Abstract

The main objective of our research was to develop a new bankruptcy prediction model for the Czech economy. For that purpose we used the logistic regression and 150,000 financial statements collected for the 2002—2016 period. We defined 41 explanatory variables (25 financial ratios and 16 dummy variables) and used Bayesian model averaging to select the best set of explanatory variables. The resulting model has been estimated for three prediction horizons: one, two, and three years before bankruptcy, so that we could assess the changes in the importance of explanatory variables and models' prediction accuracy. To deal with high skew in our dataset due to small number of bankrupt firms, we applied over- and under-sampling methods on the train sample (80% of data). These methods proved to enhance our classifier's accuracy for all specifications and periods. The accuracy of our models has been evaluated by Receiver operating characteristics curves, Sensitivity-Specificity curves, and Precision-Recall curves. In comparison with models examined on similar data, our model performed very well. In addition, we have selected the most powerful predictors for short- and long-term horizons, which is potentially of high relevance for practice.

JEL Classification	C11, C51, C53, G33, M21
Keywords	Bankruptcy prediction, Logistic regression, Bayesian model averaging, Receiver operating characteristics curves