Przegląd Statystyczny. Statistical Review, 2020, vol. 67, 1, 5–32 DOI: 10.5604/01.3001.0014.1783

Evaluation of the financial condition of companies after the announcement of arrangement bankruptcy: application of the classical and Bayesian logistic regression

Barbara Pawełek,[a] Jadwiga Kostrzewska,[b] Maciej Kostrzewski,[c] Krzysztof Gałuszka[d]

Abstract. The aim of this paper is to present the results of an assessment of the financial condition of companies from the construction industry after the announcement of arrangement bankruptcy, in comparison to the condition of healthy companies. The logistic regression model estimated by means of the maximum likelihood method and the Bayesian approach were used. The first achievement of our study is the assessment of the financial condition of companies from the construction industry after the announcement of bankruptcy. The second achievement is the application of an approach combining the classical and Bayesian logistic regression models to assess the financial condition of companies in the years following the declaration of bankruptcy, and the presentation of the benefits of such a combination. The analysis described in the paper, carried out in most part by means of the ML logistic regression model, was supplemented with information yielded by the application of the Bayesian approach. In particular, the analysis of the shape of the posterior distribution of the repeat bankruptcy probability makes it possible, in some cases, to observe that the financial condition of a company is not clear, despite clear assessments made on the basis of the point estimations.

Keywords: company, arrangement bankruptcy, financial condition, Maximum Likelihood Method, Bayesian approach, logistic regression

JEL: C11, C25, G33

1. Introduction

Company bankruptcy is an important issue for economic sciences. The establishment of new companies and the termination of business activity by some existing companies are natural phenomena in a free-market economy. Nevertheless, the situation where a company announces bankruptcy, and, consequently, discontinues its business activity has been the object of research carried out by scientists, economic practitioners, and financial institutions. This increased interest might be explained by serious social and economic consequences of bankruptcies, therefore there is a need for developing methods of bankruptcy prediction.

[a] Cracow University of Economics, Department of Statistics, e-mail: barbara.pawelek@uek.krakow.pl (corresponding author), ORCID: https://orcid.org/0000-0002-9589-6043.

[b] Cracow University of Economics, Department of Statistics, e-mail: jadwiga.kostrzewska@uek.krakow.pl, ORCID: https://orcid.org/0000-0002-6142-9156.

[c] Cracow University of Economics, Department of Econometrics and Operational Research, e-mail: maciej.kostrzewski@uek.krakow.pl, ORCID: https://orcid.org/0000-0003-1544-4221.

[d] University of Economics in Katowice, Department of Public Finance, ORCID: https://orcid.org/0000-0002-5851-1300.

The paper by Sun et al. (2014) contains a review of about 140 publications dated 1966–2014 on bankruptcy prediction problems, with respect to e.g. the definition of disadvantageous financial condition of a company and company bankruptcy. The authors emphasised the diversity of definitions of the above-mentioned phenomena. They also demonstrated that theoretical arguments define three levels of a disadvantageous financial condition, while empirical studies are usually limited to the analysis of just two conditions: a healthy company and a bankrupt company.

Bankruptcy is awarded to a company by a court. In the period from 1 October 2003 to 31 December 2015, a court could decide on two types of bankruptcy with regard to insolvent companies: bankruptcy open to arrangements and liquidation bankruptcy.[1] The former enables a company to continue its business activity on condition that it complies with procedures enabling it to pay off as much of its debt as possible.

In the meantime, the company bankruptcy law in Poland was amended to the effect that now, each bankruptcy is announced as business liquidation. However, the new regulation,[2] in force since 1 January 2016, foresees, instead of arrangement bankruptcy, four new procedures leading to an arrangement between the debtor and the creditors, thus giving debtors more opportunities to solve their problems. Additionally, the legislator introduced the possibility of an arrangement in the case of insolvency, which is applied in order to give a company a chance to survive in a situation where opening or continuing a restructuring procedure is impossible. The aim of the new restructuring law is to increase the efficiency of procedures leading to an arrangement between the debtor and the creditors.

Predicting bankruptcy is a frequent topic in literature describing the application of multidimensional statistical analysis to the business sector. However, only few papers focus on the statistical assessment of the financial condition of companies in the years following the declaration of bankruptcy. Examining the process of overcoming insolvency issues might become a valuable source of information that is useful in the assessment of a success probability in the case of the accomplishment of restructuring proposals by subsequent bankrupts.

The aim of this paper is to present the results of the assessment of the financial condition of companies from the construction industry after the announcement of arrangement bankruptcy in comparison to the condition of healthy companies. The logistic regression model estimated by means of the maximum likelihood method and the Bayesian approach were used for this purpose. The research hypothesis claims that the application of the Bayesian approach to the logistic regression model

[1] Ustawa z dnia 28 lutego 2003 r. – *Prawo upadłościowe*, Dz.U. 2003 nr 60 poz. 535.
[2] Ustawa z dnia 15 maja 2015 r. – *Prawo restrukturyzacyjne*, Dz.U. 2015 poz. 978.

guarantees the enrichment of conclusions on a company's post-arrangement-bankruptcy financial condition drawn on the basis of the logistic regression model estimated by means of the maximum likelihood method. The first achievement of our study is the assessment of the financial condition of companies from the construction industry in relation to which arrangement bankruptcy has been announced. This means they try to deal with their solvency issues following the bankruptcy arrangements affirmed by a court. The second achievement involves applying a combination of the classical and Bayesian logistic regression models in the process of assessing the financial condition of bankrupt companies in the years following the announcement of bankruptcy, and presenting the benefits of such a combination. The results of our pilot research for the years 2005 and 2009 were previously presented in Kostrzewska et al. (2016).

The remaining part of the paper is organised in the following way: Section 2 presents the state of research on bankrupt companies, Section 3 describes the process of data preparation, Section 4 details the methodology of the empirical examination, Section 5 shows the results of calculations, their interpretation and graphical presentation, Section 6 discusses the received results in the light of knowledge on further histories of the analysed bankrupt companies, Section 7 summarises the reflections on the topics covered in the paper. The list of literature referred to in the work is included at the end of the paper.

2. State of research

Scientists have been researching efficient bankruptcy prediction methods since as early as the 20th century. The most popular bankruptcy prediction methods include for instance: the linear discriminant function (e.g. Altman, 1968; García et al., 2019; Lee and Choi, 2013), the logit model (e.g. Ohlson, 1980; Li and Wang, 2014; Tseng and Hu, 2010), the classification tree (e.g. Frydman et al., 1985; Abellán and Castellano, 2017; Tsai et al., 2014), the neural network (e.g. Odom and Sharda, 1990; López et al., 2015; Tkáč and Verner, 2016), the support vector machine (e.g. Liang et al., 2015; Sun et al., 2017; Zhou et al., 2015), the hazard model (e.g. Beaver et al., 2005; Beaver et al., 2012; Shumway, 2001), and the ensemble method (e.g. Ekinci and Erdal, 2017; Pawełek, 2019; Zhou and Lai, 2017). The above-mentioned types of analyses are based on sets of financial data obtained from healthy and bankrupt companies dated usually a year or two before the bankruptcy occurred. The Bayesian approach in forecasting bankruptcy was also applied and described in literature (e.g. Sarkar and Sriram, 2001; Sun and Shenoy, 2007; Trabelsi et al., 2015), whereas the Bayesian logistic regression has not yet been used in the presented context, as far as we know.

Scientific literature on company bankruptcies also discusses issues relating to insolvent companies' activity ensuing their bankruptcy. For instance, Eberhart et al. (1999) examined reactions of the capital market to bankruptcies of listed companies. In their analysis, the authors applied, for example cumulative abnormal returns to analyse changes in share prices of bankrupt companies in a period of 200 days following the bankruptcy announcement.

Another important issue is the assessment of the risk of repeat bankruptcy of companies (e.g. Platt and Platt, 2002; Altman and Branch, 2015). Analyses are conducted on the basis of financial data of bankrupt companies. In order to predict the threat of repeat bankruptcy, Platt and Platt (2002) used the logit model. The research set consisted of 51 bankrupt companies, 9 out of which were subject to bankruptcy procedures for the second time. The model included three explanatory variables: annual abnormal return, net sales one year after re-emerging from bankruptcy, and the number of months spent in the first bankruptcy process. The classification accuracy of the model was high, i.e. 90.2% in total, 88.9% in the group of repeat bankrupts and 90.5% in the group of bankrupts continuing their business activity. On the basis of the estimated model, it was demonstrated that a decrease in annual abnormal return, a decrease in net sales one year after re-emerging from bankruptcy, and an increase in the number of months spent in the first bankruptcy process might increase the probability of repeat bankruptcy *ceteris paribus*. Platt and Platt (2002) presented the similarities and differences between their research and the results presented in the work of Hotchkiss (1995).

Altman and Branch (2015) measured the usefulness of the Altman Z-Score formula intended for predicting the success or failure of a company's post--bankruptcy business activity. The linear discriminant function was determined on the basis of the following financial indicators: Current Assets – Current Liabilities / Total Assets, Retained Earnings / Total Assets, EBIT / Total Assets and Book Value of Equity / Total Liabilities. In their research, the authors analysed two groups of companies – one group that consisted of companies which experienced problems with continuing their business activity only once, and the other comprising businesses which had become bankrupt twice. The research set consisted of 148 bankrupt companies, 61 of which had difficulty in continuing their business activity. According to the authors, both courts and persons responsible for restructuring of companies should employ statistical methods to predict whether a company is prone to repeat bankruptcy as a form of supplementing the traditional analysis. Such methods may facilitate the assessment of the restructuring plan and the monitoring of the post-bankruptcy condition of a company in order to adjust the plan accordingly. A return to court means that the restructuring failed in terms of the concept, moreover generating social and economic costs. The authors emphasised the importance of an early-warning system which would reduce the probability of the occurence of repeat bankruptcy, often preceded by a long and costly restructuring process.

An alternative approach to the above-mentioned concepts is the statistical evaluation of the financial condition of companies in the years following the announcement of arrangement bankruptcy compared against the condition of healthy companies (e.g. Kostrzewska et al., 2016; Pawełek et al., 2017). Within this approach, analyses are based on financial data of both healthy companies and companies which have announced bankruptcy. Thus, this solution makes it possible to research methods of solving bankruptcy problems by companies when data sets on bankrupt companies are not comprehensive enough. This approach might prove helpful in the selection of an appropriate recovery programme for companies experiencing solvency problems. However, caution is recommended when interpreting the results. It is because provisions of the tax law, policies of financial institutions, etc. are likely to influence the assessment of the economic condition of a company and its further existence on the market. Additionally, differences in accounting regulations, particularly in terms of the interpretation of international accounting standards, make it difficult – or even impossible – to compare data between countries. As a consequence, the examination of the financial condition of companies after arrangement bankruptcy must be conducted for each country separately.

Moreover, no or limited comparability may occur within one country, as some companies are subject to audits of their financial statements, whereas others are not obliged to do so. In Poland there are companies with bad financial indicators that do not go bankrupt, but are doing well due to tax provisions allowing the use of financial losses when a company merges with another company with losses. On the other hand, financial institutions are distrustful of companies with bad financial indicators and seek to receive their funds quickly by means of arrangement bankruptcy. Therefore, a company in a poorer financial condition may survive while a company in a better financial condition may go bankrupt if the latter has higher debts in financial institutions. What was said above shows the difficulties in assessing the financial condition of companies after arrangement bankruptcy is announced.

3. Data processing

3.1. Database

The data used for the analysis have been collected from the Emerging Markets Information Service (EMIS).[3] The research covered 369 construction companies in

[3] http://www.emis.com.

Poland, including five companies undergoing arrangement bankruptcy (B_1–B_5). The court awards were issued between November 2003 and August 2004. The research used all 14 financial ratios available in the EMIS database which have been grouped as follows: liquidity ratios (R_{01}–R_{03}), liability ratios (R_{04}–R_{06}), profitability ratios (R_{07}–R_{10}), and productivity ratios (R_{11}–R_{14}) (Table 1).

The financial data come from the years 2005–2009. The same legal bankruptcy and recovery provisions were in force in Poland in the period when companies under research were announcing arrangement bankruptcy and in the years for which companies' financial statements were available.

Table 1. Financial ratios

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

Note. For interested readers, descriptive statistics of financial ratios for both analysed groups are available upon request.
Source: based on information from the Emerging Markets Information Service.

The empirical research was conducted on the basis of an unbalanced set. Sets of this kind, more often than balanced sets, are characterised by low accuracy of classification of bankrupt companies in the considered bankruptcy prediction methods. The above-mentioned phenomenon may result from a small proportion of bankrupts in the examined sets as well as, for instance, from the existence of untypical objects among healthy companies (e.g. Pawełek et al., 2015). An untypical company is understood by the authors as an object with outlying values of financial ratios. The studies conducted so far showed that the removal of these outliers (in this case the untypical objects among healthy companies) from a data set before performing estimations raises the classification accuracy of the logistic regression model (e.g. Pawełek et al., 2015).

Thus, the preparation of the data for analysis includes also the detection of outliers. An outlier is an element that seems to be considerably different from other elements of a set in which it is included (e.g. Grubbs, 1969; Barnett and Lewis, 1994; Hodge and Austin, 2004; John, 1995). An outlier is often an element at an incomparably larger distance from other values in the set than the distances between other elements of the set (Johnson and Wichern, 1992). Papers on the examination of the financial condition of companies, especially those focusing on bankruptcy prediction, might discuss the presence of outliers. The proposed solutions of this problem include ignoring it (e.g. Spicka, 2013), replacement or deletion of outliers (e.g. De Andrés et al., 2011; Shumway, 2001; Wu et al., 2010), or the application of robust methods (e.g. Hauser and Booth, 2011).

Untypical healthy companies, defined as above, may be characterised by both a very good and a bad financial condition, which, with regard to numerous indicators, is similar to the condition of companies in announced bankruptcy. The detection and deletion of untypical healthy companies in/from a set of objects also has a substantive justification. Economic practice shows that companies in a bad financial condition (i.e. whose financial ratios are of disadvantageous values) might not be able to fulfil the prerequisites necessary to initiate bankruptcy proceedings or the obligation to file a bankruptcy petition. If no such petition is filed by creditors either, such companies exist on the market and affect the condition of the entire sector.

3.2. Detection of outliers

Due to the above, outliers were deleted from the data sets prior to the estimation of the parameters of logistic regression models. To detect outliers, two one-dimensional methods were used, based on the quantile analysis or the Tukey criterion (Tukey, 1977) and a multidimensional method based on the projection depth function (Zuo, 2003). Each method was applied on the basis of all 14 discussed financial ratios, or – in the case of the one-dimensional methods – in combination with the discriminatory power analysis (e.g. Yu et al., 2014), which uses financial ratios with stronger discriminatory power than others. Altogether, five methods were applied to detect untypical healthy companies. When interpreting the results of the analysis, it should be remembered that the financial ratios of typical healthy companies reflected the financial condition of the construction industry in a given year, which depended, for instance on the economic situation in Poland.

The procedure based on the Tukey criterion required calculations of the first quartile ($Q_{0.25}$) and the third quartile ($Q_{0.75}$) for every financial ratio in the group of healthy companies, and then the calculation of the interquartile range ($Q = Q_{0.75} - Q_{0.25}$). Values outside of the range of $<Q_{0.25} - 1.5Q, Q_{0.75} + 1.5Q>$ were considered outliers. A healthy company was considered untypical if at least one financial ratio value was an outlier.

The detection of an outlier by means of the quantile analysis was performed in the following way: the values of quantile $Q_{0.10}$ (in the case of strong left-sided asymmetry of the financial ratio distribution), quantile $Q_{0.90}$ (in the case of strong right-sided asymmetry) or quantiles $Q_{0.05}$ and $Q_{0.95}$ (if there was no strong asymmetry) were defined for every financial ratio in the group of healthy companies. If a company had values of a given ratio lower than $Q_{0.10}$, higher than $Q_{0.90}$ or falling outside the range defined by quantiles $Q_{0.05}$ and $Q_{0.95}$, depending on the type of asymmetry observed, it was considered an outlier. As in the case of the Tukey criterion, a healthy company was considered untypical if at least one value of the financial ratios was an outlier.

The projection depth function (Zuo, 2003) was applied to detect outliers in a multi-dimensional space. The concept of data depth is related to the nonparametric robust multi-dimensional statistical analysis developed within the scope of exploratory data analysis. This enables the determination of a linear order of multi-dimensional observations with the use of a multi-dimensional median defined as a multi-dimensional centre of a set of observations (Zuo and Serfling, 2000). There are numerous depth functions available which assign a positive number to each observation derived from a certain distribution, determining its distance from the centre (e.g. Kosiorowski, 2008). We assumed arbitrarily that outlier healthy companies are the 10% of items located at the greatest distance from a multi--dimensional centre. Here, it has to be remembered that the projection depth function used to detect outliers is a method which indicates items at a large distance from the centre of the data set, regardless of the direction of these items' location (i.e. the group of the above-mentioned outlier companies may include both companies in a very good financial condition and companies experiencing serious financial problems).

3.3. Discriminatory power of financial ratios

In order to establish which ratios are best in signalling a company's deteriorating financial condition, their discriminatory power was determined. The first approach was based on ratio distribution quantiles (e.g. Yu et al. 2014). The number of bankrupts belonging to the 10-percent range of extreme values for healthy companies (in the right or left tail of the distribution) was assumed as the criterion. When values adopted by the bankrupts were present in both tails of the distribution for healthy companies, then the percentage of bankrupts was checked in the two-sided 10-percent range defined by quantiles $Q_{0.05}$ and $Q_{0.95}$. The higher the value of this criterion, the higher the discriminatory power of a given ratio. If in one of the areas defined by quantiles for healthy companies the ratio values were defined for at least two out of five bankrupt companies (which was a very mild criterion), then the ratio was considered as having a discriminatory power. The sets of financial ratios with a discriminatory power differed throughout the analysed years (Table 2).

In particular, at the beginning of the examined period, that is in 2005, as many as 12 ratios had a discriminatory power (calculated according to the adopted criterion), while in the three subsequent years – only 9 to 4 ratios. It means that the values of the financial ratios of construction companies shortly after the announcement of arrangement bankruptcy were in the tails of the distributions of ratios for healthy companies. The longer the time passed since the bankruptcy announcement, the fewer ratios had the discriminatory power and, consequently, the fewer bankrupt companies were characterised by extreme values of the financial ratios. The number of ratios with a discriminatory power increased from 4 variables in 2008 to 7 variables in 2009, which may have resulted from the global financial crisis. In Poland, its repercussions included the deterioration of conditions for business activity within the construction industry in 2009.

Table 2. Financial ratios with higher discriminatory power than other discussed ratios – the analysis based on the asymmetry and quantiles of distribution

Year	Financial ratios	m
2005	R_{01} R_{02} R_{04}* R_{05} R_{07} R_{08} R_{09}** R_{10} R_{11} R_{12} R_{13} R_{14}**	12
2006	R_{04}* R_{05} R_{06} R_{07}** R_{08}** R_{10}** R_{11} R_{12} R_{14}*	9
2007	R_{05} R_{06}** R_{09} R_{14}*	4
2008	R_{03}** R_{05} R_{06}** R_{09}	4
2009	R_{03} R_{07} R_{08} R_{09}**R_{10} R_{11} R_{14} *	7

Note. m – number of ratios, * – discriminatory power in the right tail, ** – discriminatory power in both tails, other – discriminatory power in the left tail.
Source: authors' calculation.

The financial ratios with a higher discriminatory power than the other examined ratios (calculated according to the adopted criterion – Table 2) in individual years usually include representatives of three out of four groups. The only exception is the year 2005, i.e. shortly after the announcement of arrangement bankruptcy, when selected ratios from all four examined groups had a discriminatory power. Liquidity ratios did not occur in sets specified for the years 2006 and 2007. Liability ratios were missing for 2009, while productivity ratios were not specified for 2008. Selected profitability ratios were included in all sets.

The second approach was based on the Tukey criterion with the hinge distance factor equal to 1.5. The number of bankrupts with ratio values belonging to the specified ranges of extreme values for healthy companies was the adopted criterion. The higher the value of this criterion, the bigger the discriminatory power of a given ratio. The financial ratios with discriminatory power were those for which the values of the ratios were recorded for at least two out of five bankrupt companies in areas defined by the Tukey criterion with the hinge distance factor equal to 1.5 for healthy companies. At the beginning of the examined period, i.e. in 2005, only six financial ratios had a discriminatory power (calculated in accordance with the adopted criterion). In the following years, the number of ratios with a discriminatory power decreased from three variables in 2006 and 2007 to one variable in 2008 and 2009. The approach based on the Tukey criterion with the hinge distance factor equal to 1.5 was more restrictive than the approach based on the quantiles of ratio distribution.

Table 3. Financial ratios with higher discriminatory power than other discussed ratios – the analysis based on the Tukey criterion

Year	Financial ratios	m
2005	R_{05} R_{07}** R_{08}** R_{09}** R_{10} R_{14}**	6
2006	R_{05} R_{06} R_{08}**	3
2007	R_{05} R_{066}** R_{09}**	3
2008	R_{06}**	1
2009	R_{14}*	1

Note. See Table 2.
Source: authors' calculation.

There were no liquidity ratios with a higher discriminatory power than other analysed ratios (Table 3) in the examined years in the case of the Tukey criterion. Liability ratios did not occur in the set specified for 2009. Profitability ratios were missing in 2008 and 2009, while the productivity ratios were not specified for the years 2006–2008.

3.4. Sets of companies used in logistic regression models

To determine outliers for individual years, the one-dimensional analysis method based on the Tukey criterion and the quantile analysis, as well as the multidimensional method based on the projection depth function were used. The methods were indicated with their first letters (T meant the analysis based on the Tukey criterion, Q – the quantile analysis, D – the projection depth function), which were followed by a two-digit symbol of a year (05, 06, 07, 08, and 09) and the number of financial ratios used to detect outliers (m = 4, 7, 9, 12, 14 in the analysis based on quantiles; m = 1, 3, 6, 14 in the analysis based on the Tukey criterion). If m does not equal 14, it means that the discriminatory power of financial ratios was taken into account in a given method.

In further analyses, sets established by means of the above-mentioned methods were examined. The following variants were adopted:

- variant I – sets including all companies in a given year (complete database),
- variant II – sets including all bankrupts and healthy companies other than untypical items for a given method in a given year.

The number of elements in the sets in variant II for individual years are presented in Table 4.

Table 4. Numbers of elements in the sets in variant II depending on the method applied and the number of financial ratios (m)

Method	Variant II				
	2005	2006	2007	2008	2009
$T.yy.14$	188 ($m = 14$)	205 ($m = 14$)	197 ($m = 14$)	176 ($m = 14$)	188 ($m = 14$)
$T.yy.m$	292 ($m = 6$)	320 ($m = 3$)	309 ($m = 3$)	321 ($m = 1$)	353 ($m = 1$)
$Q.yy.14$	174 ($m = 14$)	190 ($m = 14$)	167 ($m = 14$)	169 ($m = 14$)	179 ($m = 14$)
$Q.yy.m$	213 ($m = 12$)	247 ($m = 9$)	269 ($m = 4$)	272 ($m = 4$)	268 ($m = 7$)
$D.yy.14$	333 ($m = 14$)	333 ($m = 14$)	333 ($m = 14$)	333 ($m = 14$)	333 ($m = 14$)

Note. $T.yy.14$ ($T.yy.m$) – the method based on the Tukey criterion applied for data from the year $20yy$ and concerning 14 (m) financial ratios, $Q.yy.14$ ($Q.yy.m$) – the method based on quantiles applied for data from the year $20yy$ and concerning 14 (m) financial ratios, $D.yy.14$ – the method based on the projection depth function applied for data from year $20yy$ and concerning 14 (m) financial ratios.
Source: authors' calculation.

In the case of the one-dimensional analysis based on the Tukey criterion or the quantile analysis including/not including information on the discriminatory power of financial ratios for different years, various numbers of elements in variant II (Table 4) were received, as, due to the values of different ratios, the outliers do not necessarily need to be the same. In the case of the multi-dimensional analysis based on the projection depth function, sets containing the same numbers of elements in variant II (333 companies) were obtained for different years. However, in the case of both the one-dimensional methods (depending on whether the discriminatory power of ratios was taken into account) and the multi-dimensional analysis, the same items do not necessarily have to be outliers. Thus, differences may occur in logistic regression models assessed using knowledge based on outliers that have been obtained through various methods.

4. Research methodology

To assess the financial condition of construction companies in Poland, the following logistic regression model was applied:

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

where:
y_i – the dependent variable for the i-th object,
$\mathbf{x_i}$ – the vector of explanatory variables for the i-th object,
$\boldsymbol{\beta}$ – the vector of parameters,
ε_i – the error term.

On this basis, the companies were classified into two groups: a group of entities in a financial condition typical for healthy companies (i.e. non-bankrupt – 'NB' Group) and a group of entities in a financial condition typical for companies with announced arrangement bankruptcy (i.e. bankrupt – 'B' Group). Classification into the 'B' Group in the case of bankrupt companies meant that they were prone to repeat bankruptcy, while companies classified into the 'NB' Group were in a financial condition similar to that of healthy companies.

To estimate the parameters of the logistic regression model, the maximum likelihood method and the Bayesian approach were used. Explanatory variables were selected by means of the backward stepwise regression method. As a result, the maximum likelihood logistic regression model (ML logistic regression model) and the Bayesian logistic regression model were built on the same sets of explanatory variables. These approaches were applied jointly in order to supplement the results obtained in the classical approach with the information obtained through the Bayesian approach.

As far as the ML logistic regression model is concerned, the authors made point and interval estimations of the probability of a company falling into a group of items in a financial condition typical for companies with announced arrangement bankruptcy, i.e. a group of items threatened with first or repeat bankruptcy (hereinafter referred to as 'bankruptcy probability'). For each company, point estimates of bankruptcy probability were assessed using respective values of financial ratios within the ML logistic regression model. Interval estimates of bankruptcy probability for every company were obtained in compliance with the methodology described in the paper by Neter et al. (1989). The logit transformation was used for the calculations.

A company was classified into the group of entities exposed to the risk of bankruptcy if the point estimate of bankruptcy probability calculated on the basis of the ML logistic regression model was higher than 0.5. In the case of 95% confidence intervals for bankruptcy probability, a company was considered prone to bankruptcy if the lower bound of the confidence interval was above 0.5. If the upper bound of the confidence interval was below 0.5, a company was considered to be in a financial condition typical for healthy companies in a given year. However, if the lower bound of the confidence interval was below 0.5, while the upper was above 0.5, the financial condition of a given entity was considered ambiguous.

The Bayesian inference is based on the posterior distribution, which combines prior knowledge and information provided by data within a mathematical model. The approach can be applied regardless of the size of the data set. Moreover, it takes into account the uncertainty regarding unknown parameters. Therefore, the Bayesian approach provides the distribution of an unknown quantity, unlike a single point or interval estimate, which is the case for the maximum likelihood method.

In the Bayesian approach, in order to express the lack of prior knowledge for the individual parameters of the logistic regression model, relatively uninformative prior independent normal distributions $N(0, 10)$ were assumed. The expected value equal to zero corresponded to the assumption concerning the statistical insignificance of the model parameters. In other words, neither negative nor positive signs of the parameters were priorly preferable, while the standard deviation of 10 made the prior distribution diffuse. The random walk Metropolis-Hastings algorithm was used to sample from the posterior distribution of the parameters (Gamerman and Lopes, 2006). The starting points of the algorithm were assumed at the level of the maximum likelihood estimates. The acceptance rates of the numerical algorithm were high and exceeded 40 percent. The presented results of the Bayesian inference were based on 100,000 MCMC draws, preceded by 50,000 burn-in Markov chain cycles.

It was assumed that if the point estimate of the median of the posterior distribution of the bankruptcy probability was higher than 0.5, such a company was classified into the group of items threatened with first or repeat bankruptcy. The classification on the basis of the expected value of the posterior distribution led to the same conclusions. The examination of the financial condition of companies was also based on a graphical analysis of the posterior distribution of the probability of a company belonging to the 'B' Group, presented by means of histograms.

The classification accuracy of the estimated logistic regression models was assessed by means of the following measures (e.g. Birdsall, 1973; Fawcett, 2006; Krzanowski and Hand, 2009): *sensitivity* – calculated as a percentage of companies in announced arrangement bankruptcy that were classified in the 'B' Group; *specificity* – calculated as a percentage of healthy companies that were classified in the 'NB' Group; *AUC measure* – the area under the ROC curve. Due to a high percentage of healthy companies and a small percentage of bankrupts, the *sensitivity* and *AUC measures* were chosen as the classification accuracy measures, recommended in literature for imbalanced sets (García et al., 2015; Kostrzewska et al., 2016). The measures of the classification accuracy were calculated on the basis of a full data set (variant I), i.e. including the outliers. Therefore, it was possible to maintain the comparability of the calculated measures between individual models built on the basis of various data sets cleaned by means of various methods of outlier detection.

5. Empirical results

Following the removal of outliers from the data set by means of any of the above-mentioned methods, in order to classify the companies according to the financial condition either into the ‘NB’ Group (i.e. of entities whose financial condition resembles that of a healthy company) or the ‘B’ Group (i.e. of entities in a financial condition typical for companies in announced arrangement bankruptcy), the authors estimated 25 ML logistic regression models and 25 corresponding Bayesian logistic regression models (5 ML models and 5 Bayesian models for every year in the period 2005–2009). As mentioned before, the reduction of the input set of explanatory variables was performed by means of the backward stepwise regression method. The list of variables left in the logistic regression model estimated for individual years in the period 2005–2009 is presented in Table 5.

Table 5. Financial ratios in the logistic regression models after reduction of the input set of explanatory variables

Model	Explanatory variables in the logistic regression model				
	2005	2006	2007	2008	2009
$MT.yy.14$ $MTB.yy.14$	R_{04}	R_{05} R_{09}	R_{05} R_{06}	R_{05} $R_{06}R_{10}$ R_{14}	R_{06} R_{14}
$MT.yy.m$ $MTB.yy.m$	R_{10} R_{14}	R_{05}	R_{05} R_{06}	R_{06}	R_{14}
$MQ.yy.14$ $MQB.yy.14$	R_{04}	R_{04} R_{05} R_{08}	R_{05} R_{06}	R_{05} $R_{06}R_{10}$ R_{14}	R_{14}
$MQ.yy.m$ $MQB.yy.m$	R_{04}	R_{04} R_{05} R_{08}	R_{05} R_{06}	R_{05} R_{06}	R_{14}
$MD.yy.14$ $MDB.yy.14$	R_{04}	R_{04} R_{05} R_{08}	R_{05} $R_{06}R_{08}$ R_{14}	R_{05} $R_{06}R_{11}$ R_{14}	R_{09}

Note. $MT.yy.14$ ($MTB.yy.14$) – the ML logistic regression model (the Bayesian logistic regression model) built on the basis of data cleaned with a method based on the Tukey criterion dated $20yy$ and referring initially to 14 financial ratios, $MT.yy.m$ ($MTB.yy.m$) – the ML model (the Bayesian model) built on the basis of data cleaned with a method based on the Tukey criterion dated $20yy$ and referring initially to m financial ratios, $MQ.yy.14$ ($MQB.yy.14$) – the ML model (the Bayesian model) built on the basis of data cleaned with a method based on quantiles dated $20yy$ and referring initially to 14 financial ratios, $MQ.yy.m$ ($MQB.yy.m$) – the ML model (the Bayesian model) built on the basis of data cleaned with a method based on quantiles dated $20yy$ and referring initially to m financial ratios, $MD.yy.14$ ($MDB.yy.14$) – the ML model (the Bayesian model) built on the basis of data cleaned with a method based on the projection depth function dated $20yy$ and referring initially to 14 financial ratios.
Source: authors’ calculation using the *STATISTICA* software.

The analysis of the set of financial ratios with a statistically significant impact on the probability of belonging to the group of entities in a financial condition typical for companies in announced arrangement bankruptcy led to the conclusion that the role of liability ratios was of particular importance – mainly in the years 2005–2008.

The liquidity ratios did not have a significant impact on the bankruptcy probability in the analysed group of companies and in the reviewed period. Profitability ratios were present every year but for different variants of models. The role of productivity ratios was of particular importance for the assessment of bankruptcy risk in the years 2008 and 2009.

When predicting bankruptcies on the basis of unbalanced sets, it is advisable to use the *sensitivity* measure followed by the AUC measure first, due to a high percentage of healthy companies and a low percentage of bankrupts. The high values of the *specificity* measure may result from a large share of healthy companies in the sample. Tables 6 and 7 present the values of *sensitivity* and AUC measures calculated for the logistic regression model prepared by means of the maximum likelihood method and the Bayesian approach.

Table 6. Sensitivity measure for the ML and the Bayesian logistic regression models estimated on the basis of sets without outliers

Model	Sensitivity				
	2005	2006	2007	2008	2009
$MT.yy.14$	**0.8**	**0.6**	**0.6**	**0.8**	0.4
$MTB.yy.14$	**0.8**	**0.6**	**0.6**	0.4	0.2
$MT.yy.m$	0.2	0.4	**0.6**	0.2	0.4
$MTB.yy.m$	0.4	0.4	**0.6**	0.2	0.2
$MQ.yy.14$	**0.8**	**0.8**	**0.6**	**0.6**	**0.6**
$MQB.yy.14$	**0.8**	**0.8**	**0.6**	0.4	0.2
$MQ.yy.m$	**0.8**	**0.8**	0.4	0.2	**0.6**
$MQB.yy.m$	**0.8**	**0.8**	0.4	0.2	0.4
$MD.yy.14$	**0.8**	**0.6**	0.4	0.2	0.2
$MDB.yy.14$	**0.8**	**0.6**	0.4	0.2	0.2

Note. See Table 5, *specificity* measures higher than 0.5 are shown in bold.
Source: authors' calculation using the *STATISTICA* software.

On the basis of information presented in Tables 6 and 7, it may be stated that the values of the *sensitivity* and AUC measures for the ML logistic regression models and Bayesian logistic regression models are usually similar. The largest differences were observed for the *sensitivity* measures in the years 2008 and 2009, when the assessment of bankruptcy risk calculated by means of the Bayesian model was milder than in the case of the ML model.

Table 7. *AUC* measure for the ML and the Bayesian logistic regression models estimated on the basis of sets without outliers

Model	*AUC*				
	2005	2006	2007	2008	2009
$MT.yy.14$	**0.945**	**0.904**	0.764	**0.903**	**0.857**
$MTB.yy.14$	**0.945**	**0.891**	0.764	**0.879**	**0.852**
$MT.yy.m$	0.706	0.744	0.757	0.607	**0.824**
$MTB.yy.m$	0.726	0.747	0.762	0.629	**0.826**
$MQ.yy.14$	**0.945**	**0.911**	0.759	**0.908**	**0.822**
$MQB.yy.14$	**0.945**	**0.900**	0.764	**0.874**	**0.826**
$MQ.yy.m$	**0.945**	**0.901**	0.770	0.792	**0.833**
$MQB.yy.m$	**0.945**	**0.895**	0.770	0.790	**0.826**
$MD.yy.14$	**0.942**	**0.888**	**0.898**	**0.932**	0.772
$MDB.yy.14$	**0.945**	**0.885**	**0.817**	**0.927**	0.772

Note. See Table 5, *AUC* measures higher than 0.8 are shown in bold.
Source: authors' calculation using the *STATISTICA* software.

Out of the considered logistic regression models (i.e. estimated on the basis of data sets with outliers deleted by means of various methods), the authors selected models with the highest *sensitivity* and AUC measures for further analysis. Values of the *specificity* measure were also taken into consideration in the analysis. The examination of these measures led to the conclusion that taking account of the discriminatory power of financial ratios usually did not improve the classification accuracy measured by the *sensitivity* and AUC measures. Among the models estimated on sets constructed without the use of the discriminatory power analysis, the highest values of the *sensitivity*, *specificity* and AUC measures were usually observed in the case of models built on the basis of the quantile analysis. The advantage of this solution was minor compared to the solution based on the Tukey criterion. Therefore, further analysis was conducted on the basis of the models *MQ.yy*.14 and *MQB.yy*.14, where *yy* = 05, 06, 07, 08, 09, i.e. the ML logistic regression and the Bayesian logistic regression models estimated on the sets cleaned by means of the quantile analysis.

The aim of the study was to assess the financial condition of five bankrupt companies in the years 2005–2009, i.e. shortly after they announced arrangement bankruptcy. Therefore, the classification of healthy companies was ignored in the discussion on the results of the analysis.

Table 8 presents, for each bankrupt company B_1–B_5, the point estimates of the probability of belonging to the 'B' Group of entities (whose financial condition is typical for companies with announced arrangement bankruptcy) and 95% confidence intervals for the repeat bankruptcy probability determined by means of the ML logistic regression model (*MQ.yy*.14).

Table 8. Point and interval (in round brackets) estimates of the probability of belonging to a group of entities in a financial condition typical for companies with announced arrangement bankruptcy, calculated for each bankrupt company B_1–B_5 by means of the ML logistic regression model

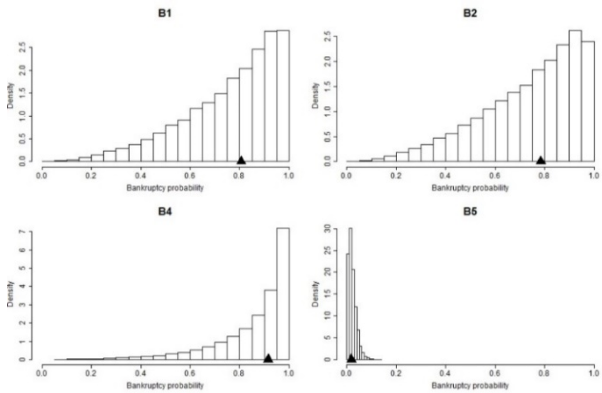
Year	Company with announced arrangement bankruptcy				
	B_1	B_2	B_3	B_4	B_5
2005	0.9408	0.9301	1.0000	0.9823	0.0123
	(0.1401, 0.9994)	(0.1338, 0.9991)	(0.8651, 1.0000)	(0.1894, 0.9999)	(0.0015, 0.0961)
2006	0.9101	1.0000	0.9235	0.9885	0.0045
	(0.0202, 0.9998)	(0.9999, 1.0000)	(0.0271, 0.9998)	(0.0355, 1.0000)	(0.0005, 0.0400)
2007	0.9996	0.6787	0.0192	0.6691	0.0024
	(0.0297, 1.0000)	(0.0440, 0.9898)	(0.0051, 0.0702)	(0.0525, 0.9866)	(0.0003, 0.0225)
2008	0.9959	0.8137	0.0102	0.9706	0.4152
	(0.1329, 1.0000)	(0.0518, 0.9972)	(0.0010, 0.0964)	(0.0005, 1.0000)	(0.0542, 0.8979)
2009	0.9953	0.0064	0.0170	0.6553	0.7272
	(0.6086, 1.0000)	(0.0011, 0.0369)	(0.0043, 0.0651)	(0.1593, 0.9502)	(0.1824, 0.9695)

Source: authors' calculation using the *STATISTICA* software.

Considering the point estimates of the probability of bankrupt companies B_1–B_5 belonging to the 'B' Group of entities, as determined by means of the ML logistic regression model for the years 2005–2009 (Table 8), it may be stated, in accordance with the adopted criterion, that the financial condition typical for healthy companies was observed for company B_2 in 2009, B_3 in 2007–2009 and B_5 in 2005–2008. In general, 95% confidence intervals confirmed these conclusions. The assessment of the financial condition of bankrupt company B_5 in 2008 was the only exception as the interval estimation indicated an unclear situation of this company. Furthermore, the risk of repeat bankruptcy established on the basis of the point estimations of the bankruptcy probability (17 cases – see Table 8) was confirmed by the interval estimations with respect to only three bankrupts, i.e. B_1 in 2009, B_2 in 2006 and B_3 in 2005. In the remaining 14 cases, the interval estimates indicated an unclear financial condition of the analysed bankrupts.

An unclear assessment of the financial condition of bankrupts on the basis of the point and interval estimations of repeat bankruptcy probability led to performing the Bayesian analysis on the 25 analysed cases. Figures 1–5 present histograms of the posterior distribution of the repeat bankruptcy probability for individual bankrupt companies B_1–B_5, with medians yielded by the Bayesian logistic regression model (*MQB.yy*.14).

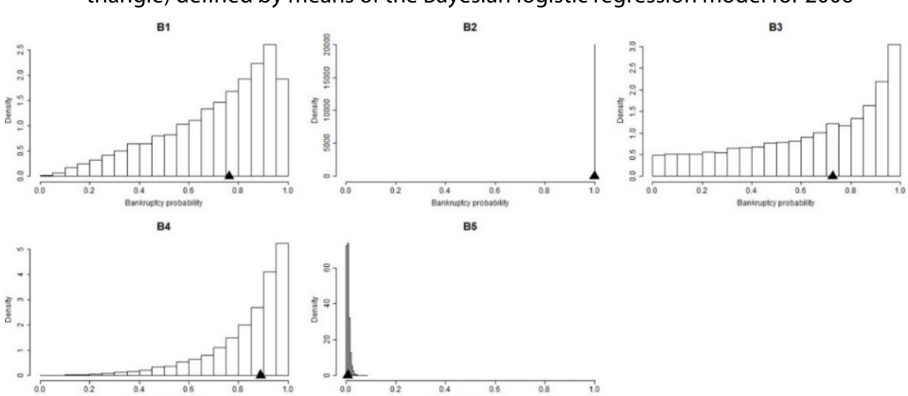
Figure 1. Histograms of the posterior distribution of the probability of bankrupt companies B_1–B_5 belonging to the group of entities in a financial condition typical for companies with announced arrangement bankruptcy, and median values (marked with a black triangle) defined by means of the Bayesian logistic regression model for 2005



Source: authors' calculation using the *MCMCpack* package in R.

The analysis of the medians and the shape of the histograms of the posterior distribution of the probability of bankrupt companies belonging to the group of entities in a financial condition typical for companies with announced arrangement bankruptcy (Figure 1) in 2005 shows the threat of repeat bankruptcy of bankrupts B_1–B_4. The shapes of the histograms obtained for bankrupts B_1 and B_2 indicate that the risk of repeat bankruptcy for these companies is lower than in the case of bankrupt B_4. As far as company B_5 is concerned, the median value and the histogram show that the financial condition of this bankrupt in 2005 was similar to the condition typical for healthy companies.

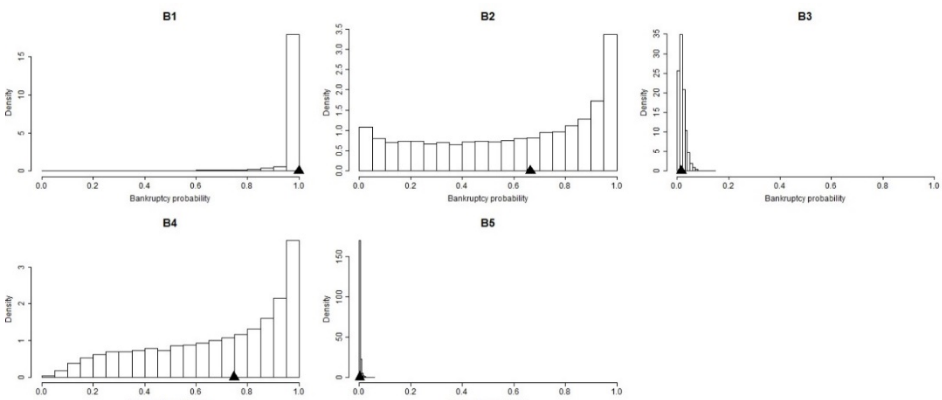
Figure 2. Histograms of the posterior distribution of the probability of bankrupt companies B_1–B_5 belonging to the group of entities in a financial condition typical for companies with announced arrangement bankruptcy, and median values (marked with a black triangle) defined by means of the Bayesian logistic regression model for 2006



Source: authors' calculation using the *MCMCpack* package in R.

The medians and histograms of the posterior distribution of bankruptcy probability calculated for 2006 (Figure 2) show the threat of repeat bankruptcy of bankrupts B_1–B_4. In the case of company B_5, the median indicates the similarity of this bankrupt's financial condition to the condition typical for a healthy company. The histograms of the posterior distribution of the repeat bankruptcy probability for companies B_1, B_2, and B_4 strengthen the conclusions made on the basis of the medians, and bankrupt B_2 seems to be more threatened with repeat bankruptcy than bankrupts B_1 and B_4. It is difficult to defend the conclusion regarding the threat of repeat bankruptcy of bankrupt B_3 obtained on the basis of the medians, as the histogram for this company shows that the assessment of its financial condition is unclear. On the basis of the shape of the histogram obtained for bankrupt B_5, it is possible to uphold the conclusion made on the basis of the medians.

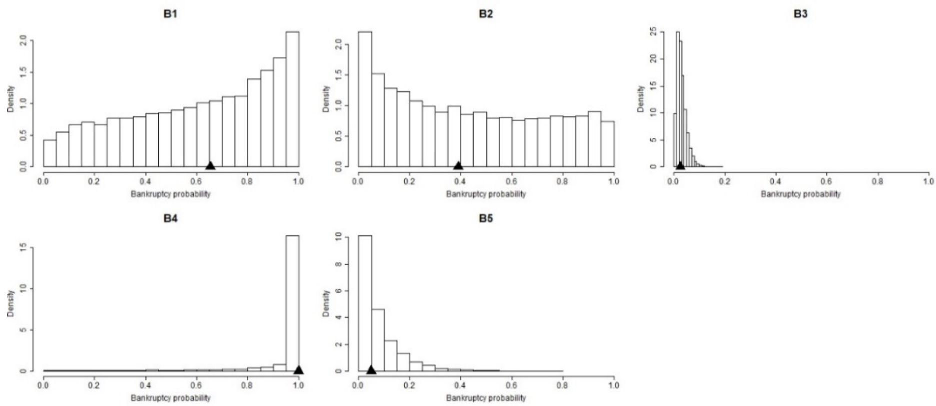
Figure 3. Histograms of the posterior distribution of the probability of bankrupt companies B_1–B_5 belonging to the group of entities in a financial condition typical for companies with announced arrangement bankruptcy, and median values (marked with a black triangle) defined by means of the Bayesian logistic regression model for 2007



Source: authors' calculation using the *MCMCpack* package in R.

The medians of the posterior distribution of bankruptcy probability calculated for 2007 (Figure 3), demonstrate that for bankrupts B_1, B_2, and B_4 there is a threat of repeat bankruptcy. The histogram of the posterior distribution of bankruptcy probability for bankrupt B_1 definitely justifies the conclusion made on the basis of the median. The shapes of the histograms in the case of bankrupts B_2 and B_4 allow only for a relatively unclear assessment of their financial condition. The medians calculated for companies B_3 and B_5 suggest the similarity between their financial conditions and that of a healthy company. The analysis of the histograms reinforces these conclusions.

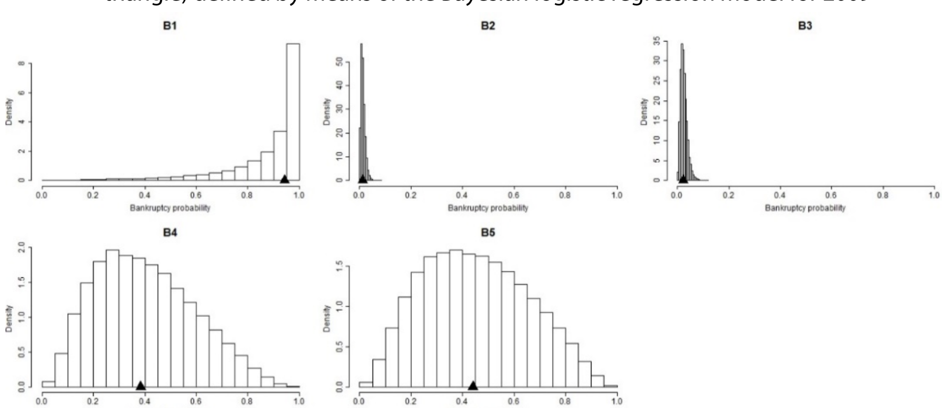
Figure 4. Histograms of the posterior distribution of the probability of bankrupt companies B_1–B_5 belonging to the group of entities in a financial condition typical for companies with announced arrangement bankruptcy, and median values (marked with a black triangle) defined by means of the Bayesian logistic regression model for 2008



Source: authors' calculation using the *MCMCpack* package in R.

The analysis of the medians of the posterior distribution of the bankruptcy probability calculated for 2008 (Figure 4) leads to the conclusion that bankrupts B_1 and B_4 are threatened with repeat bankruptcy, while bankrupts B_2, B_3, and B_5 are in a financial condition typical for healthy companies. The shape of the histograms of the posterior distribution of the bankruptcy probability confirm only the conclusions relating to bankrupts B_3–B_5. The shape of the histograms in the case of companies B_1 and B_2 indicate an unclear assessment of those companies' financial condition.

Figure 5. Histograms of the posterior distribution of the probability of bankrupt companies B_1–B_5 belonging to the group of entities in a financial condition typical for companies with announced arrangement bankruptcy, and median values (marked with a black triangle) defined by means of the Bayesian logistic regression model for 2009



Source: authors' calculation using the *MCMCpack* package in R.

On the basis of the medians of the posterior distribution of the bankruptcy probability (Figure 5), only bankrupt B_1 was threatened with repeat bankruptcy. This conclusion is confirmed by the shape of this company's histogram of the posterior distribution of the bankruptcy probability. The shapes of the histograms also confirm that bankrupts B_2 and B_3 belong to the group of companies in a financial condition typical for healthy companies, while the histograms of the posterior distribution of the bankruptcy probability of companies B_4 and B_5 show an unclear assessment of their financial condition.

6. Discussion

Table 9 presents the results of the classification of the bankrupt companies under two groups: the group of entities in a financial condition typical for healthy companies ('NB' Group) and the group of entities in a financial condition typical for companies with announced arrangement bankruptcy ('B' Group), on the basis of the performed analyses and taking into account the assumptions defined in Section 3.

Table 9. Results of the classification of bankrupt companies in 'B' and 'NB' groups on the basis of the point and interval estimations of the bankruptcy probability received by applying the ML logistic regression model, and medians and the shape of the histograms of the posterior distributions of the bankruptcy probability received in the Bayesian approach

Bankrupt	2005				2006				2007				2008				2009			
	ML		*Bayes*		*ML*		*Bayes*		*ML*		*Bayes*		*ML*		*Bayes*		*ML*		*Bayes*	
	P	*CI*	*M*	*H*	*P*	*CI*	*M*	*H*	*P*	*CI*	*M*	*H*	*P*	*CI*	*M*	*H*	*P*	*CI*	*M*	*H*
B_1	B	–	B	B	B	–	B	B	B	–	B	B	B	–	B	–	B	B	B	B
B_2	B	–	B	B	B	B	B	B	B	–	B	–	B	–	NB	–	NB	NB	NB	NB
B_3	B	B	B	B	B	–	B	–	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB
B_4	B	–	B	B	B	–	B	B	B	–	B	–	B	–	B	B	B	–	NB	–
B_5	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB	NB	–	NB	NB	B	–	NB	–

Note. *ML* – ML logistic regression model, *Bayes* – Bayesian logistic regression model, *P* – point estimation, *CI* – 95% confidence interval, *M* – posterior median, *H* – histogram of the posterior distribution, B – bankrupt, NB – non-bankrupt.
Source: authors' calculation.

The point and interval estimations of bankruptcy probability performed on the basis of the ML logistic regression model for the year 2005 indicate a threat of repeat bankruptcy in the case of company B_3, and the similarity of the financial condition of company B_5 to a healthy company's condition (Table 9). The medians and the shape of the histograms of the posterior distributions determined on the basis of the Bayesian approach confirm the above-mentioned conclusions. Moreover, the analysis of the medians and histograms of the posterior distribution indicates that there is a threat of repeat bankruptcy for companies B_1, B_2, and B_4. These conclusions comply only with the point estimations based on the maximum likelihood method.

By means of the ML logistic regression model estimated for 2006 (Table 9), company B_2 was assigned – on the basis of the point and interval estimations – to the group of companies threatened with repeat bankruptcy, while the financial condition of company B_5 was found to be similar to that of healthy companies. The above conclusions were confirmed by the analysis of the medians and histograms of the posterior distribution defined as a result of the Bayesian approach. The use of the Bayesian approach yields a conclusion that companies B_1 and B_4 were threatened with repeat bankruptcy. This conclusion complies only with the point estimations based on the maximum likelihood method.

The results of the analysis of the point and interval estimations received on the basis of the ML logistic regression model estimated for the data for 2007 (Table 9) indicate that the financial condition of companies B_3 and B_5 is similar to that of a healthy company, which is confirmed by the analysis based on the Bayesian approach. In addition, the median and the shape of the histogram indicate a threat of repeat bankruptcy for company B_1. This suggestion is confirmed only by the point estimation based on the maximum likelihood method.

Analysing the year 2008 (Table 9), the point and interval estimations lead to a conclusion that the financial condition of company B_3 was similar to the financial condition of a healthy company. This conclusion complies with the results obtained by the Bayesian approach. The results of the Bayesian approach also suggest company B_4 is threatened with bankruptcy and the financial condition of company B_5 is similar to that of healthy companies. The point estimations of bankruptcy probability obtained by means of the ML logistic regression model leads to similar conclusions.

The results received for 2009 according to the ML logistic regression model allow a statement that company B_1 was threatened with repeat bankruptcy, while the financial condition of companies B_2 and B_3 are similar to those of healthy companies. The analysis of the medians and the shape of the histograms of the posterior distributions defined on the basis of the Bayesian approach confirmed the above findings.

The results obtained for company B_3 in the course of this study merit emphasis. In four out of five cases similar conclusions were drawn on the basis of the ML logistic regression model and the Bayesian logistic regression model. Analysing the year 2005, this company was classified in the group of entities threatened with repeat bankruptcy. In the analysis for the year 2006, the point estimation of the bankruptcy probability (ML logistic regression model) and the median of the posterior distribution (Bayesian logistic regression model) again indicated that the company was in danger of repeat bankruptcy. The interval estimation of the bankruptcy probability (ML logistic regression model) and the shape of the histogram of the posterior distribution (Bayesian logistic regression model, on the other hand, suggest an unclear assessment of the company's financial condition. From 2007 bankrupt B_3 was assigned to the group of entities with a financial condition typical for healthy companies. It should be noted that in March 2007, a court confirmed that bankrupt company B_3 fulfilled the arrangements agreed upon with its creditors in December 2005.

Company B_5 was another bankrupt for which conclusions drawn on the basis of the ML logistic regression model and the Bayesian logistic regression model were similar in the majority of cases (three out of four). In the years 2005–2007, this bankrupt was classified in the group of entities whose financial condition was typical for healthy companies. In the years 2008 and 2009, the assessment of its financial condition was unclear – the finding which was reinforced by the interval estimation obtained on the basis of the ML logistic regression model and the histogram of the posterior distribution in the Bayesian approach, in the case of 2009. A court completed the bankruptcy procedure for the bankrupt company B_5 in March 2005. With regard to that company, the court also stated that any enforcement proceedings or proceedings to secure claims conducted against the bankrupt in order to satisfy claims subject to the arrangements were discontinued, and all execution and enforcement titles became invalid.

In the case of bankrupt B_2, the conclusions made on the basis of the ML logistic regression model and the Bayesian logistic regression model were in accordance twice: in 2006, when the company was found to be threatened with repeat bankruptcy, and in 2009, when it was assigned to the group of entities in a financial condition typical for healthy companies. In May 2006, a court announced the bankruptcy procedure for this company finalized. This situation took place in April 2012 again – the court then announced bankruptcy of company B_2, including the liquidation of its assets. Taking into account the poor financial condition of bankrupt B_2 in the years 2006–2008, that is after the bankruptcy procedure was completed and its condition started improving in 2009, the reasons for this company's repeat bankruptcy in 2012 may be related to the so-called 'second wave' of the global financial crisis of 2007–2008. The impact of the macroeconomic environment on bankruptcies of companies has been discussed in literature (e.g. Platt and Platt, 2002).

In the case of bankrupts B_1 and B_4, what was mainly observed were the threats of repeat bankruptcy or an unclear financial condition. As far as company B_1 is concerned, we have not found any court records of its activities after the announcement of arrangement bankruptcy. However, in the case of bankrupt B_4, the court ruled the finalisation of the bankruptcy procedure.

7. Conclusions

In the light of the obtained results and further history of the studied bankrupt companies, the information on the financial condition of companies B_3 and B_5 prove very valuable. In the case of these bankrupts, similar results were obtained by two approaches, both based on the ML logistic regression model and the Bayesian logistic regression model. Thus, decisions made in these two companies may be a source of valuable information on how restructuring processes may improve the financial condition of companies facing insolvency problems. Nevertheless, these companies should be subject to further review, as bankruptcy may reoccur in the following years. The analysis of such cases as bankrupt B_2 provides invaluable information on recovery strategies which ended in failure and so should be avoided by enterprises undergoing bankruptcy.

In the paper, the analysis performed in the main part on the basis of the ML logistic regression model was supplemented with information yielded by the application of the Bayesian approach. This facilitated performing a broader analysis than just the assessment of the financial condition by the classification of companies into two groups, namely companies threatened with repeat bankruptcy and entities in a financial condition as that of healthy companies in a given year. In particular, the analysis of the shape of the posterior distribution of bankruptcy probability, for instance, from the perspective of probability mass distribution and the level of dispersion of distribution, in some cases enabled an observation that the financial condition of a company is not clear, despite clear assessments based on the point estimations. Such information is provided by the confidence intervals established on the basis of the ML logistic regression model, although to a smaller extent than the results received through the Bayesian approach.

In the authors' opinion, the analysis of the process of overcoming insolvency problems by companies with announced arrangement bankruptcy may provide information useful for the assessment of the success probability of different restructuring solutions, and thus valuable for future bankrupts. The logistic regression model, estimated by means of the maximum likelihood method, and the Bayesian approach prove effective in the assessment of the financial condition of companies undergoing arrangement bankruptcy in comparison to the condition of healthy companies. Models built for the years following the announcement of arrangement bankruptcy may constitute an element of the assessment system of the restructuring proposals prepared by companies under the threat of bankruptcy.

The research carried out in this paper confirms the authors' previous observations, namely that the deletion of outliers from a data set before conducting an estimation of logistic regression model parameters may improve the classification accuracy. The model estimation was based on the maximum likelihood method and the Bayesian approach. The authors' intention is to further extend the research, for example to the bootstrap method. In the case of the Bayesian approach, a scarcely informative prior distribution was adopted. In any further research, we suggest that expert knowledge should be taken into account when selecting the prior distribution.

Acknowledgement. The research behind this publication was financed from the funds granted to the Cracow University of Economics.

References

Abellán J., Castellano J. G., (2017), A comparative study on base classifiers in ensemble methods for credit scoring, *Expert Systems with Applications*, 73, 1–10. DOI: 10.1016/j.eswa.2016.12.020.

Altman E. I., (1968), Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy, *The Journal of Finance*, 23(4), 589–609. DOI: 10.1111/j.1540-6261.1968.tb00843.x.

Altman E. I., Branch, B., (2015), The Bankruptcy System's Chapter 22 Recidivism Problem: How Serious is It?, *The Financial Review*, 50(1), 1–26. DOI: 10.1111/fire.12058.

Barnett V., Lewis, T., (1994), *Outliers in Statistical Data*, John Wiley & Sons, Chichester.

Beaver W. H., Correia, M., McNichols, M. F., (2012), Do differences in financial reporting attributes impair the predictive ability of financial ratios for bankruptcy?, *Review of Accounting Studies*, 17(4), 969–1010.

Beaver W. H., McNichols M. F., Rhie J. W., (2005), Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy, *Review of Accounting Studies*, 10(1), 93–122. DOI: 10.1007/s11142-004-6341-9.

Birdsall T. G., (1973), *The Theory of Signal Detectability: ROC Curves and Their Character*, Michigan, The University of Michigan, Ann Arbor, Cooley Electronics Laboratory.

De Andrés J., Sánchez-Lasheras F., Lorca P., De Cos Juez F. J., (2011), A Hybrid Device of Self Organizing Maps (SOM) and Multivariate Adaptive Regression Splines (MAR) for the Forecasting of Firms Bankruptcy, *Accounting and Management Information Systems*, 10(3), 351–374.

Eberhart A. C., Altman E. I., Aggarwal R., (1999), The Equity Performance of Firms Emerging from Bankruptcy, *The Journal of Finance*, 54(5), 1855–1868. DOI: 10.1111/0022-1082.00169.

Ekinci A., Erdal H. I., (2017), Forecasting Bank Failure: Base Learners, Ensembles and Hybrid Ensembles, *Computational Economics*, 49(4), 677–686. DOI: 10.1007/s10614-016-9623-y.

Fawcett T., (2006), An introduction to ROC analysis, *Pattern Recognition Letters*, 27(8), 861–874. DOI: 10.1016/j.patrec.2005.10.010.

Frydman H., Altman E. I., Kao D., (1985), Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress, *The Journal of Finance*, 40(1), 269–291. DOI: 10.1111/j.1540-6261.1985.tb04949.x.

Gamerman D., Lopes H. F., (2006), *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall/CRC, Boca Raton.

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. DOI: 10.1007/s10844-014-0333-4.

García V., Marqués A. I., Sánchez J. S., Ochoa-Domínguez H. J., (2019), Dissimilarity-Based Linear Models for Corporate Bankruptcy Prediction, *Computational Economics*, 53(3), 1019–1031. DOI: 10.1007/s10614-017-9783-4.

Grubbs F. E., (1969), Procedures for Detecting Outlying Observations in Samples, *Technometrics*, 11(1), 1–21. DOI: 10.1080/00401706.1969.10490657.

Hauser R. P., Booth D., (2011), Predicting Bankruptcy with Robust Logistic Regression, *The Journal of Data Science*, 9(4), 565–584. DOI: 10.6339/JDS.2011.09(4).946.

Hodge V. J., Austin J., (2004), A Survey of Outlier Detection Methodologies, *Artificial Intelligence Review*, 22(2), 85–126. DOI: 10.1023/B:AIRE.0000045502.10941.a9.

Hotchkiss E. S., (1995), Postbankruptcy Performance and Management Turnover, *The Journal of Finance*, 50(1), 3–21. DOI: 10.1111/j.1540-6261.1995.tb05165x.

John G. H., (1995), Robust Decision Trees: Removing Outliers from Databases, in: U. M. Fayyad, R. Uthurusamy, (ed.), *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, 174–179.

Johnson R., Wichern D. W., (1992), *Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey, Upper Saddle River.

Kosiorowski D., (2008), Robust Classification and Clustering Based on the Projection Depth Function, in: P. Brito, (ed.), *COMPSTAT 2008. Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, 209–216.

Kostrzewska J., Kostrzewski M., Pawełek B., Gałuszka K., (2016), The classical and Bayesian logistic regression in the research on the financial standing of enterprises after bankruptcy in Poland, in: M. Papież, S. Śmiech, (ed.), *Proceedings of the 10th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Foundation of the Cracow University of Economics, Cracow, 72–81.

Krzanowski W. J., Hand D. J., (2009), *ROC curves for continuous data*, CRC Press, Boca Raton.

Lee S., Choi W. S., (2013), A multi-industry *bankruptcy* prediction model using back-propagation neural network and multivariate discriminant analysis, *Expert Systems with Applications*, 40(8), 2941–2946. DOI: 10.1016/j.eswa.2012.12.009.

Li S., Wang S., (2014), A financial early warning logit model and its efficiency verification approach, *Knowledge-Based Systems*, 70, 78–87. DOI: 10.1016/j.knosys.2014.03.017.

Liang D., Tsai Ch.-F., Wu H.-T., (2015), The effect of feature selection on financial distress prediction, *Knowledge-Based Systems*, 73, 289–297. DOI: 10.1016/j.knosys.2014.10.010.

López Iturriaga F. J., Pastor Sanz I., (2015), Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks, *Expert Systems with Applications*, 42(6), 2857–2869. DOI: 10.1016/j.eswa.2014.11.025.

Neter J., Wasserman W., Kutner M. G., (1989), *Applied linear regression analysis*, 2nd ed. Homewood, IL, Irwin.

Odom M. D., Sharda R., (1990), A Neural Network Model for Bankruptcy Prediction, in: *Proceedings of IEEE International Conference on Neural Networks*, San Diego, 2, 163–168.

Ohlson J., (1980), Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, 18(1), 109–131. DOI: 10.2307/2490395.

Pawełek B., (2019), Extreme Gradient Boosting Method in the Prediction of Company Bankruptcy, *Statistics in Transition: new series*, 20(2), 155–171. DOI: 10.21307/stattrans-2019-020.

Pawełek B., Gałuszka K., Kostrzewska J., Kostrzewski M., (2017), Classification Methods in the Research on the Financial Standing of Construction Enterprises After Bankruptcy in Poland, in: F. Palumbo, A. Montanari, M. Vichi, (ed.), *Data Science: Innovative Developments in Data Analysis and Clustering*, Springer, Switzerland, 29–42.

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

Platt H. D., Platt M. B., (2002), A Re-examination of the Effectiveness of the Bankruptcy Process, *Journal of Business Finance* & *Accounting*, 29(9–10), 1209–1237. DOI: 10.1111/1468-5957.00467.

Sarkar S., Sriram R. S., (2001), Bayesian models for early warning of bank failures, *Management Science*, 47(11), 1457–1475. DOI: 10.1287/mnsc.47.11.1457.10253.

Shumway T., (2001), Forecasting Bankruptcy More Accurately: A Simple Hazard Model, *The Journal of Business*, 74(1), 101–124. DOI: 10.2139/ssrn.171436.

Spicka J., (2013), The financial condition of the construction companies before bankruptcy, *European Journal of Business and Management*, 5(23), 160–169.

Sun J., Fujita H., Chen P., Li H., (2017), Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble, *Knowledge-Based Systems*, 120, 4–14. DOI: 10.1016/j.knosys.2016.12.019.

Sun J., Li H., Huang Q.-H., He K.-Y., (2014), Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches, *Knowledge-Based Systems*, 57, 41–56. DOI: 10.1016/j.knosys.2013.12.006.

Sun L., Shenoy P. P., (2007), Using Bayesian networks for bankruptcy prediction: Some methodological issues, *European Journal of Operational Research*, 180(2), 738–753. DOI: 10.1016/j.ejor.2006.04.019.

Tkáč M., Verner R., (2016), Artificial neural networks in business: Two decades of research, *Applied Soft Computing*, 38, 788–804. DOI: 10.1016/j.asoc.2015.09.040.

Trabelsi S., He R., He L., Kusy M., (2015), A comparison of Bayesian, Hazard, and Mixed Logit model of bankruptcy prediction, *Computational Management Science*, 12(1), 81–97. DOI: 10.1007/s10287-013-0200-8.

Tsai Ch.-F., Hsu Y.-F., Yen D. C., (2014), A comparative study of classifier ensembles for bankruptcy prediction, *Applied Soft Computing*, 24, 977–984. DOI: 10.1016/j.asoc.2014.08.047.

Tseng F.-M., Hu Y.-Ch., (2010), Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks, *Expert Systems with Applications*, 37(3), 1846–1853. DOI: 10.1016/j.eswa.2009.07.081.

Tukey J. W., (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

Ustawa z dnia 28 lutego 2003 r. – *Prawo upadłościowe*, Dz.U. 2003 nr 60 poz. 535.

Ustawa z dnia 15 maja 2015 r. – *Prawo restrukturyzacyjne*, Dz.U. 2015 poz. 978.

Wu Y., Gaunt C., Gray S., (2010), A comparison of alternative bankruptcy prediction models, *Journal of Contemporary Accounting* & *Economics*, 6(1), 34–45. DOI: 10.1016/j.jcae.2010.04.002.

Yu Q., Miche Y., Séverin E., Lendasse A., (2014), Bankruptcy prediction using Extreme Learning Machine and financial expertise, *Neurocomputing*, 128, 296–302. DOI: 10.1016/j.neucom.2013.01.063.

Zhou L., Lai K. K., (2017), AdaBoost Models for Corporate Bankruptcy Prediction with Missing Data, *Computational Economics*, 50(1), 69–94. DOI: 10.1007/s10614-016-9581-4.

Zhou L., Lu D., Fujita H., (2015), The performance of corporate financial distress prediction models with features selection guided by domain knowledge and data mining approaches, *Knowledge-Based Systems*, 85, 52–61. DOI: 10.1016/j.knosys.2015.04.017.

Zuo Y., (2003), Projections-based depth functions and associated medians, *The Annals of Statistics*, 31(5), 1460–1490. DOI: 10.1214/aos/1065705115.

Zuo Y., Serfling R., (2000), General Notions of Statistical Depth Functions, *The Annals of Statistics*, 28(2), 461–482. DOI: 10.1214/aos/1016218226.