SORT 47 (1) January-June 2023, 1-22 DOI:

Optimal threshold of data envelopment analysis in bankruptcy prediction

Michaela Staňková[1] and David Hampel[2]

Abstract

Data envelopment analysis is not typically used for bankruptcy prediction. However, this paper shows that a correctly set up a model for this approach can be very useful in that context. A superefficiency model was applied to classify bankrupt and actively manufactured companies in the European Union. To select an appropriate threshold, the Youden index and the distance from the corner were used in addition to the total accuracy. The results indicate that selecting a suitable threshold improves specificity visibly with only a small reduction in the total accuracy. The thresholds of the best models appear to be robust enough for predictions in different time and economic sectors.

MSC: *90C08, 90C90, 90B50, 90B90.*

Keywords: *Bankruptcy prediction, Data envelopment analysis, ROC curve, Threshold optimization, Validation.*

1. Introduction

Evaluating the financial health of companies has been a substantial topic for decades in corporate finance. A company's financial situation is an important guideline not only for the creditors, shareholders and top management of a company in their decision-making but also for the government because the financial distress and bankruptcy of companies (in particular when a larger number of companies go bankrupt in the same period) bring about serious problems such as unemployment. Therefore, there is a constant demand for an ever more accurate and stable tool for forecasting a company's financial situation.

[1] Corresponding author. Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

[2] Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
Received: January 2022
Accepted: March 2023

In the area of financial health assessment, the most frequent topic is the prediction of bankruptcy. Although other situations can be predicted (detection of financial distress or other risks as in Uddin et al. (2020), Petropoulos et al. (2020) or Peláez, Cao and Vilar 2021)), company bankruptcy is a clearly defined situation. Since the second half of the last century, more attention has been given to predicting the financial situation of a company. Many bankruptcy models were developed, which differ both in the method used and the variables used. All bankruptcy models are based on the assumption that companies have some specific symptoms for some time before bankruptcy. These symptoms (i.e., problems) will be reflected in the company's financial statements. Based on these statements, a large number of financial indicators can be defined, making the forecasting of bankruptcy even more difficult.

1.1. Methods used for bankruptcy prediction

The assessment of the financial health of businesses is based on the simple idea of dividing units into two groups: active (healthy) and bankruptcy. There are many methods for dividing companies into two (or more) groups. The earliest known models, such as Beaver (1966) and Altman (1968), were based on multiple discriminant analysis. Later, the logistic regression (logit) model (Ohlson, 1980) and probit model (Zmijewski, 1984) were used in this research area. In addition to these traditional statistical methods, other approaches are also widely applied today. For example, Chen and Du (2009) adopted neural networks to construct a bankruptcy model. Decision trees or the support vector machine method have also been applied; see Klepáč and Hampel (2016) and Li et al. (2018). The application possibilities and especially the predictive abilities of individual methods are still being discussed and researched (see, for example, Klepáč and Hampel (2018) or Staňková and Hampel (2018)). According to Alaka et al. (2018), a total of eight methods can be considered to be suitable for applications in practice. Namely, these are two representatives of statistical approaches (multiple discriminant analysis and logistic regression) and six artificial intelligence tools (artificial neural network, support vector machines, rough sets, case-based reasoning, decision tree and genetic algorithm). The authors conclude that "no single tool is predominantly better than other tools".

For joint stock companies, other options can be used to predict bankruptcy. Campbell, Hilscher and Szilagyi (2011) addresses logit models and includes variables such as excess stock returns and stock volatility. Eisdorfer (2008) used real options techniques, Hillegeist et al. (2004) introduced their own BSM-Prob model based on the Black-Scholes-Merton option-pricing model, and Xu and Zhang (2009) provided an overview of existing approaches with applications to Japanese listed companies. Wu, Gaunt and Gray (2010) presented a new model based on Altman (1968), Ohlson (1980), Zmijewski (1984), Shumway (2001) and Hillegeist et al. (2004). A comprehensive model based on a multiperiod logit model overperforms the original techniques. Attention is given here to the correct selection of variables, where Tian, Yu and Guo (2015) addresses variable selection by the LASSO method and confirms the variables used by Campbell et al.

(2011). Another direction of research is given by Jones (2017) and involved the gradient boosting model, which is capable of using a large number of predictors.

In recent years, efforts have been made to use the data envelopment analysis (DEA) method in financial health assessment. The DEA method typically serves to evaluate the efficiency of decision-making units (DMUs). In this context, DMUs are divided into two groups – efficient (i.e., DMUs that lie on the efficiency frontier) and inefficient (i.e., DMUs that do not lie on the efficiency frontier). However, it is possible to look at efficiency from the other side and focus on finding very inefficient units, which cannot keep up with competition in the longer term and go bankrupt over time. For this reason, the DEA method can be used as not typical tool (e.g., not included in the list of Alaka et al. (2018)) but as a possible tool for predicting bankruptcy.

When employing the DEA method in bankruptcy prediction, the basic option is to calculate the relative efficiency using the DEA model and then use those values in another classification algorithm. For example, Li, Crook and Andreeva (2014) used this approach. Using the radial variable returns to scale model, they calculated the value of the efficiency of Chinese industrial companies and then used those values in the logistic regression model for bankruptcy prediction. A similar procedure can be found, for example, in the studies by Xu and Wang (2009) and Psillaki, Tsolas and Margaritis (2010). Although these studies suggest interesting results, in this paper, we will focus on the possibilities of classification directly through DEA models.

Currently, two different groups of DEA models are developed as a tool for the classification of bankruptcy and active companies. The common idea of both approaches is to estimate the "bankruptcy frontier". One possibility is to use the Azizi and Ajirlu (2010) approach, where the so-called optimistic view of the efficiency frontier changes to a pessimistic view – the original maximization of the objective function is changed to a minimized criterion (i.e., the so-called bounded DEA model). In this case, two frontiers are estimated, which makes it possible to limit the interval at which the production units are located. Another possibility, similar to Janová, Vavřina and Hampel (2012), is to use "standard" DEA models, where input variables are minimized and output variables are maximized with the difference that the variables will be split into inputs and outputs, with the result that the least performing companies heading for bankruptcy appear on the frontier. Active companies should then be within the set of feasible solutions, i.e., not on the "bankruptcy frontier". This approach will be examined in more detail in this work.

Various studies using the DEA method in the field of bankruptcy prediction actually appear; see, for example, Štefko, Horváthová and Mokrišová (2020), Rowland and Krulicky (2020), Chang et al. (2019), Horváthová and Mokrišová (2018), Li, Crook and Andreeva (2017) and Mendelová and Bieliková (2017), but clear application of the selected DEA model is presented there without further investigation or validation.

Janová et al. (2012) used the additive DEA model to predict bankruptcy for agricultural companies. Their models are based on the financial data of the 75 companies obtained from the Amadeus database (54 bankrupt and 21 active companies). This study shows promising results for using the DEA method for predicting bankruptcy, because overall 75% of companies were correctly classified using this procedure. Staňková and Hampel (2019) also examined the classification capabilities of the additive DEA model for the period of one to three years before bankruptcy. In contrast to the study of Janová et al. (2012), Staňková and Hampel (2019) dealt with a more realistic ratio of active and bankrupt companies in the dataset – 95% active and 5% bankrupt companies. In this case, even for the period of three years before bankruptcy, the DEA models have a total accuracy of over 86%, but at the expense of the error rate of classification of bankrupt companies (error Type II was almost 60%).

Among others, Premachandra, Bhabra and Sueyoshi (2009) focused on the impact of the size constraint on the quality of prediction. They set the ratio of bankrupt/non-bankrupt companies from 0.25 to 1. These changes in settings did not affect the error rate for bankruptcy companies, but they changed the error rate for non-bankrupt companies. At a 1:1 ratio, the overall model error rate was reduced to 14%. Premachandra et al. (2009) found that the DEA model outperforms the logit model in evaluating bankruptcy out-of-sample based on total accuracy. Furthermore, the DEA method does not need the large sample size for bankruptcy evaluation that is usually required by such statistical and econometric approaches. This feature was used in financial evaluation, for example, in Staňková and Hampel (2020).

Premachandra, Chen and Watson (2011) focused on finding a possible proper discriminating or assessing function (based on efficient and inefficient frontiers from the additive superefficiency model) as essential if DEA is used in classification problems such as predicting corporate failure. They started with the idea from logistic regression, where the probability of potential insolvency is calculated and the value of 0.5 is then taken as the classification boundary. Their results show that better results are achieved with lower thresholds (recommendation threshold of 0.1).

It is visible that only a few studies have been conducted to find optimal thresholds in DEA models employed for bankruptcy prediction. Farooq and Qamar (2019) declares the existence of a literature gap about thresholds in general, not only for a case of the DEA method but also for typical data-mining approaches. Several researchers, such as Iparragirre et al. (2022) and Staňková (2022), tried to fill this gap, at least in the case of the logistic regression method. Both mentioned studies used the so-called receiver operating characteristic (ROC) curve to optimize the threshold in the logit model.

Analogically to the case of the logit model, where a probability from the interval $\langle 0;1 \rangle$ that will divide bankrupt and active companies is sought, it is possible to set the threshold value for a particular DEA model. The typical output of a standard DEA models is an efficiency score assigned to each unit, which is compared to a frontier. Units lying on a given frontier are considered efficient, or – in the bankruptcy context – active. For such units, the score is typically equal to one. This principle can also be used in the construction of the "bankruptcy frontier", where companies headed for bankruptcy lie on the frontier (i.e., they have the score equal to one). A score of one can therefore be considered as a threshold where units with a score equal to or above this threshold will be classified as bankrupt (or inefficient). In the case of bankruptcy assessments, the question is whether such an approach is too strict and whether a different threshold setting would lead to a better bankruptcy prediction success rate. In the case of common DEA models, this threshold is taken from the interval $\langle 0;1\rangle$. If the superefficiency model is used, it is also possible to consider values higher than 1 as a potential threshold.

1.2. Motivation and contribution

To date, the DEA method as a tool for constructing the bankruptcy frontier for the purpose of classifying bankrupt companies has rarely been addressed. In the abovementioned publications, attention is typically paid to only one model, usually without further justification of the choice of a specific model. In contrast to these studies, in this paper, we will focus on different model settings to find the most suitable model settings for bankruptcy prediction. In addition, in this study, we will also address the issue of imbalance in the number of active and bankrupt companies in the dataset. In all sectors of the economy, there is naturally an imbalance between the ratio of active and bankrupt companies. Models built on datasets reflecting the real distribution of companies on the market then tend to prefer correct classification in the majority group of active companies, which of course makes them more difficult for real applications. Due to this aspect, this article presents a comprehensive view of the investigated issue.

The main aim of this article is to evaluate and validate the optimal setting of superefficiency DEA models with an optimized threshold for bankruptcy prediction. For these purposes, DEA models are estimated with different settings regarding the measurement method, returns to scale, and orientations. Since we assume that the usual approach of the DEA method, where the value 1 is used as a classification threshold, will not be suitable due to the imbalance of the dataset, attention will also be paid to the identification of a threshold that would allow more balancing of the error rate in both groups of companies. When searching for a threshold, various criteria will be used (especially criteria derived from ROC curves). Different criteria will also be used during the actual evaluation of the classification capabilities of the proposed models for up to three years in advance. The proposed procedure will also be verified on other datasets, and the results of the DEA method will be compared with the competitive statistical method of logistic regression.

The structure of this paper is as follows: Section 2 describes the datasets, variables, models and procedures used. The results are then presented, in Section 3 and the best models are validated and compared with the results of a competing logit model. Finally, the empirical results are discussed, and brief conclusions are provided.

2. Materials and methods

Financial (annual accounting) data on engineering companies (NACE Code 28 – manufacture of machinery and equipment) from 2011 to 2013 were collected from the Orbis database. To achieve a more homogeneous dataset, only small- and medium-sized companies were included. To obtain an adequate number of bankrupt companies, it was necessary to include companies from across the European Union. The dataset includes 953 companies – 902 active and another 51 companies that in 2014 changed their status to bankruptcy. This dataset (including selected variables) has already been used in the article of Staňková and Hampel (2018), where a suitable setting of standard methods was sought. The use of this dataset will therefore allow a direct comparison with a competing method for bankruptcy prediction.

In their previous research, Staňková and Hampel (2018) identified a group of 19 financial indicators that are suitable for predicting the bankruptcy of engineering companies. They verified this group of variables using three different methods, not including all variables in the models, but letting the method perform the elimination. However, the DEA method itself (unlike, for example, logit or decision tree methods) does not include a mechanism for variable elimination. The involvement of a large number of variables causes several problems in DEA models, such as the instability of the bankruptcy frontier and the dimensionality. For this reason, not all 19 recommended variables will be used in this article. With regard to financial theory, nine characteristics representing the four basic groups of financial ratios (i.e., solvency ratios, profitability ratios, liquidity ratios and turnover ratios) were chosen. Table 1 shows the variables used in the analysis.

Table 1. *Overview of the financial variables used, including their formulas.*

Type	Financial indicator	Formula
Input	Current ratio	$\frac{\text{Current assets}}{\text{Current liabilities}}$
Input	Cash flow liquidity	$\frac{\text{Cash flow}}{\text{Current liabilities}}$
Input	Net working capital (mil. EUR)	Current assets – Current liabilities
Input	Return on assets (%)	$\frac{\text{P/L for period (net income)}}{\text{Total assets}}$
Input	EBIT Margin (%)	$\frac{\text{EBIT}}{\text{Operating revenue (turnover)}}$
Input	Stock turnover	$\frac{\text{Operating revenue (turnover)}}{\text{Stock}}$
Input	Interest cover	$\frac{\text{EBIT}}{\text{Interest paid}}$
Output	Credit period (days)	$\frac{\text{Creditors}}{\text{Operating revenue (turnover)}}$
Output	Debt ratio (%)	$\frac{\text{Noncurrent and current liabilities}}{\text{Total assets}}$

There is a certain risk in the DEA method and in working with ratios. Emrouznejad and Amin (2009) stated that one of the main assumptions related to the typical efficiency frontier in the standard DEA model is the assumption of convexity. When using ratios, it is problematic not to violate this assumption. Despite a certain risk of possible devaluation of the results, however, ratios will be used, because financial ratios are typical for this type of analysis. Other assumptions regarding the production possibility set of options according to Cooper, Seiford and Tone (2007) can be considered to be fulfilled.

In general, it can be assumed that bankrupt companies have a problem in keeping up with the competition. Their products (services) are more difficult to sell, and therefore, companies have a sales problem. This fact is also negatively reflected in the company's financial statements. Production companies can then accumulate stocks; their turnover is reduced, and so on. Given this fact (and because the output variables are those that are maximized in the DEA model), two financial ratios were selected as output variables: the debt ratio and credit period. Bankrupt companies can be expected to have a higher level of indebtedness (more precisely total debt, especially liabilities), and as a result, the debt ratio will increase. The bankruptcy of the company will also negatively impact the operating cash flow, and companies will not be more likely to repay their own liabilities, thus the credit period will be prolonged.

The remaining seven financial indicators represent input variables. This group of variables contains representatives from all recommended groups, i.e., profitability ratios, liquidity ratios, solvency ratios and turnover ratios. In contrast, it can be expected that the value of these variables in the case of bankrupt companies should be lower than in the case of active companies. It can be assumed that healthy companies will be able to sell their stock and will have a higher level of profitability in all respects. Non-bankrupt companies are expected to have more efficiently adapted internal processes and to be sufficiently liquid and able to pay their obligations.

All of the described procedures are performed in MATLAB R2020b and DEA Solver-Pro version 15.

2.1. Employed bankruptcy prediction DEA models

Due to the nature of the analysis, superefficiency models were selected to compare the resulting score for units that appeared on or above the frontier. All models were estimated separately for the period of one, two and three years prior to bankruptcy. Within each period, 22 superefficiency models with different settings were constructed; see Table 2. Both oriented (input and output orientation) models and nonoriented models were considered. Models were estimated under constant and variable returns to scale. Four models were of a radial nature, and the remaining models were slack-based measure models (SBM models). Since bankrupt companies often have negative financial indicators, we decided to take into account the adjustment of standard DEA models into so-called negative data DEA models (ND models). In such models, according to Cooper et al. (2007), financial ratios are adjusted to a required value greater than zero. Furthermore, attention was given to the so-called SBM Max models. The SBM models typically report the worst efficiency scores for inefficient units. This circumstance means that the projected point is the farthest point on the associated frontier. In contrast to standard SBM models, SBM Max models look for the nearest point on the associated bankruptcy frontier. Hence, the efficiency score is maximized here; for details, see Tone (2017).

Table 2. *Overview of DEA models, including their setup.*

Type	Orientation	Returns to scale	Name
Radial (CCR)	Input	Constant	Model 1
Radial (CCR)	Output	Constant	Model 2
Radial (BCC)	Input	Variable	Model 3
Radial (BCC)	Output	Variable	Model 4
SBM	Non-oriented	Constant	Model 5
SBM	Non-oriented	Variable	Model 6
SBM	Input	Constant	Model 7
SBM	Output	Constant	Model 8
SBM	Input	Variable	Model 9
SBM	Output	Variable	Model 10
SBM Max	Non-oriented	Constant	Model 11
SBM Max	Non-oriented	Variable	Model 12
SBM Max	Input	Constant	Model 13
SBM Max	Output	Constant	Model 14
SBM Max	Input	Variable	Model 15
SBM Max	Output	Variable	Model 16
SBM ND	Non-oriented	Variable	Model 17
SBM ND	Input	Variable	Model 18
SBM ND	Output	Variable	Model 19
SBM ND Max	Non-oriented	Variable	Model 20
SBM ND Max	Input	Variable	Model 21
SBM ND Max	Output	Variable	Model 22

2.2. Characteristics of the model quality

To evaluate the success of the model classification, we follow a number of active/bankrupt companies that are on the frontier and not on the frontier. Based on these characteristics, we can calculate the total accuracy as a percentage of correctly classified subjects for all entities. Instead of the overall misclassification rate of the model, we will focus on the Type I and II errors. A Type I error evaluates the number of active companies that were falsely identified as bankrupt companies to all active companies. A Type II error shows how many bankrupt companies were incorrectly classified as active companies in ratio to all bankrupt companies. More details on these calculations can be found, for example, in Klepáč and Hampel (2018).

Based on the values of Type I and Type II errors, the ROC curve can be constructed. The ROC curve is a useful tool for evaluating classifiers based on their performance. In this context, we will deal with so-called sensitivity, defined as one minus Type I error, and specificity, which equals one minus Type II error. The area under the ROC curve

(AUC) criterion is an alternative single-number measure for evaluating the predictive ability of a model. It was proven in Ling, Huang and Zhang (2003) that the AUC value is a better measure than the total accuracy when evaluating and comparing classifiers. The resulting AUC value is between 0.5 and 1, where higher values indicate a more successful predictive ability for a model.

2.3. Optimal threshold determination

It is possible that some incorrectly classified bankrupt companies could be located near the frontier, and a shift of the bankruptcy frontier as expressed by the threshold value could improve the classification abilities of the DEA model. For this purpose, all theoretically possible thresholds are evaluated (i.e., thresholds from 0 to the maximum value of "bankruptcy score" in the individual model with 0.01 step). To find a suitable threshold, total accuracy maximization and two criteria based on ROC curves were selected.

Similar to Chen and Wu (2016), we use the Youden index, which can be represented as the difference between the probability of a sample predicted as positive when it is truly positive and the probability of the sample predicted as positive when it is not positive. A higher Youden index indicates a better ability to avoid failure in binary classification. Practically, for a particular model, we determine specificity and sensitivity values for all the possible thresholds. The Youden index is then calculated as $J = \max(sensitivity + specificity - 1)$.

Another possibility is to measure the distance from a "perfect" model with zero Type I error as well as Type II error (point [0; 1] on the ROC curve) to the nearest point of the ROC curve of the assessed model. This approach produces so-called distance-to-corner characteristics, which correspond to a suitable threshold.

2.4. Validation of the results

For decisions about the possible systematic behaviour of threshold setting, it is necessary to check the stability of the optimal threshold for particular models. For this purpose, we employ additional datasets coming from different time ranges. These consist of company data from 2013, 2014 and 2015, where some companies became bankrupt in 2016. The first dataset consists of companies from the sector NACE Code 28, i.e., the sector used for establishing optimal thresholds. In addition to threshold validation based on data from different time ranges, we employ two datasets from different sectors: the manufacture of basic metals (NACE Code 24) and the manufacture of fabricated metal products, except for machinery and equipment (NACE Code 25). These two sectors are chosen not only for their comparability with the manufacture of machinery and equipment sector but also for the existence of a sufficient number of bankruptcies with available data. The composition of the validation datasets in particular years is presented in Table 3.

Validation will be performed for DEA bankruptcy models with the best classification capabilities within the original dataset. When optimal thresholds based on NACE Code 28 sector data from 2011–2013 will give reasonably good results not only for the same

Table 3. *Composition of validation datasets.*

	Active companies (in %)			Bankrupt companies (in %)		
Year	NACE 24	NACE 25	NACE 28	NACE 24	NACE 25	NACE 28
2013	3180 (99.07)	1485 (92.47)	3187 (97.05)	30 (0.93)	121 (7.53)	97 (2.95)
2014	2784 (98.93)	1472 (92.99)	3136 (97.42)	30 (1.07)	111 (7.01)	83 (2.58)
2015	2810 (99.01)	1479 (94.44)	3188 (97.82)	28 (0.99)	87 (5.56)	71 (2.18)

sector in different time ranges but also for other sectors, we can have good faith that the optimal thresholds found will be applicable in general.

To correctly evaluate the qualification capabilities of the DEA method, the results of the best models will be compared with the results of logistic regression. For this purpose, the already tested model from Staňková and Hampel (2018) will be used. More precisely, it is a model constructed by means of forward stepwise regression – the starting model in this case contains only a constant. Logit models will be estimated for both the original and validation datasets.

3. Results

As the first part of the evaluation of the prediction capabilities of the model, we apply the common approach, where we used the value of one as the threshold for the classification of active and bankrupt companies. It was found that in such a case, the estimated models (except for Model 3) typically have a very low error rate in the group of active companies (values lower than 1%) but a very high error rate in the group of bankrupt companies (typically approximately 80 to 90%). Such models cannot be considered to be models that can be used in practice. It was also found that one of the models had problems with the superefficiency calculation. For Model 22, we were unable to obtain results in any of the three reporting periods. It can be assumed that this model does not have the appropriate settings due to the problem and data being studied.

In the case of a typical dividing point of 1, it was also found that Model 3 is different from the others. Model 3 showed the smallest error rate in the classification of the minority bankruptcy group (i.e., had the lowest value of Type II error in all three periods). However, it lags behind in terms of overall accuracy. Because of its setting, Model 3 (compared to other models) has a large number of companies on or above the bankruptcy frontier. For example, in the period of three years before bankruptcy, there were 298 companies (38 bankrupt and 260 active units). In all other models, only a few units or tens of units appeared on or above the bankruptcy frontier in this period. These specific features of Model 3 caused a dramatic reduction in the overall accuracy of the model to a value of approximately 72% (in all periods), and thus, in terms of total accuracy, Model 3 was the least suitable model. In addition, in all periods, the AUC values of Model 3 exceed the overall accuracy, which indicates that it is advisable to look for a threshold other than the value threshold. In the period of one year before bankruptcy, this fact applies not only to Model 3 but also to the eight other models.

Generally, the standard threshold equal to 1 does not reach the maximum value of the total accuracy. When choosing a model with a threshold selected to maximize the total accuracy, typically more than one appropriate threshold was found within the model. It was also found that when maximizing the overall accuracy, it is advisable very often to use thresholds higher than one. However, threshold values founded by the criterion of maximum accuracy can be described as inappropriate. For this specific threshold value, we obtain a model where the Type I error is very low but at the expense of correct classification of the less frequent companies that went bankrupt during the observed period, which results in a very low specificity value (and thus a high Type II error).

In the case of characteristics derived from ROC curves, typically one point was found relating to the given criterion. It can be stated that for the period of three years before bankruptcy, the thresholds found by the distance to corner and the Youden index are relatively consistent. Only in 8 cases (9 cases in the period of two years before the bankruptcy) did the identified thresholds differ according to these two criteria. The thresholds found typically differed by only a few hundredths, but for Models 3 and 14, the difference was 0.59 and 0.89 points in the one-year period before bankruptcy.

3.1. The results of the best models

To select the best models, the AUC values were first monitored, while the ranking of the models according to the AUC values in individual years was averaged to create the resulting average ranking of the models for the entire monitored period. Type I and II errors were monitored as a second criterion. From the group of radial models (i.e., Models 1 to 4), Model 4 was selected as the best model. Radial models in terms of AUC values had the most fluctuating results. According to the average values, Model 4 was the best, but when changing the orientation (i.e., changing to Model 3), according to the average AUC values, we obtain one of the worst models (18th in the ranking). From the group of "basic" SBM models, i.e., Models 5 to 10, Model 6 was selected as the best. In general, however, these SBM models achieved very good results (in the average ranking according to AUC, it was the 3rd to 9th position). The other SBM models fared worse in terms of average AUC values. Of the group of SBM MAX models, i.e., Models 11 to 16, Model 12 performed best. From the group of models with special treatment for negative data, i.e., Models 17 to 21, Model 17 can be identified as the best.

Several links can be found among the selected best models from each group. It was found that among the best models are models with variable returns to scale, and in three out of four cases, it is a nonoriented model. Furthermore, these models show that the value of the optimal threshold decreases with the onset of bankruptcy. It can be assumed that as the time of bankruptcy approaches, active companies become more different from bankrupt companies and therefore move away from the bankruptcy frontier, and therefore, the threshold decreases. A detailed view of the thresholds found for the four selected models according to the criterion of maximum accuracy (C1), distance to corner (C2) and Youden index (C3) for the period of one to three years before bankruptcy can be seen in Table 4. Using the thresholds found, we can make some generalizations. The Youden index and the distance to the corner show the same thresholds, with the exception of certain models for data three years before bankruptcy. Maximizing the overall accuracy gives thresholds that are substantially larger than the other criteria, and the difference decreases with increasing time to bankruptcy. Notable is Model 6, which has optimal thresholds that are mostly very similar across different times to bankruptcy and different criteria.

Table 4. *Thresholds found by the criteria C1 (maximum of total accuracy), C2 (distance to corner) and C3 (Youden index) for selected models. For criterion C1, the same accuracy values were often achieved for different thresholds, so the threshold leading to maximum accuracy is represented in the table as an interval of threshold values. This is a consequence of using an empirical ROC curve which is piecewise constant.*

Period	3 years before bankruptcy			2 years before bankruptcy			1 year before bankruptcy		
Criterion	C1	C2	C3	C1	C2	C3	C1	C2	C3
Model 4	$\langle 3.32, 5.74\rangle$	0.58	0.58	$\langle 0.57, 0.58\rangle$	0.27	0.27	0.14	0.10	0.10
Model 6	$\langle 1.25, 1.27\rangle$	0.07	0.07	0.10	0.01	0.01	0.07	0.01	0.01
Model 12	$\langle 1.25, 1.27\rangle$	0.32	0.49	$\langle 0.67, 0.68\rangle$	0.21	0.23	$\langle 0.24, 0.25\rangle$	0.05	0.05
Model 17	$\langle 1.31, 1.41\rangle$	0.15	0.24	$\langle 1.03, 1.49\rangle$	0.24	0.23	0.09	0.03	0.04

The predictive abilities of the best models from each group described above are depicted in Figure 1. This figure shows four evaluation criteria having a maximization character: the value of area under ROC curve (AUC), total accuracy (ACC), specificity (SPE) and sensitivity (SEN). The results show that the typically used point of one (magenta) as well as the threshold according to the criterion of maximum accuracy (green) lead to models where the emphasis is placed on the correct specification in the group of active companies (i.e., high sensitivity value) at the expense of correct classification in the group bankrupt companies (i.e., low value of specificity). However, if the Youden index and the distance to corner criteria are used, the results in all four evaluation areas are balanced. In cases where these two criteria did not agree on the same dividing point, the Youden index (blue) tended to have a higher sensitivity at the expense of specificity than the distance to corner (black) criterion. For all four selected models (regardless of the specific threshold), it can also be observed that the overall quality of the model decreases as the time since bankruptcy increases.

3.2. Results via validation datasets

For the four DEA models selected above, the classification capabilities of these models were verified on the other three datasets. In those cases where there was no agreement between the distance to corner criterion and the Youden index during the optimization of the threshold, only the distance to corner criterion was uniformly used, which identified

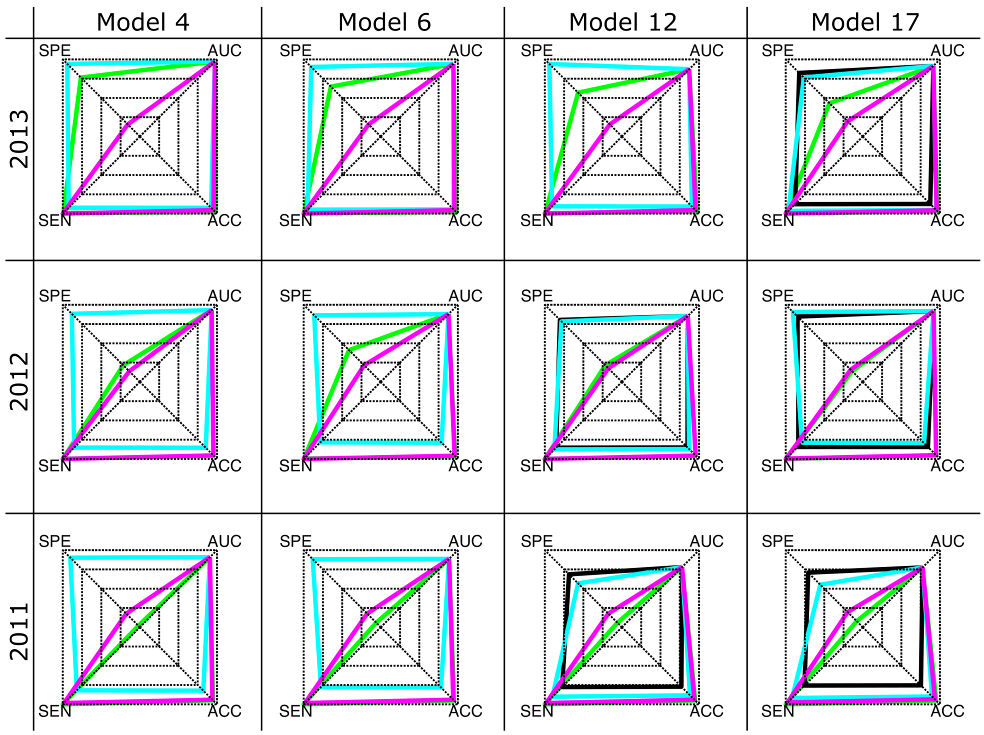

Figure 1. *Results of the area under ROC curve (AUC), total accuracy (ACC), specificity (SPE) and sensitivity (SEN) with a typical threshold equal to one (magenta) and thresholds found by the Youden index (blue), distance to corner (black) and maximum accuracy (green) for selected models. Because the thresholds coincided in some cases, not all characteristics (colours) are visible in the picture.*

thresholds that better balanced both types of errors in the original dataset. The results of total accuracy (ACC), specificity (SPE) and sensitivity (SEN) and AUC for the original dataset and for three validation datasets for one year before bankruptcy (red), two years (green) and three years before bankruptcy (blue) are shown in Figure 2. The last column in Figure 2 then presents the results for the competing logit model. As seen, the logit model based on the original dataset achieved similar results to the DEA models for the period of one year before bankruptcy. However, for other time periods and other sectors, the logit model lags significantly in the specificity values.

If we focus on the evaluation of individual DEA models, then in the case of the first validation set (NACE Code 28), Model 17 failed visibly. The classification abilities of the other models are still very good. Model 6 has the most comparable results to the original dataset. In the case of Model 4 and Model 12, very good results were achieved in the period of one year before bankruptcy, but for periods longer than one year before bankruptcy, a decrease in the specificity values can be seen for both of these models. In the case of the second validation dataset (NACE code 24), the best results were achieved

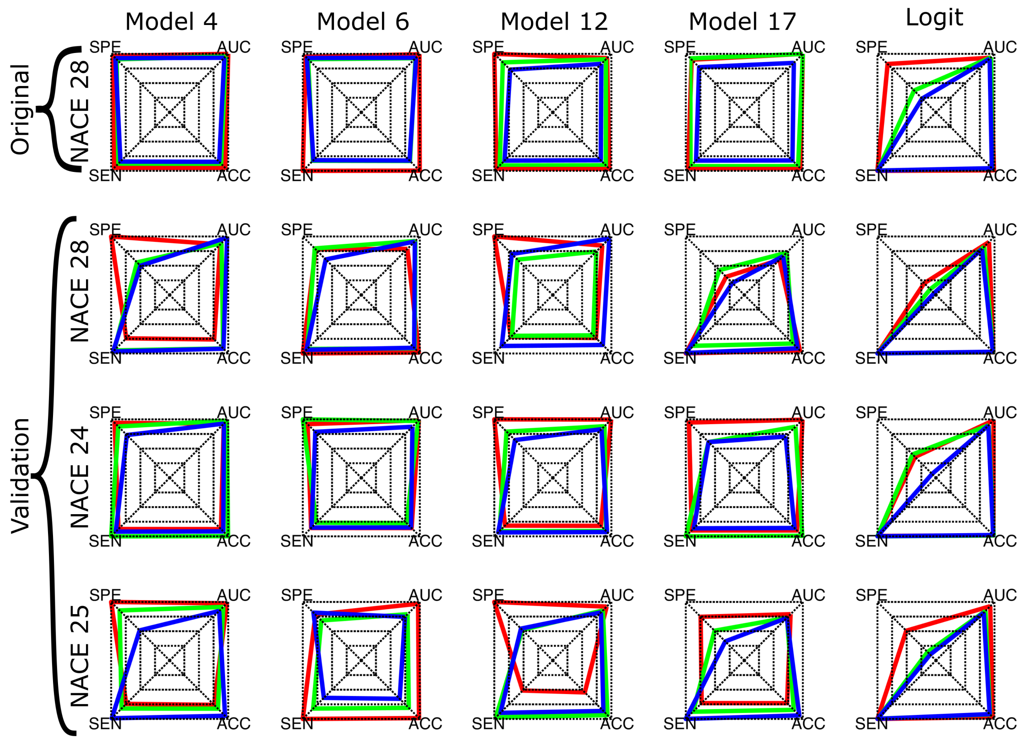

Figure 2. *Validation results for Models 4, 6, 12 and 17 together with the logit model. Original dataset characteristics with optimal threshold for selected models (the first line, NACE Code 28, bankruptcy in 2013), i.e., the area under the ROC curve (AUC), total accuracy (ACC), specificity (SPE) and sensitivity (SEN) are compared to the characteristics of the validation datasets: NACE Code 28, bankruptcy in 2016 (the second line), NACE Code 24, bankruptcy in 2016 (the third line) and NACE Code 25, bankruptcy in 2016 (the fourth line). Depicted are results for one year before bankruptcy (red), two years (green) and three years before bankruptcy (blue).*

with Models 4 and 6 throughout the observed period. Models 12 and 17 were less able to classify bankrupt companies in the period of two or more years before bankruptcy. Even in the case of the third dataset (NACE Code 25), one can see the fluctuation in the results of the models with respect to the remaining time to bankruptcy. If we were to take the classification of bankrupt companies as a priority (i.e., the specificity results), Model 6 was identified as the best.

Given the validation results, it can be stated generally that the thresholds for Model 6 appear to be stable regardless of the sector chosen and the different time periods. Model 4 can be called the second best, and Models 12 and 17 performed the worst during validation. One can assume that in these worst-case models, the optimal threshold will be more influenced by the specific characteristics of bankrupt companies in the sector. It can be said that with increasing time since the bankruptcy of a company, the ideal thresholds are more affected by other influences (industry specificities or directly by the characteristics of bankrupt companies).

4. Discussion

Generally, there is a strictly limited number of research papers dealing with the validation of DEA bankruptcy models, and subsequent threshold optimization is rarely resolved. The DEA method is not yet a broadly accepted method for the area of bankruptcy prediction (Alaka et al., 2018). However, when comparing the empirical results with common logit models (in Figure 2), the proposed DEA models have more potential for practical application. In addition, the DEA method has one advantage over conventional statistical approaches, because it does not require large datasets. This aspect allows the application of threshold-optimized DEA models in relatively small economic sectors or in the case of oligopolies.

The usefulness of threshold optimization enabled by using superefficiency models can be demonstrated by comparison with the results of Janová et al. (2012), Premachandra et al. (2009) and Staňková and Hampel (2019), where additive models are used with a standard threshold corresponding to zero slack values. Threshold optimization using the Youden-like approach of the additive DEA model is elaborated in Štefko et al. (2020). Since the proportion of active companies to bankrupt companies is not balanced in these datasets, not only the total accuracy but also the error rates for both active and bankrupt companies must be accounted for to prevent the loss of error margin classification of the less frequent companies that went bankrupt during the observed period. The characteristics of bankruptcy prediction in the abovementioned sources are summarized in Table 7.

Table 5. *Results of Janová et al. (2012), Premachandra et al. (2009), Staňková and Hampel (2019) and Štefko et al. (2020) and our results for the case one year before bankruptcy. Abbreviation ACC means total accuracy, TIE Type I error and TIIE Type II error.*

Source	ACC	TIE	TIIE
Janová et al. (2012)	0.746	0.003	0.805
Premachandra et al. (2009)	0.686	0.011	0.872
Staňková and Hampel (2019)	0.940	0.029	0.490
Štefko et al. (2020)	0.593	0.446	0.180
The best model via original dataset	0.946	0.051	0.098
The best model via validation NACE 28 dataset	0.834	0.152	0.338
The best model via validation NACE 24 dataset	0.801	0.200	0.143
The best model via validation NACE 25 dataset	0.897	0.090	0.287

It is obvious that the total accuracy of our best model results and the results reached in Staňková and Hampel (2019) are visibly higher than in Janová et al. (2012) and Premachandra et al. (2009). Inter alia, this finding can be given by the different variables used. It can be stated that the identified primary group of ratio indicators in Staňková and Hampel (2018) is suitable not only for the methods of logistic regression, support vector machines and decision trees but also to serve as a basic set for the DEA method, because both Staňková and Hampel (2019) models and our models achieved good classification results through these variables.

The research of Premachandra et al. (2009) addresses the problem of bankrupt companies' share in the dataset. They show that it is easier for the DEA method to address balanced data files (increase of total accuracy from 75% for the original dataset to 86% for the balanced dataset). We can assume that threshold optimization does not bring a serious advantage in the case of a balanced dataset, but this situation is not real. The strongly unbalanced data truly reflect the situation in today's market, which is populated far more densely by active companies than by those that are on the brink of bankruptcy. Therefore, for such datasets and especially for periods longer than one year before bankruptcy, we do not consider a threshold that is equal to or greater than one to be an optimal setting.

In Štefko et al. (2020), the authors address predicting bankruptcy in the heating industry in Slovakia. The additive DEA model and logit model are employed for this purpose. Threshold optimization based on maximization of the sum of sensitivity and specificity is provided. As in this paper, a set of 9 financial indicators with no strong correlations is used. The dataset consists of 343 companies, of which 50 were bankrupted in 2016. A relatively low total accuracy of 56% is reached, and the type II error is close to our best results, but the type I error is high. In accordance with our approach, the usefulness of threshold optimization is visible.

If we optimize the threshold in our proposed DEA models, we will not achieve the maximum total accuracy of the model, but we will obtain models where both types of errors are more balanced. From this point of view, the models proposed by us are therefore more applicable in practice than, for example, the models by Štefko et al. (2020) and Premachandra et al. (2009). With respect to the identified thresholds and classification capabilities in the original as well as the validation datasets, nonoriented SBM models proved to be the best. In general, better results were achieved by models with the assumption of variable returns to scale; however, in the case of nonoriented SBM models, the change in this setting had no significant effect on the results of the models.

Empirical results showed, among other things, that in the case of criteria derived from ROC curves, it is not advisable to use thresholds higher than 1. There is therefore no need to distinguish between companies that form the bankruptcy frontier. In practice, this means that it is possible to estimate models in their basic form (i.e., without the need to calculate superefficiency scores).

Due to the empirical shape of the estimated ROC curves, the optimal values of thresholds given by the Youden index and distance to corner do not always match exactly. However, these suitable thresholds are usually not very far away from each other. In the event that these two criteria did not provide the same values, the model based on the threshold according to the distance to corner usually balanced both types of errors slightly better. When selecting the criterion, the purpose of the models must be accounted for. If the user of the model (for example, a bank) is more interested in the correctness of the classification of a minority group of bankrupt companies, we can recommend thresholds given by the distance to corner. A model with a slightly lower overall accuracy but higher specificity will be obtained in this manner.

5. Conclusions

Given the results, it can be stated that threshold optimization can visibly improve the quality of a DEA model's bankruptcy prediction. The selection of a given threshold is individual for each type of DEA model and for the period. However, nonoriented SBM models showed that they generally have relatively low ideal thresholds according to ROC curves in the range of 0.01 to 0.07. Therefore, these models were also marked as the best. These models are the most robust in the sense of the method for optimal threshold determination and the type of returns on scale, and furthermore, these models are stable in the sense of optimal threshold for different periods before bankruptcy. Validation proved that the high quality of nonoriented SBM models' bankruptcy prediction persisted for different sector companies' data. Although we assume that the results obtained will be stable both over time and for different sectors of the economy, it will be useful in the future to check the validity of the results under different circumstances, namely in a different time frame, sector, and country. Future research will also focus on different estimation methods of ROC curves, where we can assume that smooth ROC curves will provide more stable threshold estimates.

Conflict of interest

The authors report that they have no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Author Contributions: Michaela Staňková: investigation, data curation, formal analysis, visualization, writing-original draft; David Hampel: conceptualization, formal analysis, review and editing. All authors have read and agreed to the published version of the manuscript.

References

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O. and Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94:164–184.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Azizi, H. and Ajirlu, S. (2010). Measurement of overall performances of decision-making units using ideal and anti-ideal decision-making units. *Computers & Industrial Engineering*, 59(3):411–418.

Beaver, W. (1966). Financial ratios as predictors of failure. *The Journal of Accounting Research*, 4:71–102.

Campbell, J. Y., Hilscher, J. and Szilagyi, J. (2011). Predicting financial distress and the performance of distressed stocks. *Journal of Investment Management*, 9(2):14–34.

Chang, T. M., Lin, S. J., Chen, F. H. and Hsu, M. F. (2019). Integrated multiple DEA specifications and visualization technique for advanced management analysis and decision. In *2019 20th International Conference on Mobile Data Management (MDM 2019)*, pages 491–496, Hong Kong. IEEE.

Chen, W. S. and Du, Y. K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36(2):4075–4086.

Chen, Y. and Wu, W. (2016). A prospecting cost-benefit strategy for mineral potential mapping based on ROC curve analysis. *Ore Geology Reviews*, 74:26–38.

Cooper, W. W., Seiford, M. L. and Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. Springer Science & Business Media, New York.

Eisdorfer, A. (2008). Empirical evidence of risk shifting in financially distressed firms. *The Journal of Finance*, 63(2):14–34.

Emrouznejad, A. and Amin, G. (2009). DEA models for ratio data: Convexity consideration. *Applied Mathematical Modelling*, 33(1):486–498.

Farooq, U. and Qamar, M. A. J. (2019). Predicting multistage financial distress: Reflections on sampling, feature and model selection criteria. *Journal of Forecasting*, 38(7):632–648.

Hillegeist, S. A., Keating, E. K., Cram, D. P. and Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1):5–34.

Horváthová, J. and Mokrišová, M. (2018). Risk of bankruptcy, its determinants and models. *Risks*, 6(4):117.

Iparragirre, A., Irantzu, B., Jorge, A. and Inmaculada, A. (2022). Estimation of cut-off points under complex-sampling design data. *SORT-Statistics and Operations Research Transactions*, 46(1):137–158.

Janová, J., Vavřina, J. and Hampel, D. (2012). DEA as a tool for bankruptcy assessment: the agribusiness case study. In *Proceedings of the 30th International Conference Mathematical Methods in Economics 2012*, pages 379–383, Karviná. Silesian University in Opava.

Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting Studies*, 22(3):1366–1422.

Klepáč, V. and Hampel, D. (2016). Prediction of bankruptcy with SVM classifiers among retail business companies in eu. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 64(2):627–634.

Klepáč, V. and Hampel, D. (2018). Predicting bankruptcy of manufacturing companies in EU. *E&M Ekonomie a Management*, 21(1):159–174.

Li, H., Hong, L., Mo, Y., Zhu, B. and Chang, P. (2018). Restructuring performance prediction with a rebalanced and clustered support vector machine. *Journal of Forecasting*, 37(4):437–456.

Li, Z., Crook, J. and Andreeva, G. (2014). Chinese companies distress prediction: an application of data envelopment analysis. *Journal of the Operational Research Societ*, 65(3):466–479.

Li, Z., Crook, J. and Andreeva, G. (2017). Dynamic prediction of financial distress using malmquist DEA. *Expert Syst Appl*, 80:94–106.

Ling, C. X., Huang, J. and Zhang, H. (2003). AUC: A better measure than accuracy in comparing learning algorithms. In Xiang, Y. and Chaib-draa, B., editors, *Advances in Artificial Intelligence. Canadian AI 2003. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, volume 2671. Springer, Berlin, Heidelberg.

Mendelová, V. and Bieliková, T. (2017). Diagnosing of the corporate financial health using DEA: an application to companies in the Slovak republic. *Politicka Ekonomie*, 65(1):26–44.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1):109–131.

Peláez, R., Cao, R. and Vilar, J. M. (2021). Nonparametric estimation of the probability of default with double smoothing. *SORT-Statistics and Operations Research Transactions*, 45(2):93–120.

Petropoulos, A., Siakoulis, V., Stavroulakis, E. and Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, 36(3):1092–1113.

Premachandra, I. M., Bhabra, G. S. and Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research*, 193(2):412–424.

Premachandra, I. M., Chen, Y. and Watson, J. (2011). DEA as a tool for bankruptcy assessment: A case of bankruptcy assessment. *Omega*, 39(6):620–626.

Psillaki, M., Tsolas, I. E. and Margaritis, D. (2010). Evaluation of credit risk based on firm performance. *European Journal of Operational Research*, 201(3):873–881.

Rowland, Z. and Krulicky, T. (2020). Using DEA as a useful tool for bankrupty assessment in Romanian's enterprises. In *7th International Conference on Education and Social Sciences (INTCESS 2020)*, pages 1213–1217, Istanbul. International Organization Center of Academic Research.

SAITECH Inc. (2020). DEA SolverPRO version: 15f.

Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business*, 74(1):101–124.

Staňková, M. (2022). Threshold moving approach with logit models for bankruptcy prediction. *Computational Economics*.

Staňková, M. and Hampel, D. (2018). Bankruptcy prediction of engineering companies in the EU using classification methods. *Acta Universitatis agriculturae et silviculturae Mendelianae Brunensis*, 66(5):1347–1356.

Staňková, M. and Hampel, D. (2019). Bankruptcy prediction based on data envelopment analysis. In *Mathematical Methods in Economics 2019: Conference Proceedings*, pages 31–36, České Budějovice. University of South Bohemia in České Budějovice.

Staňková, M. and Hampel, D. (2020). Efficiency assessment of the UK travel agency companies - data envelopment analysis approach. In *Mathematical Methods in Economics 2020: Conference Proceedings*, pages 550–556, Brno. Mendel University in Brno.

The MathWorks Inc. (2020). MATLAB version: 9.13.0 (r2020b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Tian, S., Yu, Y. and Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52:89–100.

Tone, K. (2017). *Advances in DEA Theory and Applications: With Extensions to Forecasting Models*. John Wiley and Sons, Hoboken.

Uddin, M. S., Chi, G. T., Habib, T. and Zhou, Y. (2020). An alternative statistical framework for credit default prediction. *Journal of Risk Model Validation*, 14(2):65–101.

Štefko, R., Horváthová, J. and Mokrišová, M. (2020). Bankruptcy prediction with the use of data envelopment analysis: An empirical study of Slovak businesses. *Journal of Risk and Financial Management*, 13(9):212.

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

Xu, M. and Zhang, C. (2009). Bankruptcy prediction: The case of japanese listed companies. *Review of Accounting Studies*, 14(4):534–558.

Xu, X. and Wang, Y. (2009). Financial failure prediction using efficiency as a predictor. *Expert Systems with Applications*, 36(1):366–373.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22:59–86.

Appendices

Table 6. *Median and average values for the used variables separately for active and bankrupt companies.*

	Median/average values for					
	Active companies			Bankrupt companies		
Variable	2011	2012	2013	2011	2012	2013
Current ratio	1.09/1.21	1.01/1.11	0.75/1.04	1.52/1.88	1.54/1.96	1.57/1.93
Cash flow liquidity	0.03/-0.03	-0.03/-0.13	-0.31/-1.59	0.14/0.24	0.15/0.22	0.15/0.23
Net working capital	0.15/0.34	0.01/-0.34	-0.43/-1.03	15.03/42.57	15.76/49.02	16.19/50.80
Return on assets	-0.13/-5.08	-3.64/-21.63	-38.03/-58.68	3.95/4.63	4.18/4.54	3.82/4.72
EBIT Margin	2.76/-4.71	-1.73/-30.34	-57.68/-1468.64	5.07/5.52	4.99/5.48	5.17/5.75
Stock turnover	3.16/5.11	2.75/4.31	2.57/15.37	5.61/9.34	5.75/10.49	5.86/10.74
Interest cover	1.11/7.00	-0.78/-1.32	-9.49/-26.15	6.90/31.17	6.97/30.27	7.82/39.82
Credit period	97.94/131.83	114.01/232.88	230.94/2611.99	46.86/53.32	42.93/49.07	43.03/48.82
Debt ratio	93.14/94.28	97.45/114.32	138.58/182.67	66.88/65.77	64.90/64.48	63.64/63.14

Table 7. *Correlation coefficients for the variables used in individual years (2011/2012/2013).*

	Current ratio	Cash flow liquidity	Net working capital	Return on assets	EBIT Margin	Stock turnover	Interest cover	Credit period	Debt ratio
Current ratio	1/1/1	.75./.21/-.15	.12/.21/.12	.19/.09/.02	.18/.05/.00	.01/.01/-.02	.07/.07/.15	-.21/-.08/.00	-.36/-.26/-.26
Cash flow liquidity		1/1/1	.06/.06/.02	.48/.54/.48	.46/.20/.02	.03/.01/-.02	.02/.29/.08	-.22/-.11/-.02	-.30/-.44/-.37
Net working capital			1/1/1	.05/.05/.04	.08/.03/.01	-.02/-.02/-.04	.01/.01/.00	-.08/-.04/-.01	-.13/-.09/-.06
Return on assets				1/1/1	.20/.09/0.06	.04/.00/-.07	.35/.23/.21	-.26/-.08/-.06	-.35/-.48/-.64
EBIT Margin					1/1/1	.02/.01/.01	.27/.08/.02	-.20/-.29/-.48	-.34/-.08/-.06
Stock turnover						1/1/1	.02/.00/.00	-.10/-.04/-.01	.00/.02/.06
Interest cover							1/1/1	-.08/-.04/-.02	-.21/-.23/-.20
Credit period								1/1/1	.24/.13/.07
Debt ratio									1/1/1