THE SELECTION OF FINANCIAL RATIOS AS INDEPENDENT VARIABLES FOR CREDIT RISK ASSESSMENT

Vytautas Boguslauskas1, Ričardas Mileris2, Rūta Adlytė3

1 Kaunas University of Technology, Lithuania, vytautas.boguslauskas@ktu.lt
2 Kaunas University of Technology, Lithuania, ricardas.mileris@ktu.lt
3 Kaunas University of Technology, Lithuania, ruta.adlyte@stud.ktu.lt

Abstract

When developing the internal credit risk assessment models in banks it is very important to find variables that allow to evaluate the credit risk of a company accurately. The classification results depend on the appropriate data characteristics and the appropriate analysis algorithm for a selected data set. A goal of variables selection is to find the informative variables that are necessary for successful analysis. If not enough features are selected, there is a risk that the information content in the set of features is low. If too many features are used for analysis, the effect of noise in the information can appear. The analysis of scientific publications about credit risk estimation models indicated that the most often financial ratios are used in credit risk models as independent variables. The dependent variable in such models is the class of a company: default or not default. For the developers of credit risk assessment models the correctly selected set of initial variables can improve their models performance. The analysis of scientific publications about credit risk allowed to find 15 financial ratios that recur in more than 10% of analyzed models. Also 5 variables were added to this set arguing their usability from the financial standpoint. The classification models were developed by applying these methods: discriminant analysis, logistic regression and artificial neural networks. The models were developed for the classification of companies into 2 groups: default and not default. Further the actual variables in developed models were analyzed. The analysis of scientific publications about credit risk assessment models and the developed classification models allowed to compile the set of informative financial ratios for the estimation of credit risk. The financial ratios described in this research allow to assess credit risk of companies successfully.

Keywords: bank, bankruptcy, classification, credit risk, financial ratios.

Introduction

Credit risk is one of the major threats that financial institutions face, and it is essential to be able to assess this risk properly (Lin, 2009). Banks always implement a credit risk analysis before making new loans (Inderst, Mueller, 2008). A potential client’s credit risk level in banks can be evaluated by the internal credit risk assessment models. The main aim of these models is to determine whether an applicant has the capacity to repay the loan. This is normally done using historical data and statistical techniques (Emel, Oral, Reisman, Yolalan, 2003). So often the primary issue of credit risk research is to determine what variables significantly influence the probability of default. A second important issue is the construction of credit scoring model (Marshall, Tang, Milne, 2010).

The object of this paper is the independent variables of statistical credit risk assessment models.

The aim of this paper is to find the most actual independent variables for the development of the internal credit risk assessment models in banks.

The tasks of this paper:
- To find the most informative financial ratios in credit risk assessment models.
- To develop the statistical credit risk assessment models analyzing financial data of Lithuanian companies.
- To analyze the performance and the results of developed statistical models.
- To analyze the financial ratios that are the most actual for the credit risk assessment.

The methods were applied:
- The analysis of scientific publications about internal credit risk assessment models in banks.
- The development of statistical credit risk assessment models.
- The statistical analysis of financial ratios.

For the developers of credit risk assessment models the correctly selected set of initial variables can improve their models performance.
The assessment of credit risk in banks

Credit risk is the risk of loss due to a debtor’s non-payment of a loan: the principal or interest (or both). Default occurs when a debtor has not fulfilled his legal obligations according to the debt contract, or has violated a loan covenant (condition) of the debt contract (Chen, Wang, Wu, 2010). The banks hold a portfolio of N credits \( A = (A_1, A_2, \ldots, A_n) \). Each credit \( A_i \) has a specific size, a time to repay, a default probability (PD), loss given default (LGD), and interest rate \( C_i \) (Alessandri, Drehmann, 2010). Credit risk assessment instruments allow banks to insure themselves against loan losses precisely in those states that signal monitoring (Chiesa, 2008). The credits are extended to those with a high probability of paying it back (Lieli, White, 2010). Also the estimated changes of credit risk in loan portfolio lead to the changes in the bank’s capital adequacy (Drehmann, Sorensen, Stringa, 2010). When developing internal credit risk assessment models it is necessary to construct a scale of credit ratings. It can be similar to the rating scales of Moody’s and S&P: Aaa/AAA, Aa1/AA+, ..., C/D (Guttler, Wahrenburg, 2007). The probability of default increases as the credit quality and rating decrease (Stefanescu, Tunaru, Turnbull, 2009). So credit scoring always need high accuracy to avoid bad debts (Xu, Zhou, Wang, 2009).

Due to financial crises and regulatory concerns of the Basel Committee on Banking Supervision, a regulatory requirement was made for the banks to use sophisticated credit scoring models for enhancing the efficiency of capital allocation. The Basel Committee formulated broad supervisory standards and guidelines for banks to implement. The Committee published a revised framework as the new capital adequacy framework, also known as Basel II (Khashman, 2010). Clearly, the internal ratings based (IRB) approach was a major innovation of the New Accord: bank internal assessments of key risk drivers are primary inputs to the capital requirements. For the first time, banks were permitted to rely on their own assessments of a borrower’s credit risk. The close relationship between the inputs to the regulatory capital calculations and banks’ internal risk assessments will facilitate a more risk sensitive approach to minimum capital. Changes in a client’s credit quality will be directly reflected in the amount of capital held by banks (Angelini, Tollo, Roli, 2008).

Credit risk assessment models and independent variables

When assessing the risk related to credit products, different problems arise, depending on the context and the different types of borrowers. Paleologo, Elisseeff & Antonini (2010) summarize the different kind of scoring as follows:

1. Application scoring: it refers to the assessment of the creditworthiness for new applicants.
2. Behavioral scoring: it involves principles that are similar to application scoring, with the difference that it refers to existing customers.
3. Collection scoring: collection scoring is used to divide customers with different levels of insolvency into groups, separating those who require more decisive actions from those who do not need to be attended to immediately.
4. Fraud detection: fraud scoring models rank the applicants according to the relative likelihood that an application may be fraudulent (Paleologo, Elisseeff, Antonini, 2010).

Bonfim (2009) identifies three different groups of models:

1. Models which rely mostly on accounting variables. These models mostly are based on statistical methods in order to solve the classification problems, such as linear discriminant analysis, logistic regression, multivariate adaptive regression splines, classification and regression tree, case based reasoning, artificial neural networks (Chuang, Lin, 2009), linear probability and multivariate conditional probability models, the recursive partitioning algorithm, multi-criteria decision-making (MCDM), mathematical programming approaches have been proposed to support the credit decision (Min, Lee, 2008).
2. Models which use mostly market information often include Merton-type approaches to credit risk modelling. The major drawback of such models is that, as they rely on market information, usually they can only be applied to quoted companies (Bonfim, 2009). The distance to default in Merton model is defined by the expression:

\[
\gamma = \frac{\ln(V_A / D) + \left( r - \frac{1}{2} \sigma_A^2 \right) (T-t)}{\sigma_A^2 \sqrt{T-t}},
\]

where \( V_A \) is the market value of the firm’s assets;
\( D \) is the total amount of the firm’s debt;
The credit market is a market with incomplete information where banks try to minimize the risk of the loan (Pang, Wang, 2008). Understanding the determinants of credit risk is a major issue for financial stability of banks (Bonfim, 2009). When developing the credit risk estimation models the researchers confront with two main problems: how to choose the optimal input feature subset for the classifier and how to select the best classifier (Hsieh, Hung, 2010). Data characteristics (noise, missing values, complexity of distribution of data, instance selection, etc.) significantly affect the resulting performance of most algorithms. Having selected an appropriate algorithm for a given data set, it can be shown that performance can further be improved by appropriate data characteristics (Piramuthu, 2006). If not enough features are selected, there is a risk that the information content in this set of features is low. If too many features are used for analysis, the effect of noise in the information can appear (Piramuthu, 2006). Mostly the input variables are the financial ratios calculated from the financial statements. They reflect the financial structure, solvency, profitability and the cash flow of a company (Chen, Shih, 2006). Also the changes (increase and decrease) in financial ratios can be analyzed (Zhang, Hardle, 2010). Grunert, Norden & Weber (2005) affirm that banks not only have to consider quantitative but also qualitative factors, for example, the availability of audited financial statements, depth and skill of management, the position within the industry and future prospects. Banks collect a full history of internal credit related data for all debtors, including the unique, government provided, firm identification number, the internal risk rating, the credit type, the amount of credit granted, actual exposure, payment status and industry code (Jacobson, Linde, Roszbach, 2006).

Informative financial ratios for credit risk assessment models

In order to find the most frequent financial ratios in the credit risk assessment models, 194 classification models in the scientific publications were analyzed. The most frequent 15 financial ratios were selected (Table 1), that recur in more than 10% models (L, %). According to percentage of frequency the ranks (R1) were attributed for every ratio.

<table>
<thead>
<tr>
<th>No.</th>
<th>Ratio</th>
<th>Group*</th>
<th>L, %</th>
<th>R1</th>
<th>M, %</th>
<th>R2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Current ratio (BPK)</td>
<td>L</td>
<td>87.0</td>
<td>1</td>
<td>79.2</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>2.</td>
<td>Total debt to equity (SNK)</td>
<td>F</td>
<td>57.4</td>
<td>2</td>
<td>62.5</td>
<td>18.5</td>
<td>20.5</td>
</tr>
<tr>
<td>3.</td>
<td>Sales to total assets (TA)</td>
<td>A</td>
<td>51.9</td>
<td>3</td>
<td>66.7</td>
<td>13.5</td>
<td>16.5</td>
</tr>
<tr>
<td>4.</td>
<td>Working capital to total assets (GST)</td>
<td>L</td>
<td>50.0</td>
<td>4</td>
<td>87.5</td>
<td>5.5</td>
<td>9.5</td>
</tr>
<tr>
<td>5.</td>
<td>Total liabilities to total assets (IK)</td>
<td>F</td>
<td>48.1</td>
<td>5</td>
<td>91.7</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>6.</td>
<td>Net profit to total assets (TP)</td>
<td>P</td>
<td>44.4</td>
<td>6</td>
<td>95.8</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>7.</td>
<td>Equity to total assets (NKT)</td>
<td>F</td>
<td>38.9</td>
<td>7</td>
<td>62.5</td>
<td>18.5</td>
<td>25.5</td>
</tr>
<tr>
<td>8.</td>
<td>Net profit to equity (NKp)</td>
<td>P</td>
<td>35.2</td>
<td>8</td>
<td>66.7</td>
<td>13.5</td>
<td>21.5</td>
</tr>
<tr>
<td>9.</td>
<td>Quick ratio (SPK)</td>
<td>L</td>
<td>33.3</td>
<td>9</td>
<td>77.1</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>10.</td>
<td>Net profit margin (GP)</td>
<td>P</td>
<td>31.5</td>
<td>10</td>
<td>81.3</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>11.</td>
<td>Unappropriate balance to total assets (NP/T)</td>
<td>O</td>
<td>18.5</td>
<td>11</td>
<td>87.5</td>
<td>5.5</td>
<td>16.5</td>
</tr>
<tr>
<td>12.</td>
<td>Working capital to sales (GAK/P)</td>
<td>O</td>
<td>16.7</td>
<td>12.5</td>
<td>64.6</td>
<td>16.5</td>
<td>29</td>
</tr>
<tr>
<td>13.</td>
<td>Gross profitability (BP)</td>
<td>P</td>
<td>16.7</td>
<td>12.5</td>
<td>70.8</td>
<td>10.5</td>
<td>23</td>
</tr>
<tr>
<td>14.</td>
<td>Current assets to total assets (TT/T)</td>
<td>O</td>
<td>14.8</td>
<td>14</td>
<td>66.7</td>
<td>13.5</td>
<td>27.5</td>
</tr>
<tr>
<td>15.</td>
<td>EBIT to total assets (TVP/T)</td>
<td>O</td>
<td>-</td>
<td>18</td>
<td>95.8</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>16.</td>
<td>EBIT to sales (TVP/P)</td>
<td>O</td>
<td>-</td>
<td>18</td>
<td>95.8</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>17.</td>
<td>Cash to current liabilities (PGP)</td>
<td>L</td>
<td>-</td>
<td>18</td>
<td>56.3</td>
<td>20</td>
<td>38</td>
</tr>
<tr>
<td>18.</td>
<td>Sales to long term assets (ITA)</td>
<td>A</td>
<td>-</td>
<td>18</td>
<td>66.7</td>
<td>13.5</td>
<td>31.5</td>
</tr>
<tr>
<td>19.</td>
<td>Long term debt to equity (ISK)</td>
<td>F</td>
<td>-</td>
<td>18</td>
<td>64.6</td>
<td>16.5</td>
<td>34.5</td>
</tr>
</tbody>
</table>

Also 5 financial ratios were added to this set (variables 16 – 20) arguing their usability from the financial standpoint. EBIT to total assets characterize the efficiency of using assets. The higher profit is earned using particular quantity of assets, the higher is the efficiency of a company. Also the efficiency is influenced on the ability to earn particular profit using less assets. EBIT to sales indicates the proportion of sales that is earned as EBIT. The higher this ratio, the better the performance of a company. Cash to current liabilities indicates the proportion of current liabilities that a company is able to refund by cash immediately. This is very important feature of company’s solvency. Sales to long term assets indicates the efficiency of using long term assets. It is the ability of a company to get more income using less long term assets. Long term debt to equity indicates the relation between these partitions in balance-sheet. The higher is equity and the lesser is long term debt, the better is financial structure of a company in point of creditors. Because these 5 financial ratios in analyzed classification models were used relatively rarely, the coherent ranks „18“ were attributed for these ratios (R1).

When the set of 20 initial variables was compiled, the credit risk assessment model was developed which allows to attribute credit ratings for Lithuanian companies (Figure 1). The reduction of initial variables was accomplished by analysis of variance (ANOVA), Kolmogorov-Smirnov (K-S) test, factor analysis and ranks of importance in neural networks. Discriminant analysis, logistic regression and artificial neural networks were applied and 15 models were developed for the classification of Lithuanian companies into 2 groups: default or not default. Financial reports of 100 Lithuanian companies were analyzed. By every method 5 models were developed that analyze financial data from 1 to 5 years (Table 2).

Table 2. The developed classification models and their total accuracy

<table>
<thead>
<tr>
<th>Analysis method and period</th>
<th>1 year</th>
<th>2 years</th>
<th>3 years</th>
<th>4 years</th>
<th>5 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminant analysis</td>
<td>DA1 (77.0)</td>
<td>DA2 (84.0)</td>
<td>DA3 (84.0)</td>
<td>DA4 (82.0)</td>
<td>DA5 (83.9)</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>LR1 (83.0)</td>
<td>LR2 (92.0)</td>
<td>LR3 (97.0)</td>
<td>LR4 (89.9)</td>
<td>LR5 (87.4)</td>
</tr>
<tr>
<td>Artificial neural networks</td>
<td>ANN1 (85.9)</td>
<td>ANN2 (92.2)</td>
<td>ANN3 (95.5)</td>
<td>ANN4 (85.3)</td>
<td>ANN5 (86.2)</td>
</tr>
</tbody>
</table>

The total accuracy, sensitivity and specificity were measured for every model. The highest total accuracy (97%) was reached by logistic regression model which analyzes data of 3 years. When the new financial data of 100 companies was available, the change of total accuracy was measured. It decreased by 14.77%, so the logistic regression coefficients were recalculated. The LR3 model update increased the total accuracy by 11.2%. Further the rating model was developed which for companies, by updated logistic regression model classified as not default, attributes ratings AAA – D2. According to 7 financial ratios and the individual possibility of default (P) estimated by the updated LR3 model, scores {0; 1; 2; 3; 4; 5; 6; 7} were attributed for companies. The sum of scores determined the credit rating (Figure 1). Also rating D1 was attributed for companies which by logistic regression model were classified as default.

1. Initial variables: X₁ – X₂₀ (1 year), X₁ – X₄₀ (2 years), X₁ – X₆₀ (3 years), X₁ – X₈₀ (4 years), X₁ – X₁₀₀ (5 years)
2. Reduction of variables: ANOVA, K-S test, factor analysis, ranks of importance.
5. Selection of classification model: Logistic regression model (3 years data analysis). Total accuracy = 97%.
6. Analysis of changes in classification accuracy: Total accuracy of model LR3 decreased: 82.23% - 97% = -14.77%.
7. Model’s update: Recalculation of logistic regression coefficients. Total accuracy increased: 93.43% - 82.23% = 11.2%.

Figure 1. Credit risk assessment model’s development process

In order to find the most actual variables, the developed 16 classification models (15 initial and 1 updated) which classify companies into default and not default groups were analyzed (stage 3 in Figure 1). The highest possible frequency of each variable is 48 times regardless of the data period in the models. The frequencies of financial ratios in developed models were calculated (column M, % in Table 1). The ranks were attributed for financial ratios according to their frequency in the developed models (column R2 in Table
1). Then sum of ranks (R1 + R2) was calculated for every variable (column Sum in Table 1). According to these results, the final rank was attributed for financial ratios (Figure 2). The highest rank is 1, the least rank is 20. These ranks reflect the importance of variables in analyzed scientific publications and developed 16 classification models.

Figure 2. Ranks attributed for financial ratios in credit risk assessment

The most actual 10 variables for credit risk assessment are: net profit to total assets, current ratio, total liabilities to total assets, working capital to total assets, sales to total assets, unappropriate balance to total assets, net profit margin, quick ratio, EBIT to total assets, EBIT to sales. The differences between average financial ratios in groups of not default (μ₁) and default (μ₂) companies are illustrated in Figure 3. The average values were calculated without exceptions (E ≠ [μ ± 3σ]). The exceptions were considered as values that distant from the average more than by 3 standard deviations (σ). All profitability ratios of default companies are negative one year before bankruptcy. Also the working capital is negative of these companies. The average net profit margin of not default companies is 8.7%. The debt ratio indicates that average debt of not default companies does not exceed 0.5 of total assets. The analysis highlighted the differences between current ratio and quick ratio. The liquidity is about 3 times higher in group of not default companies.

Figure 3. The average financial ratios of companies

Further the capability to discriminate default and not default companies by every financial ratio separately was estimated. 8 classification and regression trees were developed for this purpose (Table 3).

Table 3. Characteristics of developed classification and regression trees (CRT)

<table>
<thead>
<tr>
<th>Tree number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>19</td>
<td>19</td>
<td>13</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Resubstitution cost</td>
<td>0.030</td>
<td>0.032</td>
<td>0.041</td>
<td>0.049</td>
<td>0.069</td>
<td>0.090</td>
<td>0.115</td>
<td>0.250</td>
</tr>
</tbody>
</table>

The tree was selected for analysis with the least resubstitution cost (Figure 4).
The ID in tree graph is the node number, N is the size of node, Mu is the node mean, Var is the node variance. The financial ratio and the classification threshold are in brackets. The end nodes are highlighted in grey. The total accuracy of developed CRT is 95%, sensitivity – 98%, specificity – 92%. This means that overall 95% of companies were classified correctly. Also the developed CRT allowed to classify 98% of default and 92% of not default companies correctly.

Table 4. Predictor importance in the classification and regression tree analysis

<table>
<thead>
<tr>
<th>Rank</th>
<th>NP/T</th>
<th>TP</th>
<th>IK</th>
<th>GP</th>
<th>TVP/T</th>
<th>GST</th>
<th>TVP/P</th>
<th>BPK</th>
<th>SPK</th>
<th>TA</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>98</td>
<td>91</td>
<td>88</td>
<td>78</td>
<td>76</td>
<td>72</td>
<td>70</td>
<td>53</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

The ranks of importance were calculated for every variable in the CRT analysis. The interval of these ranks is [0; 100]. These ranks confirmed the importance of analyzed financial ratios in credit risk assessment.

Conclusions

1. The analysis of scientific publications about credit risk assessment models and the developed classification models allowed to compile the set of informative financial ratios for the estimation of credit risk.
2. The highest total accuracy reached by developed models is 97%. The accuracy of other developed models vary from 77% to 95.5%.
3. The financial ratios described in this research allow to assess credit risk of companies successfully. The results can help for the developers of credit risk assessment models to compose the initial set of financial ratios.
4. The developed classification and regression tree showed the ability of analyzed financial ratios to discriminate bank clients into default and not default groups. The calculated ranks of importance also confirmed the relevance of analyzed financial ratios in credit risk assessment.
References