Determinants of Corporate Bankruptcy in Romania- A Comparison Approach

Ana-Maria Sândică,
The Bucharest University of Economic Studies, Romania.
E-mail: anamaria.sandica@gmail.com

Abstract

The aim of this paper is to estimate the probability of default of companies based on a data set from the Romanian economy. The financial statements for 15264 companies were used and the scoring function was estimated using logistic regression. The default definition is according with Basel II definition, 90 days overdue. The proportion of “bad” firms in this portfolio is 32.84%. One first step was to establish the list of potential indicators and to check the correlation with the trigger binary variable of default. The indicators used in analysis were from different areas like profitability, liquidity, solvency and different rotations indicators. Based on observed correlations and averages for each group, a list of financial indicators was selected to enter the model. Taking into account a model with significant coefficients, p-values lower than 5%, the final model considered is explained by Return on equity, finance horizon, capital ratio, cash liquidity, current liquidity, Days of inventories on hand and operational return on assets. In order to test the accuracy of the models it was calculated the confusion matrix, ROC curve and Gini coefficient. For a cut-off of 0.5, the model accuracy obtained was 75% observing a higher detection rate for logistic regression comparing with neural networks.

Key Words: logistic regression, neural networks, bankruptcy, probability of default, financial indicators

JEL Classification: C 19, G13, G14
1. Introduction

Ability to classify companies into different predefined groups is an important business research issue, which can be utilized as a strong risk management tool. Default prediction has been an important area of business interest for many researchers, from the theoretical and practical aspect as it is an integral part of the credit risk, which is considered to be one of the most important risks. The increase in the bankruptcies in the last years emphasize the risks involved in corporate liability and bankruptcy prediction has become a very important concern for the various stakeholders in firms, including shareholders, managers, creditors and business partners, as well as government institutions accountable for maintaining stability of financial markets. Therefore an increase in research in ways to minimize default risk has been observed in the recent years. The bankruptcies not only affect the creditors but also create a vulnerable environment for the economy as a whole, investors, other firms, consumers as well as the industry involved. The increase of default occurrence can be linked to the latest global financial crisis and appropriate credit risk management.

The aim of this paper is to estimate the probability of default for a Romanian portfolio companies comparing logistic regression with neural networks. The paper contains the sections of literature review on the main research, the methodology used for estimation and assess the model while in the last section are presented the empirical results and conclusions.

2. Literature Review

Prediction of corporate financial distress and bankruptcy is a topic which has gained a lot of interest by researchers in finance starting in the late 1960s when Beaver (1966), who developed a dichotomous classification test based on a simple t-test in a univariate framework. He used individual financial ratios from 79 failed and non-failed companies that were grouped by industry and assets size from 1954 to 1964 and identified a single financial ratio – Cash flow/Total Debt as the best predictor of corporate bankruptcy. Beaver’s study was then followed by Altman (1968), who suggested a multivariate technique, known as Multivariate Discriminant Analysis (MDA). By using 33 bankrupt companies and 33 non-bankrupt companies over the period 1946 – 1964, five variables were considered to be most predictive in estimating bankruptcy. These five were: Working Capital to Total Assets, Retained Earnings to Total Assets, Earnings before Interest and Taxes to Total Assets, Market Value of Equity to Book Value of Total Debt and Sales to Total Assets. Z-Score was determined and those companies with a score greater than 2.99 fell into the non-bankrupt group, while those companies having a Z-Score below 1.81 were in the bankrupt group. The area between 1.81 and 2.99 is defined as the zone of ignorance or the grey area. Logit analysis was made popular in the financial distress prediction problem by Ohlson (1980). He used 105 bankrupt companies and 2058 non-bankrupt companies from 1970 to 1976. The results
showed that size, financial structure (Total Liabilities to Total Assets), performance and current liquidity were significant when predicting bankruptcy. In another research study, Abdullah, Halim, Ahmad and Rus (2008) compared three methodologies of identifying financially distressed companies in Malaysia that are: multiple discriminant analysis (MDA), logistic regression and hazard model. In a sample of 52 distressed and non-distressed companies with a holdout sample of 20 companies, the predictions of hazard model were accurate in 94.9% of the cases examined. This was a higher accuracy rate than generated by the other two methodologies. However, when the holdout sample was included in the sample analysed, MDA had the highest accuracy rate of 85%. Among the ten predictors of corporate performance examined, the Ratio of Debt to Total Assets was a significant determinant of corporate distress regardless of the methodology used. Additionally, Net Income Growth was another significant predictor in MDA, whereas the Return on Assets was an important predictor when the logistic regression and hazard model methodologies were used. Their analysis was similar to the studies of Low, Fauzias and Ariffin (2001), Mohamed and Sanda (2001), Zulkarnain, Mohamad Ali, Annuar and Abidin (2001). As pointed out by Jones et al. (2015) different statistical learning models involve a trade-off between flexibility and interpretability. More flexible classifiers such as neural networks are generally known to predict better, but are far less interpretable. Neural networks are classic ‘black box’ models – they are designed to capture all nonlinear effects and interactions in a dataset but the price of this greater complexity is lack of interpretability. On the other hand, a logit model is highly interpretable by comparison but has less capacity to capture nonlinear effects and interactions, which generally limits the predictive power of the model. As a summary of different researchers on logistic regression the following studies are representative: Keasey and Watson (1987), Peel and Peel (1987), Franks (1998), Mramor and Valentincic (2003), Hall (1994), Pindado and Rodrigues (2004), Bhimani et al. (2010) and Altman et al. (2014).

3. Methodology

Early neural model-based approach dates back to 1943, once the first appearance of the neuron model, proposed model of neurophysiology W.S McCulloch and mathematician W. Pitts. Particular interest to the neuron model was observed after the first appearance of works in mathematical modelling of learning processes. A first occurrence of this kind took place in 1947, is represented by the model of learning of D.O. Hebb, who opened unsuspected directions in neural calculations. Another important step on the road neural development approach was made in 1957, with the appearance of Frank Rosenblatt's work, dedicated to a simplified neural model probabilistic nature, known as the perceptron. Fundamental element of any neural network is an artificial neuron. Neurons that are part of neural networks, have
different functions, they are specialized in performing certain types of activities. From this viewpoint, a neural network contains three basic types of neurons:
• input units, acquiring the input variables values or standard values of input variables, this means that the input neurons have no own computer functionality itself, but an interface role, the input neurons form the so-called input layer or the input;
• Neurons intermediaries are brain cells are located between the input layer and output layer having a function purely computer;
• output neurons, which calculates predicted values by neural network and comparing these values with specific target values or reference values, depending on the outcome comparisons, weights or connections are not updated.
Each elementary unit of a neural network, i.e. each neuron has one or more, an internal state and an exit. Functionality of a neuron consists in that it produces a single output, represented by a single numeric value, depending on the nature or status of such units, determined based on state information that the neuron input.
Each value of $X_1, X_2, \ldots, X_p$ is a variable and the weights, also known as synaptic weights are written in the order $(k,p)$ where $k^2$ indicates the neuron to which the weight applies and $p$ indicates the variable.

\[ u_k = w_{k0}x_3 + w_{k1}x_1 + \cdots + w_{kp}x_p = \sum_{q=0}^{p} w_{kq}x_q, \quad (1) \]

\[ y_k = F(u_k) \quad (2) \]

The $u_k$ value is then transformed using an activation function known as transfer function. Various alternative activation functions have been used:
• Threshold Function

\[
F(u) = \begin{cases} 
1, & \text{if } u \geq 0, \\
0, & \text{if } u < 0 
\end{cases} \quad (3)
\]

• Logistic Function:

\[ F(u) = \frac{1}{1 + e^{-\alpha u}} \quad (4) \]

• Hyperbolic tangent :

\[ \phi(u) = \tanh(u) \]

1 If the sign is positive then the weights are known as excitatory because they would increase the corresponding variable and if is negative they would reduce the value of $u_k$ for positive variables are known as inhibitory.
2 If the architecture is a single layer neuron then k is 1
In order to apply neural network technique the problem of specifying the weights that are used in the architecture built and this task is accomplished by the learning algorithm which trains the network and iteratively modifies those weight until a condition is satisfied, especially when the error between the desired output and the one produced by the model is minimal.

There are three typologies of learning mechanism for neural networks: supervised, unsupervised and reinforced learning. The training set is used in order to offer the desired output and in this manner to adjust the weights. In comparison with this the second mechanism, unsupervised learning is using a set without the desired output and the weights are adjusted based on self-organizing. The reinforced learning mechanism assumes that the best method to adjust the weights is to introduce prizes and penalties as a function of network response.

A multilayer perceptron is composed by an input of layer of signals, an output layer and a number of layers of neurons between, called hidden layers. The weights apply in the input neurons may differ from the weights applied on hidden layers. A three layer network is shown below.

![Figure 1: Multilayer Perceptron](image-url)

\[
F(u) = 1 - \frac{2}{1 + e^{2u}}
\]  

(5)

\[
\gamma_k = \frac{1}{r} \left( \sum_{q=0}^{p} w_{kj} x_q \right)
\]  

(6)
Where the subscript 1 in equation (6) indicates the fact that is the first layer and $y_k$ are the outputs from the first hidden layer and the output of one layer is the input for the following layer the relation became:

$$z_v = F_2 \left( \sum_{k=1}^{r} K_{vk} y_k \right) = F_2 \left( \sum_{k=1}^{r} K_{vk} F_1 \left( \sum_{q=0}^{p} w_{kq} x_q \right) \right)$$ \hspace{1cm} (7)$$

Where $z_v$ is the output of neuron v in the output layer, $v=1...s$, $F_2$ is the transfer function the output layer and the weight applied to the $y_k$ layer is $K_{vk}$.

The method of calculation these weights are also known as training process, and most frequently method is the back-propagation algorithm, that looks for the minimum of the error function in weight space using the method of gradient descent. The solution of the learning problem is the combination of weights which minimizes the error function.

First all weights are equal to some randomly chosen numbers and a training pair is selected, the forward pass is ending when $z_v$ is calculated. The backward pass consists of distributing the error between known value $a_v$ and calculated one, $z_v$, through the network proportionally with the contribution made by each weight. After that a second pair is selected and both forward and back pass are calculated this process is known as epoch the repeated process ends when a stopping criterion is fulfilled.

Defining the error $e_v(t)$ as

$$e_v(t) = a_v(t) - y_v(t)$$ \hspace{1cm} (8)$$

Where $a_v(t)$ is the observed outcome for case t in neuron v and $y_v(t)$ is the predicted outcome? The purpose is to choose a vector of weights that minimizes the average value over all training cases of:

$$E(t) = \frac{1}{2} \sum_{t=1}^{s} e^2_v(t)$$ \hspace{1cm} (9)$$

where $s$, is the number of neurons in the output layer.

For any neuron $v$ in any layer $c$ the relations could be written as follows:

$$u_v^{[c]} = \sum_{k=0}^{r} w_{vk} y_k^{[c-1]}$$ \hspace{1cm} (10)$$

$$y_v^{[c]} = F(u_v^{[c]})$$ \hspace{1cm} (11)$$

Writing the partial derivative of $E(t)$ with respect to weight $w_{vk}(t)$ and splitting into a chain rule:

$$\frac{\partial E(t)}{\partial w_{vk}(t)} = \frac{\partial E(t)}{\partial z_v(t)} \cdot \frac{\partial z_v(t)}{\partial y_v(t)} \cdot \frac{\partial y_v(t)}{\partial u_v(t)} \cdot \frac{\partial u_v(t)}{\partial w_{vk}(t)}$$ \hspace{1cm} (12)$$

From equation (9):
From equation (8):

\[
\frac{\partial e_{v}(t)}{\partial y_{v}(t)} = -1
\]  \hspace{1cm} (14)

From equation (11)

\[
\frac{\partial y_{v}(t)}{\partial u_{v}(t)} = F'(u_{v}(t))
\]  \hspace{1cm} (15)

From equation (10)

\[
\frac{\partial u_{v}(t)}{\partial w_{vk}(t)} = y_{k}(t)
\]  \hspace{1cm} (16)

Substituting equations (13)-(16) in equation (12) the result is:

\[
\frac{\partial E(t)}{\partial w_{vk}(t)} = -\alpha_{v}(t)F'(u_{v}(t))y_{k}(t)
\]  \hspace{1cm} (17)

Between forward pass and backward pass is therefore:

\[
\Delta w_{vk}(t) = -\eta \frac{\partial E(t)}{\partial w_{vk}(t)} = \eta \delta_{v}(t)y_{k}(t)
\]  \hspace{1cm} (18)

Where

\[
\delta_{v}(t) = e_{v}(t)F'(u_{v}(t))
\]  \hspace{1cm} (19)

\(\eta\)=training rate coefficient.

Smaller values for this training rate coefficient improve accuracy but extend the training time. The equation (18) is known as “Delta Rule” and was developed by Widrow and Hoff, is one of the most commonly used learning rules. For a given input vector, the output vector is compared to the correct answer. If the difference is zero, no learning takes place; otherwise, the weights are adjusted to reduce this difference.

If the neuron \(v\) is in the output layer then the value \(e_{v}(t)\) is directly observable but if it is in the hidden layer \(o_{v}(t)\) is not observable in this case the formula for \(\delta_{v}(t)\) is calculating different. In general this might be done by this formula:

\[
\delta_{v}[c-1] = \frac{1}{c-1} \sum_{m=1}^{c} \delta_{v}[m] w_{vk}[c]
\]  \hspace{1cm} (20)

From (18) and (20) the change in weight becomes:

\[
\Delta w_{vk}[c] = \eta \delta_{v}[c] y_{k}[c-1]
\]  \hspace{1cm} (21)

For a giving training set the weights in the network are the only parameters that can be modified to make the quadratic error \(E\) as low as possible. This can be minimizing by using an iterative process of gradient descent for which the gradient is:
\[
\Delta E = \begin{pmatrix}
\frac{\partial E(1)}{\partial w_{nk}(1)} & \frac{\partial E(2)}{\partial w_{nk}(2)} & \cdots & \frac{\partial E(t)}{\partial w_{nk}(t)}
\end{pmatrix}
\]

The whole learning problem now reduced to the questions of calculating the gradient of a network function with respect to its weights, minim of the error function, where \( \Delta E = 0 \).

In order to assess the quality of the model, some tests and statistics were used once the probabilities are calculated. Confusion matrix reveals information about actual and predicted classifications done by a logistic model.

<table>
<thead>
<tr>
<th>Actual Condition</th>
<th>Total Population</th>
<th>Actual Positive</th>
<th>Actual Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Positive</td>
<td>True Positive (TP)</td>
<td>False Positive (FP) (Type I error)</td>
<td></td>
</tr>
<tr>
<td>Predicted Negative</td>
<td>False Negative (FN) (Type II error)</td>
<td>True Negative (TN)</td>
<td></td>
</tr>
</tbody>
</table>

The matrix is a specific table layout with two rows and two columns and that reports the number of false positives, false negatives, true positives, and true negatives.

- False positive: Actual value is negative, but model predicts it as positive
- False negative: Actual value is positive, but model predicts it as negative
- True positive: Actual value is positive, model predicts it correctly
- True negative: Actual value is negative, model predicts it correctly

Sensitivity = TPR (true positive rate or Hit Rate [HR]) = \( \frac{TP}{TP + FN} \)

Specificity = TNR (true negative rate) = \( \frac{TN}{FP + TN} \)

Precision = PPV (positive predictive value) = \( \frac{TP}{TP + FP} \)

1 - Specificity = FPR (false positive rate or False alarm Rate [FAR]) = \( \frac{FP}{FP + TN} \)

\( F - \text{measure} = -2 \frac{\text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \)

Receiver Operating Characteristic (ROC) is a visual tool that can be easily constructed by plotting for every probability calculated the Sensitivity and 1-Specificity. One of the summary indices of ROC, the ROC measure (or Area under the Curve, AUC), is a linear transformation of the Accuracy Ratio mentioned above.

\( AR = 2(AUC) - 1 \)
An ROC curve demonstrates several things:

1. It shows the trade-off between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
2. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
3. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
4. The slope of the tangent line at a cut-point gives the likelihood ratio (LR) for that value of the test.
5. The area under the curve is a measure of test accuracy. The area A is 0.5 for a random model without discriminative power and it is 1.0 for a perfect model. It is between 0.5 and 1.0 for any reasonable model in practice.

3.1 Data

The database used contains 12078 companies having financial statements and loss and profit statements in 2013. According to the definition given by Basel Committee on Banking Supervision credit default occurs when one or more of the following takes place:

- “It is determined that the obligor is unlikely to pay its debt obligations (principal, interest or fees) in full;
- A credit loss event associated with any obligation of the obligor, such as charge-off, specific provision, or distressed restructuring involving the forgiveness or postponement of principal, interest, or fees;
- The obligor is past due more than 90 days on any credit obligation or
- The obligor has filed for bankruptcy or similar protection from creditors.”

Using the Basel definition, the observation period was for 12 months and all firms that in 2014 had more than 90 days overdue were considered as “bad” and the variable Trigger was marked as 1. The proportion of “bad” firms in this portfolio is 32.84%.
4. Results and Discussion

In the process of selecting variables were taken into account more qualitative aspect in terms of expert judgment and quantitatively. The first stage was to start from a list of 19 financial indicators and based on quantitative analysis the final list of variables to be adjusted.

<table>
<thead>
<tr>
<th>Financial Indicator</th>
<th>Good</th>
<th>Bad</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROA</td>
<td>9%</td>
<td>1%</td>
<td>7%</td>
</tr>
<tr>
<td>OROA</td>
<td>16%</td>
<td>3%</td>
<td>12%</td>
</tr>
<tr>
<td>ROE</td>
<td>30%</td>
<td>13%</td>
<td>24%</td>
</tr>
<tr>
<td>Dummy</td>
<td>0.982</td>
<td>0.764</td>
<td>0.909</td>
</tr>
<tr>
<td>Current liquidity</td>
<td>2.37</td>
<td>1.31</td>
<td>2.01</td>
</tr>
<tr>
<td>Quick Ratio</td>
<td>1.70</td>
<td>0.83</td>
<td>1.41</td>
</tr>
<tr>
<td>Cash Ratio</td>
<td>57%</td>
<td>15%</td>
<td>43%</td>
</tr>
<tr>
<td>DIR</td>
<td>140.89</td>
<td>106.27</td>
<td>129.24</td>
</tr>
<tr>
<td>CCR</td>
<td>93.45%</td>
<td>88.42%</td>
<td>91.76%</td>
</tr>
<tr>
<td>DSO</td>
<td>97.84</td>
<td>99.16</td>
<td>98.28</td>
</tr>
<tr>
<td>DIH</td>
<td>54.44</td>
<td>70.62</td>
<td>59.89</td>
</tr>
<tr>
<td>DPO</td>
<td>128.81</td>
<td>213.34</td>
<td>157.26</td>
</tr>
<tr>
<td>CCC</td>
<td>23.47</td>
<td>-43.56</td>
<td>0.91</td>
</tr>
<tr>
<td>Net Profit</td>
<td>6.22%</td>
<td>-2.15%</td>
<td>3.40%</td>
</tr>
<tr>
<td>EBITDA</td>
<td>8.35%</td>
<td>0.39%</td>
<td>5.67%</td>
</tr>
<tr>
<td>Debt Ratio</td>
<td>56.75%</td>
<td>79.86%</td>
<td>64.53%</td>
</tr>
<tr>
<td>Finance Horizon</td>
<td>80.53%</td>
<td>79.35%</td>
<td>80.13%</td>
</tr>
<tr>
<td>Assets Ratio</td>
<td>31.64%</td>
<td>33.96%</td>
<td>32.42%</td>
</tr>
<tr>
<td>Capital Ratio</td>
<td>40.55%</td>
<td>17.54%</td>
<td>32.80%</td>
</tr>
</tbody>
</table>

Correlation matrix of indicators with the binary variable Trigger already indicates that certain variables are correlated to a greater extent, for example indebtedness or equity rate of 44% while other variables do not correlate at all. Moreover, the sign of the correlation coefficient is also important. If for the debt ratio is a positive one because it indicates a high probability of default, the variable ROA has a negative sign because a high value of this indicator decreases the probability of default of the company. Using the same principle, for each indicator it was calculated the average within each group. Basically, the expectations are that the indicators of group companies "good" to have better figures for indicators of liquidity, profitability and solvency, while indicators of indebtedness and rotations to have lower averages than the group of companies "bad". It can be observed that liquidity indicators all comply with the above condition, which is a good prerequisite for entry into the model. In terms of profitability indicators, the differences are even more notable for the group of companies "bad" recorded negative values. The variables positively correlated with probabilities of default, such as indebtedness and rotations are also in line with the
assumption. With the establishment of financial ratios for estimation, the initial sample was partitioned into two sub-samples: training and testing. The role of these subsamples is to estimate parameters on a sub-sample and verify the accuracy on the other sub-sample. This method is useful for out-of-sample back testing. The initial sample was randomly divided into sub-sample to 80% training and 20% for the test. The estimation was done using logistic regression on relevant indicators and only those variables that were not statistically significant or whose coefficients did not have the right sign, were eliminated. For example, ROA and quick liquidity showed a positive coefficient, which led to their removal from the final model.

Table 2: LR estimated parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2.7330</td>
<td>0.1572</td>
<td>302.3626</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>OROA</td>
<td>1</td>
<td>-0.6393</td>
<td>0.2208</td>
<td>8.3811</td>
<td>0.0038</td>
</tr>
<tr>
<td>ROE</td>
<td>1</td>
<td>-0.5182</td>
<td>0.0663</td>
<td>61.1666</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Profit</td>
<td>1</td>
<td>-2.0965</td>
<td>0.1170</td>
<td>321.1996</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Current Liquidity</td>
<td>1</td>
<td>-0.2248</td>
<td>0.0300</td>
<td>56.1748</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cash Liquidity</td>
<td>1</td>
<td>-0.3901</td>
<td>0.0935</td>
<td>17.4155</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DIR</td>
<td>1</td>
<td>-0.00135</td>
<td>0.000318</td>
<td>18.0971</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CCR</td>
<td>1</td>
<td>-0.2617</td>
<td>0.1006</td>
<td>6.7646</td>
<td>0.0093</td>
</tr>
<tr>
<td>DIH</td>
<td>1</td>
<td>0.00159</td>
<td>0.000363</td>
<td>19.0756</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AIMO</td>
<td>1</td>
<td>-0.2044</td>
<td>0.1096</td>
<td>3.4800</td>
<td>0.0621</td>
</tr>
<tr>
<td>Capital Ratio</td>
<td>1</td>
<td>-1.5294</td>
<td>0.1269</td>
<td>145.3157</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Association of Predicted Probabilities and Observed Responses

<table>
<thead>
<tr>
<th>Percent Concordant</th>
<th>Somers’ D</th>
<th>0.476</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Discordant</td>
<td>Gamma</td>
<td>0.476</td>
</tr>
<tr>
<td>Percent Tied</td>
<td>Tau-a</td>
<td>0.212</td>
</tr>
<tr>
<td>Pairs</td>
<td>c</td>
<td>0.738</td>
</tr>
</tbody>
</table>

All remaining parameters are statistically significant and corresponding coefficients have the signs of economic theory. Likelihood Ratio with ten degrees of freedom is 5171.38, which indicates that the associated probabilities $\chi^2$ test is 0.00. Thus the null hypothesis that the model coefficients are zero is rejected.

As defined in the methodology section, the ROC curve actually illustrates the evolution of a classifier without considering the distribution of classes or the cost that it involves error, thus decoupling the classification performance of these two factors. ROC curve for this model estimated for the training sub-sample indicates an accuracy of 74%.
The confusion matrix for the entire sample indicates that 7883 firms were correctly assessed as being “good” and 1173 firms as being “bad” and the accuracy on the test sample is 76%.

### Table 3: LR Confusion Matrix

<table>
<thead>
<tr>
<th>Confusion Matrix</th>
<th>Realized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**Neural Networks**

The first step was to split the data into Training, Validation and Testing samples. The training sample of 8454 firms is used for preparing the network. The validation sample, 1812 firms, is used to measure network generalization and to halt training when generalization stops improving. The last 15% sample of testing has no effect on training and provides an independent measure of network performance during and after training. The next step is to create a neural network that will learn to classify the firms. Two-layer feed forward neural networks can learn any input-output relationship given enough neurons in the hidden layer. Layers which are not output layers are called hidden layers. In this exercise the first attempt was with 10 neurons for the hidden layer.
Once the network is ready to be trained the training continues as long as the network continues improving on the validation set. In order to see how the network performed during the training the mean squared error is calculated and shown in log scale. It can be observed that the performance decreases while the network is trained. Once the best performance is hit by the validation sample performance the version of this network is kept.

The training confusion matrix show a detection of 75.1% while for all sample the detection is 74.7%. The number of “good” companies that were correctly assigned as “good” is 7881 while for the “bad” companies the number is 1147.
With obtained parameters it was calculated for each firm, probability of default and the observed default rate is 32.84% while the estimated default rate is 32.6% for neural networks and 33.28% is for logistic regression. Therefore the first conclusion is that neural networks underestimate the default rate while the logistic regression overestimates it.

Table 4: PD Average NN vs LR

<table>
<thead>
<tr>
<th>Default</th>
<th>PD-NN</th>
<th>PD-logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.266410851</td>
<td>0.270436633</td>
</tr>
<tr>
<td>1</td>
<td>0.783273094</td>
<td>0.808083687</td>
</tr>
<tr>
<td>Total</td>
<td><strong>0.326364817</strong></td>
<td><strong>0.332801554</strong></td>
</tr>
</tbody>
</table>

www.globalbizresearch.org
5. Conclusions and Recommendations

The purpose of this paper was to estimate the probability of default for Romanian companies using logistic regression. When analysis only the financial ratios of year 2012 to predict the default of the company 1 year ahead the results showed that the best financial predictors are Capital Ratio, ROE, OROA, Own capital share, Current and cash liquidity, finance Horizon, DIR and DIH.

The first conclusion of this paper is that the results of this study are not only useful for any company to survive and to take early actions as a precaution, but also for any bank, investor and regulatory authority. The inconvenience however of these prediction models is that they highly depend on the data used in the analysis and perhaps, in case of a larger sample of data the model might behave differently. Even so, the conclusions are quite encouraging. Since the out-of-sample forecast accuracy of the estimated model of this study is high, it indicates that logistic regression used for estimating the probability of default for the Romanian companies overestimate de default rate while the neural networks underestimates it. When analysing the confusion matrix, the logistic regression detected 7883 while neural networks 7881 of the good firms and 1173 comparing with 1147 of the bad firms for the entire sample. Even though the accuracy of the models is similar when counting the detection for each class, the logistic regression obtained a better accuracy. The accuracy of “good” firms of logistic regression is 97.18% comparing with 97.16% for neural networks while for the “bad” firms the detection rate is 29.56% while for neural networks is 28.9%.

This work has been done through the Partnerships Program in Priority Areas - PNII, developed with the support of MEN-UEFISCDI, project no. 334/2014, project code PN-II-PT-PCCA-2013-4-0873, project title „A New Model for Corporate Risk Assessment: a Scientific Tool for Knowledge Based Management”.

References


Brier, G. W., Monthly Weather Review, 75 (1952), 1—3 (Brier Score)


Tasche, D., A traffic lights approach to PD validation, Deutsche Bundesbank, preprint (A traffic lights approach to PD validation)