***Explanatory note to the repository***

This work was developed as part of a study of the Russian corporate bond defaults that occurred in the period from 2008 to 2018. The idea was to develop a mathematical model capable of predicting an issuer's default based on its financial statements. The motivation for the study was the fact that there is no source of financial statements of Russian companies issuing bonds and relevant works researching the Russian debt securities market. Author investigated corporate bonds that were listed on Russian stock exchanges in rubles. All the necessary information was collected in the Spark system.

At the first step of the study, it was necessary to analyze all types of defaults and their frequency to select the most relevant financial indicators of companies. The study did not include issuers representing the financial sector, since the reporting of such companies differs from the reporting of other industries. Based on the collected statistics, it turned to be obvious that the lack of funds became the most frequent reason both in terms of the number of companies and the amount (in RUB) of unfulfilled obligations. The statistics of the types of defaults are presented in Table 1.

Table 1. Statistics of defaults of Russian bonds issuers for 2008-2018

|  |  |  |
| --- | --- | --- |
| **The reason for default** | **Number of companies** | **The amount of unfulfilled obligations (RUB)** |
| Restructuring | 14 | 9 371 699 074,79 |
| Restriction on satisfying claims | 12 | 4 190 431 330,75 |
| Lack of funds | 249 | 188 478 274 256,83 |
| Bankruptcy | 44 | 21 240 119 318,33 |
| Impossibility of conducting operations | 21 | 4 489 407 941,20 |
| No information | 21 | 7 355 820 313,24 |

In the work, it was assumed that the default of the issuing company can be indicated by financial indicators calculated for the year preceding the year of default. For example, if a company defaulted in 2015, then financials for 2014 may indicate an impending issuer default. For non-default companies, this method of calculating financial indicators is not suitable, therefore, financial indicators were calculated for all years of the bond's life (except for the first year of life, since no payments may occur during this period (for example, zero-coupon bonds)), and then the year was chosen in which the values ​​of financial indicators were the lowest. For example, if a company's bond was listed on the stock exchange in the period from 2010 to 2015, then the financial indicators are calculated for 2011-2014, and the year with the lowest indicator is selected. It is assumed that if the company, even with its lowest financial performance, was able to remain stable and prevent default, then such indicators are the minimum necessary to classify the company as default or non-default. As a result of collecting data, the sample consisted of 300 non-default companies and 56 default companies. Final database consisting of 210 companies and 31 financial records. The number of companies in the sample is less than the number of bond issues during this time. This was due to the following factors:

1. It is impossible to find annual report of the company.

2. It was not possible to find annual report of the company for the required year.

3. Annual report for the required year is empty.

Lack of funds and bankruptcy, as the most common causes of defaults, suggest that the cause of defaults should have been sought in the incorrect management of the company's liquidity, as well as in the inability to service its own debt. Based on this, a list of used financial multipliers was selected, and it is presented in table 2.

Table 2. Financial multipliers

|  |  |  |
| --- | --- | --- |
| **№** | **Multiplier name** | **Calculation method** |
| 1 | Current liquidity ratio | Current assets / Current liabilities |
| 2 | Quick liquidity ratio | (Short-term receivables + short-term financial investments + cash) / Short-term liabilities |
| 3 | Absolute liquidity ratio | (Short-term financial investments + cash) / Short-term liabilities |
| 4 | Long-term debt / Equity | Long-term debt / Equity |
| 5 | Financial leverage effect | ROE-ROA |
| 6 | ROA | Net Profit / Balance |
| 7 | ROE | Net Profit / Equity |
| 8 | Debt concentration ratio | (Short-term liabilities + long-term liabilities) / Balance |
| 9 | Coefficient of provision with own circulating assets | (Equity - Non-current assets) / Current assets |
| 10 | Operating profit margin | Operating profit / Revenue |
| 11 | Financial stability | Equity / Balance |
| 12 | Interest coverage ratio | EBIT / Interest Payable |
| 13 | Revenue growth rate | Revenue (n-year) / Revenue (n-1 year), where n - year before default |

Due to the lack of values in the data, 2 methods of processing the collected database were sequentially applied:

1) Removing companies whose share of missing values exceeds 30%

2) Filling in the missing data using the method of means (filling in the missing data with more complex methods, such as bagging, does not improve the accuracy of the estimated models). This method was applied separately for default and non-standard companies.

3) The data was divided into training and test samples in an 80/20 ratio

Further, the following models were calculated:

***1) Logit regression***

15% of companies were randomly selected from the sample for the test. The remaining 85% of the companies constituted the training sample. Significant financial multipliers (at 10% significance level) were х2 - Quick liquidity ratio; х11 - Financial stability; х12 - Interest coverage ratio (see picture 1). The accuracy of the model on the training and test samples was 70.95% and 73.33%, respectively.

Figure 1. Logit regression

Изображение выглядит как текст

Автоматически созданное описание

***2)*** ***Probit regression***

For this type of model, the results were similar. As expected, the financial multipliers from point 1 were significant in the probit regression. The accuracy of the model on the training and test set was 73.65% and 73.33%, respectively.

Figure 2. Probit regression

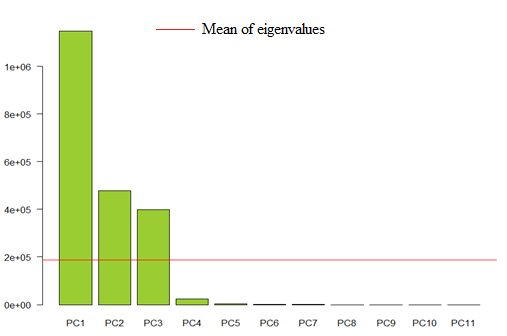
Изображение выглядит как текст

Автоматически созданное описание

***3) Logit-regression on a sample transformed using the PCA method.***

In order to reduce the dimension of dataset, possible lack of information content and duplication of information by strongly correlated variables, the PCA method was used. To select the optimal number of principal components, the Kaiser-Gutmann test was applied, which assumes to leave only those principal components whose eigenvalues exceed their average (the results in Figure 3).

Figure 3. Results of the Kaiser-Gutmann test



The results of the constructed logit regression on the transformed dataset can be seen in Figure 4. The accuracy of the model on the training and test set was 95.27%, respectively. and 97.1%.

Figure 4. Logit regression on the transformed data using the PCA method

Изображение выглядит как стол

Автоматически созданное описание

***4) SVM with different kernels.***

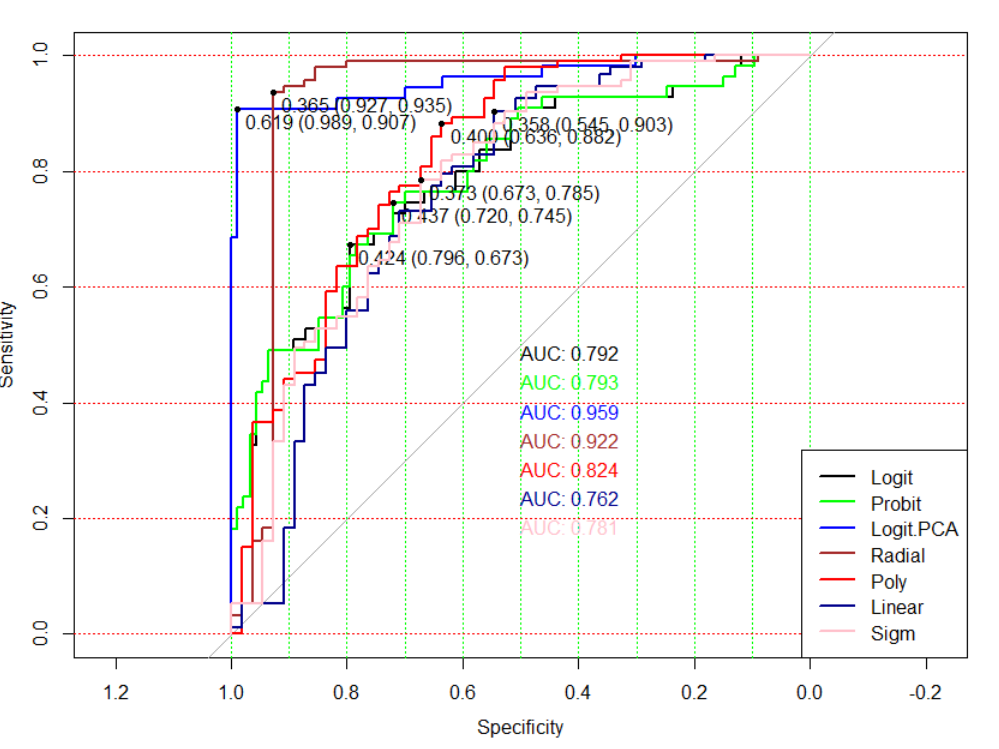
SVM models were built with different kernels such as Linear kernel, Polynomial kernel with power p, Gaussian kernel, Sigmoid kernel. For each model, cross-validation was used, dividing the sample into 10 equal parts, and intervals were set for the parameters to be optimized. The results of calculating SVM models are presented in Table 3.

Table 3. Results of SVM models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Range of investigated parameters** | **Best model parametrs** | **Training set accuracy** | **Testing set accuracy** |
| Linear kernel | - | - | 73,47% | 87% |
| Polynomial kernel | - | - | 89,80% | 93,10% |
| Sigmoid kernel | sigma=2^[-10,10],  cost=2^[-10:10] | sigma=0.0625, cost=4 | 74,15% | 80% |
| Gaussian kernel | cost=2^[-10:10], gamma=2^[-10,10] | cost=256, gamma=8 | 97,90% | 98% |

To compare the quality of the models, the ROC-curve was used (Figure 5). It can be concluded that the most effective models turned out to be logit-regression on the transformed data of the PCA method and the SVM method with a radial kernel.

Figure 5. ROC curves for estimated models



It can be concluded that these types of models is applicable to forecasting the default of corporate bonds in the Russian market. Due to the incompleteness of data on companies, it is impossible to consider non-quantitative indicators of companies, for example, information about the owners of the companies, etc. However, it would be useful to use the dynamics of the proposed financial multiplies and use this information, for example, as adjustment coefficients for the original predictors.