### Национальный исследовательский университет

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

Международный Институт Экономики и Финансов

##### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**по образовательной программе высшего профессионального образования, направление 080100.68 Экономика**

на тему: Моделирование вероятности дефолта российских банков

 (Probability of Default Modeling for Russian Banks)

 Работа выполнена на английском языке

Студент 2 курса магистратуры:

Костров Александр Владимирович

Научные руководители:

 Слонимчик Фабиан,

PhD

Карминский Александр Маркович,

профессор, д.э.н., д.т.н.

**МОСКВА, 2014**

**Аннотация**

Данная магистерская диссертация посвящена актуальной теме создания системы раннего предупреждения для российской банковской системы. Моделирование осуществлялось в соответствии с рекомендациями Базельского комитета по банковскому надзору и ориентировано на внедрение IRB-подхода в банковскую практику.

В данной работе построена адекватная модель вероятности дефолта российского банка, и потенциально она может быть использована на практике, как регулятором, так и контрагентами российских банков, чьи средства не попадают под действие системы страхования вкладов (ССВ).

Исследование проводилось в течение всего периода обучения в магистратуре МИЭФ и состоит из двух частей:

1. В течение первого года обучения в магистратуре была построена модель вероятности дефолта российского банка. Была проведена серьезная работа по сбору данных для исследования и поиску объясняющих переменных для построения эконометрических моделей дефолта банка, обзору литературы по теме исследования. При построении моделей вероятности использовались современные эконометрические методы, а также адекватные критерии отбора наилучших моделей. В магистерской диссертации построена и описана модель вероятности дефолта российских банков, ежеквартально предсказывающая более 60% дефолтов российских банков.
2. В течение второго года обучения полученные результаты были использованы для подготовки статьи к публикации в журнале Eurasian Economic Review, Springer; а также в Журнале Новой экономической ассоциации (ВАК). Кроме того, исследование продолжается. С использованием ранее построенной модели, исследуются изменения в политике регулятора по надзору за банковским сектором РФ в 2013-2014 гг.

Данная магистерская диссертация является статьей, опубликованной в журнале Eurasian Economic Review. В ней осуществлено сравнение нескольких моделей вероятности дефолта российского банка с использованием национальной банковской статистики за период с 1998 по 2011 гг. Мы обнаружили, что модель бинарного выбора типа логит с квази-панельной структурой данных демонстрирует наиболее хорошие результаты. Также мы обнаружили квадратичную U - образную зависимость между достаточностью капитала банка и его вероятностью дефолта. Кроме того, макроэкономические, институциональные и временные факторы существенно улучшают качество модели. Эти результаты полезны для национальных регуляторов в финансовой сфере, как и для риск-менеджеров в коммерческих банках.

**Probability of Default Modeling for Russian Banks (The Probability of Default in Russian Banking†)**

**Alexander M. Karminsky\* and Alexander Kostrov\*\***

*Abstract*: We compare several models for estimating the default probabilities of Russian banks using national statistics from 1998 to 2011, and find that a binary logit regression with a quasi-panel data structure works best. The results indicate that there is a quadratic U-shaped relationship between a bank's capital adequacy ratio and its probability of default. In addition, macroeconomic, institutional, and time factors significantly improve model accuracy. These results are useful for national financial regulatory authorities, as well as for risk-managers in commercial banks.

*Keywords:* Probability of Default, Banks, Risk-Management, Default Classification

*JEL Classification:* G21, G24, G32

**1. Introduction**

Achieving sustainable development of financial institutions has been at the forefront of the world agenda since the end of the recent economic crisis in 2008-2009. Prudent supervision of the banking sector is important to reach this goal. Russia ranks third in the number of banks, after the United States and Germany: there are about 900 operating banks in Russia. At the same time, the Central Bank of the Russian Federation (the Bank of Russia, CBR) lacks the necessary resources to organize regular field inspections of a large number of its banks. This is why the CBR needs a remote system to monitor the national-banking sector. The most vulnerable banks should be identified and properly supervised to improve banking sector stability.

The probability-of-default (PD) model is a possible instrument to address this problem. The PD shows the likelihood of a bank failure over a fixed assessment period. This paper reviews binary choice models that attempt to describe, predict, and prevent defaults of Russian banks with regard to national banking sector peculiarities. We have used the experience of PD model creation for emerging economies, i.e., BRICS and Eastern Europe. The majority of the existing work on the Russian banking experience examines the collapse of the Russian banking system in 1998. However, the rules of the game and the economic environment have dramatically changed since that time.

In addition, the PD model offers the advantage of being able to be utilized by other parties: banks, bank creditors and business partners. First, banks can use it to predict and prevent hardships. The model meets the requirements of Basel II and takes the dependence of credit risk on the business cycle into account. This can be exploited for internal financial monitoring in commercial banks. Second, information about a bank's financial stability (its ability to survive in hard times and meet its financial obligations) is essential for bank creditors and business partners. The rest of the paper is organized as follows:

The second section presents an overview of the Russian banking sector. Section 3 provides a brief literature review of PD-model development. Section 4 describes the database and sources used. Section 5 explains the methodology of PD model creation; a comparison of the derived model with alternatives is addressed, as well. Section 6 discusses the model estimation results. The final section contains a discussion and conclusions.

**2. The Russian Banking System: Stability Issues**

In the late 1980s, commercial banking experienced resurgence in Russia. More than 2,500 banks had been launched by 1995. Currently about 3,500 charters of incorporation have been issued by the CBR since 1980s.

Two periods of mass license withdrawals in the Russian banking sector can be distinguished: the first, from 1996 to 2000; and the second, between 2008 and 2010, which were related to the financial crises of 1998 and 2008, respectively. Table 1 provides information on the number of banks competing in the market during these periods. By the 2000s, as a result of the 1998 financial crisis, the surge in the number of credit institutions came to an end.

Table 1. The number of banks in the Russian banking system

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1998 | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 |
| Operating banks | 1,447 | 1,274 | 1,282 | 1,249 | 1,143 | 1,058 | 965 | 898 |
| 100% foreign-owned banks | 18 | 22 | 27 | 33 | 52 | 77 | 80 | 74 |
| with over 50% state ownership | 18 | 22 | 26 | 32 | 33 | 63 | 82 | 77 |

**Source:** The Bank of Russia

Since 2000, there has been intensive growth in the Russian banking system. 35 Russian banks were among the Top 1,000 World Banks (by total assets) in 2008; this is substantially higher than at the beginning of the millennium, when there were only 20 Russian banks in that list. Although close in size to bank assets in BRICS countries (by 0.5% of the Top 1,000 World Banks' total assets each, except for Brazil with 5%), the Russian banking industry (by total assets) is tiny in comparison to any of the Top 20 World Banks.

The Russian banking sector has passed through two stages of development with crises forming their boundaries (Figure 1).

**Stage 1. Formation (1989-1999):** This stage is characterized by unsystematic development, an excessive number of banks, and many regulatory loopholes.

**Stage 2. Rapid development (2000-2008):** Rapid growth of quantitative and some qualitative measures are typical for this period. There was an upward trend in the Bank assets-to-GDP ratio which was 36% in 2000 and jumped to 67% in 2008. A similar pattern is evident in the Credit-to-GDP ratio, which was under 20% in 2000 and rose to 50% by the end of 2008. Additionally, some systematic problems in the Russian banking system were resolved such as, implementation of deposit insurance and Basel I compliance.

|  |
| --- |
|  |

Figure 1. Size of the Russian banking system

This drastic growth in the sector, however, caused overheating with a resultant upturn in bad debts when a recession shook the world economy in 2008.

**Stage 3. Sustainable growth (2009 to present):** Throughout this period, a great emphasis has been placed on the proportional development of the Russian banking system. It is vital to reinforce the capital base of the system and supervise Russian financial institutions prudently. The latter issue is also urgent at the international level. Owing to the adoption of Basel II in Western Europe, national economic shocks are now less sharp, although not unavoidable. This means that the implementation of up-to-date risk-management technologies is an important issue for Russian regulators.

Russia still has about 900 banks, although the number of banks in Russia has dropped significantly since 1998. However, it is still impossible to regularly conduct field inspections.

The PD model is the cornerstone of an effective prospective remote supervision system and it can improve the quality of some of the regulations. For example, the proper estimation of risks in the banking sector can help regulators smooth the pro-cyclicality of capital requirements.

**3.** **PD-Model Development Background: Literature Review**

Approaches to developing Early Warning Systems (EWS) for the banking system, as well as determinants of bank financial distress in the developed world, have been investigated in numerous papers and are summarized in Bluhm *et al.* (2010). However, doing business in developing countries has a lot of peculiarities (Yigit and Behram, 2013). Mainly, this review will cover the experience of developing countries and Russia, and take the national banking industry peculiarities into consideration.

Foreign and Russian authors have addressed the issue of PD-model development for Russia. Among the most distinguished papers are Karminsky *et al.* (2005), Peresetsky (2010), and Peresetsky *et al.* (2011), whose ideas have been developed in this study.

Generally speaking, balance sheet structure and other financial characteristics of banks, such as bank size and capitalization, are the most meaningful predictors of defaults (Peresetsky *et al.* 2011). The former is usually measured as a bank's total assets on a logarithmic scale. There are many points in support of the crucial importance of this factor:

1. Bank’s total assets significant in the majority of models (Karminsky *et al.* 2005; Peresetsky *et al.* 2011).

2. In emerging markets, larger banks are better able to sustain and control the credit risk of long-term lending (Chernykh and Theodossiou, 2011).

3. Ceteris paribus, large banks have a higher insolvency risk than the small ones (Fungacova and Solanko, 2009).

4. Large banks with complex balance sheets are not always adequately disciplined, engaging in operations that are too complex for the CBR (Claeys and Schoors, 2007).

The second variable is capitalization and used to assess the capital adequacy of banks; it is calculated as a proportion of a bank's capital to its total assets. This index defines a bank's coverage of risks against its own resources. This is why a low capital adequacy ratio is suspicious from a regulator's point of view: banks with moderate capitalization shift future potential losses to clients and have a free-hand in taking excessive risk. However, debt financing has its merits: overcapitalized banks are usually run inefficiently, which may result in more non-working assets (Tabak *et al.* 2011). Consequently, a U-shaped relationship between PD and the bank capital adequacy ratio is expected. According to the EWS for the Russian banking sector in Lanine and Vennet (2006), greater capitalization of a bank diminishes its PD; whereas, bank size has no significant impact on its PD.

The next determinant of PD is a bank's liquidity position. Liquid assets are required to meet deposit outflows when they occur. Consistent with Lanine and Vennet (2006), the positive effect of liquidity exhaustion on the odds of default is theoretically and empirically confirmed. To capture liquidity risk, a ratio of non-government securities-to-bank assets was employed. The problem is to test whether or not a very liquid position worsens a bank's financial statement, as a result of lower profitability or higher market risks incurred.

Z-score is another well-founded predictor of PD. Z-score measures a slump in returns sufficient to erode a bank's equity. In Fungacova and Solanko (2009), this index was interpreted as an integrated proxy for a bank's insolvency probability.

In line with the literature, the inclusion of macroeconomic and institutional factors improves the model’s performance (Karminsky *et al.* 2005). Quarterly GDP growth rates and the Consumer Price Index are often used to take the macroeconomic features of a bank's operating environment, which are early predictors of a banking crisis, into account. In Bock and Demyanets (2012), the authors examined the determinants of non-performing loans in developing countries with panel data analysis. Their results underscored the significance of the GDP growth rate for empirical research in banking. In Mannasoo and Mayes (2009), this variable was regarded as a forward-looking factor for future bank insolvency. Parameters reflecting the stage of the economic cycle were discussed in Karminsky *et al.* (2005). They came to the conclusion that GDP growth rate, export-to-import ratio, and conditions of trade, are among the most reliable predictors of bank failure in the long-run.

**Figure 2. Number for foreign-owned banks and ones with significant foreign ownership**

The other important factor is institutions; many researchers define bank ownership type (Figure 2) as the dominant factor for its performance. Hanafi (2013) finds that banks with high ownership concentration are relatively more stable. While Fungacova and Solanko, 2009 concluded that foreign-owned banks show relatively high PDs, Micco *et al.* (2007) showed that foreign banks achieve better operational results than others in developing countries.

Clarke *et al.* (2005) reveal three principal factors that negatively-affect banks' stability scores in the developing world. Firstly, an agency problem is inevitable in the case of governmental bank management. In addition, politicians often interfere with internal procedures of banks to influence the economy in a desirable way, particularly before elections. Moreover, state banks lack a competitive market; this means that they are artificially protected from pure competition.

Furthermore, Micco *et al.* (2007) explained that state banks hire excess employees, carry vast administrative expenses, and are less profitable than the others. Nevertheless, governments always support a state bank in cases of financial distress. These banks also traditionally enjoy wider access to the interbank lending market. Unfortunately, it is impossible to quantify the effect of state ownership on PD since there are no defaults of banks with considerable government participation in capital (more than 50% owned by the government, according to Vernikov, 2011). 100% foreign-owned banks are in a similar situation: not a single default occurred over the period from 1998 to 2011. The slight drop in the number of entirely foreign-owned banks after 2008 was due to consolidation in the sector, and not bank failures.

Yet, in conditions of growing political, social and economic instability worldwide, foreign banks are not in the clear. Presumably, the solution is to broaden the definition of a foreign bank, as banks with more than 50% of equity owned by non-residents. Unfortunately, relevant statistics are not available.

Russian banks changed their risk profile after the introduction of the Deposit Insurance System (DIS) in 2004 (Fungacova and Solanko, 2009). As a result, depositor motivation to monitor bank performance was discouraged, and banks adopted risky investment policies to increase profits and offered higher interest rates on savings to attract clients. This led to the problem of moral hazard in banking.

Another essential institutional factor of a bank’s PD is competition. Russia is a country with steep competitive heterogeneity (Anzoategui *et al.* 2012). Using a panel data model Fungacova and Weill (2009) found supporting evidence that banks with higher market power are more financially stable. The authors also claim that in a competition profile, the location of a banking business might also influence default occurrences.[[1]](#footnote-1)

**4. Data and Model**

*4.1. Default Definition*

The initial step to develop a PD model is to define default; however, there is no common opinion in the literature. In practice, the revocation of a bank license does not always mean a default due to a weak financial position; illegal financial operations, or “book cooking,” can also cause license withdrawals. Modeling license revocations due to legal reasons is examined in Peresetsky (2010).

The primary target of this paper is to address bank defaults due to poor financial performance and an inability to cover debts. So, in this paper, the following events are the indicators of default:

1. A bank's capital sufficiency level falls below 2%.

2. The value of a bank's internal resources drops below the minimum established at the date of registration.

3. A bank fails to reconcile the size of the charter capital and the amount of internal resources.

4. A bank is unable to satisfy creditors' claims and make compulsory payments.

5. A bank is subject to sanitation by the Deposit Insurance Agency (Bank Restructuring Agency) or another bank.

The next step was to create a list of the banks which defaulted over the period from 1998 to 2011. This information was collected from the official website of the CBR and other public sources.

*4.2. Sources of Bank-Specific Financial Information*

Three sources of financial data relevant to the Russian banking system have been analyzed: the Interfax database, Bureau van Dijk’s “BankScope” database, and Mobile's "Banks and Finances" database.[[2]](#footnote-2) Preliminary analysis confirmed the preeminence of the last database due to its wider time-horizon and satisfactory coverage of the Russian banking industry.

*4.3. Database Description*

We constructed a quarterly bank-specific financial database based on Mobile's information from 1998 to 2011. Earlier data seem to be spoiled by numerous tiny, fake (so-called "sleeping") banks, and chaos in the Russian banking system (Karminsky, 2010). In addition, monthly banking statistics might be less reliable than quarterly statistics, for accounting reasons. Raw data from "Banks and Finances" were collected in accordance with Russian accounting standards.

Over the 14-year period considered, there were 894 license revocations, 464 of which were defaults in compliance with the classification in Section 4.1; additionally, there were 37 bank sanitations. The bar chart in Figure 3 demonstrates the distribution of registered license revocations and defaults.

A quasi-panel data structure was employed to carry out the analysis; it can be easily transformed for building panel models. The set of raw variables used, with a description and summary statistics, is presented in Table 2.

To facilitate the data analysis, the sample was split into two parts. The first comprised observations before the beginning of 2010; this was used to develop models. The predictive power of the final PD models was tested on the second sample, with observations for the years 2010 and 2011.

**Figure 3. Withdrawals of licenses and bank defaults in the Russian banking sector: Q11998 – Q32011**

**Notes:** 1 – (Jan. 2000) – effect of 1998-1999 crisis has almost expired; 2 – (Jan. 2004) – the Deposit Insurance System was launched in Russia; 3 – (Sep. 2008) – 2008-2009 financial crisis started in Russia.

**Table 2. Raw variables from "Banks and Finances" Database and summary statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Var. symbol | Variable name | Obs. | Mean | Std. Dev. |
| BP | Balance profit | 59270 | 139057.6 | 2749515 |
| TA | Total assets | 59391 | 1.39e+07 | 1.90e+08 |
| NP | Net profit | 58461 | 102745.2 | 2242601 |
| GO | National and local government obligations | 32556 | 1000707 | 1.66e+07 |
| LE | Total loans to the economy | 57608 | 6798499 | 1.02e+08 |
| LA | Liquid assets | 59276 | 2814943 | 3.15e+07 |
| NGS | Non-government securities | 48854 | 1104642 | 1.1e+07 |
| OR\_q | Operational revenues | 40005 | 1092664 | 1.88e+07 |
| TCA | Turnover on correspondent accounts in commercial banks and the CBR | 55853 | 3.53e+07 | 3.15e+08 |
| OC\_q | Operational costs | 39753 | 1121457 | 2.08e+07 |
| RR | Required reserves in the CBR | 39845 | 146057.9 | 1839618 |
| DL | Demand liabilities | 59365 | 4172650 | 6.65e+07 |
| NPL | Non-performing loans | 47611 | 369878.6 | 5511744 |
| WA | Working assets | 58448 | 9228285 | 1.27e+08 |
| RES | Loss reserves | 58342 | 597177.8 | 9699533 |
| PE\_q | Personnel expenses | 40004 | 42522.98 | 552660.7 |
| Eq | Capital | 59629 | 1596060 | 2.29e+07 |
| DI\_30 | Less than 30 days of deposits of individuals | 54222 | 551249.4 | 9284596 |

**Note:** mean values are expressed in thousands of Russian rubles; in accordance with Russian accounting standards.

*4.4. Binary Choice Model: Logit Analysis*

This section briefly illustrates how to use a logit model to predict defaults. The binary dependent variable default equals one if an observation is classified as insolvency, and zero if otherwise. The model is able to estimate a bank's PD directly in the form:

$$P\left(default=1\right)=Λ (x\*β),$$

*s. t.*

$$\left\{\begin{array}{c}P(default=1) \rightarrow 1 if x\*β\rightarrow \infty \\P\left(default=1\right)\rightarrow 0 if x\*β\rightarrow -\infty \end{array}\right.$$

where $Λ(x\*β)$= $\frac{exp⁡(x\*β)}{1+exp⁡(x\*β)}$ is a function taking values between 0 and 1; *x* is a vector of *n* regressors (i.e., $x\*β=β\_{0}+β\_{1}\*x\_{1}+β\_{2}\*x\_{2}+…+β\_{n}\*x\_{n}$).[[3]](#footnote-3)

**5. Empirical Model Estimation**

*5.1. Choice of Financial Regressors for the Initial Model*

The database created could contain some measurement errors or inaccurate observations (Table 2). To mitigate the problem, a clearing algorithm was implemented, which excluded suspicious observations for the class of “alive” banks:[[4]](#footnote-4)

1. Observations in the database with negative values for total loans to the economy (LE), total assets (TA) and capital (EQ).

2. The 1st and 99th percentiles of observations for each of the relative variables, listed in Table 3, to avoid statistical outliers. These ratios seem to be significant to determine a bank's PD, as indicated by the literature review and common sense.

An automatic classification of the independent variables was applied to test the separating power of the variables. To avoid statistical problems, variables with high correlation (over 0.3 in absolute value) were excluded from those with prominent separating power, as well as factors with insufficient or unequally-distributed observations in order to mitigate the multicollinearity problem.

**Table 3.** **An automatic classification: one-way analysis of variance for relative variables to determine whether means for both “alive” and “insolvent” groups are equal**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **№** | **Variable name** | **Var.****symbol** | **The formula for calculation** | **P–value in ANOVA test**  |
| 1 | Capital-to-Total assets ratio | eq\_ta | eq/ta | 0,031 |
| 2 | Balance profit-to-Total assets ratio | bp\_ta | bp/ta | 0,042 |
| 3 | Liquid assets-to-Total assets ratio | la\_ta | la/ta | 0, 007 |
| 4 | Non-government securities-to-Total assets ratio | ngs\_ta | ngs/ta | 0,004 |
| 5 | Non-performing loans-to-Total loans to the economy ratio | npl\_le | npl / le | 0,008 |
| 6 | Turnover on correspondent accounts-to -Total assets ratio | ln\_tca\_ta | ln (tca / ta) | 0,065 |
| 7 | Liquid assets-to-Demand liabilities ratio | la\_dl | la / dl | 0,109 |
| 8 | Logarithm of Total assets  | lnca | ln(ta) | 0,079 |
| 9 | Operational revenues-to-Net profit ratio | or\_ np | or / np | 0,165 |
| 10 | Net profit-to-Total assets ratio | сp\_ta | сp / ta | 0,078 |
| 11 | Liquid assets-to-Non-government securities ratio  | la\_go | la / go | 0,123 |
| 12 | National and local government obligations-to-Total assets ratio | go\_ta | go / ta | 0,324 |
| 13 | Working assets-to-Total assets ratio | wa\_ta | wa / ta | 0,168 |
| 14\* | Less than 30 days of deposits of individuals-to-Deposits of individuals ratio | vdfl30\_dfl | di\_30 / dfl | 0,069 |
| 15 | Personnel expenses-to-Operational costs ratio | pe \_oc | pe / oc | 0,654 |
| 16\* | Non-performing loans-to-Required reserves in the CBR ratio | npl\_rr | npl / rr | 0,098 |
| 17 | Loss reserves-to-Total assets ratio | res\_ta | res / ta | 0,023 |

**Note:** Variables with p-value in ANOVA test less than 0.1 are of high separating power.[[5]](#footnote-5)

\* indicates unevenly-distributed variables with a small number of observations.

In the next stage, we carried out a factor selection process to cover the main risks of banking in Russia.

1) Business Risk

Financial troubles immediately result in a sharp decline in a bank's capital. Furthermore, fluctuations in regulatory capital ratios for banks are monitored to check if they fall below the minimum requirement. This is why the Capital-to-Total assets ratio was exploited as a factor in the PD model. This ratio also shows how much “skin in the game” the bank has.

Profitability creates the economic value of a bank; no commercial company with permanent losses can be successful in the long-run. A typical profitability measure for all businesses, a *Balance profit-to-Total assets* ratio, was used.

The proportion of *Turnover on correspondent accounts-to-Total assets* reflects the level of economic activity in a bank. Lower values might indicate a bank’s inability to make transactions and incentives of managers to curtail business. A natural logarithm transformation was used to mitigate any possible multicollinearity problem.

2) Credit Risk

Asset quality is a dominant factor of future profits and losses calculated by the ratio of *Non-performing loans-to-Total loans* *to the economy*.

3) Market and Liquidity Risks

We used Non-government securities-to-Total assets ratio to assess both liquidity and market risks management carried in a bank.

Also, the *natural logarithm of Total assets* is a good measure of bank size. So, the initial model with ambiguous lag length of *n* quarters is:

|  |  |
| --- | --- |
| $$P(default=1)=Λ (eq\\_ta\_{t-n}; ln\\_ta\_{t-n}; npl\\_le\_{t-n} ;ln\\_tca\\_ta\_{t-n} ;bp\\_ta\_{t-n}; ngs\\_ta\_{t-n})$$ | ((1) |

The database used is highly unbalanced, which is destructive for logistic analysis. A class imbalance problem arises when one group of observations (in this case, "insolvency" observations) is underrepresented compared to another (in this case, "alive" observations). There are many methods to solve the problem, which are discussed in He and Edwardo (2009).

In this paper, we followed Hosmer and Lemeshow (2000). Firstly, one thousand subsamples were selected. Each contained all "insolvent" observations and enough “alive” observations to balance subsets. Then, a logistic model was estimated on a random subsample. Finally, derived coefficients were averaged and checked to be stable for the previously-created subsamples.

*5.2. Modeling*

Our main goal is to predict a bank's default. For this reason, using independent lagged-variables is appropriate. In this research, statistical criteria have been employed to assess the goodness-of-fit of the current models: the significance level of coefficients, pseudo R-squared value, the ROC curve comparison, specificity, sensitivity, and the proportion of correctly classified outcomes.

We found that the use of high lag length *n* in the initial model (eq. 1) decreases the model quality. Nevertheless, one-quarter lag is unacceptable for our goals: in that case the obtained PD would be the likelihood of a bank failure in a three-month period. The user of such model has no enough time to prevent the upcoming default event and its negative consequences. So two-quarter lags were applied. The basic model is a special case of initial model (eq. 1) with lag length of two quarters. Its specification is:

|  |  |
| --- | --- |
| $$P\left(default=1\right)= Λ (eq\\_ta\_{t-2}; ln \\_ta\_{t-2}; npl\\_le\_{t-2}; ln \\_tca\\_ta\_{t-2};$$$$bp\\_ta\_{t-2}; ngs\\_ta\_{t-2}).$$ | (2) |

In the next stage, we considered possible nonlinearity in the relationship by expanding the model by including all the factors from the previous model in powers up to three. Then, insignificant coefficients were dropped. As a result, a basic model with nonlinearities appeared:

|  |  |
| --- | --- |
| $$P\left(default=1\right)=$$$$Λ (eq\\_ta\_{t-2}; (eq\\_ta\_{t-2})^{2};ln⁡ \\_ta\_{t-2}; (ln⁡\\_ta\_{t-2})^{2}; npl\\_le\_{t-2}; $$$$ln⁡\\_tca\\_ta\_{t-2}; bp\\_ta\_{t-2};(bp\\_ta\_{t-2})^{2};ngs\\_ta\_{t-2}).$$ | (3) |

Next, a model improvement process was initiated as we described the process in Figure 4.

|  |  |  |
| --- | --- | --- |
| **Stage** | **Means of improvement** | **Output model name** |
| 1 | **Time factor:** use of two groups of time dummies for quarters (I-IV) *d\_qx* and years (1998-2009) *d\_xx.* | Basic with nonlinearities and time factor. |
|  |
| 2 | **Macroeconomic parameters:** use of quarterly GDP growth rates *dgp\_gr* and Consumer Price Index *cpi* in order to account for the effect of the macroeconomic environment on bank performance. | Basic with nonlinearities, time factor, and macro parameters. |
|  |
| 3 | **Institutional parameters:** - use of the Lerner index *lindex* to consider the impact of monopoly power of the firm on its default probability.- use of a dummy variable *region* indicating Moscow location of a bank’s headquarters.- use of a dummy variable on a bank’s participation in the Deposit Insurance System. | Basic with nonlinearities, time factor, macro parameters, and institutional variables. |

**Figure 4. A model improvement process: time dummies, macroeconomic variables, and institutional factors**

An LR-test confirmed the adequacy of inclusion for every set of factors. After each iteration, the most insignificant variables were excluded from the model.

Table 4 provides the estimation results. If there is insufficient information to develop the final model, a potential user may successfully employ one of these.

*5.3 Testing Model Stability and Comparison with Alternatives*

The model is said to be stable when the estimated coefficients are not sensitive to minor changes in observations used. Additionally, the results should be approximately the same across smaller subsamples. The final model (eq. 4), is stable enough to avoid the problem of over-fitting, according to estimates in Table 5:

|  |  |
| --- | --- |
| $$P\left(default=1\right)=$$$$Λ (eq\\_ta\_{t-2} ; (eq\\_ta\_{t-2})^{2};ln⁡\\_ta\_{t-2} ; (ln⁡\\_ta\_{t-2})^{2} ;npl\\_le\_{t-2}; $$$$ln⁡\\_tca\\_ta\_{t-2}; bp\\_ta\_{t-2};\left(bp\\_ta\_{t-2}\right)^{2}; ngs\\_ta\_{t-2}; d\_{2009}; d\_{q1};$$ | (4) |

$$gdp\\_gr\_{t-2};cpi\_{t-2}; l\_{index};region)$$

|  |  |  |  |
| --- | --- | --- | --- |
| Model name | (eq. 2)Basic | (eq. 3)Basic with nonlinearities | (eq. 4)Basic with nonlinearities, time factor, macro parameters, and institutional variables |
| $$eq\\_ta\_{t-2}$$ | -0.55 | -9.88\*\*\* | -12.24\*\*\* |
| $$(eq\\_ta\_{t-2})^{2}$$ |  | 14.55\*\*\* | 16.03\*\*\* |
| $$ln \\_ta\_{t-2}$$ | -0.13\*\* | -1.15\* | -1.86\*\* |
| $$(ln \\_ta\_{t-2})^{2}$$ |  | 0.04\* | 0.06\*\* |
| $$bp\\_ta\_{t-2}$$ | -11.5\*\*\* | -70\*\*\* | -57\*\*\* |
| $$(bp\\_ta\_{t-2})^{2}$$ |  | 964\*\*\* | 1031\*\*\* |
| $$ngs\\_ta\_{t-2}$$ | 3.99\*\*\* | 4.54\*\*\* | 3.11\*\*\* |
| $$npl\\_le\_{t-2}$$ | 6.38\*\*\* | 4.52\*\*\* | 5.17\*\*\* |
| $$ln \\_tca\\_ta\_{t-2}$$ | -1.19\*\*\* | -1.06\*\*\* | -1.2\*\*\* |
| $$d\\_2009$$ |  |  | 2.41\*\*\* |
| $$d\\_q1$$ |  |  | -1.51\*\*\* |
| $$gdp\\_gr\_{t-2}$$ |  |  | 0.12 |
| $$cpi\_{t-2}$$ |  |  | 0.11\*\*\* |
| $$l\\_index$$ |  |  | -2.51\*\*\* |
| $$region$$ |  |  | 2.91\*\*\* |
| Pseudo R2 | 0.5219 | 0.59 | 0.6403 |
| ROC area | 0.8936 | 0.9157 | 0.9697 |
| Sensitivity | 72.30% | 75.90% | 79.14% |
| Specificity | 97.20% | 97.68% | 96.96% |
| Correctly classified | 92.67% | 93.72% | 93.72% |

**Table 4. Estimation Results**

**Note:** Asterisks indicate the level of significance as [\*\*\*] – 1%; [\*\*] – 5%; and [\*] – 10%.

In addition, we tested the robustness of the model (4) to gaps in the data. Our results indicate that gaps in the data impact the banks' default probabilities insignificantly. Coefficients for the final model were averaged on 1,000 subsamples. Tables 5 and 6 shows the estimation and predictive accuracy results for model (4), dependent on the threshold. As mentioned previously, the out-of-sample prediction power was estimated in a subsample with observations for the years 2010 and 2011. To evaluate the quality-of-prediction, two criteria were applied: the quarterly average number of banks in a risk group (predicted to face insolvency within half a year) and the number of correctly-predicted defaults. Overall, 19 banks collapsed in 2010 and 2011.

The out-of-sample prediction performance of the final model is prominent. According to our results, the level of 30% is an optimal threshold value: the size of a risk group is reasonable, and 12 out of 19 bank failures were correctly predicted. In practice, the choice of a separating criterion is based on a user's capacity to accurately inspect banks in the risk group.

**Table 5. Estimation results for the model (4) on 1,000 subsamples. Testing the problem of over-fitting**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Sign of coefficient** | **Level of significance** |
| 1% | 5% | 10% | >10% |
| $$eq\\_ta\_{t-2}$$ | - | 599 | 396 | 5 | 0 |
| $$(eq\\_ta\_{t-2})^{2}$$ | + | 188 | 785 | 27 | 0 |
| $$ln \\_ta\_{t-2}$$ | - | 258 | 627 | 115 | 0 |
| $$(ln \\_ta\_{t-2})^{2}$$ | + | 38 | 295 | 667 | 0 |
| $$bp\\_ta\_{t-2}$$ | - | 1000 | 0 | 0 | 0 |
| $$(bp\\_ta\_{t-2})^{2}$$ | + | 205 | 623 | 169 | 3 |
| $$ngs\\_ta\_{t-2}$$ | + | 2 | 83 | 670 | 245 |
| $$npl\\_le\_{t-2}$$ | + | 700 | 213 | 86 | 1 |
| $$ln \\_tca\\_ta\_{t-2}$$ | + | 1000 | 0 | 0 | 0 |
| $$d\\_2009$$ | + | 1000 | 0 | 0 | 0 |
| $$d\\_q1$$ | - | 119 | 667 | 187 | 27 |
| $$gdp\\_gr\_{t-2}$$ | + | 0 | 0 | 11 | 989 |
| $$cpi\_{t-2}$$ | + | 46 | 937 | 17 | 0 |
| $$l\\_index$$ | - | 121 | 634 | 245 | 0 |
| $$region$$ | + | 1000 | 0 | 0 | 0 |

**Table 6. Out-of-sample predictive accuracy of the model (4)**

|  |  |  |
| --- | --- | --- |
| Threshold: a bank with PD over *x* is a candidate to fail | Quarterly average size of a risk group | Number of correctly-predicted defaults, of 19 (proportion). |
| *x* = 10% | 54 | 16 (84%) |
| *x* = 20% | 34 | 12 (63%) |
| *x* = 30% | 30 | 12 (63%) |
| *x* = 40% | 28 | 10 (52%) |

In this study, we have explored the benefits of panel data analysis. The sample of Russian banks used is close to the population. This is why a fixed-effect logit model with the final specification (eq. 4) was applied. Surprisingly, no predominance of the panel logit model was observed. Moreover, a Hausman test for fixed-effect logit versus simple logit confirmed our findings.

An attempt to improve the final specification with the Z-score measure also failed. Two of the three main components of this index, a bank's return on assets and capitalization, are presented in the specification as the Balance profit-to-Total assets and Capital-to-Total assets ratios, respectively. Presumably, this is the reason for the insignificance of the Z-score.

**6. Estimation Results**

As described, the final model’s estimated coefficients are given in Table 5. It is important to note that the regression coefficient sign is useful to judge the influence of the relevant variable on a probability of default:

$\frac{∂Λ\left(f(x)\right)}{∂xi}\*∆x=λ\left(f\left(x\right)\right)\*f'\left(x\right)\*∆x, where λ\left(f\left(x\right)\right)>0$. (5)

|  |  |
| --- | --- |
|  |   |

So, we now interpret the estimation results for each of the factors, ceteris paribus.

*6.1. Financial Bank-specific Ratios*

Capitalization: Capital to total assets ratio

According to the estimation results, over- and under-capitalized banks exhibit higher default probabilities (Figure 5) which is consistent with the expectations.

Profitability: Profit to total assets ratio

Banks with extremely high or low profitability score higher default rate risks. Naturally, poor banks lack sufficient funds to meet obligations. A bank with unusually high earnings could take excessive risk, which leads to an increase in PD. Moreover, in efficient markets, it is impossible to maintain outstanding profitability without bearing commensurate financial risk.

**Figure 5. Impact of eq\_tat-2 ratio on default probability:**

**f(eq\_tat-2)=**$ -12.2\*$ **eq\_tat-2+ 16.0\*(eq\_tat-2)2**

Bank size: Natural logarithm of total assets

Small, as well as large, banks have a higher risk of insolvency. So, the "too big to fail" thesis does not hold in our paper. It is important for researchers to bear in mind that, without nonlinearity in the final model, the factor is not significant at all.

Credit quality: Non-performing loans to total loans in the economy (economy (given to individuals, industry, financial organizations etc.) ratio

Banks with a considerable amount of bad debt are less stable, as supposed.

Operational activity: Natural logarithm ofturnover on correspondent accounts to total assets ratio

Our main regression results demonstrate a negative correlation between а PD and а bank's operational activity. Hypothetically, lower Turnover on correspondent accounts in comparison with Total assets indicates a bank's inability to process payments and incentives for managers to curtail business.

Liquidity and market risks: Non-government securities to total assets ratio

Banks with a higher proportion of corporate securities in assets carry a higher risk of a price slump in the market. Indeed, substantial investments in non-government securities might have no relation to liquidity management; it is probably the result of an aggressive investment policy, which causes higher PDs.

*6.2. Time Factors*

Quarterly dummies: The only significant quarterly dummy variable indicates that, on average, the PD is lower in the first quarter of a year. Our conjecture is that this is closely associated with the regulator's incentive to finish current investigations by the end of a year and start a new cycle from January.

Annual dummies: The developed model underestimates default probabilities for the year 2009. On the one hand, the result reveals some unrecorded channels that significantly increased risks in the period of the recent financial crisis in 2009, e.g., the dependence of the Russian banking sector on funding from abroad. On the other hand, the model is adequate for the banking crisis in Russia in 2004, and even for the recession in 1998-1999. In other words, the model is able to predict crises.

*6.3. Macroeconomic Parameters*

Quarterly GDP growth rates: Unexpectedly, this variable is not significant. It is likely that we should have applied not quarterly, but annual GDP growth rates. Additionally, the second macroeconomic variable could distort the estimation results (Karminsky *et al.* 2005). Alternatively, the financial ratios could have absorbed the impact of a business cycle on the default probability.

Consumer price index: A growing consumer price index, which accelerates inflation, increases a bank's default probability. Inflation reduces the real returns on loans. At the same time, depositors are able to withdraw money, and put it into the bank again at a higher interest rate or spend it; consequently, banks suffer.

*6.4. Institutional Variables*

Lerner index: In line with the literature review, banks with higher monopoly power are more financially stable compared to others, due to lower market pressure.

Location dummy: Moscow-based banks have higher PDs on average, which contradicts the findings in Fungacova and Solanko (2009). According to Claeys and Schoors (2007), Russian banking regulators are reluctant to withdraw licenses of banks outside the Moscow region, so as not to weaken competition in local markets.

We found evidence that bank participation in the Deposit Insurance System does not influence its PD. The explanation for this is that the set of Deposit Insurance System participants is too diversified.

**7. Conclusion**

In this paper, we developed a PD model for the Russian banking system. We confirmed that bank-specific financial statistics, together with macroeconomic and institutional data, provide invaluable information to predict the PD of a bank. We found that banks with a significant proportion of overdue loans or commercial securities in their portfolio have a higher PD. In addition, there is an optimal level of profitability and capitalization of Russian banks from the point of view of stability. Moscow banks face higher PDs, on average, as a result of severe competition and tougher supervision by regulators. Banks with higher monopoly power have a lower PD. Also, high inflation harms the stability of Russian banks.

Over 60% of bank failures were correctly classified in the out-of-sample prediction power tests for the years 2010 and 2011. We believe that the model meets the expectation of the CBR and can be used to identify suspicious banks for proper monitoring. Banks with a weak balance structure, operating in hostile macroeconomic and institutional environment, should draw the regulator’s attention.

A potential user should be aware of three pitfalls of our approach. First, as shown in Table 6, the size of the identified “risk group” of banks rises much faster than the number of correctly-predicted defaults by our model. This means that any attempt to identify more defaults will cause an exponential growth of expenses to monitor the expanded group of the weakest banks in the sector. The potential user should compare losses from an unpredicted default event to the costs of better supervision and find a balance between them. Second, state and foreign banks have never failed in Russia. An assessment of their financial stability should be carried out separately, and is beyond the scope of our research. Third, there is still significant information to predict bank defaults that is not addressed in our paper, such as the conflict of bank stockholders or relationships with local government officers. It is almost impossible to collect data for the entire Russian banking sector. A possible solution is the use of bank ratings to estimate its financial position (Hainsworth, *et al.* 2013). This is a promising direction for our future research.

**References**

Anzoategui, D., Peria, M., and Melecky, M., 2012. Bank competition in Russia: An examination at different levels of aggregation. *Emerging Markets Review*, 13(1), pp.52-53.

Bluhm, C., Overbeck, L., and Wagner C., 2010. *Introduction to credit risk modeling*. London: Chapman and Hall/CRC.

Bock, R. and Demyanets, A., 2012. Bank asset quality in emerging markets: Determinants and spillovers. *IMF Working Papers,* 12/71.

Chernykh, L. and Theodossiou, A., 2011. Determinants of bank long-term lending behavior: Evidence from Russia. *Multinational Finance Journal*. 15(3/4), pp.193-216.

Claeys, S. and Schoors, K., 2007. Bank supervision Russian style: Evidence of conflicts between micro- and macro-prudential concerns. *Journal of Comparative Economics*, 35(3), pp.630-657.

Clarke, G., Cull, R., and Shirley, M., 2005. Bank privatization in developing countries: A summary of lessons and findings. *Journal of Banking and Finance*, 29(8/9), pp.1905-1930.

Fungacova, Z. and Solanko, L., 2009. Risk-taking by Russian banks: Do location, ownership and size matter? *BOFIT Discussion Papers,* 21/2008.

Fungacova, Z. and Weill, L., 2009. How market power influences bank failures: Evidence from Russia. *BOFIT Discussion Papers,* 12/2009.

Greene, W., 2007. Econometric analysis. Sixth Edition. New Jersey: *Prentice Hall.*

Hainsworth, R., Karminsky, A., Solodkov, V., 2013. Arm's length method for comparing rating scales. *Eurasian Economic Review,* 3(2), 2013, pp. 114-135

Hanafi, M., Santi, F., Muazaroh, 2013. The impact of ownership concentration, commissioners on bank risk and profitability: evidence from Indonesia. *Eurasian Economic Review,* 3(2), 2013, pp. 183-202

He, H. and Edwardo, A., 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp.1263-1284.

Hosmer, D. and Lemeshow, S., 2000. *Applied logistic regression*. New York: John Wiley and Sons.

Karminsky, A., Peresetsky, A., and Petrov, A., 2005. *Рейтинги в экономике: методология и практика* [*Ratings in economics: Methodology and practice*]. Moscow: Finansi i Statistika.

Karminsky, A., 2010. Повышение устойчивости и эффективности банковской системы [Improving stability and efficiency of the Russian banking system], in V. Polterovich Стратегия модернизации российской экономики [Modernization strategy for Russian economy], Saint Petersburg, Aletheia.

Lanine, G. and Vennet, R., 2006. Failure prediction in the Russian bank sector with logit and trait recognition models. *Expert Systems with Applications*, 30(3), pp.463-478.

Mannasoo, K. and Mayes, D., 2009. Explaining bank distress in Eastern European transition economies. *Journal of Banking and Finance*, 33(2), pp.244-253.

Micco, A., Panizza, U., and Yanez, M., 2007. Bank ownership and performance. Does politics matter? *Journal of Banking and Finance*, 31(1), pp. 219-241.

Peresetsky, А., 2010. *Модели причин отзыва лицензий российских банков* [*Modeling license revocations of Russian banks*] / *Preprin*t #WP/2010/085. New Economic School.

Peresetsky, A., Karminsky, A., and Golovan, S., 2011. Probability of default models of Russian banks. *Economic Change and Restructuring*, 44(4), pp.297-334.

Tabak, B., Craveiro, G., and Cajueiro, D., 2011. Bank efficiency and default in Brazil: Causality tests. *The Central Bank of Brazil Working Paper Series,* 253.

Totmyanina, K., 2011. Обзор моделей вероятности дефолта [Review of Probability of Default Models]. *Journal of Financial Risk- Management*, 1(25), pp.3-17.

Vernikov, A., 2011. Government banking in Russia: Magnitude and new features. *IWH Discussion Papers*, 13.

Yigit, I., Behram N., 2013. The relationship between diversification strategy and organizational performance in developed and emerging economy. *Eurasian Business Review*, 3(2), 2013, pp. 121-136.

1. A detailed overview of PD model types is beyond the scope of the literature review. Please see Totmyanina (2011) for a more detailed explanation. [↑](#footnote-ref-1)
2. Interfax database: http://www.spark-interfax.ru, “BankScope” database: https://bankscope2.bvdep.com, "Banks and Finance" database has no official description in the Internet. [↑](#footnote-ref-2)
3. For additional details on this topic, please see Greene (2007) [↑](#footnote-ref-3)
4. There were no corrections for observations of the "default" class. In this case, extraordinary values could be caused by a weak financial position of a bank. [↑](#footnote-ref-4)
5. Analysis of variance (ANOVA) test is used to analyze the differences between group means and variances for two groups: in our case of healthy and insolvent banks. A test determines statistical significance of some variable to distinguish between two classes of observations. Higher level of significance in the test for some variable means its higher separating power (ability to distinguish between healthy and insolvent banks). [↑](#footnote-ref-5)