Problem of Outliers in Corporate Bankruptcy Prediction

Barbara Pawełek, Józef Pociecha, Jadwiga Kostrzewska, Mateusz Baryła and Artur Lipieta

Abstract The results of financial condition analysis are used in the research on bankruptcy prediction of companies. The assessment of financial data quality involves also the detection of outliers. In the literature on bankruptcy prediction one can find deliberations on the problem of outliers. The proposals for solving this problem range from not taking any actions, through replacing or removing the outliers, to applying robust methods. Therefore, in the empirical research, some doubts concerning the choice of an appropriate approach to the outliers appear. The aim of the article is to present the outcomes of empirical research on the usefulness of selected techniques for identifying outliers in

Barbara Pawełek
Cracow University of Economics, 27 Rakowicka Street, 31-510 Cracow, Poland,
✉ barbara.pawelek@uek.krakow.pl

Józef Pociecha
Cracow University of Economics, 27 Rakowicka Street, 31-510 Cracow, Poland,
✉ jozef.pociecha@uek.krakow.pl

Jadwiga Kostrzewska
Cracow University of Economics, 27 Rakowicka Street, 31-510 Cracow, Poland,
✉ jadwiga.kostrzewska@uek.krakow.pl

Mateusz Baryła
Cracow University of Economics, 27 Rakowicka Street, 31-510 Cracow, Poland,
✉ mateusz.baryla@uek.krakow.pl

Artur Lipieta
Cracow University of Economics, 27 Rakowicka Street, 31-510 Cracow, Poland,
✉ artur.lipieta@uek.krakow.pl

ARCHIVES OF DATA SCIENCE, SERIES A (ONLINE FIRST)
KIT SCIENTIFIC PUBLISHING
Vol. 2, No. 1, 2017
DOI 10.5445/KSP/1000058749/19
ISSN 2363-9881

bankruptcy forecasting. In the study, both one-dimensional (based on centiles and Tukey's criterion) and multi-dimensional (projection depth function) procedures of detecting outliers were considered. So as to assess the classification accuracy of chosen bankruptcy prediction methods for a test set, the sensitivity, specificity, accuracy and AUC measure were used. The analysis was based on data concerning manufacturing companies in Poland.

1 Introduction

The predicted worsened financial standing of enterprises which may lead to bankruptcy is an important issue in social and economic sciences. Bankruptcy is an inherent element of the market economy. It constitutes the system of purging the economy by eliminating economically ineffective enterprises or those unable to operate in the market. The possibility of bankruptcy forecasting allows the company board to undertake repair procedures which can prevent the company from going bankrupt. The issue of bankruptcy forecasting draws the attention of the following entities: companies management, banks, auditors, investors and financial analysts, government institutions and economic organizations (e.g. Pawełek and Pociecha, 2012). The methods of bankruptcy forecasting enjoy great popularity among scientists, business practitioners and financial institutions. The accuracy of these forecasts is the main criterion for evaluating the suitability of these methods for predicting bankruptcy. The research is based on sets of enterprises containing entities which declared bankruptcy and entities which did not declare bankruptcy during the analysed period. One of the two elementary approaches used in the research of this kind involves the construction of a balanced set with the use of the pairing method or independent random choice (e.g. Baryła et al, 2016). This solution is criticised for the fact that the actual structure of the enterprise population by the criterion of declared bankruptcy is not balanced. Another approach to the forecasting of enterprise bankruptcy is, therefore, based on unbalanced sets. In the case of this approach, the problem of low accuracy of the classification of bankrupt companies within the analysed methods of enterprise bankruptcy forecasting is more frequent than in the case of analysing balanced sets. The cause of this phenomenon is mainly attributed to the low share of bankrupt companies in the examined sets.

This paper is based on the thesis that the accurate classification of bankrupt companies is affected by the appearance of atypical objects among financially sound enterprises. An atypical financially sound enterprise is understood by the authors as an object with outlying values of financial ratios. Enterprises defined in this way may be characterised by both a very good financial standing and a very poor financial standing, close in terms of various ratios to the situation of companies going bankrupt in the future. The presence of such objects in sets which are the basis for constructing classification models and rules makes it more difficult to obtain an effective tool used to forecast enterprise bankruptcy. Due to the generally lower accuracy of the classification of forecasting methods in the case of the analysis of unbalanced sets than balanced sets (e.g. Branco et al, 2015), in the opinion of the authors it is a good idea to analyse the problem of outliers when forecasting the bankruptcy of enterprises based on the sets of the structure of bankrupt companies and non-bankrupt companies, similar to the one occurring in the real economy.

Papers devoted to the forecasting of enterprise bankruptcy present considerations related to the occurrence of outliers among the data. Proposed solutions to this problem oscillate from ignoring (e.g. Spicka, 2013), through substitution or removal of outliers (e.g. De Andrés et al, 2011; Pawełek et al, 2015; Pociecha et al, 2014; Shumway, 2001; Wu et al, 2010), to the use of robust methods (e.g. Hauser and Booth (2011) used the robust logistic regression model with a three-fold cross validation method to investigate the classification performance). Empirical research is full of doubts concerning the choice of the right approach to the issue of outliers. Should outliers be detected or not? If they are to be detected, how should they be detected and how should the knowledge of outliers be used?

This paper is based on the assumption that an outlier is an observation which is considerably different from other elements of the set in which it occurs (e.g. Barnett and Lewis, 1994; Hodge and Austin, 2004). The literature presents various classifications of detection methods of outliers. One distinction is into the following types of methods: one-dimensional (e.g. Ben-Gal, 2005; Tukey, 1977), multi-dimensional (e.g. Zuo and Serfling, 2000; Zuo, 2003).

This paper is aimed at presenting the results of empirical research on the suitability of selected methods of detecting outliers in the forecasting of enterprise bankruptcy based on an unbalanced set of objects.

The following research questions have been formulated: [1]

1. Is the detection of atypical objects among financially sound enterprises in an unbalanced set of objects conducive to the improved classification accuracy of the enterprise bankruptcy forecasting methods?
2. Does the choice between a one-dimensional and multi-dimensional approach to the detection of outliers affect the improved classification accuracy of the analysed methods?

The literature presents works related to the forecasting of enterprise bankruptcy based on an unbalanced sample. In one of the review papers (García et al, 2015), the following criteria for the comparative analysis of papers have been adopted: database type, approach used in the evaluation of classification accuracy, measures of classification accuracy, application of statistical tests. The following characteristics of the research of this kind have been indicated after analysing more than 140 papers from the period of 2000-2013: the actual data concerning specific economies of states (65% of papers) is the basis for analysis; analyses are based on one database (69%); databases contain up to 1000 objects (i.e. they are small – 54%; medium-sized databases with 1000 to 10000 objects occur in approximately 36% of papers); division into training dataset and testing dataset (35%; V-fold cross-validation appears in approximately 31% of papers); division into training dataset and testing dataset in the ratio of 80/20 (the next division is 70/30); the following classification accuracy measures are used: accuracy (88%), type I and II error (41%), AUC measure (10%), cost (5%); statistical tests are not used (68%).

Paper authors (García et al, 2015) noted, for example, the appearance of atypical objects, asymmetry in the costs of I and II type errors (i.e. wrong classification of a bankrupt enterprise and wrong classification of a financially sound enterprise, respectively) and proper selection of classification accuracy measures in the tests of the risk of enterprise bankruptcy based on an unbalanced set (the importance of sensitivity measure and AUC measure were emphasized).

Empirical studies presented in further parts of the paper have most of the characteristics of research published on the issue analysed herein.

To ignore the aspect of application fields, the review of papers containing research based on unbalanced sets due to the research approach used can be found

[1] Additionally, the authors made an attempt to answer the following research question: Will the removal of atypical financially sound enterprises from the training dataset have an influence on the choice of the final set of financial ratios in the considered methods? Due to the length limitation of this paper, the results of the conducted analysis connected with the posed question will be presented in a separate paper.

in the paper by Branco, Torgo, and Ribeiro (2015). The research covers approximately 190 papers published mainly during the period of 2000-2013. The authors proposed the classification of approaches used in case of unbalanced samples. They distinguished, for example, the method of early data processing (including the data cleaning method). The method of k-nearest neighbours (with k=1) has been given as an example of such a procedure in the case of using such classification techniques as a classification tree or a neural network. No example was given of a data cleaning method which would be used in the case of, for example, a logit model. The research approach proposed in further parts of the paper, which uses detection methods of outliers, can be classified in the group of data cleaning methods from the preliminary data processing class in cases of estimating statistical models based on unbalanced datasets.

2 Data and research procedure

The research uses two unbalanced sets of objects: set S_1 (used to forecast bankruptcy one year in advance) and set S_2 (used to forecast bankruptcy two years in advance). Each set comprised 5435 enterprises from the group of manufacturing companies operating in Poland. Approximately 0.9% of the enterprises were companies which went bankrupt during the period of 2007-2010. Financial data for the years 2005-2009 has been taken from the Emerging Markets Information Service. Each enterprise was described by 32 financial ratios divided into groups: liquidity (4 ratios), liability (10 ratios), profitability (7 ratios) and productivity (11 ratios).

The research covered two classical methods of forecasting enterprise bankruptcy: linear discriminant function and logit model. The analysis considered linear discriminant function due to the role played in the development of enterprise bankruptcy forecasting. Results obtained for this forecasting model should be, however, treated with caution due to the fact that the assumptions of multivariate normal distribution of the considered financial ratios were not met.

The analysed sets of enterprises S_1 and S_2 were randomly divided thirty times into the training set and testing set in the 80/20 ratio (S_1_8 : 2 and S_2_8 : 2 type sets) and 60/40 ratio (S_1_6 : 4 and S_2_6 : 4 type sets). In the obtained 120 pairs of training and testing sets, the ratio between the financially sound and bankrupt enterprises occurring in the input set was kept.

In order to indicate atypical financially sound enterprises, the research used the following methods of detecting outliers:

- one-dimensional quantile-based method,
- one-dimensional method based on Tukey's criterion,
- multi-dimensional method based on a projection depth function.

The quantile-based procedure (Pawełek et al, 2015) looked as follows:

- For each financial ratio, in each set of S_1 or S_2 type and for each division into the training and testing part, the asymmetry coefficient was calculated (in the following version: the third central moment divided by the standard deviation raised to the third power) as well as the following centiles: fifth, tenth, ninetieth and ninety-fifth. The analysis used financial ratios of financially sound enterprises assigned to the training set.
- Based on the values of the asymmetry coefficient, it was decided which centile would be used to show outliers.
  - If the value of the asymmetry coefficient was larger than 1, values greater than the 90^{th} centile were considered outliers.
  - If the value of the asymmetry coefficient was lower than -1, values smaller than the 10^{th} centile were considered outliers.
  - If the value of the asymmetry coefficient lay in the range from -1 to 1, then values lower than the 5^{th} centile or greater than the 95^{th} centile were considered outliers.
- A financially sound enterprise was thought to be atypical if at least one value of financial ratios was considered to be outlying.

The procedure based on Tukey's criterion (Tukey, 1977) had the following stages:

- The first and the third quartile and quartile deviation were calculated for each financial ratio, in each S_1 or S_2 set and for each division into the training and testing part. The analysis uses financial ratios of financially sound enterprises assigned to the training set.
- Outliers were the values from beyond the following range: $<Q_1–1.5Q, Q_3+1.5Q>$, where Q means quartile deviation.
- A financially sound enterprise was thought to be atypical if at least one value of financial ratios was considered to be outlying.

The procedure based on projection depth function [2] looked as follows:

- Calculations with the use of the projection depth function were made separately for each S_1 or S_2 set and for each division into the training and testing part.
- In case of using the depth function, 10% of all financially sound enterprises in a given training set, which lay furthest from the multi-dimensional centre designated for financially sound enterprises in the analysed training set, were treated as atypical financially sound enterprises.

When using Tukey's criterion and the function of projection depth for the detection of outliers, it should be remembered that these methods indicate objects distant from the dataset centre without considering the direction of "outlying" (i.e. atypical enterprises may include both enterprises characterised by very good financial standing and enterprises coping with serious financial problems). The application of the quantile-based method, in turn, generally leads to the indication of those of financially sound enterprises whose financial standing is close to the situation of enterprises bankrupt in the future.

Having used the above methods of detecting outliers, 360 extra training sets were constructed. 120 training sets were created upon removing atypical financially sound enterprises chosen with the use of a quantile-based method ($S_1\_8:2\_Q$, $S_2\_8:2\_Q$, $S_1\_6:4\_Q$, $S_2\_6:4\_Q$), 120 training sets were created upon removing atypical financially sound enterprises indicated with the use of the method based on the Tukey's criterion ($S_1\_8:2\_T$, $S_2\_8:2\_T$, $S_1\_6:4\_T$, $S_2\_6:4\_T$) and 120 training sets were created upon removing atypical financially sound enterprises indicated with the use of the method based on the projection depth function ($S_1\_8:2\_D$, $S_2\_8:2\_D$, $S_1\_6:4\_D$, $S_2\_6:4\_D$).

The following measures were used to evaluate the classification accuracy of the analysed methods:

- sensitivity (percentage of bankrupt companies which have been properly classified),
- specificity (percentage of financially sound enterprises which have been properly classified),

[2] The concept of data depth is an issue of non-parametric robust multi-dimensional statistical analysis, developed as part of exploratory data analysis (e.g. Kosiorowski, 2008). It 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 an observation set (Zuo and Serfling, 2000). There are many proposed functions, known as depth functions (e.g. Euclidean depth function, the Mahalanobis depth, the Tukey depth, projection depth, the Student depth), which assign to each observation from a certain distribution a positive number which is a measure of its deviation from the centre, based on the distribution.

- accuracy (percentage of enterprises which have been properly classified),
- AUC measure (space under the ROC curve, where the ROC curve presents sensitivity as function 1-specificity).

When predicting bankruptcies on the basis of unbalanced sets, it is good to use the sensitivity measure as the first one, and then only the values of the AUC measure due to a high percentage of healthy companies and low percentage of bankrupts. The high values of specificity and accuracy measures may result from a large share of healthy companies in the sample.

R environments, Stata and Statistica have been used in calculations and result presentations.

3 Results of empirical research

In order to find answers to the first research question, the comparison was made of the classification accuracy of the considered methods of forecasting enterprise bankruptcy obtained on the testing set [3] after having built a model based on the training set cleansed of atypical financially sound enterprises (sets: $S_1\_8:2\_j$, $S_2\_8:2\_j$, $S_1\_6:4\_j$, $S_2\_6:4\_j$, where $j = Q,T,D$) and uncleansed training set (sets: $S_1\_8:2$, $S_2\_8:2$, $S_1\_6:4$, $S_2\_6:4$). The number of financially sound enterprises in training sets is presented in Table 1. The set of financial ratios was reduced with the use of backward stepwise analysis.

Should the classification accuracy of the forecasting model built based on the cleansed training set be greater than in the case of using an uncleansed training set, number 1 is assigned to the method of detecting outliers for a given division (one of thirty analysed ones). Otherwise, the considered method of detecting outliers has number 0 assigned. The classification accuracy was evaluated based on the value of four measures, i.e. accuracy, sensitivity, specificity and AUC. Due to a given measure of classification accuracy, the analysed method of detecting outliers after summing up the numbers allocated in subsequent set divisions into the training and testing part might have taken the maximum number of 30. Values larger than 15 (highlighted in bold in Table 2) mean that the use of a given method of detecting outliers more frequently led to the improved classification accuracy measured by a specific measure than to the

[3] Testing sets comprised 1077 financially sound enterprises and 10 bankrupts in the case of data division in the 80/20 ratio, and 2154 financially sound enterprises and 20 bankrupts in the case of data division in the 60/40 ratio.

Table 1 The number of financially sound enterprises in training sets.

	Training set							
Division	S_1	S_1_Q	S_1_T	S_1_D	S_2	S_2_Q	S_2_T	S_2_D
80/20	4309	1331*	1864*	3878	4309	1270*	1803*	3878
60/40	3232	977*	1395*	2909	3232	934*	1349*	2909

S_1 (S_2) – set used to forecast bankruptcy one year (two years) in advance, S_1_Q (S_2_Q) – set S_1 (S_2) after using the quantile-based method, S_1_T (S_2_T) – set S_1 (S_2) after using the method based on Tukey's criterion, S_1_D (S_2_D) – set S_1 (S_2) after using the method based on a projection depth function, * – the average number of financially sound enterprises in 30 training sets.

deterioration of the analysed accuracy. Calculation results are included in Table 2.

The Kruskal-Wallis test and post-hoc analysis for mean ranks of all pairs of groups were also conducted. [4] The significance of the difference between mean ranks for a given pair of groups, confirmed by p-value for two-sided test including the Bonferroni correction at the $\alpha = 0.10$ or $\alpha = 0.05$ level, was marked with an asterisk beside the number of relations in the testing set. Indicated cases are characterized by higher values of the mean and median for a given measure of classification accuracy after using a given method of detecting outliers rather than relying on an uncleansed training set (Tables 2 – 4).

For both analysed set divisions, i.e. the ratio of 80/20 and 60/40, in the case of the linear discriminant function, the general increase in the value of the sensitivity measure was noted after the application of both considered one-dimensional methods of detecting outliers. In most cases, an increase was obtained of the value of the sensitivity measure for the logit model after applying all analysed methods of detecting outliers (Table 2).

In an attempt to find an answer to the second research question, values of basic descriptive statistics of the empirical distribution of the analysed measures of classification accuracy were obtained (Tables 3–5). The presentation of results ignores values of the accuracy measure due to their similarity to the values obtained for the specificity measure. This is caused by the considerable advantage of financially sound enterprises over bankrupt companies in the analysed sets of objects.

For both divisions of sets into the training and testing part, one can note that the values of sensitivity measure median for the linear discriminant function

[4] ANOVA was not employed in the study because the tests indicated that the assumptions concerning both normality of distribution in groups and variances equality in groups were not met. Additionally, the condition of equal number of observations in groups was not met either.

Table 2 The comparison of classification accuracy of selected forecasting models based on the training set which contains or does not contain outliers.

				Number of relations in the testing set			
Division	Model	Horizon	Relation	Accuracy	Sensitivity	Specificity	AUC
80/20	LDA	1 year	$S_1 < S_1\_Q$	0^{**}	**19***	0^{**}	6^{**}
			$S_1 < S_1\_T$	0^{**}	**21***	0^{**}	9
			$S_1 < S_1\_D$	0^{**}	11	0^{**}	12
		2 years	$S_2 < S_2\_Q$	0^{**}	**27****	0^{**}	3^{**}
			$S_2 < S_2\_T$	0^{**}	**21****	0^{**}	3^{**}
			$S_2 < S_2\_D$	**25**	12	**24**	9
	LR	1 year	$S_1 < S_1\_Q$	0^{**}	**23****	0^{**}	4^{**}
			$S_1 < S_1\_T$	0^{**}	**30****	0^{**}	5^{**}
			$S_1 < S_1\_D$	0^{**}	**23***	0^{**}	**18**
		2 years	$S_2 < S_2\_Q$	0^{**}	**30****	0^{**}	5^{**}
			$S_2 < S_2\_T$	0^{**}	**28****	0^{**}	5^{**}
			$S_2 < S_2\_D$	0^{**}	**18***	0^{**}	**16**
60/40	LDA	1 year	$S_1 < S_1\_Q$	0^{**}	**28****	0^{**}	2^{**}
			$S_1 < S_1\_T$	0^{**}	**26****	0^{**}	2^{**}
			$S_1 < S_1\_D$	1^{**}	13	1^{**}	15
		2 years	$S_2 < S_2\_Q$	0^{**}	**27****	0^{**}	0^{**}
			$S_2 < S_2\_T$	1^{**}	**25***	1^{**}	2^{**}
			$S_2 < S_2\_D$	**19**	12	**19**	12
	LR	1 year	$S_1 < S_1\_Q$	0^{**}	**16****	0^{**}	5
			$S_1 < S_1\_T$	0^{**}	**22****	0^{**}	6
			$S_1 < S_1\_D$	0^{**}	**26**	0^{**}	14
		2 years	$S_2 < S_2\_Q$	0^{**}	**30****	0^{**}	3^{**}
			$S_2 < S_2\_T$	0^{**}	**29****	0^{**}	2^{**}
			$S_2 < S_2\_D$	0^{**}	**20***	0^{**}	**19**

For example, $S_1 < S_1\_Q$ means that for a given division variant into the training and testing part (80/20 or 60/40) the classification accuracy measure in the testing set was greater when the model was based on the training set S_1_Q than in case of set S_1. Maximum value in the cell = 30.
LDA – the linear discriminant function, LR – the logistic regression, * (**) – p-value for two-sided test for mean ranks including the Bonferroni correction is smaller than 0.10 (0.05).

based on the cleansed training set with the use of the analysed one-dimensional methods of detecting outliers are greater than the median of the same measure for the model estimated on the uncleansed training set. Upon applying the multi-dimensional method of detecting outliers, a decrease was noted in the range of sensitivity measure values, with the preservation of the median at a level similar to the median for the model estimated on an uncleansed training set. A considerable improvement in the classification accuracy of bankrupt companies was obtained after using one-dimensional quantile-based method on the set comprising bankrupt companies one year before they actually declared

Table 3 The sensitivity measure for linear discriminant function and logit model in the testing set.

Division	Model	Horizon	Training set	Testing set Mean	St.Dev.	Min	Me	Max
80/20	LDA	1 year	S_1	0.420	0.165	0.000	0.400	0.700
			S_1_Q	0.533	0.165	0.200	0.550	0.900
			S_1_T	0.523	0.141	0.200	0.500	0.800
			S_1_D	0.440	0.152	0.100	0.400	0.800
		2 years	S_2	0.193	0.148	0.000	0.200	0.600
			S_2_Q	0.440	0.169	0.000	0.400	0.700
			S_2_T	0.330	0.178	0.000	0.300	0.700
			S_2_D	0.221	0.132	0.000	0.200	0.500
	LR	1 year	S_1	0.260	0.097	0.000	0.300	0.400
			S_1_Q	0.804	0.093	0.600	0.800	1.000
			S_1_T	0.680	0.142	0.500	0.650	1.000
			S_1_D	0.413	0.143	0.100	0.400	0.700
		2 years	S_2	0.007	0.025	0.000	0.000	0.100
			S_2_Q	0.327	0.123	0.100	0.300	0.600
			S_2_T	0.233	0.103	0.000	0.200	0.400
			S_2_D	0.080	0.066	0.000	0.100	0.200
60/40	LDA	1 year	S_1	0.367	0.097	0.200	0.350	0.600
			S_1_Q	0.488	0.113	0.250	0.500	0.750
			S_1_T	0.460	0.118	0.300	0.450	0.750
			S_1_D	0.390	0.094	0.250	0.350	0.650
		2 years	S_2	0.198	0.148	0.000	0.150	0.600
			S_2_Q	0.383	0.125	0.100	0.400	0.600
			S_2_T	0.293	0.129	0.050	0.300	0.600
			S_2_D	0.137	0.077	0.000	0.150	0.300
	LR	1 year	S_1	0.257	0.091	0.100	0.250	0.450
			S_1_Q	0.706	0.095	0.550	0.700	0.950
			S_1_T	0.652	0.096	0.450	0.650	0.800
			S_1_D	0.352	0.093	0.200	0.350	0.550
		2 years	S_2	0.008	0.019	0.000	0.000	0.050
			S_2_Q	0.280	0.116	0.050	0.300	0.500
			S_2_T	0.190	0.087	0.000	0.200	0.350
			S_2_D	0.060	0.052	0.000	0.050	0.200

LDA – the linear discriminant function, LR – the logistic regression, S_1 (S_2) – set used to forecast bankruptcy one year (two years) in advance, S_1_Q (S_2_Q) – set S_1 (S_2) after using the quantile-based method, S_1_T (S_2_T) – set S_1 (S_2) after using the method based on Tukey's criterion, S_1_D (S_2_D) – set S_1 (S_2) after using the method based on a projection depth function, Mean – average, St.Dev. – standard deviation, Min – the minimum value, Me – median, Max – the maximum value.

bankruptcy (Table 3). This improvement meant that the median of sensitivity measure exceeded 50%.

Similar results were also noted for the logit model in the case of both analysed divisions into the training and testing part (Table 3). The values of sen-

Table 4 The specificity measure for linear discriminant function and logit model in the testing set.

Division	Model	Horizon	Training set	Testing set Mean	St.Dev.	Min	Me	Max
80/20	LDA	1 year	S_1	0.986	0.005	0.973	0.986	0.993
			S_1_Q	0.878	0.028	0.831	0.873	0.934
			S_1_T	0.928	0.021	0.883	0.925	0.967
			S_1_D	0.973	0.005	0.964	0.972	0.983
		2 years	S_2	0.904	0.041	0.812	0.914	0.950
			S_2_Q	0.623	0.031	0.571	0.625	0.687
			S_2_T	0.715	0.038	0.629	0.712	0.784
			S_2_D	0.927	0.017	0.882	0.928	0.955
	LR	1 year	S_1	0.999	0.001	0.994	0.999	1.000
			S_1_Q	0.805	0.051	0.686	0.819	0.890
			S_1_T	0.865	0.028	0.817	0.864	0.916
			S_1_D	0.986	0.005	0.975	0.986	0.994
		2 years	S_2	1.000	0.000	0.999	1.000	1.000
			S_2_Q	0.812	0.027	0.722	0.817	0.863
			S_2_T	0.861	0.019	0.823	0.860	0.898
			S_2_D	0.993	0.003	0.986	0.994	0.998
60/40	LDA	1 year	S_1	0.986	0.005	0.973	0.987	0.995
			S_1_Q	0.887	0.035	0.822	0.893	0.943
			S_1_T	0.925	0.024	0.860	0.932	0.962
			S_1_D	0.974	0.006	0.961	0.975	0.985
		2 years	S_2	0.911	0.058	0.774	0.936	0.974
			S_2_Q	0.663	0.039	0.591	0.672	0.729
			S_2_T	0.731	0.048	0.625	0.731	0.819
			S_2_D	0.943	0.020	0.905	0.950	0.971
	LR	1 year	S_1	0.999	0.001	0.996	0.999	1.000
			S_1_Q	0.861	0.030	0.814	0.869	0.910
			S_1_T	0.888	0.028	0.826	0.893	0.929
			S_1_D	0.984	0.004	0.974	0.984	0.991
		2 years	S_2	1.000	0.001	0.997	1.000	1.000
			S_2_Q	0.823	0.034	0.753	0.821	0.912
			S_2_T	0.861	0.030	0.794	0.863	0.910
			S_2_D	0.993	0.005	0.981	0.995	0.999

LDA – the linear discriminant function, LR – the logistic regression, S_1 (S_2) – set used to forecast bankruptcy one year (two years) in advance, S_1_Q (S_2_Q) – set S_1 (S_2) after using the quantile-based method, S_1_T (S_2_T) – set S_1 (S_2) after using the method based on Tukey's criterion, S_1_D (S_2_D) – set S_1 (S_2) after using the method based on a projection depth function, Mean – average, St.Dev. – standard deviation, Min – the minimum value, Me – median, Max – the maximum value.

sitivity measure median for the logit model based on the cleansed training set are higher than the median of this measure for the model estimated on an uncleansed training set. After applying the analysed one-dimensional methods of detecting outliers, median values are definitely higher than after applying

Table 5 AUC measure for linear discriminant function and logit model in the testing set.

Division	Model	Horizon	Training set	Testing set Mean	St.Dev.	Min	Me	Max
80/20	LDA	1 year	S_1	0.886	0.059	0.732	0.894	0.974
			S_1_Q	0.822	0.079	0.628	0.830	0.938
			S_1_T	0.848	0.066	0.707	0.854	0.952
			S_1_D	0.876	0.065	0.700	0.891	0.957
		2 years	S_2	0.671	0.068	0.551	0.667	0.805
			S_2_Q	0.566	0.092	0.344	0.575	0.695
			S_2_T	0.582	0.090	0.368	0.594	0.729
			S_2_D	0.644	0.089	0.452	0.652	0.787
	LR	1 year	S_1	0.914	0.041	0.815	0.930	0.986
			S_1_Q	0.879	0.042	0.762	0.884	0.967
			S_1_T	0.878	0.059	0.708	0.875	0.974
			S_1_D	0.921	0.034	0.854	0.929	0.982
		2 years	S_2	0.717	0.071	0.587	0.717	0.829
			S_2_Q	0.640	0.069	0.417	0.646	0.749
			S_2_T	0.652	0.070	0.501	0.658	0.757
			S_2_D	0.716	0.062	0.605	0.727	0.833
60/40	LDA	1 year	S_1	0.880	0.043	0.789	0.875	0.974
			S_1_Q	0.799	0.042	0.698	0.807	0.865
			S_1_T	0.817	0.046	0.669	0.822	0.899
			S_1_D	0.874	0.045	0.761	0.880	0.959
		2 years	S_2	0.651	0.056	0.497	0.656	0.753
			S_2_Q	0.539	0.064	0.381	0.547	0.666
			S_2_T	0.546	0.058	0.363	0.542	0.642
			S_2_D	0.616	0.049	0.513	0.615	0.739
	LR	1 year	S_1	0.891	0.042	0.802	0.894	0.970
			S_1_Q	0.876	0.027	0.838	0.872	0.931
			S_1_T	0.876	0.035	0.802	0.881	0.942
			S_1_D	0.896	0.034	0.824	0.906	0.968
		2 years	S_2	0.705	0.041	0.594	0.708	0.788
			S_2_Q	0.623	0.079	0.373	0.632	0.750
			S_2_T	0.612	0.080	0.378	0.603	0.735
			S_2_D	0.712	0.044	0.610	0.721	0.791

LDA – the linear discriminant function, LR – the logistic regression, S_1 (S_2) – set used to forecast bankruptcy one year (two years) in advance, S_1_Q (S_2_Q) – set S_1 (S_2) after using the quantile-based method, S_1_T (S_2_T) – set S_1 (S_2) after using the method based on Tukey's criterion, S_1_D (S_2_D) – set S_1 (S_2) after using the method based on a projection depth function, Mean – average, St.Dev. – standard deviation, Min – the minimum value, Me – median, Max – the maximum value.

a multi-dimensional method. A considerable improvement in the accuracy of bankrupt companies classification was obtained after using one-dimensional methods on the set comprising bankrupt companies one year before they actually declared bankruptcy. This improvement meant that the median of sensitivity measure exceeded 50%.

Interesting conclusions come from the analysis of the values obtained for the specificity measure (Table 4). In the case of linear discriminant function, for both divisions of sets into a training and testing set, improved classification accuracy of financially sound enterprises is noted, with the use of the analysed one-dimensional methods of detecting outliers. The application of the projection depth function led both to the deterioration in the forecasting of bankruptcies one year in advance and a slight improvement in the situation of forecasting bankruptcies two years in advance (Table 4).

In the case of the logit model, for both analysed divisions, considerable deterioration is noted in the classification accuracy of financially sound enterprises after using the analysed one-dimensional methods of detecting outliers. Slight deterioration, in turn, occurred when applying the projection depth function (Table 4). The obtained results make it more difficult to evaluate the suitability of the analysed methods of detecting outliers in the forecasting of enterprise bankruptcy in the light of the results obtained for the sensitivity measure (Table 3).

The study of the influence of the use of the considered detection methods of outliers on the AUC measure in the case of linear discriminant function gave different results of the analysed divisions into the training and testing part. In the case of division in the ratio of 80/20, negative consequences of using one-dimensional methods have been noted. The use of the projection depth function led to the preservation of classification accuracy typical of the model estimated based on an uncleansed training set in the situation of forecasting bankruptcy one year in advance. The reduction of the AUC measure median occurred in consequence of forecasting bankruptcy two years in advance. For division in the 60/40 ratio, negative consequences were noted after applying all considered detection methods of outliers in the case of forecasting bankruptcy two years in advance. One year before the bankruptcy, median of a level similar to the value noted for the model estimated based on an uncleansed training set was noted when using the projection depth function; a decrease in the median value took place when using one-dimensional methods (Table 5).

In the case of a logit model for both analysed divisions into the training and testing part, tests were conducted to indicate negative consequences of using one-dimensional methods and non-deteriorated classification accuracy in case of using a multi-dimensional method (Table 5).

4 Conclusion

The paper raises the problem of bankruptcy forecasting, which is crucial for the development of an economy. Enterprise bankruptcy has many adverse implications which are social costs of bankruptcy, such as the loss of work places or at least a part of earnings in a bankrupt company. It also involves company owners' or shareholders' losses. Moreover, it means a failure for the company board as long as they have not deliberately led the enterprise to the bankruptcy. All these negative consequences of bankruptcy imply that the knowledge of threats to the company's existence are urgently needed by business practice.

The main conclusion of the conducted research is the statement that the detection and removal of atypical financially sound enterprises from the unbalanced training sets may be conducive to improved classification accuracy of the methods of forecasting bankruptcies of enterprises on testing sets.

The detection of atypical financially sound enterprises in unbalanced training sets with the use of one-dimensional methods may bring greater improvement in the classification accuracy of the methods of forecasting bankruptcies than multi-dimensional methods in the case of sensitivity measure for the linear discriminant function and the logit model. The superiority of the multi-dimensional method may be evident in the classification accuracy measured with the accuracy and specificity measures in the case of linear discriminant function two years before bankruptcy and AUC measure in case of the logit model.

In further research, authors are planning to incorporate other methods of detecting outliers in the analysis. They also intend to consider other approaches (e.g. V-fold cross-validation) in the verification of the obtained results, to increase the number of divisions into the training and testing part, and to incorporate also other methods of forecasting enterprise bankruptcies (including the robust methods) in the research.

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

Barnett V, Lewis T (1994) Outliers in Statistical Data. Wiley

Baryła M, Pawełek B, Pociecha J (2016) Selection of balanced structure samples in corporate bankruptcy prediction. In: Wilhelm AF, Kestler HA (eds) Analysis of Large and Complex Data, Springer International Publishing, Cham, pp 345–355, DOI 10.1007/978-3-319-25226-1_30

Ben-Gal I (2005) Outlier detection. In: Maimon O, Rokach L (eds) Data Mining and Knowledge Discovery Handbook, Springer US, Boston, MA, pp 131–146, DOI 10.1007/0-387-25465-X_7

Branco P, Torgo L, Ribeiro R (2015) A Survey of Predictive Modelling under Imbalanced Distributions. arXiv preprint arXiv:150501658

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

García V, Marqués AI, Sánchez JS (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

Hauser RP, Booth D (2011) Predicting Bankruptcy with Robust Logistic Regression. Journal of Data Science 9(4):565–584

Hodge VJ, Austin J (2004) A Survey of Outlier Detection Methodologies. Artificial Intelligence Review 22(2):85–126, DOI 10.1007/s10462-004-4304-y

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

Pawełek B, Pociecha J (2012) General SEM model in researching corporate bankruptcy and business cycles. In: Pociecha J, Decker R (eds) Data Analysis Methods and Its Applications, Warsaw: CH Beck, pp 215–231

Pawełek B, Kostrzewska J, Lipieta A (2015) The Problem of Outliers in the Research on the Financial Standing of Construction Enterprises in Poland. In: Papież M, Śmiech S (eds) Proceedings of the 9th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-economic Phenomena, Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków, pp 164–173

Pociecha J, Pawełek B, Baryła M, Augustyn S (2014) Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej. Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków

Shumway T (2001) Forecasting bankruptcy more accurately: A simple hazard model. The Journal of Business 74(1):101–124, DOI 10.1086/209665

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

Tukey JW (1977) Exploratory Data Analysis. Addison-Wesley

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

Zuo Y (2003) Projection-based depth functions and associated medians. The Annals of Statistics 31(5):1460–1490, URL http://www.jstor.org/stable/3448384

Zuo Y, Serfling R (2000) General notions of statistical depth function. The Annals of Statistics 28(2):461–482, DOI 10.1214/aos/1016218226