indicators,
- activity indicators,
- debt indicators,
- profitability indicators,
- market value indicators.

As we already mentioned, financial indicators play the role of predictors in financial distress modelling, but in many occasions, some of them are also used to define the financial status of companies and thus they are used to create a response variable, e.g. status of a financial health, as well. In the case of Slovak companies, such an approach was already adopted for example in (Úradníček et al., 2016) and we apply it also in our paper.

There exist many similar studies, e.g., (Boďa & Úradníček, 2016; Balcean & Ooghe, 2006; Brezigar-Masten, 2012) which are mainly based on static classification models constructed using various statistical methods, e.g. discriminant analysis, logistic regression, decision trees. Based on (Kráľ et al., 2014; Stachová et al. 2015) we believe, that time dynamic or time gap incorporated into these well-known static models can improve their predictive accuracy. In the paper, we restricted ourselves to a simple approach where predictors and a response variable in our model are from different years and the time gap between them ranges from one to three years. We estimate models utilizing classification tree algorithm (CART algorithm) and random forest algorithm. The resulting models can be seen as updates of models presented in (Kráľ et al., 2016).

1 Data and Methodology

The data sets used in the paper contains data of 87971 Slovak enterprises covering economic activities 10110–33200, 41100 – 43990, 45110 - 47900 according to SK NACE classification and recognized as C Manufacturing (15702 enterprises), F Contruction (22501 enterprises), and G Wholesale and retail trade, repair of motor vehicles and motorcycles (49768 enterprises)). The data set involved all the four legal forms of enterprises common in Slovakia (i.e. v.o.s. – general partnership, k.s. – limited partnership, s.r.o. – private limited company, a.s. – joint-stock company) and related to a range of 5 fiscal periods: from 2013-2017. As in (Boďa & Úradníček 2016), the data set contains five financial ratios (predictors), mimicking those in the Altman's model, and one binary response variable – a status of a company –taking values "being in distress" and "not being in distress".

The predictors are X1 – working capital / total assets, X2 – retained earnings / total assets, X3 – earnings before interest and taxes / total assets, X4 – book value of equity / book value of debt, and X5 – sales / total assets (Boďa & Úradníček 2016).

Although there exist multiple valid alternatives how to define financially distressed enterprise, for simplicity and comparability of results we adopt here an identical definition of financially distressed enterprises as the one utilized in (Boďa & Úradníček 2016) where the status variable has been defined as follows. "An enterprise was considered financially distressed if

a) its equity was negative,

b) its EAT (earnings after tax) was negative.

c) its current ratio attained a value lower than 1.

All the three conditions had to be satisfied in order for an enterprise to be considered financial distressed".

The models are created separately for each of economic activities assuming the following six scenarios:

- predictors from the year 2013 and the status variable from the year 2014,
- predictors from the year 2013 and the status variable from the year 2015,
- predictors from the year 2013 and the status variable from the year 2016,
- predictors from the year 2014 and the status variable from the year 2015,

- predictors from the year 2014 and the status variable from the year 2016,
- predictors from the year 2015 and the status variable from the year 2016.

For each scenario a simple random sample of one thousand distressed and one thousand non-distressed enterprises is used as a balanced training set.

Each model is evaluated based on its overall accuracy computed using a tenfold cross-validation (Hastie et al.) on balanced training data during the model fitting and then based on overall accuracy, specificity and sensitivity computed directly on the available test data. Finally, the models are discussed in the view of naïve thresholds for years with test data available.

When fitting the models, we restricted ourselves to classification trees based on CART algorithm and classification random forests (Hastie et al., 2001).

All analyses are performed in statistical system R (R Core Team, 2016) using package caret (Kuhn, 2019).

2 Results

The numbers and percentage of enterprises considered as financially distressed in the period 2013-2017 are listed in Tab. 1:

Tab. 1: The number of financially distressed companies in the period 2013-2017

NACE	2013	2014	2015	2016	2017
C	1737 (16.4%)	1801 (14.9%)	1600 (12.9%)	1520 (11.8%)	813 (9.5%)
F	2481 (18.8%)	2626 (16.6%)	2225 (13.5%)	1909 (11.1%)	857 (8.0%)
G	6686 (21.2%)	7086 (19.5%)	6439 (17.6%)	6149 (16.3%)	2909 (13.0%)

In Tab. 1 we can see that the data are quite imbalanced with the percentage of financially distressed companies ranging from 8% to approximately 21%. Based on that we can establish naïve thresholds for overall accuracy of predictive models ranging from approximately 79% to 92%.

Tab. 2 lists overall accuracy of models computed using ten-fold cross-validation on balanced training data. We can see that the models are best for the sector of economic activities C, followed by F and G. In all cases, they provide us with much better predictions than random guess in balanced data. Moreover, models with one-year gap seem to be preferable to those utilizing the gap of two or three years, regardless the type of economic activities. Finally, models based on classification trees and classification random forest provides us with very similar results. Interpretation of Tab. 2 is further supported by test data results as overall accuracy on test data (Tab. 3 and Tab 4.) shows similar patterns to those observed in cross-validated overall accuracy and all models demonstrate reasonable stability through the time period with available data. However, as the test data sets (contrary to training data sets) are highly unbalanced towards companies not being in distress, the computed overall accuracies are not very favorable when compared to naïve overall accuracies based on Tab. 1. On the other hand, values of a sensitivity indicate that both models based on classification trees and classification random forest are quite satisfactory in identification of enterprises with possibility of financial distress.

Tab. 2: Cross-validated overall accuracy of models

	Classification trees			**Classification random forests**		
	C	F	G	C	F	G
Predictors 2013 vs status 2014	0.822	0.787	0.801	0.815	0.787	0.805
Predictors 2013 vs status 2015	0.786	0.720	0.751	0.793	0.717	0.745
Predictors 2013 vs status 2016	0.730	0.677	0.713	0.745	0.694	0.712
Predictors 2014 vs status 2015	0.835	0.774	0.821	0.782	0.722	0.741
Predictors 2014 vs status 2016	0.779	0.737	0.749	0.790	0.715	0.745
Predictors 2015 vs status 2016	0.817	0.783	0.816	0.816	0.794	0.805

Tab. 3: Accuracy measures of models utilizing classification trees on test sets

	Predictors 2013 vs status 2014			Predictors 2013 vs status 2015		Predictors 2013 vs status 2016	Predictors 2014 vs status 2015		Predictors 2014 vs status 2016	Predictors 2015 vs status 2016
	Test 2015	Test 2016	Test 2017	Test 2016	Test 2017	Test 2017	Test 2016	Test 2017	Test 2017	Test 2017
C										
acc.	0.833	0.837	0.860	0.859	0.879	0.889	0.861	0.879	0.874	0.915
sens.	0.999	1.000	1.000	0.995	0.995	1.000	0.995	0.996	0.995	1.000
spec.	0.999	0.816	0.846	0.841	0.867	0.878	0.843	0.867	0.861	0.906
F										
acc.	0.887	0.890	0.914	0.812	0.837	0.866	0.828	0.851	0.853	0.872
sens.	1.000	1.000	1.000	0.994	0.997	1.000	0.995	0.997	0.966	0.987
spec.	0.870	0.889	0.906	0.789	0.823	0.854	0.807	0.838	0.843	0.862
G										
acc.	0.837	0.851	0.874	0.842	0.859	0.832	0.860	0.875	0.877	0.908
sens.	0.999	0.999	1.000	0.982	0.983	0.989	0.990	0.992	0.982	1.000
spec.	0.803	0.822	0.855	0.814	0.841	0.809	0.835	0.857	0.861	0.894

Notes: acc. = overall accuracy, sens. = sensitivity, spec. = specificity

Tab. 4: Accuracy measures of models utilizing random forest trees on test sets

	Predictors 2013 vs status 2014			Predictors 2013 vs status 2015		Predictors 2013 vs status 2016	Predictors 2014 vs status 2015		Predictors 2014 vs status 2016	Predictors 2015 vs status 2016
	Test 2015	Test 2016	Test 2017	Test 2016	Test 2017	Test 2017	Test 2016	Test 2017	Test 2017	Test 2017
C										
acc.	0.870	0.870	0.892	0.847	0.871	0.836	0.849	0.871	0.860	0.893
sens.	0.995	0.997	0.994	0.982	0.980	0.989	0.985	0.983	0.989	0.993
spec.	0.851	0.853	0.881	0.829	0.859	0.820	0.830	0.859	0.846	0.882
F										
acc.	0.857	0.864	0.884	0.820	0.844	0.786	0.822	0.846	0.838	0.844
sens.	0.987	0.990	0.994	0.984	0.985	0.952	0.983	0.979	0.974	0.987
spec.	0.837	0.848	0.874	0.800	0.831	0.771	0.803	0.834	0.826	0.832
G										
acc.	0.859	0.865	0.883	0.825	0.848	0.832	0.827	0.849	0.856	0.887
sens.	0.969	0.969	0.968	0.978	0.983	0.960	0.979	0.982	0.978	0.983
spec.	0.835	0.844	0.870	0.795	0.828	0.813	0.797	0.829	0.838	0.872

Notes: acc. = overall accuracy, sens. = sensitivity, spec. = specificity

Conclusion

In the paper, we present recent updates (utilizing available data from the years 2013-2017) of models based on classification tree and classification random forest models originally presented in (Kráľ et al., 2016) utilizing more recent data of Slovak enterprises covering economic activities C, F and G according to SK NACE classification. Although quite simplistic, all models demonstrate reasonable prediction abilities, especially in identification of companies in distress, comparable across assumed economic activities and quite stable across the whole time period at hand, with small edge in the case of manufacturing enterprises. Therefore, they can be used as a supporting tool for decision-making process focusing on identification of financially distressed enterprises. The models are ready for deployment in the form of R rds files and can be provided by authors upon request. In the future research, we plan to increase the range of methods utilized in our model fitting and further investigate the models via alternative accuracy measures.

Acknowledgment

The work has been supported by project VEGA, No. 1/0767/18 SMART model - SMART model – a decision support tool in management of enterprises.

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.

Balcaen, S., & Ooghe, H. (2006). 35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems.The British Accounting Review, 38(1), 63–93.

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.

Hastie, T., Friedman, J. H., Tibshirani, R. 2001. The elements of statistical learning. New York : Springer. 2001.

Kráľ, P., Fleicher, M., Stachová, M., Nedelová, G., Sobíšek., L. (2016) Corporate financial distress prediction of Slovak companies: Z-score models vs. Alternatives. In the 19th AMSE, international scientific conference, conference proceedings, Slovakia, 224-231.

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.

Kuhn, M. (2019). caret: Classification and Regression Training. R package version 6.0-84. https://CRAN.R-project.org/package=caret

Lesáková, Ľ. et al. (2015). Finančno-ekonomická analýza podniku 1. Banská Bystrica : Belianum. Vydavateľstvo Univerzity Mateja Bela v Banskej Bystrici. Ekonomická fakulta, 2015. 142 s. ISBN 978-80-557-0982-6.

Mihoková, L., Vida, M., Kádár, G. (2007). Diagnostika efektívnosti a konkurencieschopnosti podniku a prechod podniku do krízového stavu. Intercathedra, 23, 78-81.

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.

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

Úradníček, V. et al. (2016). *Variantné metódy predikcie finančného zdravia podnikov v podmienkach dynamického ekonomického prostredia*. Banská Bystrica : Belianum. Vydavateľstvo Univerzity Mateja Bela v Banskej Bystrici. Ekonomická fakulta, 136 p.

Contact

Mária Stachová

Faculty of Economics, Matej Bel University

Tajovského 10

975 90 Banska Bystrica, Slovakia

maria.stachova@umb.sk

Pavol Kráľ

Faculty of Economics, Matej Bel University

Tajovského 10

975 90 Banska Bystrica, Slovakia

pavol.kral@umb.sk