Theoretical and Applied Economics
Volume XXIV (2017), No. 1(610), Spring, pp. 5-22

Bankruptcy prediction logit model developed on Romanian paired sample

Daniel BRÎNDESCU-OLARIU
West University of Timișoara, Romania
daniel.brindescu@e-uvt.ro, www.levier.ro

Abstract. *The current research targeted to develop a methodology for bankruptcy prediction applicable for Romanian companies. Under the effect of the economic crisis, as well as of the entrance in the European Union, Romania has registered a significant increase of the yearly bankruptcy frequency, which makes the availability of bankruptcy assessment tools more important than ever before. Using the logistic regression, the study proposes a multivariate model based only on financial data easily accessible to all the stakeholders of the company.*
The target population consisted of all the companies from the Timis County with annual sales of over 10,000 lei (aprox. 2,200 Euro). The model was developed over a paired sample that included all the companies from the 2010 target population that went bankrupt by the end of 2012. The testing was performed over the entire target population from the period 2007-2010. The study has thus included 53,252 yearly financial statements from the period 2007-2010.
The proposed model will not allow the analyst to conclude with absolute certainty whether the analyzed company will go bankrupt or not within a two year horizon. Instead, it will offer the possibility of placing the companies on risk classes, thus creating the basis for a differentiated behavior of the stakeholder toward a specific company.

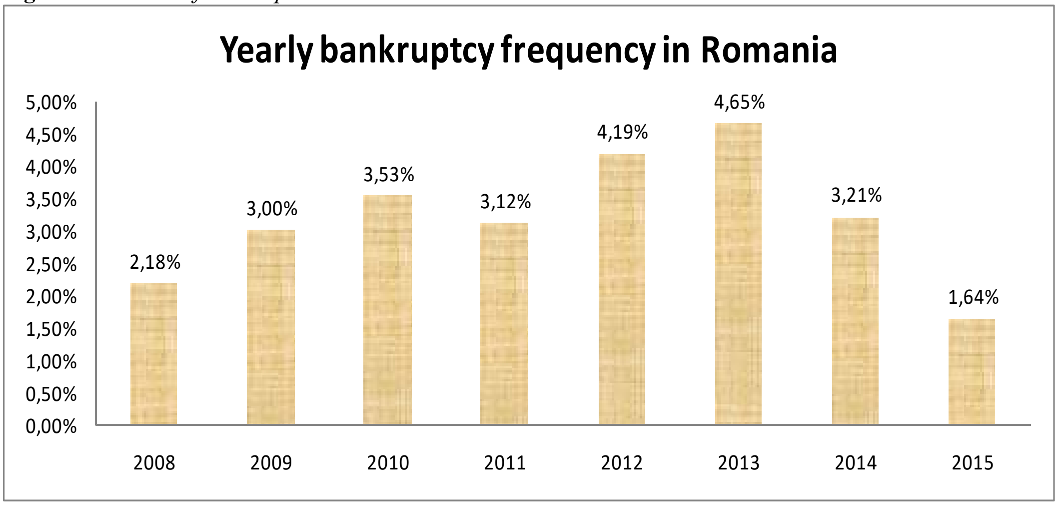
Keywords: ratio analysis, financial statements, risk, accuracy, benchmark.

JEL Classification: G33, M10.

Introduction

During the period 2008-2013, the corporate bankruptcy frequency in Romania has doubled, reaching in 2013 a peak of 4.65% (29,587 bankruptcies to a total number of companies of 636,042).

Figure 1. *Number of bankruptcies in Romania*



Source of data: calculations based on data published by the National Registry of Commerce (number of bankruptcies) and the Ministry of Public Finances (number of existing companies).

The context in which this phenomenon has manifested itself was marked by the economic crisis, the entrance of Romania in the European Union, as well as by an increase in the loan default ratio, Romania topping in this regard at the end of 2012 the 4th place within the European Union and 6th place worldwide in a ranking which included 131 countries (Brîndescu-Olariu, 2014a).

The state of bankruptcy negatively affects all the stakeholders of the company, which makes the existence of instruments for bankruptcy prediction important.

Assessment of the default risk in general and of the bankruptcy risk in particular have always been in the centre of the financial ratios analysis (Brîndescu-Olariu, 2016). Introduced in the USA in the 1860s, the financial ratios analysis only became popular in the continental Europe in the 1960s. The initial approaches of the financial ratios analysis in the prediction of bankruptcy were univariate. Some of the most important contributions in this field include (Yadav, 1986):

- the study conducted by J.R. Ramster and L.O. Foster in 1931 over a sample of 173 companies;
- the study conducted by FitzPatrick in 1932 over a sample of 38 de companies, of which 19 were bankrupt and 19 were healthy;
- the study conducted by Raymond Smith and Winakor Arthur in 1935 over a sample of 183 companies that failed over the period 1923-1931;
- the study conducted by Charles Merwin in 1942, over a sample of 900 companies.

Inspired by the models developed by Altman (1968) through discriminant analysis and later on by Ohlson (1980), through logistic regression, multivariate studies in the field of bankruptcy prediction have been performed all over the world during the last 50 years (Brîndescu-Olariu, 2016). The initial multivariate studies were performed on paired samples, an approach that still persists to great extent today.

The current study sets to create a multivariate model applicable in the Romanian practice. As differences between the profiles of the companies of different regions of Romania are expected, the focus was set on developing a model specific to only one geographical region: Timis County (largest and most western County in Romania).

The hypothesis on which the focus on a single county was based was that the differences in the characteristics of the companies make specific models for each county necessary. This hypothesis would be tested in a future research, by verifying the accuracy of the model developed in the current research on different populations (from other regions).

Population

The data employed is the same as in Brîndescu-Olariu (2016). More specifically, the target population included all companies from Timiș County that submitted yearly financial statements to the fiscal authorities during the period 2007-2010 and that registered yearly sales of at least 10000 lei (53,252 financial statements from the period 2007-2010).

The source of the data was represented by the online publications of the Ministry of Public Finances of Romania. Of the entire target population, 712 companies went bankrupt in the period 2009-2012, two years from the date of the financial statements of reference:

- of the 12,570 companies included with financial statements for 2007 in the research, 30 went bankrupt in 2009 (0.24%); the rest of the companies continued their activity under normal conditions at least until the end of 2012.
- of the 13,037 companies included with financial statements for 2008 in the research, 94 went bankrupt in 2010 (0.72%); the rest of the companies continued their activity under normal conditions at least until the end of 2012.
- of the 12,574 companies included with financial statements for 2009 in the research, 159 went bankrupt in 2011 (1.26%); the rest of the companies continued their activity under normal conditions at least until the end of 2012.
- of the 15,071 companies included with financial statements for 2010 in the research, 429 went bankrupt in 2012 (2.85%); the rest of the companies continued their activity under normal conditions at least until the end of 2012.

Methodology

In the purpose of combining the potential of individual financial ratios, the development of a multivariate model through logistic regression was targeted. The logistic regression was chosen as it is indicated by the literature as one of the most popular and accurate methods (alongside discriminant analysis).

In accordance with many of the approaches from the international literature, the model was configured based on a paired sample.

In order to build a paired-sample, each of the 712 companies that went bankrupt in the period 2009-2012 was associated with a company from the same economic field that had the closest turnover in the year of reference for the financial statements included in the analysis.

In a first step, a model was developed over the paired sample of 2010, which included 429 companies that went bankrupt in 2012 and 429 companies that continued their activity under normal conditions at least until the end of 2012. Each of the 429 companies that were included in the target population of 2010 and went bankrupt in 2012 was paired with the company from 2010 with the same SIC code that had the closest turnover in 2010 but continued its activity under normal conditions at least until the end of 2012.

In a second step, the model was tested over the paired sample from the period 2007-2009. In a final step, the model was tested over the entire target population from the period 2007 – 2010.

The data was processed by using the SPSS software. The state of the company two years from the date of the financial statements of reference was defined as the dependent variable, a binary variable that can take the following values:

- 1, for the companies that went bankrupt 2 years after the date of the financial statements of reference;
- 0, for the companies that continued their activity under normal conditions at least until the end of 2012.

In order to simplify the explanations, the companies that went bankrupt 2 years after the date of the financial statements of reference will simply be referred to as „bankrupt”, while the companies that continued their activity under normal conditions at least until the end of 2012 will be referred to as „non-bankrupt”.

When defining the target population, the companies that close their activity for other reasons than bankruptcy during the period of analysis were excluded.

As an example, the value of the variable „State” was „1” for the companies that went bankrupt in 2011 and it was associated with the financial ratios of the respective companies from 2009. These companies were not included in the analysis for the following years (for 2010 with the financial statements and for 2012 with the state variable), even if they still existed.

Financial ratios from 2 years prior to the date of the variable „State” were used as independent variables. As the objective was to build a model and a methodology of analysis accessible to all the stakeholders of the company, only financial ratios that can easily be calculated based on the public data were taken into consideration. Under these circumstances, 26 financial ratios were tested as possible explanatory variables (Table 1).

Table 1. *Financial ratios tested as explanatory variables*

No.	Financial ratios	Symbol	Formula
1	Equity working capital	Ewc	Equity+Provisions-Fixed assets
2	Labor productivity in terms of profits	LPp	Gross profits/No. of employees
3	Profitability ratio	Pr	Net profits/Sales
4	Tax to profits ratio	TPr	Income tax/Gross profits
5	Autonomy ratio	Ar	(Equity+Provisions)/Total assets
6	Debt ratio	Dr	Total debt/Total assets
7	Solvency ratio	Sr	Total assets/Total debt
8	Equity working capital to sales ratio	EwcSr	Equity working capital / Sales
9	Arrears to sales ratio	ASr	Arrears/sales
10	Equity to fixed assets ratio	EFAr	Equity/Fixed assets
11	Arrears to cash ratio	ACr	Arrears/Cash
12	Cash to total debt ratio	CTDr	Cash/Total debt
13	Current assets to total debt ratio	CATDr	Current assets/Total debt
14	Cash ratio	Cr	Cash/Total assets
15	Equity working capital to current assets ratio	EwcCAr	Ewc/Current assets
16	Receivables collection period	RCP	(Receivables/Sales)x360
17	Total assets turnover ratio	TAtr	Sales/Total assets
18	Arrears ratio	Ar	Arrears / Total debt
19	Cash to sales ratio	CSr	Cash/Sales
20	Fixed assets turnover ratio	FAtr	Sales/Fixed assets
21	Inventory conversion ratio	ICr	(Inventory/Sales)x360
22	Tax to sales ratio	TSr	Income tax/Sales
23	Fixed assets ratio	FAr	Fixed assets/Total assets
24	Labor productivity in terms of sales	LPs	Sales/No. of employees
25	Current assets ratio	CAr	Current assets/Total assets
26	Return on equity	ROE	Net profits/(Equity+Provisions)

Only a part of the 26 financial ratios are suggested in the international literature as possible predictors of bankruptcy or were at least tested by other researchers as predictors. Thus, the ratios were not selected starting from the recommendations of the scientific literature, but from the availability to the general public of the data necessary for their calculation.

Initially, in the SPSS environment, logit univariate functions were estimated for each of the 26 ratios.

Each logit function has the following form:

$$P(State=1|Z_1)=\frac{e^{Z_1}}{1+e^{Z_1}}=\frac{1}{1+e^{-Z_1}}$$

where:
Z_1 = the score associated to a company. The Z_1 score represents a linear regression in which the explanatory variables are represented by the financial ratios:
$Z_1 = a + b_1X_1 + b_2X_2 + ... + b_mX_m$. Initially, only one ratio was used for each function.
$P(State=1|Z_1)$ = the probability of bankruptcy 2 years after the date to which the Z_1 score is characteristic, estimated based on Z_1 .
e = Euler’s constant.

The value of the Z_1 score does not have superior or inferior limits. The function $P_{(State = 1|Z_1)}$ can only take values within the interval [0; 1]. In the estimation of the probability of bankruptcy in the year N+2, the variables $X_1, X_2, \ldots, X_m$ represent financial ratios that characterise the financial state of the company in year N.

Thus, the financial ratios specific to the financial statements of the companies from the target population at the end of 2010 were tested as explanatory variables for the state of the company (bankrupt or non-bankrupt) at the end of 2012. In a similar manner, the financial ratios from 2009 were correlated to the state of the company at the end of 2011, the financial ratios from 2008 were correlated to the state of the company at the end of 2010 and the financial ratios from 2007 were correlated to the state of the company at the end of 2009.

The higher the Z_1 score, the closer to 1 the value of the function $P_{(State = 1|Z_1)}$ will be, thus suggesting a high probability of bankruptcy. The probability of bankruptcy 2 years after the date of the financial statements of reference would be estimated at 0.5 in the case of a Z_1 score equal to 0.

For values of the Z_1 score that are below 0, the probability of bankruptcy 2 years after the date of the financial statements of reference would be estimated at less than 0.5, with a tendency toward 0 of the probability of bankruptcy as the Z_1 score decreases.

The performance of each univariate model developed over the paired sample of 2010 was initially evaluated by its in-sample general accuracy. In the evaluation process, each univariate model was used to classify every company from the paired sample of 2010. The cut-off value was considered to be 0.5. For each model, a company for which the probability of bankruptcy was estimated as higher than 0.5 was classified as bankrupt. All the companies with estimated probabilities of bankruptcy of 0.5 or less were classified as non-bankrupt. The percentage of companies correctly classified from the total sample represents the general accuracy of the classification.

$$\text{general accuracy} = \frac{\text{number of companies correctly classified}}{\text{total number of companies}} \times 100\%$$

The general accuracy represents the weighted average of the sensitivity and the specificity.

$$\text{general accuracy} = w_b \times \text{sensitivity} + w_{nb} \times \text{specificity}$$

where:

$$w_b = \frac{\text{number of bankrupt companies}}{\text{total number of companies}} \times 100\%$$

and

$$w_{nb} = \frac{\text{number of non-bankrupt companies}}{\text{total number of companies}} \times 100\%.$$

The sensitivity represents the accuracy of the classification of bankrupt companies.

$$\text{sensitivity} = \frac{\text{number of bankrupt companies correctly classified}}{\text{total number of bankrupt companies}} \times 100\%$$

The specificity represents the accuracy of the classification of non-bankrupt companies.

$$\text{specificity} = \frac{\text{number of non} - \text{bankrupt companies correctly classified}}{\text{total number of non} - \text{bankrupt companies}} \times 100\%$$

As the sample from 2010 used for the development of the univariate models was a paired sample, the weight of the bankrupt companies was equal to the weight of the non-bankrupt companies (50%). For such a sample, the „by chance" accuracy is 50% (by classifying all 858 companies as bankrupt, the analyst would be correct in 50% of the cases). A model is considered a useful classifier if it allows for a general accuracy at least 25% higher than the „by chance" accuracy (Chung et al., 2008).

Based on this benchmark, the univariate models would be considered as potentially useful if they would offer an in-sample accuracy (the accuracy in the classification of the companies from the development sample) of at least 62.5% (a = 50% x 125%).

The purpose of the development of univariate models was not to obtain practical univariate tools of analysis, but to observe the individual potential of the ratios, especially in terms of sensitivity and specificity. Together with information concerning the intercorrelations, the sensitivity and the specificity ensured by each ratio in the univariate approach were used as reference for grouping the ratios into multivariate models.

Thus, ratios were grouped into multivariate models, avoiding intercorrelations, searching for statistical significance at 0.01 significance level and targeting to ensure complementarity for sensitivity and specificity. Over 100 combinations were tested in the SPSS environment. In the end, the combination that ensured the best general in-sample accuracy was retained as a proposed multivariate model.

Additional in-sample tests were performed on the model using the area under the ROC Curve. Out of sample tests were performed on the paired sample from the period 2007-2009, using the general accuracy and the area under the ROC Curve as performance indicators.

In the final testing stages, the model was tested on the entire population of 2010 in terms of general accuracy and area under the ROC Curve.

The structure of the target population is significantly different from the structure of the sample, with the weight of the bankrupt firms being of only 2.85% within the target population of 2010 (2.85% of the target population of 2010 went bankrupt in 2012). Such structure differences brought in the past some criticism to the models developed based on paired samples. Nevertheless, the sampling method remains very popular in this field.

Based on the „25% rule" (Chung et al., 2008), the model should be considered a useful classifier if it could offer a general accuracy of at least 62.5% over a paired sample and of 100% over the entire target population:

- the „by chance" general accuracy over the 2010 target population would be of 94.46% ($w_{nb}^2 + w_b^2 = 97.15\%^2 + 2.85\%^2 = 94.46\%$);
- $94.46\% \times 125\% \approx 118\%$.

Most of the models available in the international literature cannot ensure even in-sample general accuracies at least equal to the „by chance" general accuracies. Of the 40 models presented, the highest reported in-sample (paired sample of 54 companies) accuracy was of 95.60% (Ugurlu and Aksoy, 2006).

Considering that the current research targets to develop a model that only employs data easily available to all stakeholders, an out of sample general accuracy close to 100% does not constitute a reasonable objective.

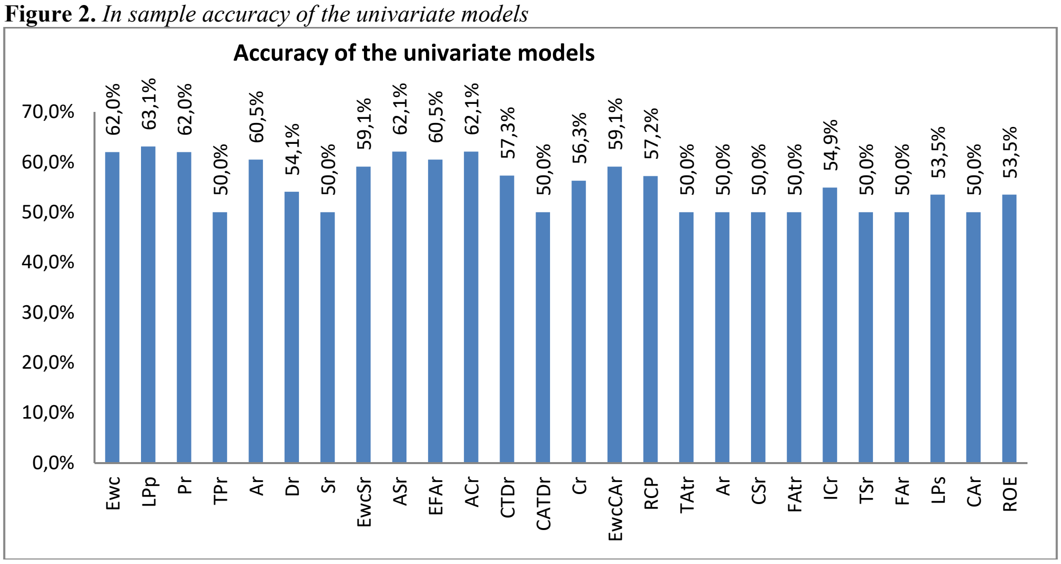
Instead, the purpose of the model would be to ensure the classification of the analysed companies on risk classes.

In practice, using the model, the analyst should expect every analysed company not to go bankrupt (the model would not indicate a probability of bankruptcy higher than 0.5), but it will be possible to evaluate the bankruptcy risk as higher or lower than the average.

Results

The in-sample general accuracies of the univariate models are showed in Figure 1. The highest in-sample general accuracy was offered by the model based on the labour productivity in terms of profits (63.1%).

Figure 2. *In sample accuracy of the univariate models*



In an attempt to increase the general accuracy of the classification, different ratio combinations were tested for the development of multivariate models. The combinations were prioritized based on the sensitivity and specificity ensured by each ratio. The increase of the general accuracy obtained by combining ratios into multivariate models was limited by the intercorrelations between the ratios. Combinations between ratios with low Pearson coefficients were granted priority from this point of view.

The ratios finally retained in the model were:

- Receivables collection period (RCP);
- Profitability ratio (Pr);
- Cash to total debt ratio (CTDr);
- Fixed assets ratio (FAr);
- Equity working capital (Ewc).

The ratios were retained within the final model based on their contribution to the in-sample general accuracy and on the significance of their correlation to the State variable.

Table 2. *Coefficients of the logit function and tests*

No.	Rate	B	Wald	Sig.
1	RCP	0,000635	14,557	0,000
2	Pr	-0,343	11,600	0,001
3	CTDr	-0,243	12,283	0,000
4	FAr	-1,185	27,958	0,000
5	Ewc	-0,000000544	21,322	0,000

All the independent variables are statistically significant at a significance level of 0.001. The receivables collection period and the equity working capital are expressed in absolute values, while the other independent variables are expressed in relative values. This leads to significantly lower coefficients for the receivables collection period and the equity working capital.

The receivables collection period is positively correlated to the Z_1 score. Higher values of the receivables collection period will influence the Z_1 score in a positive sense from the mathematical point of view, thus sustaining high levels of estimated probability of bankruptcy.

The profitability ratio, the cash to total debt ratio, the fixed assets ratio and the equity working capital are negatively correlated to the Z_1 score. Higher values of these ratios will influence the Z_1 score in a negative sense from the mathematical point of view, thus sustaining low levels of estimated probability of bankruptcy.

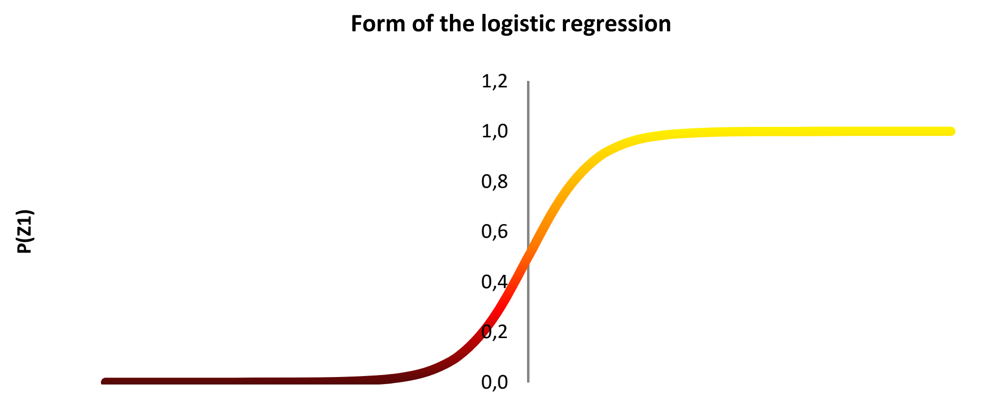
The Z_1 score has the following form:

$$Z_1 = 0{,}635\ \mathrm{RCP}\times 10^{-3} - 0{,}343\ \mathrm{Pr} - 0{,}243\mathrm{CTDr} - 1{,}185\ \mathrm{FAr} - 0{,}544\ \mathrm{Ewc}\times 10^{-6}$$

The probability of bankruptcy can be estimated based on the following function:

$$P(State = 1|Z_1) = \frac{1}{1 + e^{-(0{,}635\ \mathrm{RCP}\times 10^{-3} - 0{,}343\ \mathrm{Pr} - 0{,}243\mathrm{CTDr} - 1{,}185\ \mathrm{FAr} - 0{,}544\ \mathrm{Ewc}\times 10^{-6})}}$$

Figure 3. *Form of the logistic regression*



$$Z_1 = 0{,}635 \text{ RCP} \times 10^{-3} - 0{,}343 \text{ Pr} - 0{,}243\text{CTDr} - 1{,}185 \text{ FAr} - 0{,}544 \text{ Ewc} \times 10^{-6}$$

Using 0.5 as cut-off value, the in-sample general accuracy (the general accuracy over the 2010 paired sample) is of 66.7%. Thus, by classifying as bankrupt all the companies for which the estimated probability of bankruptcy is higher than 0.5 (estimation generated through the use of the model), 267 of the 429 bankrupt companies are correctly evaluated. At the same time, by classifying as non-bankrupt all the companies for which the probability of bankruptcy was estimated at less than 0.5, the evaluation would be correct for 305 out of the 429 non-bankrupt companies from the 2010 paired sample.

An additional performance test was made through the area under the ROC Curve (AUC), which can take values between 0 and 1 (Skalska and Freylich, 2006). An AUC of 0.5 corresponds to a “by chance” classification accuracy, while an AUC of 1 corresponds to a perfect accuracy. The ROC Curve reflects graphically the relationship between the sensitivity and the specificity for all possible cut-off values (van Erkel and Pattynama, 1998). The area under the ROC Curve thus isolates the classification performance of the logit model with no connection to a specific cut-off value, which makes it one of the most viable solutions for measuring the classification performance of a model and for comparing models (Hanely and McNeil, 1982, Faragi and Reiser, 2002).

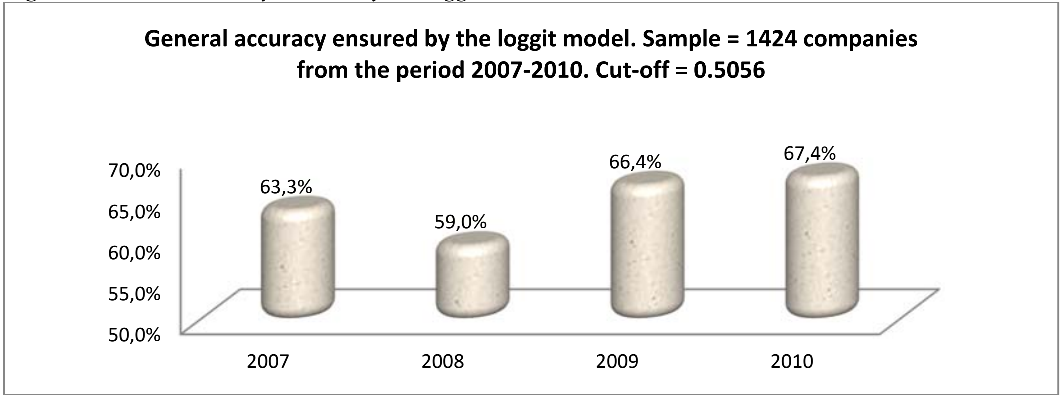
The area under the ROC Curve over the 2010 paired sample specific to the logit model was of 0.728, which can be evaluated as fair classification accuracy (Tazhibi and Bashardoost and Ahmadi, 2011).

The inspection of the ROC Curve coordinates revealed that the cut-off value that maximizes the in-sample general accuracy is of 0.5056 (which ensures a general accuracy over the 2010 paired sample of 67.4%).

The performance of the model in terms of general accuracy decreases over the period 2010-2007, as shown in Figure 4. The area under the ROC Curve also decreases over the period 2010-2007 (over the paired sample), from 0.728 in 2010 to 0.727 in 2009, 0.680 in 2008 and 0.616 in 2007. The area under the ROC Curve for the entire sample (the 1424 companies from the period 2007-2010) was of 0.714.

The general accuracy shown in Figure 4 was estimated based on the cut-off value of 0.5056 (which maximizes the general accuracy for the 2010 paired sample). The out-of-sample general accuracy (over the paired sample from the period 2007-2009) was of 63.6%.

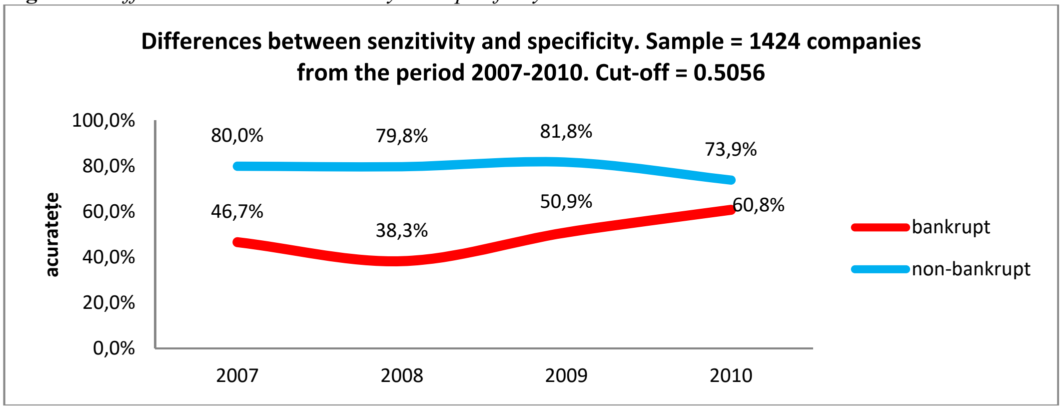
Figure 4. *General accuracy ensured by the loggit model*



The in-sample general accuracy of the multivariate logit model (as well as its out-of-sample general accuracy) overlaps the „by chance” general accuracy, as well as the in-sample accuracy of the best univariate model (63.1%).

Still, the multivariate model generates significant differences between sensitivity and specificity, as shown in Figure 5.

Figure 5. *Differences between sensitivity and specificity*

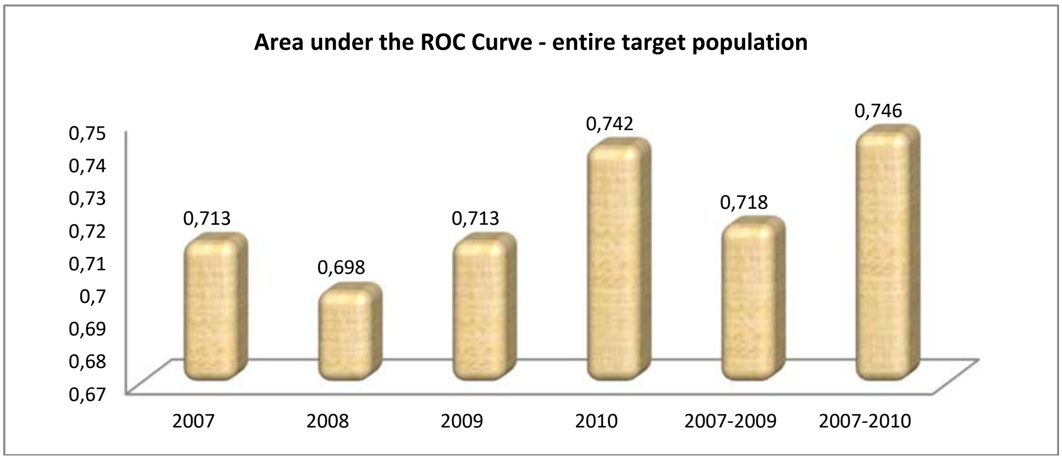


The area under the ROC Curve over the entire target population of 2010 (the population from which the development sample was extracted) was of 0.742, suggesting a fair classification performance.

The area under the ROC Curve over the entire target population of 2007-2009 was of 0.718.

As shown in Figure 6, the area under the ROC Curve over the entire target population of 2007-2010 was of 0.746.

Figure 6. *Area under the ROC Curve*



Although the general accuracy represents an easy to use criterion for evaluating the performance of a model, its relevance is diminished with the decrease of the frequency of the studied phenomenon (Metz, 1978). For any population with a low frequency of the studied phenomenon, the "by chance" general accuracy is high. For the target population of 2010, the bankruptcy frequency was of 2.85%. By simply classifying all the companies as non-bankrupt, the analyst would be accurate in 97.15% of the cases. Over such a population, the general accuracy of a model could be maximised by maximizing the specificity. Such an objective could be partially reached by simply increasing the cut-off value (which, for a logit model, would be limited between 0 and 1).

Using a high cut-off value, the analyst would reduce the sensitivity, but would increase the specificity. As the weight of the non-bankrupt companies is significantly higher, the general accuracy would increase (the specificity weights more than the sensitivity within the general accuracy). Most of the bankrupt companies would be wrongly classified (as non-bankrupt), but most of the non-bankrupt companies would be correctly classified. The analyst would thus reduce the false positive rate, but would accept an increase of the false negative rate.

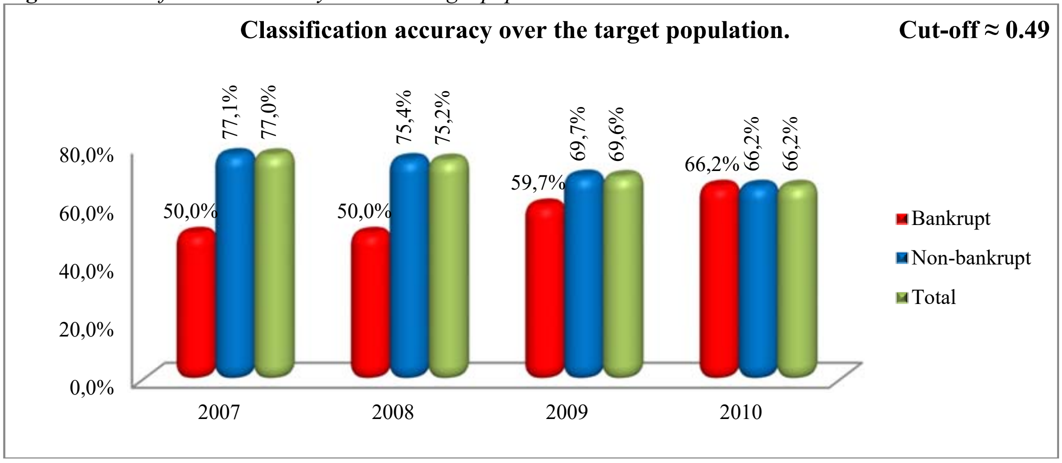
By simply using a cut-off value of 1, the analyst would classify as non-bankrupt all the companies with estimated probabilities of bankruptcy lower or equal to 1 (all the companies from the target population). Thus, no matter of the specific classification performance of the model (and no matter what model would be used), the general accuracy obtained over the 2010 target population would be of 97.15% (with 100% specificity and 0% sensitivity). The 97.15% general accuracy would present no relevance for the classification performance of the model. Instead, over populations with low frequencies of the bankruptcy cases, the classification performance of a model could be better measured through the general accuracy at the cut-off value that equalises the sensitivity and the specificity.

The cut-off value that equalises the sensitivity and the specificity for the entire 2010 target population is of approximately 0.49. By classifying all the companies from the

2010 target population that have an estimated probability of bankruptcy (based on the proposed model) of higher than 0.49 as bankrupt, the sensitivity would be equal to the specificity (66.2%).

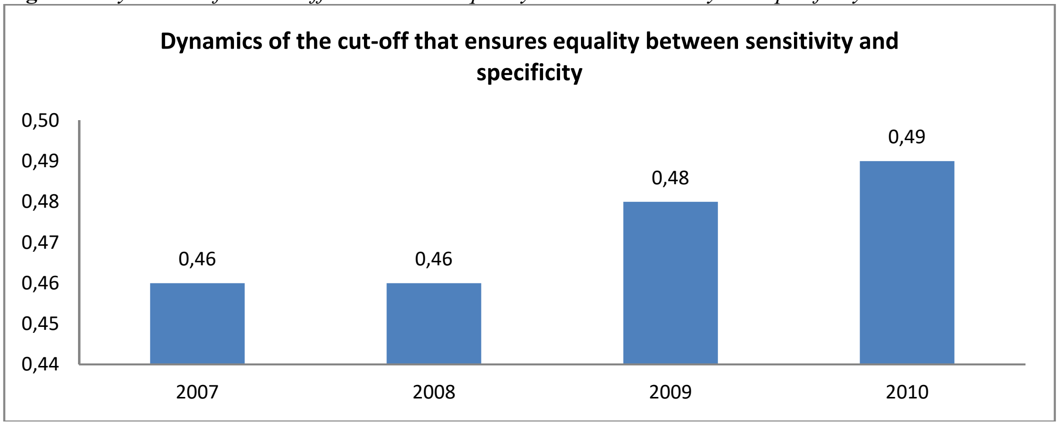
Figure 7 shows the classification performance of the model on the target population from the period 2007-2010 based on the cut-off value of 0.49 (the cut-off value that equalises the sensitivity and the specificity for the 2010 target population).

Figure 7. *Classification accuracy over the target population*



The cut-off that ensures equality between sensitivity and specificity varies between 0.46 and 0.49 for the 2007-2010 target population, as shown in Figure 8.

Figure 8. *Dynamics of the cut-off that ensures equality between sensitivity and specificity*



As the model cannot ensure a classification accuracy higher than the „by chance" accuracy (most models present in the international literature can't), the classification based on a singular cut-off is not considered a practical approach.

Instead, classifying the companies on risk classes based on the estimated probabilities of bankruptcy is considered more useful. In this sense, the companies from the target population of 2010 were grouped on ten intervals of the logit function, as shown in Table 3.

Table 3. *Bankruptcy incidence over the target population of 2010 on intervals of the estimated probability of bankruptcy*

No.	P (State = Bankrupt)	Risk index
1	P (State = Bankrupt) < 0.1	6%
2	0.1 <= P (State = Bankrupt) < 0.2	0%
3	0.2 <= P (State = Bankrupt) < 0.3	9%
4	0.3 <= P (State = Bankrupt) < 0.4	54%
5	0.4 <= P (State = Bankrupt) < 0.5	75%
6	0.5<= P (State = Bankrupt) < 0.6	140%
7	0.6 <= P (State = Bankrupt) < 0.7	283%
8	0.7 <= P (State = Bankrupt) < 0.8	546%
9	0.8 <= P (State = Bankrupt) < 0.9	464%
10	P (State = Bankrupt) >= 0.9	421%

For each interval of the estimated probability of bankruptcy (based on the proposed logit model), the risk index was calculated by dividing the actual bankruptcy frequency specific to the interval to the actual bankruptcy frequency specific to the entire target population of 2010.

Thus, a risk index higher than 100% shows an actual bankruptcy frequency higher than the average (2.85%), while a risk index lower than 100% shows an actual bankruptcy frequency lower than the average.

Based on the risk indexes shown in Table 3 for the 2010 target population, 3 risk classes are proposed, as shown in Table 4.

Table 4. *Risk classes*

No.	Estimated probability of bankruptcy	Risk index
1	High risk: P (State = Bankrupt) >= 0.5	216.4%
2	Average risk: 0.3 <= P (State = Bankrupt) < 0.5	66.9%
3	Low risk: P (State = Bankrupt) < 0.3	7.1%

As shown in Table 4, the group of companies from the 2010 target population with estimated probabilities of bankruptcy higher or equal to 0.5 had an actual bankruptcy frequency 2.164 times higher than the average. At the same time, the group of companies from the 2010 target population with estimated probabilities of bankruptcy lower than 0.3 had an actual bankruptcy frequency of only 7.1% from the average.

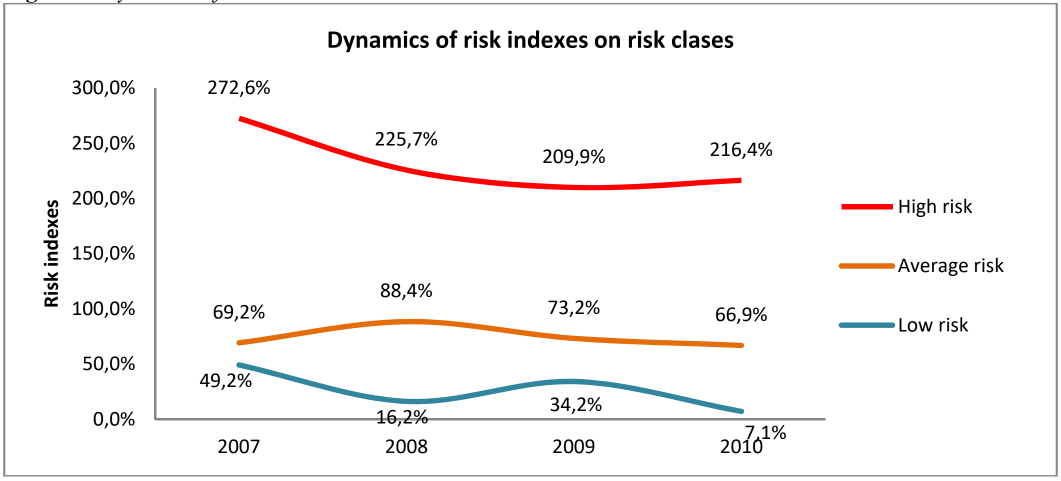
The risk indexes for the three proposed classes over the 2007-2009 target population are shown in Table 5 and Figure 9.

Table 5. *Bankruptcy incidence over the target population of 2007-2009 on intervals of the estimated probability of bankruptcy*

No.	P (State = Bankrupt)	2007	2008	2009	2010
1	P (State = Bankrupt) < 0.1	87%	0%	14%	6%
2	0.1 <= P (State = Bankrupt) < 0.2	0%	0%	22%	0%
3	0.2 <= P (State = Bankrupt) < 0.3	49%	25%	46%	9%
4	0.3 <= P (State = Bankrupt) < 0.4	36%	79%	38%	54%
5	0.4 <= P (State = Bankrupt) < 0.5	96%	96%	98%	75%
6	0.5<= P (State = Bankrupt) < 0.6	157%	195%	135%	140%
7	0.6 <= P (State = Bankrupt) < 0.7	0%	190%	476%	283%
8	0.7 <= P (State = Bankrupt) < 0.8	381%	226%	494%	546%
9	0.8 <= P (State = Bankrupt) < 0.9	1132%	433%	356%	464%
10	P (State = Bankrupt) >= 0.9	1529%	504%	213%	421%

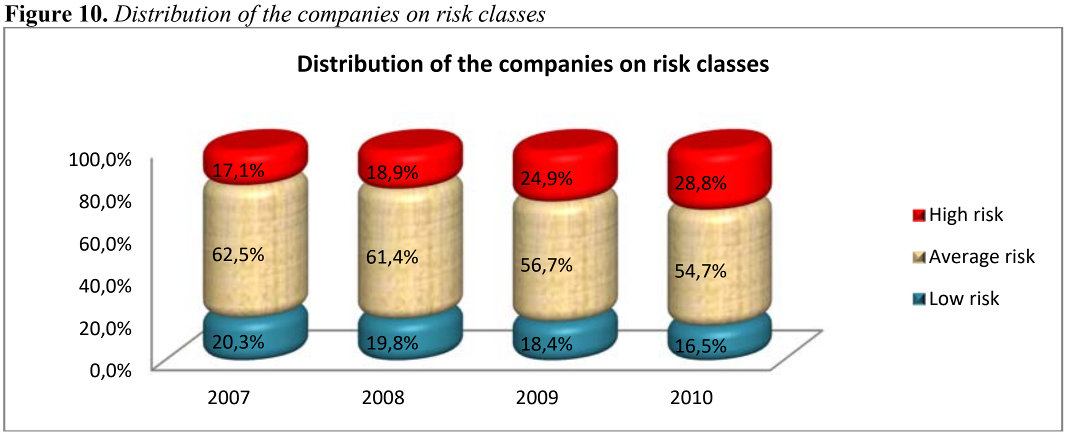
As shown in Figure 9, the risk indexes remain relatively stable over the period 2007-2010, with no intersections between the proposed risk classes.

Figure 9. *Dynamics of risk indexes on risk classes*



According to Figure 10, approximately 55-60% of the companies are positioned within the average risk class. Under these circumstances, a stakeholder could find it useful to treat with more caution the relationship with a company positioned in the high risk class, while developing on the other hand the relationships with the companies positioned in the low risk class.

Figure 10. *Distribution of the companies on risk classes*



Conclusions

Based on the results of the research, the analyst is recommended to classify the company under analysis in one of the following tree risk classes:

- high bankruptcy risk, for estimated bankruptcy probabilities of 0.5 or more;
- average bankruptcy risk, for estimated bankruptcy probabilities between 0.3 and 0.5;
- low bankruptcy risk, for estimated bankruptcy probabilities of less than 0.3.

The proposed methodology is easily accessible to all the stakeholders of the company, as it only employs publically available data that can be obtained freely online (from the website of the Romanian Ministry of Public Finance).

To evaluate the bankruptcy risk for a specific company, the analyst only has to retrieve the data, calculate the five ratios, calculate the Z1 score and estimate the probability of bankruptcy based on the logit function:

$$P(State = 1|Z_1) = \frac{1}{1 + e^{-(0{,}635\ \mathrm{RCP} \times 10^{-3} - 0{,}343\ \mathrm{Pr} - 0{,}243\mathrm{CTDr}\ - 1{,}185\ \mathrm{FAr} - 0{,}544\ \mathrm{Ewc} \times 10^{-6})}}$$

The model will not allow the analyst to conclude with absolute certainty whether the analyzed company will go bankrupt or not within a 2 – year horizon. Instead, it will offer the possibility to differentiate the companies on risk classes, thus creating the basis for a differentiated behavior of the stakeholder toward those companies.

Based on the bankruptcy frequency of the last year included in the research, the information offered by the proposed methodology is the following:

- out of 1000 companies included in the low bankruptcy risk class, approximately 2 will go bankrupt within 2 years;
- out of 1000 companies included in the average bankruptcy risk class, approximately 19 will go bankrupt within 2 years;
- out of 1000 companies included in the high bankruptcy risk class, approximately 62 will go bankrupt within 2 years;
- with no consideration to output of the model, out of 1000 observed companies, approximately 29 will go bankrupt within 2 years.

The time horizon of 2 years between the date of the financial statements of reference and the date of the predicted state of the company was considered optimal, as 1 year would have been unpractical (the financial statements are only made public after the middle of the following year) and a longer period would have unjustifiably reduced the accuracy.

Although initial univariate studies conducted within the current research showed some predictive capacities for the ratios included in the final multivariate model as independent variables, attempts should not be made to relate directly the bankruptcy probability estimated through the multivariate model to each individual ratio. The purpose of such a multivariate model is specifically to respond to the lack of prediction capabilities of any single financial ratio.

While other studies (or common sense) could suggest positive correlations between the probability of bankruptcy and the Receivables collection period (through the negative impact on cash-flows), negative correlations between the probability of bankruptcy and the Equity working capital (through its contribution in financing the operating cycle), or between the probability of bankruptcy and the Profitability ratio (through its contribution to equity), the analyst should no attempt to explain a certain state of the company by decomposing the model on individual ratios calculated on past data. The model should therefore not be employed for explanatory purposes, but only for predictions.

The current research offers a solution specific for the Romanian companies, as the Romanian practice currently employs on a large scale methodologies developed over populations from overseas, from too distant past periods or simply over Romanian samples that lack representativeness.

The research should be continued with studies on the following directions:
- development of multivariate models based on the dynamics of the ratios;
- the correlation between the estimated probability of bankruptcy and the financial risk;
- the correlation between the estimated probability of bankruptcy and the return on equity;
- the applicability of the proposed methodology for companies from other regions.

The proposed methodology should be updated annually, to ensure a permanent adaptation to any modifications of the characteristics of the target population.

References

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

Andreica, M., 2013. Modele de avertizare timpurie a firmelor neproductive. Studiul de caz al firmelor românești listate pe RASDAQ. *Economie teoretică şi aplicată*, 20 (5), pp. 4-12.

Anghel, I., 2002. *Falimentul – radiografie și predicție*. București: Editura Economică.

Armeanu, D., Vintilă, G., Moscalu, M., Filipescu, M.-O. and Lazăr, P., 2012. Utilizarea tehnicilor de analiză cantitativă a datelor pentru estimarea riscului de faliment al corporațiilor. *Economie teoretică și aplicată*, 19 (1), pp. 86-102.

Băileșteanu, Gh., 1998. *Diagnostic, risc şi eficienţă în afaceri*. Timișoara: Editura Mirton.

Bătrâncea, L., 2011. Measuring the risk of bankruptcy in the commercial sector in Romania. *Analele Universității din Oradea*, 1 (2), pp. 393-399.

Bordeianu, G.D., Radu, F., Paraschivescu, M.D. and Păvăloaia, W., 2011. Analysis models of the bankruptcy risk. *Economy Transdisciplinarity Cognition*, 14 (1), pp. 248-259.

Brîndescu-Olariu, D., 2014a. Labor productivity as a factor for bankruptcy prediction. SEA – Practical Application of Science, II, Issue 4 (6) <http://sea.bxb.ro/Article/SEA_6_5.pdf>

Brîndescu-Olariu, D., 2014b. Payment capacity sensitivity factors, available at: <http://seaopenresearch.eu/Poster/6002%20-%20Brindescu.pdf>

Brîndescu-Olariu, D., 2014c. The potential of the equity working capital in the prediction of bankruptcy, available at: <http://seaopenresearch.eu/Poster/6001%20-%20Brindescu.pdf>

Brîndescu-Olariu, D., 2016. Bankruptcy Prediction Based on the Autonomy Ratio. EuroEconomica, Vol. 35, Issue 2, pp. 61-75 <http://journals.univ-danubius.ro/index.php/euroeconomica/article/view/3286/3690>

Caracota, R., Dumitru, M. and Dinu, M., 2010. Construirea unui model de scoring pentru întreprinderile mici și mijlocii. *Economie teoretică și aplicată*, 17 (9), pp. 103-114.

Căprariu, O., 2010. The bankrupt risk in feed distribution branch in Dolj District – FDR model. *Management and Marketing Journal*, 8, pp. 156-169.

Cârciumaru, D., 2010. A Model for Evaluating the Bankruptcy Risk of the Romanian Companies. *The Young Economists Journal*, 1 (14), pp. 35-40.

Chung, K., Tan, S. and Holdsworth, D., 2008. Insolvency prediction model using multivariate discriminant analysis and artificial neural network for the finance industry in New Zealand. *International journal of business and management*, 39 (1), pp. 19-29.

van Erkel, A. and Pattynama, P., 1998. Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, *European Journal of Radiology,* 27 (2), pp. 88-94.

Hanely, J.A. and McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143 (1), pp. 29-36.

Mironiuc, M., Robu, M. and Robu I., 2012. Estimating the probability of bankruptcy risk occurrence in an emerging capital market. *The Proceedings of the 6th International Conference on Globalization and Higher Education in Economics and Business Administration GEBA 2012*, pp. 611-623.

Ohlson, J., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research,* 18 (1), pp. 109-131.

Skalska, H. and Freylich, V., 2006. Web-Bootstrap Estimate of Area under ROC Curve. *Australian Journal of Statistics,* 35 (2 and 3), pp. 325-330.

Tazhibi, M., Bashardoost, N. and Ahmadi, M., 2011. Kernel Smoothing For ROC Curve and Estimation for Thyroid Stimulating Hormone. *International Journal of Public Health Research,* Special Issue, pp. 239-242.

Ugurlu, M. and Aksoy, H., 2006. Prediction of corporate financial distress in an emerging market: the case of Turkey, Cross Cultural Management: An International Journal, 13 (4), pp. 277-295.

Vintilă, G. and Toroapă, G., 2012. Forecasting the bankruptcy risk on the example of Romanian enterprises. *Revista Română de Statistică*, pp. 377-388.

Yadav, R.A., 1986. *Financial ratios and the prediction of corporate failure*. New Delhi: Concept Publishing Company.

National Registry of Commerce <http://www.onrc.ro/index.php/ro/statistici?id=252>