The classical and Bayesian logistic regression in the research on the financial standing of enterprises after bankruptcy in Poland

Jadwiga Kostrzewska[1], Maciej Kostrzewski[2], Barbara Pawełek[3], Krzysztof Gałuszka[4]

Abstract

The problem of the enterprise's bankruptcy is an important issue of economic sciences. In the literature the applications of multivariate statistical analysis are focused more on the prediction of bankruptcy than the evaluation of the financial standing of enterprises after bankruptcy. Studying the path of recovering from insolvency can be a source of valuable information useful in assessing the probability of a success of the restructuring. The aim of the paper is to present the results of the pilot research on the financial standing of construction enterprises after bankruptcy in comparison with the situation of financially sound companies. We use the logit model to classify enterprises as "healthy" ones and the ones "after bankruptcy". The maximum likelihood estimation (classical approach) and Bayesian approach is applied. The point and interval estimators of probability of a bankruptcy risk are compared with medians and histograms of their posterior distributions. The evaluation of the classification accuracy is based on the sensitivity, specificity, and AUC measures. We apply the one-dimensional methods (quantile analysis, Tukey's criterion) and the multi-dimensional method (projection depth function) to detect outliers. We also employ a method of analyzing discriminatory power of financial ratios. The pilot research covers construction enterprises in Poland in 2005 and 2009.

Keywords: *logistic regression, maximum likelihood estimation, Bayesian approach, financial standing, after bankruptcy*

JEL Classification: C11, C250, G330

1. Introduction

The problem of enterprise bankruptcy is an important issue in economic sciences. In Poland, the law distinguishes liquidation bankruptcy and arrangement bankruptcy. Arrangement bankruptcy declared by the court gives an enterprise a chance to continue its operations. The court is guided by the principle according to which bankruptcy proceedings must be conducted in such a way that the creditors could be satisfied as far as possible, and the debtor's enterprise functioning could be preserved. A lot of attention is paid in literature on the subject to the issue

[1] Corresponding author: Cracow University of Economics, Department of Statistics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: jadwiga.kostrzewska@uek.krakow.pl.

[2] Cracow University of Economics, Department of Econometrics and Operational Research, Rakowicka 27, 31-510 Kraków, Poland, e-mail: maciej.kostrzewski@uek.krakow.pl.

[3] Cracow University of Economics, Department of Statistics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: barbara.pawelek@uek.krakow.pl.

[4] University of Economics in Katowice, Department of Public Finance, 1 Maja 50, 40-287 Katowice, Poland, e-mail: krzysztof.galuszka@ue.katowice.pl.

of forecasting bankruptcy, but few works deal with an analysis of the financial standing of enterprises after they have been declared bankrupt. An analysis of the process of getting out of the insolvency problem may be a source of valuable information, useful to other bankrupt enterprises for the assessment of the likelihood of success as a result of applying restructuring proposals. The aim of the paper is to present results of a pilot study in which a logit models obtained using the maximum likelihood method and the Bayesian approach were applied to assess the financial standing of enterprises after they have been declared bankrupt in comparison with the situation of financially sound enterprises.

2. Research methodology

The research is based on the data from the Emerging Markets Information Service. The dataset consists of 396 construction enterprises in Poland, which presented their financial statements during the years 2005-2009. The set includes 'financially sound' enterprises (non-bankrupt ones, NB), as well as bankrupt enterprises (B) with court verdicts passed within the period from 17 November 2003 to 30 August 2004, i.e. just before the examined period. Bankrupt enterprises constituted about 1.3% of all the enterprises in the database. The study of the financial standing of the companies used fourteen financial indicators compiled in Table 1. The pilot study covers the years 2005 and 2009.

LIQUIDITY RATIOS R_{01} – Current liquidity ratio R_{02} – Quick liquidity ratio R_{03} – Cash Ratio	**PROFITABILITY RATIOS** R_{07} – Gross profitability R_{08} – Net profitability R_{09} – ROE R_{10} – ROA
DEBT RATIOS R_{04} – Total Debts to Assets R_{05} – Debt to Equity R_{06} – Long-term debt to Equity	**PRODUCTIVITY RATIOS** R_{11} – Accounts Receivable Turnover R_{12} – Fixed Asset Turnover R_{13} – Total Asset Turnover R_{14} – Operation cost to sales revenues

Table 1. Financial indicators underlying the research.

To assess the financial standing of the construction companies in Poland, the following logit model was used:

$$P\left(y_i = \text{bankrupt} \mid \mathbf{x_i}\right) = \frac{\exp\left(\mathbf{x_i}\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{x_i}\boldsymbol{\beta}\right)} \tag{1}$$

where $y_i = \begin{cases} 1 & \text{if } i\text{-th enterprise had court decisions about bankruptcy} \\ 0 & \text{otherwise} \end{cases}$.

On its basis, the companies were classified into two groups: a group of financially sound companies (NB) and a group of companies at risk of bankruptcy (B). In the case of bankrupt companies, the fact that they were categorised as ‘B’ meant that they were at risk of recurrence of bankruptcy, while ‘NB’ category indicated that they were similar to financially sound companies in terms of financial standing.

In order to estimate parameters of the logit model, two kinds of approaches were applied: the maximum likelihood method (classical approach) and the Bayesian approach (see e.g. Koop and Poirier, 1993; Marzec, 2008). The selection of explanatory variables was performed with the use of the backward stepwise regression method. As a result, a maximum likelihood logit model (ML logit model) and a Bayesian logit model were constructed on the same set of explanatory variables.

In the classical approach, there were calculated point and interval estimations of the probability of bankruptcy. A company was classified into the group of companies at risk of bankruptcy if the point estimation of the probability obtained on the basis of the ML logit model was higher than 0.5. Moreover, there were constructed 95% confidence intervals for the probability of the bankruptcy risk. In this case, a company was classified into a group at risk of bankruptcy when the left boundary of the confidence interval was above the value of 0.5.

In Bayesian approach, in order to express the lack of prior knowledge, fairly diffuse prior independent N(0, 10) distributions were assumed. The random walk Metropolis algorithm was used to sample from the posterior distributions (Gamerman and Lopes, 2006). The starting points for numerical procedure equalled to the maximum likelihood estimators. Metropolis acceptance rates were (sufficiently) high and exceed 0.4. The final results and conclusions were based on 100,000 draws, preceded by 50,000 burn-in cycles.

It was assumed that if the point estimation of the median of the posterior probability distribution of bankruptcy was higher than 0.5, then the company was classified into the group at risk of bankruptcy. The classification based on the mean value of the posterior distribution led to the same conclusion. Furthermore, the study on the financial standing of the companies was also based on an analysis of the posterior distribution of the bankruptcy probability presented with the use of histograms.

The classification accuracy of the estimated logit models was assessed using the following measures (see e.g. Birdsall, 1973): the sensitivity calculated as a percentage of correctly classified bankrupts, the specificity – a percentage of correctly classified 'financially sound' enterprises (non-bankrupts) and the AUC measure – the area under the ROC curve.

The authors' previous studies suggest that cleaning a dataset of outliers (in this case of untypical objects among financially sound enterprises) before the estimation contributes to an improvement in the classification accuracy of the logit model (see Pawełek et al., 2015a, 2015b, 2015c, 2016). For this reason, before estimating parameters of the logit models, datasets were cleaned of outliers. In order to detect outliers, there were applied two one-dimensional methods based on a quantile analysis (Shumway, 2001) or Tukey's criterion (Tukey, 1977) and a multi-dimensional method based on a projection depth function (Zuo, 2003). Each of these methods was applied on the basis of all fourteen financial indicators considered or in conjunction with an analysis of discriminatory power (Yu et al., 2014), i.e. on the basis of financial indicators of a higher discriminatory power compared to others[5]. In total, six methods for detecting outliers were used.

In the following part of the article, there are discussed results obtained on the basis of sets cleaned of outliers using a quantile analysis, hence only a brief description of this method is provided here. The detection of outliers using a quantile analysis was carried out as follows. For every financial indicator from among the fourteen considered, quantile $q_{0.1}$ (in the case of a left-tailed asymmetry of distribution of a financial indicator), quantile $q_{0.9}$ (in the case of a right-tailed asymmetry) or quantiles $q_{0.05}$ and $q_{0.95}$ (in the absence of a asymmetry) were calculated in the group of financially sound enterprises. Objects (companies) with values of a given indicator lower than $q_{0.1}$, higher than $q_{0.9}$ or outside the double-sided range determined by quantiles $q_{0.05}$ and $q_{0.95}$ – depending on the type of the observed asymmetry – were considered outliers.

Measures of the classification accuracy of the estimated logit models were calculated on the complete dataset, i.e. taking into account the outliers, as well. It enabled us to maintain the comparability of the calculated measures among individual models constructed on different datasets cleaned with the use of various methods for detecting outliers.

[5] The way of using an analysis of discriminatory power, the method for cleaning a set of outliers using a quantile analysis and multi-dimensional projection depth function are discussed in Pawełek et al. (2015b). The method for cleaning a set of outliers using Tukey's criterion and multi-dimensional projection depth function are also discussed in Pawełek et al. (2015a).

When predicting bankruptcy on the basis of imbalanced datasets, it is recommended applying the sensitivity measure and the AUC measure because of a high percentage of non-bankrupts and a small percentage of bankrupts (García et al., 2015).

3. Empirical results

Once a dataset has been cleaned of outliers using one of the six methods mentioned above, in order to classify companies into the groups of 'financially sound' companies or companies 'at risk of bankruptcy', taking into account their financial standing, 12 ML logit models and 12 corresponding Bayesian logit models (6 classical models and 6 Bayesian models for each of the studied years of 2005 and 2009) were estimated.

From among the considered different logit models (i.e. estimated based on datasets cleaned of outliers using different methods), those models were selected for further analysis for which the highest sensitivity and AUC measures were recorded, taking into account the value of the specificity measure, as well. An analysis of values of these measures led to the conclusion that taking into consideration an analysis of discriminatory power of financial indicators when cleaning a set of outliers generally did not improve the classification accuracy of a model measured with the sensitivity and AUC measures. Models constructed on the basis of sets built with consideration given to a quantile analysis had, in general, the highest values of the sensitivity, specificity and AUC measures among other methods for cleaning a set of outliers. Table 2 shows values of measures of the classification accuracy for these logit models in ML approach (*MQ*.05 and *MQ*.09) and Bayesian approach (*MQB*.05 and *MQB*.09). These models were further analysed.

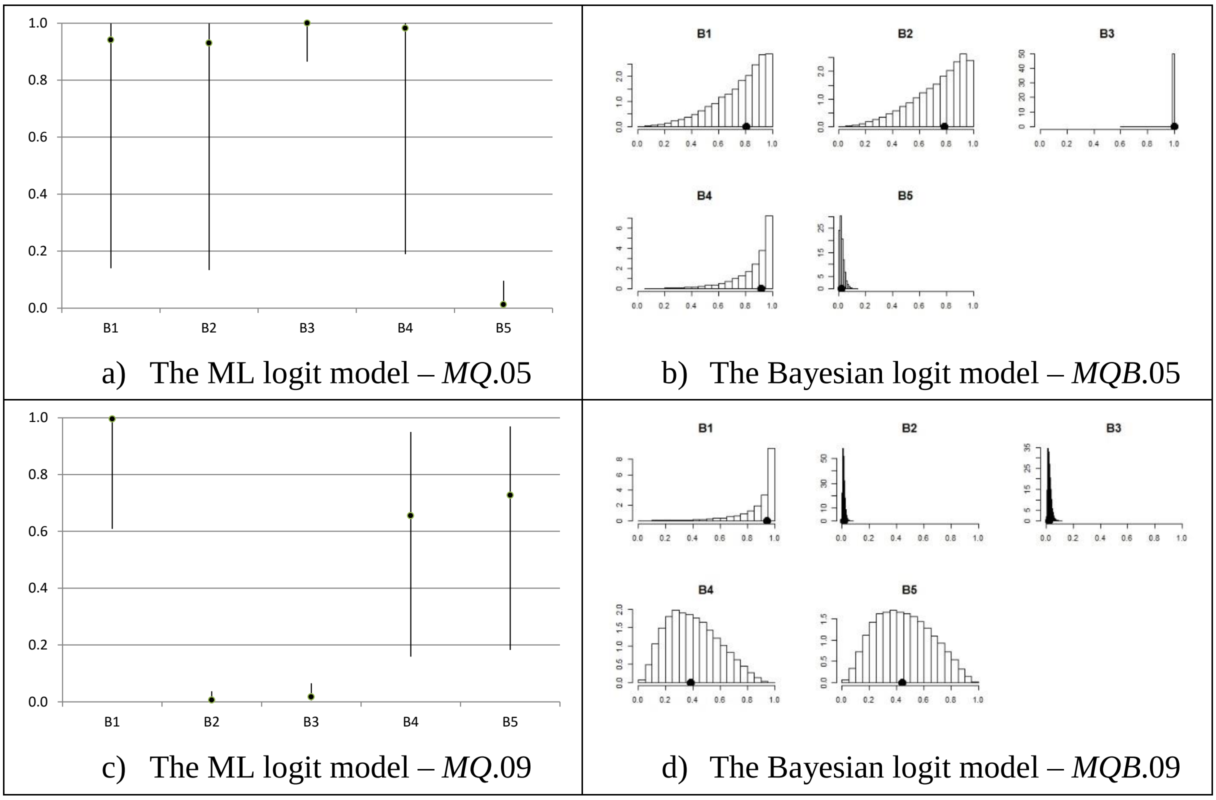
Model:	***MQ.yy***			***MQB.yy***		
Measure:	**sensitivity**	**specificity**	**AUC**	**sensitivity**	**specificity**	**AUC**
2005 (*yy*=05)	0.8	0.970	0.945	0.8	0.975	0.945
2009 (*yy*=09)	0.6	0.962	0.822	0.2	0.967	0.826

Table 2. Sensitivity, specificity and AUC measures for ML logit models and Bayesian logit models obtained on sets cleaned of outliers using a quantile analysis for the years 2005 and 2009.

On the basis of the information included in Table 2, it can be noted that the values of measures for ML logit models and Bayesian logit models are generally similar. The largest difference was observed for the sensitivity measure for the year 2009 where the company's assessment in terms of the risk of bankruptcy using the Bayesian logit model was more lenient than in the classical approach.

The purpose of the research is to assess the financial standing of five bankrupt companies in 2005, i.e. shortly after the bankruptcy, and a few years later – in 2009. Therefore, results of the classification of financially sound companies were omitted in the discussion.

Point estimations of the probability of bankruptcy and corresponding 95% confidence intervals obtained using the ML logit model were included for individual bankrupt companies B1–B5 in Fig. 1a and 1c. On the other hand, Fig. 1b and 1d include probability distributions and their medians obtained on the basis of the Bayesian logit model.



a) The ML logit model – *MQ*.05

b) The Bayesian logit model – *MQB*.05

c) The ML logit model – *MQ*.09

d) The Bayesian logit model – *MQB*.09

Fig. 1. Point estimations of the probability of bankruptcy obtained using the ML logit model for 2005 (a) or 2009 (c) and histograms of the posterior distribution of the probability of bankruptcy and their medians obtained using the Bayesian logit model for 2005 (b) or 2009 (d) for bankrupt companies B1–B5.

In 2005, point estimations of the probability obtained on the basis of the *MQ*.05 logit model showed the risk of bankruptcy in case of companies B1, B2, B3 and B4 (see Fig. 1a). This result is consistent with probability estimations at the median level of posterior distributions obtained in the Bayesian approach (*MQB*.05) (see Fig. 1b). Additional information is provided by an analysis of histograms of posterior distributions. The greatest risk of bankruptcy was recorded for company B3 (probability mass close to 1) and B4 (slightly lower concentration of the probability mass around 1). Histograms for B1 and B2 are also characterised by the probability mass shifted closer to the value of 1, but in a more dispersed manner (see Fig. 1b). Similar information can be interpreted on the basis of confidence intervals – broader, including the value of 0.5 in the case of companies B1, B2, and B4, and narrower and close to the value of 1 – in the case of B3 (see Fig. 1a).

In both approaches applied in 2005, company B5 was assigned to the group of financially sound companies in terms of the financial standing. The posterior distribution with the probability mass concentrated close to zero indicates the legitimacy of such an assessment, as well as a confidence interval with boundaries close to zero (see. Fig. 1a–1b).

Using the *MQ*.09 ML logit model estimated for 2009 (see Fig. 1c), companies B1, B4 and B5 were classified into the group of companies at risk of recurrence of bankruptcy on the basis of point estimations of probability. The constructed confidence intervals indicate decisively a high risk of recurrence of bankruptcy only in the case of company B1, whereas in the case of companies B4 and B5, they indicate their ambiguous financial standing (the intervals are rather wide and include the value of 0.5). Both point and interval estimations of the probability of recurrence of bankruptcy in the case of companies B2 and B3 indicate their similarity, in terms of financial standing, to financially sound enterprises.

Slightly different conclusions can be formulated on the basis of medians obtained as a result of using a Bayesian logit model (see. Fig. 1d). Medians of posterior distribution indicate a high risk of recurrence of bankruptcy only in the case of company B1, classifying other companies into the group of financially sound enterprises. Conclusions similar to those formulated on the basis of confidence intervals in the classical approach can be formulated when analysing histograms of posterior distributions of the probability of recurrence of bankruptcy (see Fig. 1d). The posterior distribution for company B1 with the probability mass concentrated closer to 1 corresponds to the confidence interval with the lower boundary above the value of 0.5. Posterior distributions for companies B2 and B3 with the probability mass concentrated around zero correspond to confidence intervals with lower boundaries below 0.5. Histograms of posterior distributions obtained for companies B4 and B5 are dispersed, which corresponds to the wider confidence intervals obtained in the classical approach. These results lead to assessing the financial standing of these companies as ambiguous.

Based on the obtained results, it can be noted that in 2005, four of the five companies were classified into the group at risk of recurrence of bankruptcy in both applied approaches, classical and Bayesian. Only in the case of B5, its financial standing was similar to the situation of financially sound enterprises. The situation of individual bankrupt companies in 2009 is summarised below.

In 2009, company B1 was still classified into the group at risk of bankruptcy in both approaches. The financial standing of companies B2 and B3 in 2009 was assessed as similar to financially sound enterprises. The financial standing of company B4 and, despite improvement recorded in 2005, of company B5 in 2009 was ambiguous, which was indicated by both an interval estimation in the classical approach and a histogram of the posterior distribution in Bayesian approach.

Conclusions

In view of the obtained results, the fact that in 2009, the financial standing of company B3, which had had a problem with solvency before 2005, was assessed as similar to that of financially sound enterprises is particularly valuable information. This result was obtained in both considered approaches: classical and Bayesian. Therefore, decisions made in this company may be a source of valuable information about what restructuring processes can lead to an improvement in the financial standing of a company with the problem of solvency. At the same time, it is important to remember that this company should be subject to further observation since bankruptcy may also recur in the following years.

On the other hand, decisions made especially in company B1, classified in both approaches into the group at risk of recurrence of bankruptcy, can provide valuable information on what strategies for company restructuring should be avoided.

If the analysis is complemented based on a generally applicable ML logit model with information obtained in Bayesian approach, the financial standing can be assessed not only in terms of classification into the group of companies ‘at risk of bankruptcy’ and the group of ‘financially sound’ enterprises. In particular, an analysis of the posterior distribution, including a distribution of the probability mass and a degree of the distribution dispersion, in some cases leads to the conclusion that the company’s situation is ambiguous – despite unambiguous point estimates. To a lesser degree, such information is provided by confidence intervals constructed on the basis on an ML logit model.

The conducted research confirmed previous observations of the authors as regards the effect of cleaning a dataset of outliers before estimating parameters of a logit model on an improvement in the classification accuracy.

It should be mentioned that in the obtained models, there were different sets of variables explaining the financial standing of companies in the context of the risk of bankruptcy, determined with the use of the backward stepwise regression method. An analysis of changes in the sets of variables in the subsequent years will be carried out on the basis of results for all the years of the period 2005-2009.

The presented results are based on a pilot study covering two years 2005 and 2009. The authors plan to analyse the complete path of companies after they have been declared bankrupt in the context of their financial standing in comparison with financially sound enterprises, i.e. to repeat the analysis for the years 2006-2008. The full research can provide valuable information about the process of getting out of the problem of insolvency by bankrupt companies. It seems that these first years after the declaration of bankruptcy by the court may be of particular importance in this process.

Acknowledgements

Publication was financed from the funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

References

Birdsall, T.G. (1973). The Theory of Signal Detectability: ROC Curves and Their Character. In *Technical Report*, No. 177. Ann Arbor, Michigan: Cooley Electronics Laboratory, Department of Electrical and Computer Engineering, The University of Michigan.

Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press.

García, V., Marqués, A. I., & Sánchez, J. S. (2015). An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. *Journal of Intelligent Information Systems, 44*(1), 159-189.

Koop, G., & Poirier, D. J. (1993). Bayesian analysis of logit models using natural conjugate priors. *Journal of Econometrics, 56*(3), 323-340.

Marzec, J. (2008). Bayesowskie modele zmiennych jakościowych i ograniczonych w badaniach niespłacalności kredytów. *Zeszyty Naukowe/Uniwersytet Ekonomiczny w Krakowie. Seria Specjalna, Monografie*, (188).

Pawełek, B., Gałuszka, K., Kostrzewska, J., & Kostrzewski, M. (2015a). Classification Methods in the Research on the Financial Standing of Construction Enterprises after

Bankruptcy in Poland. In A. Montanari, F. Palumbo & M. Vichi (Eds.), *Studies in Classification, Data Analysis, and Knowledge Organization*. Publication after Conference of the International Federation of Classification Societies IFCS 2015, 6-8 July 2015, Bologna, Italy, paper in review.

Pawełek, B., Kostrzewska, J., & Lipieta, A. (2015b). The Problem of Outliers in the Research on the Financial Standing of Construction Enterprises in Poland. In M. Papież& S. Śmiech, S. (eds.), *The 9th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*. Cracow: Foundation of the Cracow University of Economics, 164-173.

Pawełek, B., Pociecha, J., Kostrzewska, J., Baryła, M., & Lipieta, A. (2015c). Problem of Outliers in Corporate Bankruptcy Prediction. In *Archives of Data Science*. Publication after European Conference on Data Analysis ECDA 2015, 2-4 September 2015, Colchester, UK. Paper in review.

Pawełek, B., Pociecha, J., Kostrzewska, J., Baryła, M., & Lipieta, A. (2016). Problem wartości odstających w prognozowaniu zagrożenia upadłością przedsiębiorstw (na przykładzie przetwórstwa przemysłowego w Polsce). In K. Jajuga & M. Walesiak (Eds.), *Taksonomia. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*. Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu. In press.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business J BUS, 74*(1), 101-124.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley Pub.

Yu, Q., Miche, Y., Séverin, E., & Lendasse, A. (2014). Bankruptcy prediction using Extreme Learning Machine and financial expertise. *Neurocomputing, 128*, 296-302.

Zuo, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist. The Annals of Statistics, 31*(5), 1460-490.