The application of ensemble methods in forecasting bankruptcy*

Miklós Virág – Tamás Nyitrai

In practice, one chosen method is generally used to solve classification tasks. Although the most modern procedures yield excellent accuracy rates, international research findings show that a concurrent (ensemble) application of methods with weaker classification performance achieves comparable rates of high accuracy. This article's main objective is to compare the predictive power of the two ensemble methods (Adaboost and Bagging) most commonly used in bankruptcy prediction, using a sample consisting of 976 Hungarian corporations. The article's other objective is to compare the accuracy rates of bankruptcy models built on the deviations in specific financial ratios from industry averages to those of models built on financial ratios and variables factoring in their dynamics.

JEL-codes: C38, C49, G33

Keywords: bankruptcy prediction, ensemble methods, industry average, decision trees

1 Introduction

Increasing uncertainty surrounds the global economy in the wake of the economic crisis and growing market competition, rendering business decisions more difficult and making forecasts regarding the future survival of corporations more important than ever before (Cao, 2012). Bankruptcy can result in substantial losses for all stakeholders of corporate activity – in particular, shareholders and management, and ultimately, the entire national economy. Therefore, bankruptcy forecasting has been one of the key focus areas of the financial and accounting literature (Kim and Kang, 2012).

* The views expressed in this article are those of the author(s) and do not necessarily reflect the offical view ot the Magyar Nemzeti Bank.
Miklós Virág, Professor (Corvinus University Budapest, Department of Enterprise Finances); Tamás Nyitrai, PhD candidate, research assistant (Corvinus University Budapest, Department of Enterprise Finances).

In methodological terms, bankruptcy prediction is a simple task of classification with the purpose of distinguishing between solvent and insolvent corporations as accurately as possible. According to Du Jardin (2010), there are over 500 methodological procedures for this purpose; nevertheless, no consensus has emerged on which method to use when building bankruptcy forecasting models (Oreski et al., 2012).

The statistical and data mining methods that are suitable for classification purposes, along with developments in supporting IT, have significantly shaped the evolution of bankruptcy prediction. The past period has seen rapid progress in both areas, affecting bankruptcy forecasting as well. While the presentation of methodological comparative analyses remains the main method in this discipline today (Sánchez-Lasheras et al., 2012), a new direction of research in this field is the investigation of possible applications of method combinations (Cao, 2012).

Ensemble methods constitute a special type of method combination. They consist of drawing multiple random samples with replacement from the sample available for modelling, to which a previously selected classification procedure is applied. The model's final accuracy rate is then determined as the average of the predictions yielded by the various models, which characteristically exceeds the accuracy rate achieved by using the given method alone. One of the main objectives of this article is to compare the two most popular ensemble methods in the international literature – the Adaboost and Bagging[1] procedures – and determine whether the prediction rate they yield exceeds the accuracy achieved by using only the classification method chosen for this study.

Quotient-type financial and accounting ratios are traditionally used in models for bankruptcy prediction as the explanatory variable. However, these ratios are often used in modelling with no basis of comparison, despite the fact that even university manuals emphasise that financial ratios are not an absolute criterion and can only be objectively assessed in light of a basis of comparison (Virág et al., 2013). Platt and Platt (1990) recommend using relative industry ratios, comparing corporate indicator values to the 100-fold multiple of the industry average. This approach presents the advantage of allowing a comparison of financial ratios of corporations engaged in different industries and provides a solution to the temporal instability of bankruptcy models.[2]

Aside from these benefits, however, the use of relative industry ratios also presents some challenges. It is important to take into account the fact that several financial ratios are measured on an interval scale, which poses a problem in the case of relative industry ratios. For instance, comparing negative corporate profitability to a negative industry

1 To our knowledge, there are no accurate Hungarian equivalents for Adaboost and Bagging, so we will use the English terms when referring to these methods.

2 The temporal instability of bankruptcy models refers to the fact that the accuracy rate of models built on data for a given moment in time or period declines substantially for data located at a later time.

average could be misleading.[3] In such cases, a positive relative industry indicator does not necessarily provide a realistic picture, compared to the industry average. To address this issue, this study examines the approach consisting of corporate ratios, using the difference compared to their own industry average instead of their quotient as the explanatory variable. In addition to providing a methodological comparison, the study also scrutinises the accuracy rate achieved by deviations calculated compared to the industry average – used as explanatory variables – in a sample containing the data of 976 Hungarian corporations.

The international literature has also increasingly focused on the issue of the majority of modellers using only financial ratios from the last year preceding the crisis among the models' explanatory variables. The problem of this approach is that it ignores the nature of corporate financial management and only uses corporations' static financial situation during modelling (Chen et al., 2013). According to the article's assumption, comparing a corporation's specific financial ratio values to the equivalent values for the preceding period could impart relevant information. The article also examines raw financial ratios and their deviations from industry averages, as well as the accuracy rates that can be achieved if the correlation of specific financial ratios with the equivalent ratios of earlier periods is included as a separate variable in bankruptcy models.

The study attempts to apply these three groups of variables (raw, deviation from the industry average, dynamic) individually or in a combined manner in bankruptcy forecasting models. The study aims to determine whether these groups of variables are individually or collectively better suited to maximise the accuracy rate of bankruptcy models, which is the primary objective in developing bankruptcy models (Du Jardin, 2010).

It is important to emphasise that the analysis presented in this article is experimental. It is not based on a representative sample and it does not intend to present a bankruptcy model optimised for a specific forecast. That said, the available sample and the results achieved with it could be suitable for answering the research questions posed here, while the conclusions drawn from the findings could be the subject of further studies in the field of bankruptcy prediction and other independent areas of research.

The study features the following structure: the next part presents international publications that examine the bankruptcy prediction performance of ensemble methods in the context of comparative analyses. The third part includes a brief presentation of the Adaboost and Bagging procedures. Part four showcases the sample and the research methods applied in the empirical study. Part five covers the findings of the study. The sixth and final part summarises the conclusions drawn from the analyses and suggests possible areas for further research.

3 According to the database available for this study, several industries have been plagued by a negative industry average in the context of economic recession.

2 Theoretical bases of ensemble methods

The operation of ensemble methods presented here follows Marqués et al. (2012a). These procedures are called an ensemble method, because a given classification method is applied multiple times to a dataset. When applying ensemble methods, a sample with replacement is taken from the observations or the available input variables, and the classification method is only applied to the selected partial sample.

When applying ensemble methods, it is necessary to determine in advance the number of times that the dataset will be sampled. In the case of a partial sample consisting of p elements, the procedure will be applied p times. The results of these have the same weight in the elaboration of the final prediction of the ensemble methods; in other words, individual observations are classified into the group where they featured most frequently during the application p.

In order for the accuracy rate achieved using ensemble methods to exceed the outcomes obtained by the application of individual methods, the findings of the different model applications should differ to the greatest possible extent. This diversity can be achieved if a different partial sample is used for every application p from among the available observations. This is the principal used by the Adaboost and Bagging procedure. The latter takes p samples with replacement from the available dataset, applies the classification method chosen by the user to each of them, and takes the simple arithmetic average of their outcome to establish the final classification. The Adaboost procedure only differs in that sampling changes as a function of the results of model applications, as observations that are incorrectly classified upon application i have a greater chance of being included in sample i+1, while those classified correctly based on partial sample i have less chance of being included in sample i+1. The final classification in this case will also be into the class where the observations wound up most frequently during the application of the model p times.

The performance of classification methods is sensitive to the relevance of the independent variables used to build the model. Different input variables using the same classification method yield different accuracy rates. The diversity essential for the effectiveness of ensemble methods can also be achieved if a sample with replacement is used from the available variables instead of the available observations. In the literature, this method is referred to as “random subspace”.

The first approach is applied in this study: the case where the use of samples with replacement taken from the observations ensures diversity within the group. The article compares the aforementioned Adaboost and Bagging method using the data of Hungarian corporations, which, as far as the authors are aware, is the first of its kind in an academic publication of Hungarian economic literature.

3 An overview of the literature

The use of ensemble methods is currently topical in the international literature on bankruptcy prediction. The following section presents some of the findings of comparative analyses published over the past years, without attempting to be comprehensive.

Alfaro et al. (2008) compared the predictive power of the Adaboost method as applied to neural networks and decision trees. According to the findings of those authors, the models built using the ensemble procedure yielded far better results than neural networks and even linear discrimination analysis.

In contrast, using data from Russian manufacturing firms, Fedorova et al. (2013) did not find any significant difference between the use of neural networks alone or in conjunction with the Adaboost procedure.

Ensemble methods apply a single classification method multiple times. The performance of classification methods, however, is substantially influenced by the quality of the explanatory variables used, which highlights the importance of variable selection in bankruptcy prediction. Wang et al. (2014) posit that the accuracy of bankruptcy models may deteriorate considerably with the presence of unnecessary explanatory variables among the input variables. Accordingly, the authors attempted to integrate a variable selection procedure into the Adaboost method. The accuracy rate of bankruptcy models built using complex methods significantly exceeded those of traditional ensemble methods, as well as the most commonly used individual techniques.

The majority of studies on bankruptcy prediction present a comparative analysis of the various methods available (Sánchez-Lasheras et al., 2012). The data mining methods suitable for classification purposes and the development of supporting IT tools form the basis of this, and the main motives for such comparisons stem from the absence of a consensus on the optimal method for bankruptcy prediction (Oreski et al., 2012). The outcomes of these comparisons of various methods suggest that different methods yield the highest accuracy rates in different cases. This begs the question of whether a similar pattern can be observed for ensemble methods. Marqués et al. (2012a) attempted to find the answer, comparing the performance of seven classification methods with the six most frequently used ensemble methods based on the databases of six bankruptcy models. Their findings suggest that the C4.5 classification procedure yielded the most reliable results of the various ensemble methods. Accordingly, this article also uses that technique.

Marqués et al. (2012b) try to enter the question of whether the concurrent application of two ensemble method approaches (sampling based on observations and on variables) would improve the forecasting ability of the models. They conducted their analyses using six different bankruptcy model databases. The results suggest that the concurrent use of ensemble methods achieved better results than those obtained by individual procedures or a single ensemble method.

A special set of ensemble methods consists of applying several different classification procedures to a full set of data instead of applying a single method multiple times to a subset of observations and/or variables within the training dataset, with the results being combined to yield the final classification of observations into different groups. The work of Cao (2012) in this domain deserves a mention, as it attempts to combine the results of various methods using the data of corporations quoted on the Chinese stock market. Although Cao's findings significantly exceeded the accuracy rates achieved using individual methods, it should nevertheless be noted that, contrary to traditional ensemble methods, the author did not perform analyses; therefore, there is no empirical research evidence corroborating whether ensemble methods applying a single classification method several times or procedures incorporating the outcomes of multiple classification methods can be regarded as superior in bankruptcy prediction.

4 The examined dataset and the research methods used

4.1 Sample

In an attempt to answer the questions posed by the study, a sample comprising 976 elements compiled from a data collection of the authors and consisting of solvent and insolvent corporations (in a ratio of 51/49 per cent) was used. Accordingly, the sample is not representative by any means, which is not uncommon in bankruptcy prediction. The overrepresentation of insolvent companies is due to the fact that data mining procedures based on machine learning are prone to specialising on the attributes of the dominant group in cases of unequal distribution (Horta and Camanho, 2013), which may result in the overly low accuracy rate of insolvent corporations in bankruptcy prediction. For the purposes of sampling, corporations were considered to be insolvent if they were under bankruptcy or winding-up proceedings at the time of data collection, according to the Trade Register.

Criteria adhered to during sampling:

1. Only in the sample were observations featuring data available for at least the three past years in the website of the Ministry of Justice and Public Administration's Company Information and Electronic Company Registration Service.[4] This sampling criterion was necessary to eliminate very new businesses, which are more similar to older, insolvent corporations than functioning ones due to their initial difficulties (Du Jardin, 2010). In addition, the study also uses dynamic financial ratios that compare corporations' latest observed financial ratio values to data pertaining to the previous period, which requires the availability of financial ratio values at least for the past three years.

2. Observations for companies that failed to generate sales revenue for two consecutive years were not included in the sample. This is because such corporations do not conduct material business activity, and their inclusion in the sample could distort the outcomes of the models.

3. Observations featuring financial ratios that did not exhibit any deviation throughout the at least three-year horizon were not included in the sample. With those, the standardised values needed for the calculation of dynamic financial ratios used in the study could not be calculated, as the deviation of the observed data was used as the basis of basis of comparison for the calculation.

According to Du Jardin (2010), the common approach in bankruptcy prediction is to use the financial ratios that have proven relevant in other studies as the explanatory variable. This study also follows this approach. When selecting input variables, the ratios of the first Hungarian bankruptcy model (for more detail, see Virág and Hajdu, 1996) and our own data were used. Table 1 lists the names and calculation methods of the 17 ratios. Virág et al. (2013) discuss the contents of the ratios in detail. The different ratios were calculated on the basis of balance sheet items and the relevant profit and loss account rows at their closing date value.

Return on equity, an indicator frequently used in bankruptcy models, often poses the problem of dual negative division (Kristóf, 2008), for which the literature does not present a clearly preferred method; therefore, this indicator was not taken into account in the calculations.

4 http://e-beszamolo.kim.gov.hu/kereses-Default.aspx

Table 1
Name and calculation method of the ratios used in the empirical study

Indicator	Calculation method
Liquidity rate	Current assets/Current liabilities
Quick ratio	(Current assets-Inventories)/Current liabilities
Liquid assets ratio	Liquid assets/Current assets
Cash flow/Liabilities	(Net profit+Depreciation write-off)/Liabilities
Cash flow/Current liabilities	(Net profit+Depreciation write-off)/Current liabilities
Capital stock	(Fixed assets+Inventories)/Equity
Asset velocity	Net sales/Balance sheet total
Inventory velocity	Net sales/Inventories
Receivable velocity	Receivables/Net sales
Indebtedness	Liabilities/Balance sheet total
Equity ratio	Equity/Balance sheet total
Creditworthiness	Liabilities/Equity
Return on sales	Net profit/Net sales
ROA	Net profit/Balance sheet total
Receivables/Current liabilities	Receivables/Current liabilities
Net working capital ratio	(Current assets-Current liabilities)/Balance sheet total
Company size	Natural-based logarithm of the stock of assets

Quotient-type ratios also present another typical problem that arises when the denominator is zero. This issue is often handled by regarding such data as missing values, which is then replaced by a median observation value or an extreme percentile. It should be pointed out, however, that the former approach does not necessarily provide consistent values to bankruptcy forecasting models, while in the latter case, the substitution of the value in question could be sample-specific. This study opted for the approach of choosing 1 in cases where the denominator would be zero.

The sample based on the above criteria yielded a database containing the financial ratios of 976 Hungarian corporations and their deviations from the industry average[5] for the 2001–2012 period. The last business year observed for the observations included in the sample cover the 2009–2012 interval. Due to the random nature of the sample, it is highly heterogeneous, including micro-enterprises with small volumes of assets alongside medium-sized and large enterprises. The sample is also heterogeneous in terms of activities, and it includes the key areas of the economy (agriculture, industry, commerce, IT, etc.). Despite this strong heterogeneity, the models presented in the article feature an accuracy rate of about 80 per cent, meaning that they should be able to function as a

5 The sectoral classification of the corporations included in the sample was based on the data included in the publicly accessible online Trade Register, effective on the sampling date.

basis for a practical application of the conclusions drawn from the findings of the study and, hopefully, spur additional research.

4.2 Research methods

The primary objective of the study is to compare the forecasting performance of the Adaboost and Bagging procedures using the C4.5 classification method, a data mining procedure yielding a decision tree. We chose (p) 100 as the number of members for the ensemble methods and at least five observations were required for the C4.5 procedure to form a new branch. The last branches of the decision trees generated were classified into the group where the ratio of the specific group (solvent/insolvent) was higher.[6] Nyitrai (2014) briefly presents the methodological background of the classification technique; see Quinlan (1993) for in-depth coverage of the topic.

The predictive power of the examined methods was estimated by using a 100-fold randomised distribution of the sample. The procedure consists of distributing the available dataset into a training and testing sample in a proportion of 75/25 per cent using 100 randomly chosen distribution points. Forecasting capacity was measured as the average of the accuracy rate of the 100 testing samples. The accuracy rate of models refers to the ratio of correctly classified observations as compared to the total number of observations.

The other objective of the article is to examine the use of the deviation of financial ratios from the industry average, alongside or instead of raw financial ratios in the models, as well as dynamic financial ratios quantifying the correlation of the financial ratio of the most recently observed year with the corporation's earlier equivalent financial ratio values. This was quantified by means of the following formula:

$$\frac{X_{i,t-1} - X_{i,min_{(t-2;t-n)}}}{X_{i,max_{(t-2;t-n)}} - X_{i,min_{(t-2;t-n)}}}$$

In the formula, i refers to a specific corporation, t to the year we would like to make prediction and n the length of the time series available for the observation (number of years observed).

As the ratio time series of the corporations observed contained outliers in many cases, the financial ratio time series were standardised for all observations by using the time series

6 This approach should only be used when the two groups are present in equal proportions in the sample. Otherwise, the performance of the model should be assessed using a GINI indicator or the ROC curve, avoiding the definition of subjective cut-off values.

average and deviation,[7] then values falling outside the two deviations were replaced with the closest values still within the range[8] in order for the above formula to most accurately express the correlation of the most recently observed ratio of the corporation under review with the equivalent value of the preceding period.

The reader may why ask outliers were defined by using the two deviation ranges for the ratio time series. As a statistical rule of thumb, deviation ranges of 5, 3 and 2 are used for eliminating outliers. The “strictest” value (deviation of 2) was chosen because the time series available for calculating the average and deviation for the observations consist of 2–11 elements. Experimental calculations showed that in the case of such short time series, bankruptcy models based on dynamic ratios exhibited superior predictive performance if “stricter” rules were applied.

5 Findings of empirical studies

The research presented in the study defined two objectives:

1. To examine whether there is any significant difference between the accuracy rates of the Adaboost and Bagging procedures when applying the C4.5 classification method. Here arises the question of whether it is at all worthwhile to use ensemble methods instead of just using the C4.5 procedure alone. To examine this topic, calculations were also performed with the selected classification method alone.

2. To compare raw financial ratios, the dynamic ratios generated from them, and the use of the deviations of raw ratios from industry averages as input variables in the aforementioned models.

The research methods presented in the previous section were used in an attempt to answer these research questions. Table 2 presents the outcomes of the calculations.

7 When calculating the average and deviation for standardisation, data was used between time points t-2 and t-n of the ratio time series of the observations.

8 To allow the most accurate estimation of the financial position of the corporation under review in the most recently observed year in function of the preceding period, the data pertaining to the most recently observed year (t-1) was not replaced nor was it used for substitution.

Table 2
Accuracy rates of experimental model applications on average for the testing samples

Methodology	Input variable group							Average
	raw	industry	dynamic	raw dynamic	raw industry	raw dynamic industry	dynamic industry	
Adaboost	78.80%	76.06%	78.67%	81.33%	79.33%	81.51%	80.34%	79.43%
Bagging	79.88%	77.97%	79.91%	83.01%	79.76%	82.17%	81.29%	80.57%
Independent	74.34%	70.64%	72.37%	75.12%	73.29%	74.91%	73.31%	73.43%
Average	77.67%	74.89%	76.98%	79.82%	77.46%	79.53%	78.31%	

Based on the findings, the following conclusions can be drawn:

– The accuracy rate of the Bagging procedure surpassed that of the Adaboost in all cases; however, this difference was typically around 1 percentage point.

– When applying the C4.5 method alone and in the case of the Bagging procedure, the best performance was achieved when dynamic ratios were included alongside raw ratios; however, in the case of the Adaboost method, the best predictive performance was achieved by the model where all three variable groups were included in the group of independent variables. Among the models created, the best predictive performance was achieved using the Bagging method when dynamic ratios were included alongside raw ratios among explanatory variables.

The question of the extent to which these results stem from sampling and idiosyncrasies remains open; in other words, this means the significance of accuracy rate differences among the various methods. In view of the fact that, based on the calculations performed here, a normal distribution of accuracy rates cannot be presumed for any of the three procedures under review, the significant difference between classification accuracy was examined using the Mann-Whitney test, the non-parameter equivalent of the independent sample t-test (Du Jardin, 2010). The test yielded a significant difference in all three comparisons for all relevant significance levels, which means that the discrepancies found in Table 2 are significant. Thus, the Bagging procedure's accuracy rate in the case of the C4.5 method significantly exceeds that of the Adaboost procedure, albeit slightly; and the performance of both ensemble methods significantly exceeds the results achieved by the application of the C4.5 classification method alone. The difference was far greater in the latter case, around 6–7 percentage points.

The other objective of the study was to compare the accuracy rates of the models built using various groups of variables. The findings in Table 2 suggest that the Bagging procedure and the application of the C4.5 method alone yield a better predictive performance if the dynamic ratios generated from raw financial ratios are included alongside the latter. The Adaboost procedure forms an exception, exhibiting the best predictive performance when all three variable groups were included among the models' independent variables. The testing of the significance of the difference between the various accuracy rates obtained using different sets of variables also calls for the verification of their normality. The pertaining test rejects the assumption of a normal distribution of accuracy rates at all relevant significance levels, so a nonparameter test must be used to examine the significant differences among the outcomes yielded by the application of different variable groups. Table 3 presents the p-values of the Mann-Whitney test applied in pairs among the different independent variable groups.

Table 3
P-values of the Mann-Whitney test applied to the average accuracy rates of the variable groups under review

	raw	industry	dynamic	raw dynamic	raw industry	raw industry dynamic	industry dynamic
raw	-	0.000	0.134	0.000	0.827	0.000	0.003
industry		-	0.000	0.000	0.000	0.000	0.000
dynamic			-	0.000	0.220	0.000	0.000
raw dynamic				-	0.000	0.389	0.000
raw industry					-	0.000	0.002
raw industry dynamic						-	0.001
industry dynamic							-

In cases where p is close to zero, it can be presumed with a great likelihood that the average accuracy rates included in Table 2 differ for reasons other than sampling idiosyncrasies. This was fulfilled in most cases. Test outcomes can be interpreted in the following manner.

No significant discrepancy were seen in the following cases:

- between raw financial ratios and dynamic rates;
- between raw financial ratios and cases where deviations from the industry average were also included;
- between dynamic ratios in cases where deviations from the industry average were also included alongside raw financial ratios;
- in cases incorporating all three groups of variables and where deviations from the industry average were also included alongside raw financial ratios.

The differences in the averages included in Table 2 are significant, except in the above cases. This confirms statistically that the best predictive performance was achieved when dynamic ratios were included alongside raw financial ratios among the input variables. The case where deviations from the industry average were included among independent variables does not differ significantly from this. This suggests that the latter has no added value in terms of the models' predictive power. A surprising outcome is that the accuracy rate achievable using deviations from the industry average is significantly the weakest in all possible combinations.

6 Summary

The main reason for this study was the trend in the international literature recommending the use of ensemble methods in bankruptcy prediction instead of an application of the classification method alone. Research findings confirm that the accuracy rate of models can thus be improved significantly.

Given the fact that no similar study has yet been published in the Hungarian economic literature, the performance of the two most commonly used ensemble methods (Adaboost and Bagging) were compared with the application of the C4.5 classification method. This choice was based on the fact that the comparative analysis of Marqués et al. (2012a) revealed this procedure as yielding the greatest gains in performance when using one classification method alone.

Based on the findings, a significant improvement can be achieved when using the C4.5 procedure in combination with ensemble methods, compared to the use of the method alone, for both the Adaboost and Bagging procedures.

Three variable groups were compared in this study: raw financial ratios, the dynamic ratios generated from them (which quantify the correlation of the most recently observed ratio for a specific corporation, compared to the equivalent value for the preceding period), and the deviation of raw financial ratios compared to the industry average. The relative industry ratio formula recommended by Platt and Platt (1990) were replaced with custom industry ratios, as the interpretation of the former is problematic in the case of financial ratios measured on an interval scale. The research findings reveal that the predictive power of models based on the deviation of corporate ratios from the industry average significantly falls short of the accuracy rate achievable using raw and dynamic financial ratios in all possible combinations examined in the study.

The best predictive performance was achieved when dynamic ratios were included alongside raw financial ratios among the models' input variables. This result is suggestive of a synergy between these two groups of variables. The findings of the nonparameter test performed among the average accuracy rates of the examined variable groups suggest that the average predictive power resulting from the joint application of these two variable groups significantly exceeds the result achieved by any combination of the three groups of variables. The sole exception was the concurrent application of all three variable groups; however, the accuracy thus achieved does not differ significantly from the accuracy achieved by using the raw ratios and the dynamic ratios generated from them, which suggests that industry deviations have no added value in terms of model performance.

It is important to emphasise that these conclusions were drawn solely from an application of the C4.5 method. It may be worthwhile to examine the robustness of these findings using other classification methods as well.

While in this study the models based on the rate of deviation from the industry average did not result in a material improvement, compared to models based on raw financial ratios, this result does not mean that the use of the industry mean is unwarranted in bankruptcy prediction. In a Hungarian context, for instance, Kristóf (2005) experienced an improvement in predictive performance based on his research conducted on the first Hungarian bankruptcy model database. Kristóf attributed the improvement in part to the application of industry ratios. Therefore, a possible direction for future research could consist of examining ways of managing interpretation issues of relative industry ratios in cases of financial ratios measured on an interval scale, so that the financial ratios compared to the industry average might improve the models' predictive power. Based on the findings of the article, examining the classification performance that can be achieved by using the dynamic financial ratios calculated from industry ratios could be a relevant area of research.

Another question warranting further scrutiny is why the models' predictive powers show no improvement with the application of financial ratios compared to industry averages. One possible explanation could be that the scopes of activity of the corporations included in a sample were defined based on the principal activity "Standard Classification of All Economic Activities (TEÁOR)" listed in the Trade Register. In fact, the corporations observed may rely on other ancillary activities, on which the publicly available databases did not yield any information. It can be assumed that these ancillary activities have sufficient significance that the industry averages based on the principal activities' TEÁOR codes are less suitable for use as bases of comparison in case of financial ratios.

Finally, it is necessary to highlight another limitation of the analysis, the resolution of which could also be a subject of future research. As a statistical rule of thumb, this study used a range of two deviations to identify outliers. This "definition" was the authors' own arbitrary choice and may not necessarily be the optimal one. The analyses of this study can be conducted again, using other definitions of outliers, and the results compared.

References

Alfaro, E., García, N., Gámez, M. and Elizondo, D. (2008): "Bankruptcy forecasting: An empirical comparison of Adaboost and neural networks", *Decision Support Systems*, Vol. 45, pp. 110–122.

Cao, Y. (2012): "MCELCCh: Financial distress prediction with classifier ensembles based on firm life cycle and Choquet integral", *Expert Systems with Applications*, Vol. 39, pp. 7041–7049.

Chen, N., Riberio, B., Vieira, A. and Chen, A. (2013): "Clustering and visualization of bankruptcy trajectory using self-organizing map", *Expert Systems with Applications*, Vol. 40, pp. 385–393.

Du Jardin, P. (2010): "Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy", *Neurocomputing,* Vol. 73, pp. 2047–2060.

Fedorova, E., Gilenko, E., and Dovzhenko, S. (2013): "Bankruptcy prediction for Russian companies: Application of combined classifiers", *Expert Systems with Applications*, Vol. 40, pp. 7285–7293.

Horta, I. M. and Camanho, A. S. (2013): "Company failure prediction in the construction industry", *Expert Systems with Applications*, Vol. 40, pp. 6253–6257.

Kim, M-J. and Kang, D-K. (2012): "Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction", *Expert Systems with Applications,* Vol. 39, pp. 9308–9314.

kristóf, t. (2005): "A csődelőrejelzés sokváltozós statisztikai módszerei és empirikus vizsgálata.", *Statisztikai Szemle*, Vol. 83, No. 9, pp. 841–863.

Kristóf, T. (2008): "A csődelőrejelzés és a nem fizetési valószínűség módszertani kérdéseiről", *Közgazdasági Szemle*, Vol. 55, pp. 441–461.

Marques, A. I., Garcia, V. and Sanchez, J. S. (2012a): "Exploring the behaviour of base classifiers in credit scoring ensembles", *Expert Systems with Applications*, Vol. 39, pp. 10244–10250.

Marques, A. I., Garcia, V., And Sanchez, J. S. (2012b): "Two-level classifier ensembles for credit risk assessment", *Expert Systems with Applications*, Vol. 39, pp. 10916–10922.

Nyitrai, T. (2014): "Validációs eljárások a csődelőrejelző modellek teljesítményének megítélésében", *Statisztikai Szemle*, Vol. 92, No. 4, pp. 357–377.

Oreski, S., Oreski, D. and Oreski, G. (2012): “Hybrid system with genetic algorythm and artificial neural networks and its application to retail credit risk assessment”, *Expert Systems with Applications*, Vol. 39, pp. 12605–12617.

Platt, H. D. And Platt, M. B. (1990): “Development of a class of stable predictive variables: The case of bankruptcy prediction”, *Journal of Business Finance and Accounting*, Vol. 17, No. 1, pp. 31–44.

Quinlan, J. R. (1993): “*C4.5: Programs for Machine Learning*”, San Mateo, CA: Morgan Kaufmann.

Sánchez-Lasheras, F., De Andrés, J., Lorca, P., and De Cos Juez, F. J. (2012): “A hybrid device for the solution of sampling bias problems in the forecasting of firms’ bankruptcy”, *Systems with Applications*, Vol. 39, pp. 7512–7523.

Virág, M. and Hajdu, O. (1996): “Pénzügyi mutatószámokon alapul csődmodell-számítások”, *Bankszemle,* Vol. 40, No. 4, pp. 42–53.

Virág, M., Kristóf, T., Fiáth, A. and Varsányi, J. (2013): *Pénzügyi elemzés, csődelőrejelzés, vállalati válságkezelés*. Budapest: Kossuth Kiadó.

Wang, G., Ma, J. and Yang, S. (2014): “An improved boosting based on feature selection for corporate bankruptcy prediction”, *Expert Systems with Applications*, Vol. 41, pp. 2353–2361.