NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Alexander Karminsky, Alexander Kostrov,
Taras Murzenkov

COMPARISON OF DEFAULT PROBABILITY MODELS: RUSSIAN EXPERIENCE

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: FINANCIAL ECONOMICS
WP BRP 06/FE/2012

SERIES: FINANCIAL ECONOMICS

Alexander M. Karminsky[1], Alexander V. Kostrov[2], Taras N. Murzenkov[3]

COMPARISON OF DEFAULT PROBABILITY MODELS: RUSSIAN EXPERIENCE[4]

Under the Basel II accord, improving probability of default models is a key risk-management priority. There are four main aspects of this research: suggesting the bank default classification; using a wide time horizon (quarterly Russian banking statistics from 1998 to 2011); investigating the macroeconomic and institutional characteristics of the banking sector environment and finally, testing the accuracy of the models developed.

We have employed nonlinearity and automatic classification of the independent variables in our models, paying attention to the structure of the banking market as well as to the reliability of the models developed. We have compared several models for estimating default probabilities. From the results of this comparison, we have chosen the binary logit – regression with quasi panel data structure. Our key findings are:

- There is a quadratic relationship between bank's capital adequacy ratio and its probability of default.
- The "too big to fail" hypothesis does not hold for the Russian banking sector.
- There is a negative relationship between the Lerner index and bank's PD.
- Macroeconomic, institutional and time factors significantly improve the model quality.

We believe that these results will be useful for the national financial regulatory authorities as well as for risk-management in commercial banks. Moreover, we think that these models will be valuable for other emerging economies.

JEL classification: G21, G24, G32.

Key words: probability of default, PD, banks, Russia, risk-management, default classification.

[1] National Research University Higher School of Economics, Professor, Doc. (Ec., Tech.), karminsky@mail.ru

[2] National Research University Higher School of Economics, Graduate student, kostrov.alexander.v@gmail.ru

[3] National Research University Higher School of Economics, Graduate student, murzenkovtn@gmail.com

[4]This paper is an output of a project №05-0030 «Analysis of credit institutions insolvency» by the scientific research group "Modeling defaults of banks" at the National Research University Higher School of Economics (HSE).

Introduction

This paper reviews binary choice models that attempt to describe, predict and prevent the defaults of financial institutions, namely banks, with regard to the Russian banking industry specifically. The probability of default (PD) is the likelihood of a bank failure over a fixed assessment horizon. To forecast a bank's default probability a logistic model was applied based on CAMELS methodology together with an adequate proxy selection.

Numerous papers dedicated to the probability of default model (PD model) creation studied circumstances associated with recessions in developed countries, not in transitional ones. However logistic modeling is inherently more reliable for stable periods of economic development. The aim of this research is to propose an adequate forward-looking model, which rests on the relationship between banks' default rates and macroeconomic or financial banking statistics.

The model can be applied by three main parties in Russia:

- First, the model is valuable for the government, particularly the Central Bank of Russia. The approach developed in this study can diversify the existing estimates of bank performance. For example, by employing the model the Central Bank of Russia can identify the most vulnerable banks and supervise them.

It can also provide some ideas to improve regulations, e.g. PD model development is an adequate strategy of risk-management, which can help the Regulator to smooth procyclicality of capital requirements.

- Second, banks can use it to predict and prevent hardships. The model meets the requirements of Basel II and takes into account the dependence of credit risk on the business cycle. It can be exploited for internal financial monitoring in a commercial bank.

- Third, the information about bank's default probability is essential for bank creditors and business partners.

To cover the topic of this paper we have taken the following steps:

- The collection of financial, macroeconomic and institutional data about banks and their operational environment for the period from 1998 to 2011.
- The selection of the relevant factors to assess a bank's default probability.
- The econometric modelling.
- A statistical comparison of models.
- The exploration of the benefits of using panel data structure.
- A discussion of the results of the study concerning the relationships between PDs and independent factors used in the models.

The rest of the paper is organized as follows:

The first section contains an overview of the Russian banking sector. After that, Section 2 provides a brief literature review of PD model development. Section 3 describes the database and sources used. Section 4 reports the methodology of the PD model creation. A comparison of the derived model with alternatives is addressed as well. Then Section 5 discusses the model estimation results. The final Section concludes.

1. The Russian banking system: stability issues

In the late 1980s commercial banking revived in Russia. More than 2500 banks had been launched by 1995. Overall, about 3500 charters of incorporation have been issued by the Central Bank of Russia by the present moment.

Table 1.

Number of credit institutions (CIs) in the Russian banking system: yearly, 1997 – 2012

№	Characteristics	Time period									
		1997	1998	2000	2002	2004	2006	2008	2010	2011	2012
1	Total number of CIs registered by Bank of Russia	2552	2481	2124	1826	1516	1345	1228	1146	1117	1102
2	of which: banks	2526	2451	2084	1773	1464	1293	1172	1084	1055	1036
3	of which: wholly foreign-owned CIs	16	18	22	27	33	52	77	80	77	76
4	Total number of banks with license to conduct banking operations	1654	1447	1274	1282	1249	1143	1058	965	925	905
5	of which: with general license	262	263	244	293	311	287	298	283	273	271
6	of which: with license granting them the right to take deposits of individuals	1589	1372	1239	1202	1165	921	886	819	799	786
7	CIs whose banking license has been revoked (cancelled)	852	1004	806	491	218	155	117	132	135	136
8	Total number of CIs liquidated as legal entities	408	468	869	1238	1569	1758	1900	1991	2023	2043
9	of which: CIs liquidated due to re-organization	319	326	340	357	367	389	402	433	448	454

Notes: Data as at December 1 of each period with the exception of 2012 (August 1)
Source: Central Bank of Russia

Two periods of mass license withdrawals in the Russian banking sector can be distinguished: the first, an intensive revocation period from 1996 to 2000 and the second, a rapid one between years 2008 and 2010, which were related to the financial crises of 1998 and 2008 respectively. Table 1 provides some information to judge the scale of downturns in the Russian economy and the regulatory responses from The Central Bank of Russia. By the 2000s, the surge in the number of credit institutions had ended as a result of the financial crisis in 1998.

Since 2000 there has been intensive growth in the Russian banking system. 35 Russian banks were among the Top 1000 World Banks (by total assets) in 2008 that is substantially higher compared to the beginning of the millennium, when there were only 20. Although close in size to bank assets in BRIC countries (by 0,5% of the Top 1000 World Banks' total assets each except for Brazil with 5% recorded), the Russian banking industry (by assets) is tiny in comparison with any of the Top 20 World Banks.

The Russian banking sector has passed through two stages of development with crises bounding them. As a result, the modern Russian banking industry effectively fulfills its financial intermediary functions.

Stage 1. Formation: 1989-1999.

This stage was 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 is typical for this period. There was an upward trend in the *Bank assets to GDP* ratio with 36% recorded in 2000 and 67% by 2009. A similar pattern is evident in the *Credit to GDP* ratio which made up slightly less than 50% at the end of the period with diversification into lending to individuals. Also, most of the systematic problems in the Russian banking system were resolved.

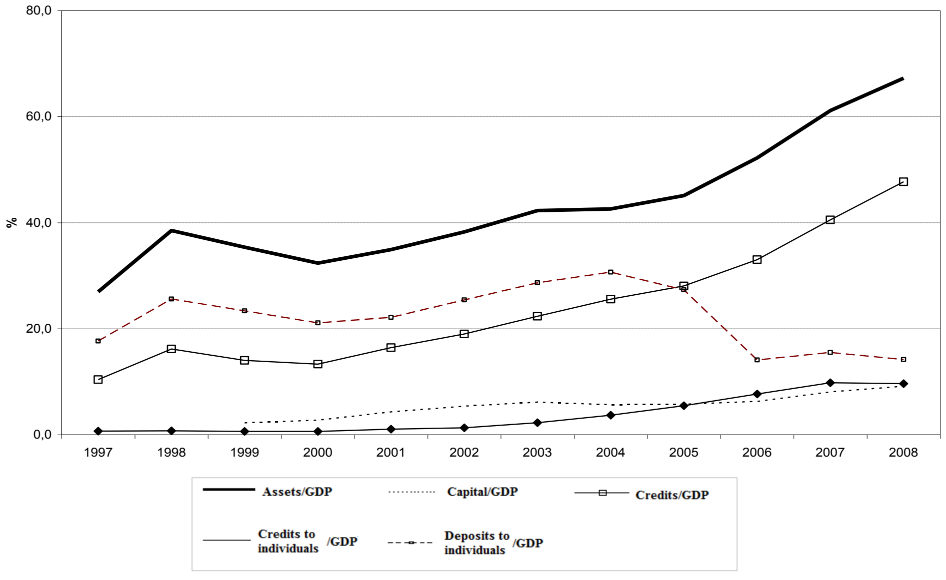

Diagram 1. Russian banking system relative to GDP

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

Stage 3. Sustainable growth: since 2010.
Throughout this period a great emphasis has been put on the proportional development of the Russian banking system. It is vital to find a way out of Dutch disease, reinforce the capital base of the system and supervise Russian financial institutions prudently. The last issue is also urgent at the international level. Owing to the adoption of Basel II in Western Europe, national economic shocks are less sharp, although not avoided. It means that the implementation of up-to-date risk-management technologies is a priority for Russian regulators, but this implementation has hit institutional, methodological and organizational barriers.

Russia still has 978 credit institutions, although the number has almost halved over the period from 1998 to 2012. Consequently, it is impossible to conduct field inspections regularly. That is why The Central Bank of the Russian Federation needs remote systems to monitor the national banking sector. The most vulnerable banks should be identified and supervised properly to make the Russian banking system more stable.

The probability of default model as an instrument of the distant supervising systems can also improve some regulations. For example, capital requirements ought to be risk-sensitive. It means that these should be stricter when risks are higher. In practice capital requirements are increased during recessions and decreased during upturns. In other words the capital requirements are procyclical. Obviously this contradicts the core idea of the Regulator. The application of an average PD in a banking system for risk measurement can help the Regulator to smooth procyclicality of capital requirements.

2. PD model development background: literature review

Approaches to developing Early Warning Systems (EWS) for the banking system as well as determinants of bank's financial distress in the developed world have been scrutinized in numerous papers and are summarized in Bluhm et al. (2010). Mainly, this review will cover experience of developing countries and Russia, thus the national banking industry peculiarities will be taken 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); 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 their defaults Peresetsky et al. (2011). The first is usually measured as the logarithm of bank's total assets. There is a number of points in support of the crucial importance of this factor:

- This variable is significant in the majority of models (see Karminsky et al. (2005); Peresetsky et al. (2011)).
- In emerging markets larger banks are better able to sustain and control the credit risk of long-term lending (see Chernykh & Theodossiou (2011)).

- Ceteris paribus, large banks a have higher insolvency risk than small ones (see Fungacova & Solanko (2009)).
- Large banks with intricate balance sheets are not always disciplined adequately, engaging in operations too complex for The Bank of Russia (see Claeys & Schoors (2007)).

The second variable is capitalization, used to assess capital adequacy of banks, calculated as a proportion of a bank's capital to its total assets. This index defines a bank's coverage of risks with its own resources. That is why a low capital adequacy ratio is suspicious from the 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 run inefficiently, which may result in more non-working assets (see Tabak et al. (2011)). Consequently U-shaped relationship between PD and bank capital adequacy ratio is expected.

According to the EWS for the Russian banking sector in Lanine & Vennet (2006), greater capitalization of a bank diminishes its PD while bank size has no significant impact.

The next determinant of PD is a bank's liquidity position. Liquid assets are required to meet deposits outflows when they take place. Consistent with Lanine & Vennet (2006), the positive effect of liquidity exhaustion on the odds of default is theoretically and empirically confirmed. To capture the liquidity risk a ratio of non-government securities to bank assets was employed. The task is to test whether 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. Statistically speaking, the Z-score shows the drop of returns measured in its standard deviation sufficient to erode bank's equity. In Fungacova & 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 performance (see Karminsky et al. (2005)). Quarterly GDP growth rates and Consumer Price Index are often used to take into account macroeconomic features of bank's operating environment, which are early predictors of a banking crisis. In Bock & Demyanets (2012) the authors examined the determinants of non-performing loans in developing countries with panel data analysis. Their results underscored the significance of GDP growth rate for empirical research in banking. In Mannasoo & Mayes (2009) this variable was regarded as a forward-looking factor of upcoming 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 a bank failure in the long-run.

The other meaningful factors are institutional ones. Many researchers define bank ownership type as the dominant factor of its performance. The study Fungacova & Solanko (2009) concludes that foreign-owned banks show relatively high PDs as well as domestic and large state-controlled banks.

In contrast, Micco et al. (2007) showed that foreign banks achieve better operational results than others in developing countries. Clarke et al. (2005) reveals three principal points that negatively affect banks' scores for stability in the developing world. Firstly, an agency problem is inevitable in the case of governmental bank management. In addition, politicians often interfere with banks internal procedures 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. Regardless of those facts, the Government always supports a state bank in cases of financial distress. Besides, these banks traditionally enjoy wider access to the interbank lending market. Unfortunately, it is impossible to quantify the effect of state ownership on PD on the grounds that there were no defaults of banks with considerable government participation in capital (more than 50% owned by the government, according to Vernikov (2011)).

The factor of absolute foreign participation in bank capital (100% of capital) is in a similar trap: not one default occurred over the period from 1998 to 2011. A slight drop in the number of entirely foreign banks after 2008 was due to consolidation in the sector, not bank failures.

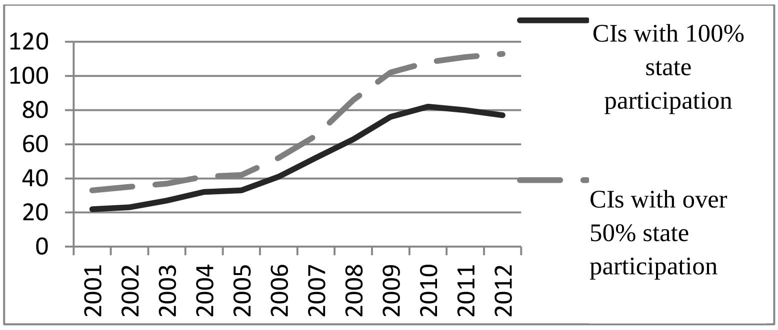

Diagram 2. The figures for credit institutions (CIs), entirely state-owned and with significant (over 50% in capital) government participation in the Russian banking sector: annually, 2001-2012

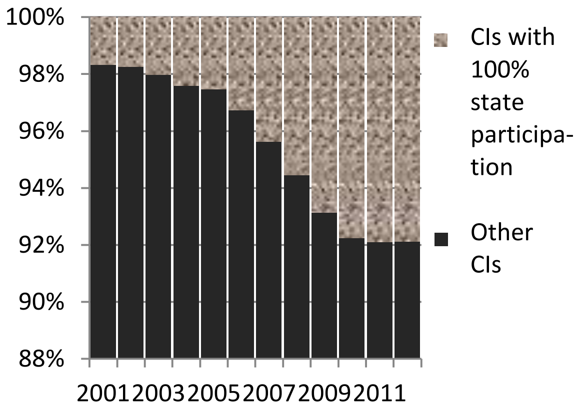

Diagram 3. Share of entirely state-owned credit institution as a percentage of total number of these in the Russian banking system

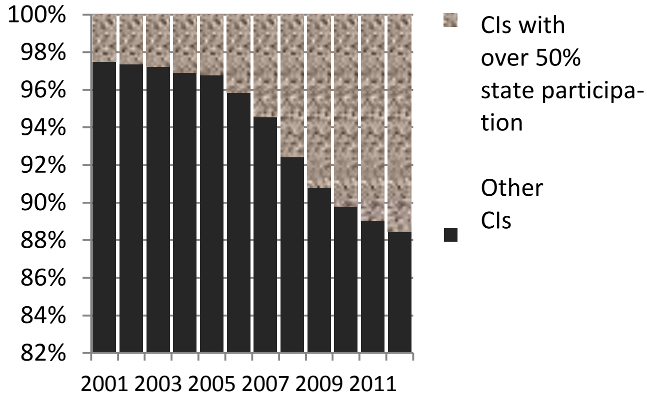

Diagram 4. Share of credit institution with significant government participation (over 50% in capital) as a percentage of total number of these in the Russian banking system

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

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

Another essential institutional factor is competition. Russia is an example of a territory with steep competitive heterogeneity (see Anzoategui et al. (2012)). As a rule, Herfindahl Hirschman index (HHI) and Lerner index (LI) are applied to convey the level of competition that exists within a market and monopoly power of a definite bank correspondingly. Fungacova & Weill (2009) with panel data analysis contributes to the idea that banks with higher market power are more financially stable. In the competition profile the location of a banking business might also determine the default occurrences.

A detailed overview of PD model types is beyond the scope of the literature review. See Totmyanina (2011) regarding this issue.

3. Data and model

3.1. Default definition

The initial step to develop a PD model is to define what default is. There is no common opinion in the literature. A bank default probably implies the recall of a license and a prohibition to operate independently as a result of insolvency. The problem is that illegal financial operations as well as some other factors often cause license withdrawals[5]. Approximately half of these have no reference to the bank's ability to cover debts[6].

In this paper the following events are indicators of default:

- A bank's capital sufficiency level falls below 2%.
- The value of bank's internal resources drops lower than the minimum established at the date of registration.
- A bank fails to reconcile the size of the charter capital and the amount of internal resources.
- Bank is unable to satisfy creditors' claims and make compulsory payments.
- A bank is subject to sanitation by the Deposit Insurance Agency (Bank Restructuring Agency) or another bank.

[5] All available causes of license withdrawals are specified in Federal Law no. 395-1 of the Russian Federation of December 2, 1990 "Concerning Banks and Banking Activities", Article 20.

[6] Modeling license revocations has its own peculiarities, see Peresetsky (2010)

The next step is to create a list of the banks which defaulted over the period from1998 to 2011. The information was collected from the official website of the Central Bank of Russia and other public sources.

3.2. Sources of bank-specific financial information

Three sources of financial data relevant to the Russian banking system have been analyzed: the “Interfax” database, the BankScope database by Bureau van Dijk company and Mobile's “Banks and Finances” database. Preliminary analysis acknowledged the preeminence of the last database due to the wider time horizon and its satisfactory coverage of the Russian banking industry.

Table 2.

Comparison of sources of bank-specific financial information for Russian banks

Database	**Coverage quality**	**Coverage period**	**Data frequency**	**Number of missings**
Interfax	high	2000 – 2012	quarter	notable
BankScope	low	1996 – 2012	quarter	moderate
Banks and Finances	high	1998 – 2012	month	notable

We had Mobile's database available and authors are grateful to Prof. Petrov for access to the database.

3.3. Database description

We constructed a quarterly bank-specific financial database based on Mobile’s information from 1998 to 2011. Earlier data seems to be spoiled by numerous tiny, fake (so-called “sleeping”) banks and total chaos in the Russian banking system. In addition, monthly banking statistics might be less reliable than quarterly 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 910 license revocations, 467 of which were defaults in compliance with classification in Section 2.2, as well as 37 bank sanitations. The bar chart below demonstrates the distribution of registered defaults and their coverage.

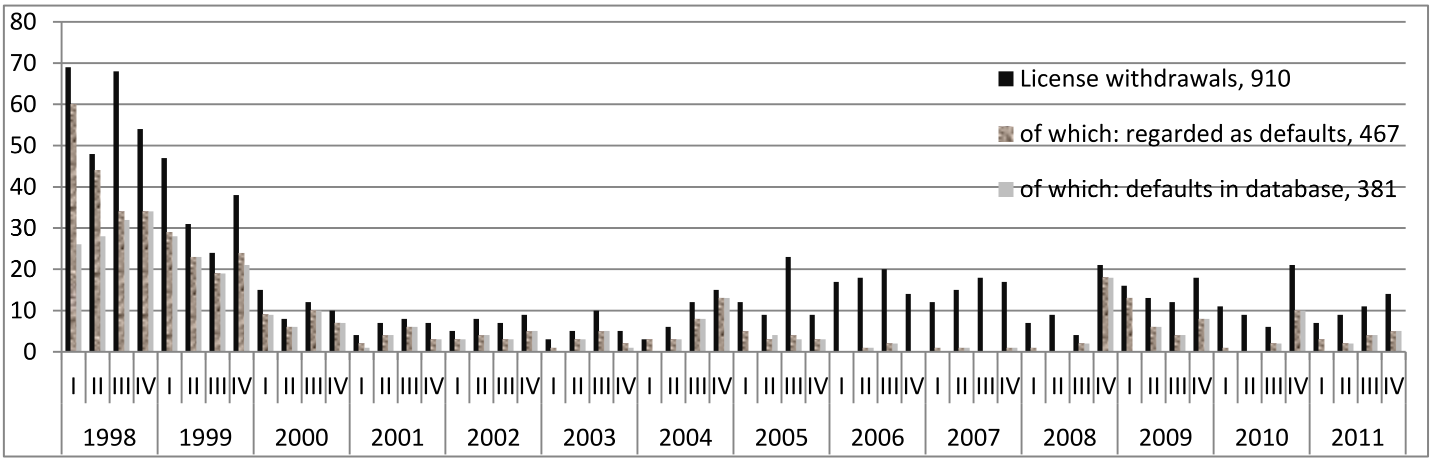

Diagram 5. Comparison of license revocations, withdrawals regarded as defaults and availability of defaults in the research database

As well as bank defaults, sanitations were unequally distributed with a sharp spike in the year 2008 (Diagram 6).

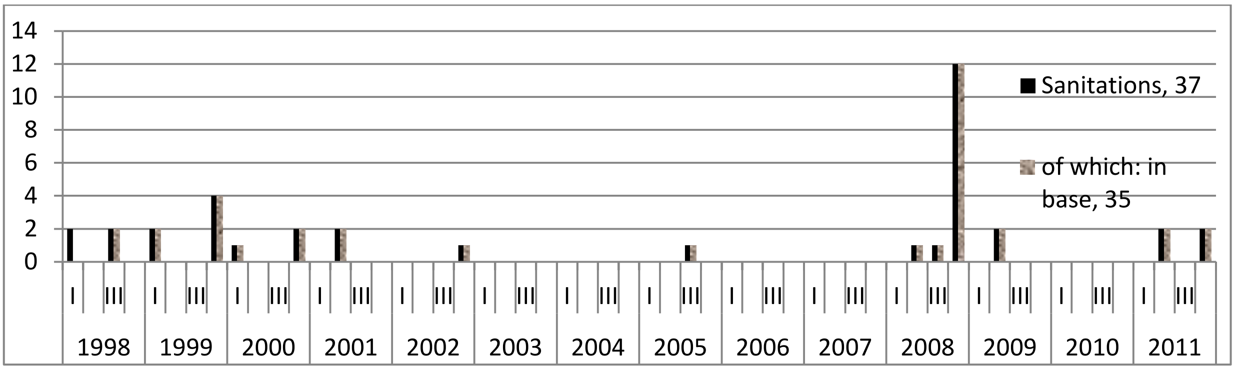

Diagram 6. Comparison of bank sanitations in the Russian banking sector and their availability in the research database.

Quasi-panel data structure was employed to carry out the analysis. It can be easily transformed for building panel models. The structure of database can be clearly seen below.

Bank's license number_period	Bank performance	A set of bank-specific variables				
				Variable names		
507_1/4/2005	default (1) or alive (0)			values		

Diagram 7. Quasi-panel data structure. A typical observation

The set of raw variables from "Banks and Finances" with a description and summary statistics is given in the Appendix. The last seven variables in the table marked with an asterisk are based on the Profit and loss statements while the remainder of the variables are based on the Balance sheet data.

To facilitate the data analysis the sample was split into two parts. The first comprised observations before the beginning of 2010. It 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.

3.4. Binary choice model: logit analysis

This section includes a brief interpretation of a logistic model application to predict defaults. The binary dependent variable *default* equals one if an observation is classified as insolvency and zero otherwise. The model is able to estimate a bank's probability of default directly in the form:

$$P(default = 1) = \Lambda\,(x * \beta),$$
$$s.\ t.$$
$$\begin{cases} P(default = 1) \ \to 1\ if \quad x * \beta \to \infty \\ P(default = 1) \to 0\ if \quad x * \beta \to -\infty \end{cases}$$

where $\Lambda(x * \beta)$=$\frac{\exp(x*\beta)}{1+\exp(x*\beta)}$ is a function taking values between 0 and 1; x is a vector of n regressors (i. e. $x * \beta = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_n * x_n$).

The basic principle of the maximum likelihood approach, used to estimate the logit-model, is to find the coefficient vector $\beta^{\wedge}$ that maximizes the likelihood of observing the bank performance states (“insolvent” or “alive”) in the sample as they are, dependent on the explanatory variables $x_1, x_2 \dots x_n$.
Formally speaking, the following task is set:

$$\ln[L] \rightarrow max_{\beta},$$

where $L = \Pi_{i=1}^{v} [\Lambda(x_i\beta)] \ * \ \Pi_{j=1}^{w} [1 - \Lambda(x_j\beta]$ for a sample with *v* and *w* observations of “insolvent” and “alive” classes respectively.

For greater details on this topic, see Greene (2007).

4. Empirical model estimation

4.1. A Choice of financial regressors for the initial model

The database created could contain some measurement errors or inaccurate observations. To allow for this a clearing algorithm was implemented, which dropped suspicious observations for the class of “alive” banks[7]:

- Lines in database with negative values for Total loans to the economy (KE), Total assets (CA) and Capital (SK).
- The 1st and 99th percentiles of observations for each of relative variables, described in table 2 to avoid statistical outliers. Those ratios seem to be significant to determine bank’s PD as provided by the literature review and common sense.

An automatic classification of the independent variables was applied to test the separating power of variables. To avoid statistical troubles 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. After that the multicollinearity problem was mitigated.

In the next stage we used the CAMELS framework to carry out a factor selection process. The idea of selecting capital, assets, management, earnings, liquidity and sensitivity variables to describe a bank’s financial position was borrowed.

- Capital-related set of variables

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 break minimum requirement. That is why the *Capital to Total assets* ratio was exploited as a factor in the PD model.

[7] There were no corrections for observations of the “default” class. In this case negative values can be caused by a weak financial position of a bank.

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	Variable symbol	The formula for calculation	P–value in ANOVA test
1	Capital to Total assets ratio	sk_ca	sk/ca	0,031
2	Balance profit to Total assets ratio	bp_ca	bp/ca	0,042
3	Liquid assets to Total assets ratio	la_ca	la/ca	0, 007
4	Non-government securities to Total assets ratio	ncb_ca	ncb/ca	0,004
5	Non-performing loans to Total loans to the economy ratio	pzs_ke	pzs / ke	0,008
6	Turnover on correspondent accounts to Total assets ratio	ln_oks_ca	ln (oks / ca)	0,065
7	Liquid assets to Demand liabilities ratio	la_ov	la / ov	0,109
8	Logarithm of Total assets	lnca	ln(ca)	0,079
9	Operational revenues to Net profit ratio	odb_ cp	odb / cp	0,165
10	Net profit to Total assets ratio	cp_ca	cp / ca	0,078
11	Liquid assets to Non-government securities ratio	la_gdo	la / gdo	0,123
12	Liquid assets to Over 1 year liabilities ratio	la_solong	la / so_long	0,243
13	National and local government obligations to Total assets ratio	gdo_ca	gdo / ca	0,324
14	Working assets to Total assets ratio	ra_ca	ra / ca	0,168
15*	Less than 30 days deposits of individuals to Deposits of individuals ratio	vdfl30_dfl	vdfl_30 / dfl	0,069
16	Personnel expenses to Operational costs ratio	rsa _orb	rsa / orb	0,654
17*	Non-performing loans to Required reserves in the Central bank of Russia ratio	pzs_orcb	pzs / orcb	0,098
18	Loss reserves to Total assets ratio	res_ca	res / ca	0,023

Notes: Shaded variables are of high separating power (with p-value for ANOVA test less than 0.1).
Asterisks indicate unevenly distributed variables with small number of observations.

- Assets-related set of variables

Asset quality is a dominant factor of future profits and losses considered with a ratio of *Non-performing loans to Total loans to the economy*. The *logarithm of Total assets* is a good measure of bank size.

- Management-related set of variables

Management is an important analytical consideration. Information about the organizational complexity of Russian banks is not available in open sources, so proxies were used instead. The proportion of *Turnover on correspondent accounts to Total assets* reflects the level of economic activity in a bank. Lower values might recognize a bank's inability to process payments and incentives of managers to curtail business. A logarithmic scale was used to avoid multicollinearity.

- Earnings-related set of variables

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

- Sensitivity & Liquidity related set of variables

In traditional CAMELS methodology sensitivity reflects vulnerability of a business to market risks. We exploited *Non-government securities to Total assets* ratio to assess both liquidity management and carried market risks in a bank.

$$P(default = 1) = \Lambda\ (sk\_ca;\ \ln\_\text{ca};\ pzs\_ke;\ \ln\_oks\_ca;\ bp\_ca;\ ncb\_ca). \tag{1}$$

It is noteworthy that the database used is highly unbalanced, which is harmful for logistic analysis. The class imbalance problem arises when one group of observations (in this case, "insolvency" observations) are underrepresented compared to another (in this case, to "alive" observations). There are many ways to tackle the problem, which are discussed in He (2007).

In this paper a modified methodology by Hosmer & Lemeshow (2000) was applied. Firstly, one thousand subsamples were done with all "insolvency" observations and enough of the other class to balance each of the one thousand 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.

4.2. Modeling

Our main goal is to predict a bank's default. For this reason independent lag variables are worth using. 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, the area under the ROC curve and the ROC curve graphical comparison, specificity, sensitivity and the proportion of correctly classified outcomes[8].

We found that model quality falls with lag climbing. Nevertheless, one-quarter lag is inadmissible for our goals: the obtained PD is the likelihood of a bank failure in a three-month period. The actual insolvency is likely to happen even earlier, so a user of the model is limited in time. To deal with this issue two-quarter lags were applied. The basic model specification is:

$$P(default = 1) = \\ \Lambda\ (sk\_ca_{lag2};\ \ln\_\text{ca}_{lag2};\ pzs\_ke_{lag2};\ \ln\_oks\_ca_{lag2};\ bp\_ca_{lag2};\ ncb\_ca_{lag2}). \tag{2}$$

At the next stage nonlinearities were considered. An expanded model included all factors from the model (2) in powers up to three. Then, insignificant coefficients were dropped. As a result, a basic model with nonlinearities appeared:

$$P(default = 1) = \\ \Lambda\ (sk\_ca_{lag2};\ (sk\_ca_{lag2})^2; \ln\_\text{ca}\_lag2;\ (\ln\_\text{ca}_{lag2})^2;\ pzs\_ke_{lag2};\ \ln\_oks\_ca_{lag2}; \\ bp\_ca_{lag2}; (bp\_ca_{lag2})^2; ncb\_ca_{lag2}). \tag{3}$$

After that a model improvement process was launched. In general the process is described by the flow chart below. After each of the iterations the most insignificant variables were excluded from a model. Also, LR-test confirmed the adequacy of inclusion for every set of factors.

---

[8] A bank with PD over 30% was regarded as a candidate to face default. That is important for calculating specificity, sensitivity and the proportion of correctly classified outcomes for models.

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_xxxx*	Basic with nonlinearities and time factor
2	Macroeconomic parameters: use of quarterly GDP growth rates *dgp_gr* and the consumer price index *cpi* in order to account for the effect of macroeconomic environment on bank performance.	Basic with nonlinearities, time factor and macro parameters
3	Institutional parameters: - use of Lerner index *l_index* to consider the impact of monopoly power of the firm on its default probability. - use of a dummy variable *region* indicating Moscow location of bank's headquarters. - use of a dummy variable on a bank participation in a Deposit insurance system.	Basic with nonlinearities, time factor, macro parameters and institutional variables.

Diagram 8. A model improvement process: time dummies, macroeconomic variables, institutional factors.

Table 4 provides information about the statistical properties of the key working models. These models should not be treated as final by the potential user, but might be employed if some data to develop more advanced models is unavailable.

4.3 Testing the model stability and comparison with alternatives

The final specification (4), the last working model, is stable enough to avoid the problem of overfitting according to estimation results in table 5 and diagram 9.

$$P(default = 1) = \Lambda\,(sk\_ca_{lag2}\,;\ (sk\_ca_{lag2})^2; \ln\_\mathrm{ca}_{lag2}\,;\ (\ln\_\mathrm{ca}_{lag2})^2\,; pzs\_ke_{lag2};\ \ln\_oks\_ca_{lag2}; bp\_ca_{lag2}; \left(bp\_ca_{lag2}\right)^2;\ ncb\_ca_{lag2};\ d_{2009};\ d_{q1}; \mathrm{gdp\_gr_{lag2}}; cpi_{lag2}; l_{index}; \mathrm{region}). \quad (4)$$

In addition, we tested robustness of the model (4) to gaps in the data. Under our results, gaps in the data impact the banks' default probabilities insignificantly.

Coefficients for the last working model were averaged on 1000 subsamples. The final model is (5).

$$P(default = 1) = \Lambda\,(-12.34 * sk\_ca_{lag2} + 15.45 * (sk\_ca_{lag2})^2 - 1.86 * \ln\_\mathrm{ca}_{lag2} + 0.06 * (\ln\_\mathrm{ca}_{lag2})^2 + 4.63 * pzs\_ke_{lag2} - 1.17 * \ln\_oks\_ca_{lag2} - 59 * bp\_ca_{lag2} + 1001 * \left(bp\_ca_{lag2}\right)^2 + 3.11 * ncb\_ca_{lag2} + 2.41 * d_{2009} - 1.51 * d_{q1} + 0.21 * \mathrm{gdp\_gr_{lag2}} + 0.11 * cpi_{lag2} \quad (5)$$

$$-2{,}34 * l\_index + 2.75 * \text{region} - 1.48).$$

Table 4.

Statistical properties for the key working models

Model name / **Variable name**	Basic _1_	Basic with nonlinearities _2_	Basic with nonlinearities and time factor _3_	Basic with nonlinearities, time factor and macro parameters _4_	Basic with nonlinearities, time factor, macro parameters and institutional variables _5_
sk_ca_lag2	-0.55	-9.88***	-10.01***	-9.23***	-12.24***
$(\text{sk\_ca lag2})^2$		14.55***	14.76	14.13***	16.03***
ln_ca_lag2	-0.13**	-1.15*	-1.5**	-1.4**	-1.86**
$(\text{ln\_ca\_lag2})^2$		0.04*	0.05**	0.04*	0.06**
bp_ca_lag2	-11.5***	-70***	-68***	-63***	-57***
$(\text{bp\_ca\_lag2})^2$		964***	1022***	949***	1031***
ncb_ca_lag2	3.99***	4.54***	5.02***	5.06***	3.11***
pzs_ke_lag2	6.38***	4.52***	4.21***	4.47***	5.17***
ln(oks_ca_lag2)	-1.19***	-1.06***	-1.06***	-1.02***	-1.2***
d_2009			1.76***	2.36***	2.41***
d_q1			-1.14***	-1.32***	-1.51***
gdp_gr_lag2				0.18	0.12
cpi_lag2				0.1***	0.11***
l_index					-2.51***
region					2.91***
Pseudo R^2	0.5219	0.59	0.6046	0.6279	0.6403
ROC area	0.8936	0.9157	0.9392	0.9422	0.9697
Sensitivity	72.30%	75.90%	77.34%	78.42%	79.14%
Specificity	97.20%	97.68%	98.16%	96.64%	96.96%
Correctly classified	92.67%	93.72%	94.37%	93.32%	93.72%

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

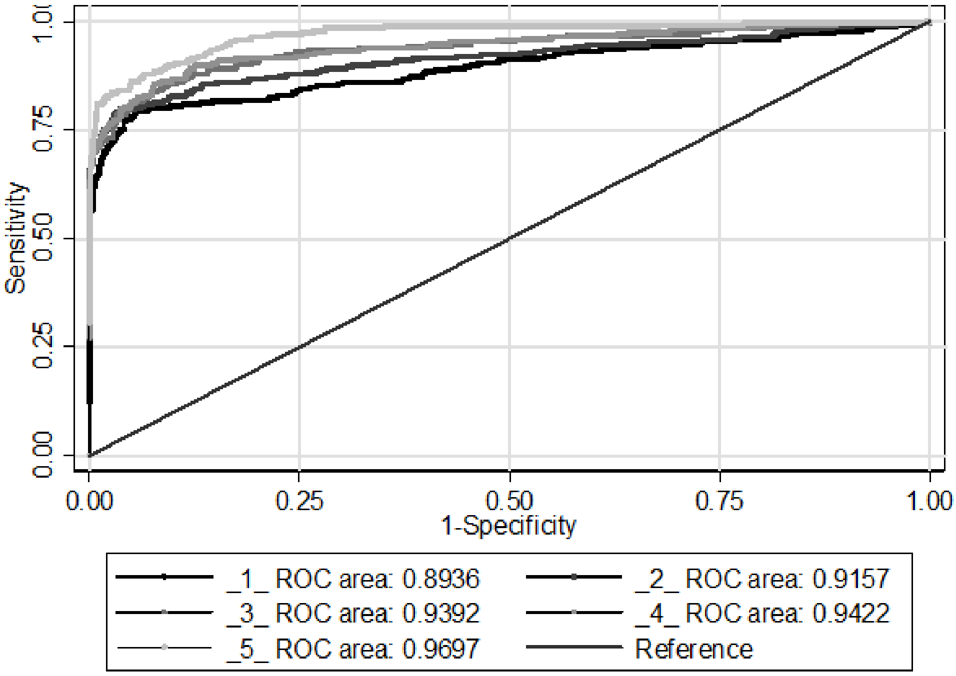

Diagram 9. ROC curve comparison for the key working models

Table 5.

Estimation results for the model (4) on 1000 subsamples. Testing the problem of overfittness

Coefficient	Sign		Level of significance			
	"+"	"-"	1%	5%	10%	>10%
sk_ca_{lag2}	0	1000	599	396	5	0
$(sk\_ca_{lag2})^2$	1000	0	188	785	27	0
ln_ca_{lag2}	0	1000	258	627	115	0
$(ln\_ca_{lag2})^2$	1000	0	38	295	667	0
bp_ca_lag2	0	1000	1000	0	0	0
$(bp\_ca_{lag2})^2$	1000	0	205	623	169	3
ncb_ca_{lag2}	1000	0	2	83	670	245
pzs_ke_{lag2}	1000	0	700	213	86	1
$ln\_oks\_ca_{lag2}$	1000	0	1000	0	0	0
d_2009	1000	0	1000	0	0	0
d_q1	0	1000	119	667	187	27
gdp_gr_{lag2}	1000	0	0	0	11	989
cpi_{lag2}	1000	0	46	937	17	0
l_index	0	1000	121	634	245	0
$region$	1000	0	1000	0	0	0

As mentioned, the out-of-sample prediction power was estimated in a subsample with observation for the years 2010 and 2011. To evaluate the quality of prediction two touch-stones 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. Table 6 shows the predictive results of the final regression (5) dependent on a separating condition.

Table 6.

Out-of-sample predictive accuracy of the final model (5)

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

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

In this study we have explored the benefits of panel data analysis. In our case, the researcher is primarily interested in the particular units (Russian banks) in the sample, which is close to the population. That is why a fixed effect logit model with the final specification (4) was applied. Surprisingly, no predominance of 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, that is the reason for insignificance of the Z-score.

5. Estimation results

As described, the final model is:

$$\begin{aligned} P(default = 1) = \\ \Lambda\,(-12.34 * sk\_ca_{lag2} + 15.45 * (sk\_ca_{lag2})^2 - 1.86 * \ln\_\mathrm{ca}_{lag2} + 0.06 \\ * (\ln\_\mathrm{ca}_{lag2})^2 \\ +4.63 * pzs\_ke_{lag2} - 1.17 * \ln\_oks\_ca_{lag2} - 59 * bp\_ca_{lag2} + 1001 * \left(bp\_ca_{lag2}\right)^2 \\ + 3.11 * ncb\_ca_{lag2} + 2.41 * d_{2009} - 1.51 * d_{q1} + 0.21 * \mathrm{gdp\_gr}_{\mathrm{lag2}} + 0.11 * cpi_{lag2} \\ -2{,}34 * l\_index + 2.75 * \mathrm{region} - 1.48). \end{aligned} \tag{5}$$

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:

$$\frac{\partial\Lambda(f(x))}{\partial xi} * \Delta x = \lambda(f(x)) * f'(x) * \Delta x, \text{ where } \lambda(f(x)) > 0. \quad (6)$$

5.1. Financial bank-specific ratios

Capitalization: *Capital to Total assets* ratio

According to the estimation results over and undercapitalized banks exhibit higher default probabilities (diagram 10). The conclusion is consistent with the expressed expectations.

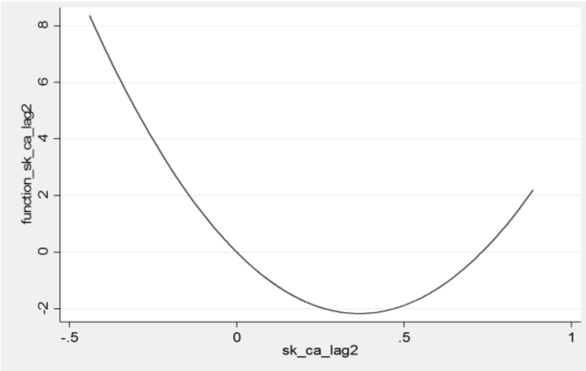

Diagram 10. Impact of $\boldsymbol{sk\_ca_{lag2}}$ ratio on default probability:
$\boldsymbol{f\ (sk\_ca_{lag2}) = -12.24 * sk\_ca_{lag2} + 16.03 * (sk\_ca_{lag2})^2}$

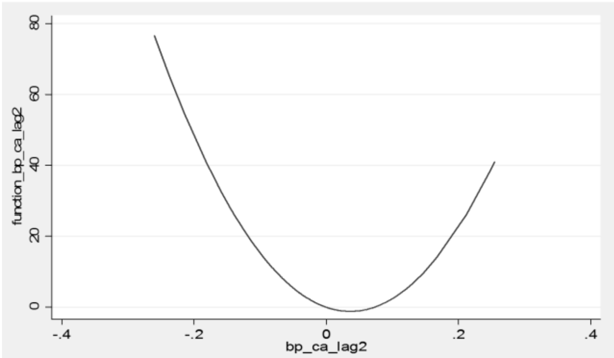

Diagram 11. Impact $\boldsymbol{of\ bp\_ca_{lag2}}$ ratio on default probability:
$\boldsymbol{f\ (bp\_ca) = -57 * bp\_ca_{lag2} + 1031 * (bp\_ca_{lag2})^2}$

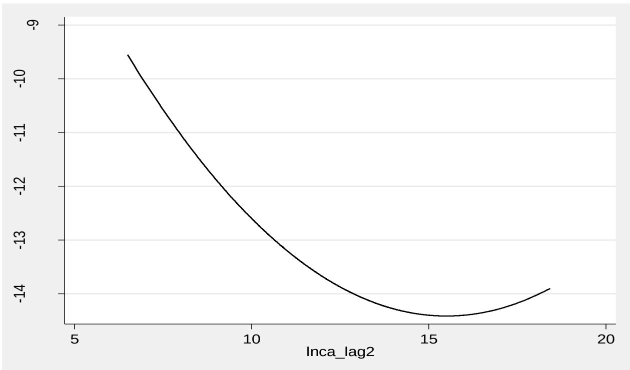

Diagram 12. Impact of $\boldsymbol{ln\_ca_{lag2}}$ ratio on default probability:
$\boldsymbol{f\ (ln\_ca_{lag2}) = -1.86 * ln\_ca_{lag2} + 0.06 * (ln\_ca_{lag2})^2}$

Profitability: *Balance profit to Total assets* ratio

Banks with extremely high and low profitability score higher default rates risks (diagram 11). Naturally, poor banks lack funds to pay the bills. Perhaps, banks with really high earnings take excessive risk, which leads to growing default probability. Moreover, in efficient markets it is impossible to maintain an outstanding profitability without bearing appropriate financial risk.

Bank size: Logarithm of *Total assets*

Small as well as large banks have higher risk of insolvency (diagram 12). 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 to the economy* ratio
Bank with considerable amount of bad debts are less stable, as supposed.

Operational activity: Logarithm of *Turnover on correspondent accounts to Total assets* ratio

Our main regression results demonstrate a negative correlation between a PD and a 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 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

5.2. Time factor

Quarterly dummies
The only significant quarterly dummy variable indicates that on the average the probability of default is lower in the first quarter of a year. Our guess is that it is closely associated with the Regulator’s bias 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, for instance, 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.

5.3. 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. Also the second macroeconomic variable could harm the estimation results Karminsky et al. (2005). Alternatively, the financial ratios could have absorbed the impact on default probability of a business cycle.

Consumer price index
A growing consumer price index, which accelerates inflation, increases 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.

5.4 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

The Moscow-based banks have higher PDs on the average, which contradicts the findings in Fungacova & Solanko (2009). According to Claeys & Schoors (2007), the Russian banking regulator is reluctant to withdraw licenses out of 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 is that the set of Deposit insurance system participants is too diversified.

Conclusion

In this paper an adequate probability of default model was developed. Over 60% of bank failures were correctly classified in out-of-sample prediction power tests.

Our main regression results revealed a quadratic relationship between a bank's capitalization, size, quarterly earnings and its probability of default. We confirmed the validity of the CAMELS framework together with macroeconomic, institutional and time factors to forecast bank's PD.

In addition, a variety of statistical criteria was employed to test the quality of the logistic model. A comparison with alternative models was carried out. However, neither panel data structure nor Z-score improved the final model performance.

There were a number of issues concerning the database: a lack of balanced; missing data, outliers and measurement errors in the raw bank-specific statistics; an overfitting problem and others. However, we have proposed solutions for these issues except for logistic model limitations to estimate the impact of a bank's ownership type on its default probability. A visible way out is the use of bank ratings to estimate PDs. This is a prospective direction for our research in the near future.

References

Anzoategui D., Pería M., Melecky M. (2012). Bank competition in Russia: An examination at different levels of aggregation // Emerging Markets Review. Vol. 13. No 1, P. 52–53.

Bluhm C., Overbeck L., Wagner C. (2010). Introduction to Credit Risk Modeling. Chapman and Hall/CRC, 2010.

Bock R., Demyanets A. (2012). Bank Asset Quality in Emerging Markets: Determinants and Spillovers, IMF Working Papers 12/71, International Monetary Fund.

Chernykh L., Theodossiou A. (2011). Determinants of Bank Long-Term Lending Behavior: Evidence From Russia // Multinational Finance Journal. 2011. No. 15: P. 193–216.

Claeys S., Schoors K. (2007). Bank supervision Russian style: Evidence of conflicts between micro- and macro-prudential concerns // Journal of Comparative Economics. 2007. Vol. 35. No. 3. P. 63–657.

Clarke G., Cull R., Shirley M. (2005). Bank privatization in developing countries: A summary of lessons and findings // Journal of Banking & Finance. 2005. Vol. 29. No. 8–9. P. 1905–1930.

Fungacova Z., Solanko L. (2009). Risk-taking by Russian banks: Do location, ownership and size matter? BOFIT Discussion Papers 21/2008. Bank of Finland. Institute for Economies in Transition.

Fungacova Z., Weill L. (2009). How market power influences bank failures: Evidence from Russia. BOFIT Discussion Papers 12/2009. Bank of Finland. Institute for Economies in Transition.

Greene W. (2007). Econometric analysis. 6th Edition. Prentice Hall, 2007.

He H., Edwardo A. (2009) Learning from Imbalanced Data, IEEE Transactions on Knowledge and Data Engineering, vol. 21, 9, 1263-1284.

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

Karminsky A., Peresetsky A., Petrov A. (2005). Ratings in Economics: Methodology and Practice. Moscow, Finansi i Statistika (Finance and Statistics Publishing House), 2005 [in Russian].

Lanine G., Vennet R. (2006). Failure prediction in the Russian bank sector with logit and trait recognition models // Expert Systems with Applications. 2006. Vol. 30, No. 3, P. 463-478.

Mannasoo K., Mayes D. (2009). Explaining bank distress in Eastern European transition economies // Journal of Banking and Finance. 2009. Vol. 33 No. 2. P. 244-253.

Micco A., Panizza U., Yanez M. (2007). Bank ownership and performance. Does politics matter? // Journal of Banking and Finance. 2007. Vol. 31. No. 1, P. 219-241.

Peresetsky A., Karminsky A., Golovan S. (2011). Probability of default models of Russian banks // Economic Change and Restructuring. Vol. 44. No. 4.

Peresetsky A. (2010) Modeling license revocations of Russian banks / Preprint #WP/2010/085 . New Economic School. 2010 [in Russian].

Tabak B., Craveiro G., Cajueiro D. (2011). Bank efficiency and Default in Brazil: Causality Tests. Working paper series 253. The Central Bank of Brazil. 2011.

Totmyanina K. (2011). Review of models of default probability. Journal of financial risk-management. 2011. №1 (25) [in Russian].

Vernikov A. (2011). Government banking in Russia: Magnitude and new features (2011). IWH Discussion Papers. Halle Institute for Economic Research.

Appendix

Raw variables from “Banks and Finances” with description and summary statistics

Variable symbol	Variable name	Obs	Mean	Std. Dev.	Min	Max
BP	Balance profit	59270	139057.6	2749515	-7.16e+07	3.18e+08
CA	Total assets	59391	1.39e+07	1.90e+08	-6.53e+07	1.02e+10
CP	Net profit	58461	102745.2	2242601	-7.16e+07	2.54e+08
DFL	Deposits of individuals	37579	4536610	8.64e+07	0	5.06e+09
GDO	National and local government obligations	32556	1000707	1.66e+07	0	9.12e+08
KE	Total loans to the economy	57608	6798499	1.02e+08	-2082	6.83e+09
KE_F	Loans to individuals and small business	39454	2093819	3.02e+07	0	1.66e+09
KE_F_12	Over 1 year loans to individuals and small business	36060	1964401	2.99e+07	0	1.58e+09
KE_Prom	Loans to industry	29410	9631169	1.10e+08	0	5.17e+09
LA	Liquid assets	59276	2814943	3.15e+07	0	2.29e+09
NCB	Non-government securities	48854	1104642	1.11e+07	0	5.14e+08
NMO	Net interbank operations	29726	-799589.2	2.22e+07	-1.37e+09	3.36e+08
OKS	Turnover on correspondent accounts in commercial banks and the Central bank of Russia	55853	3.53e+07	3.15e+08	0	1.54e+10
ORCB	Required reserves in the Central bank of Russia	39845	146057.9	1839618	0	1.00e+08
OV	Demand liabilities	59365	4172650	6.65e+07	-8719	4.48e+09
PZS	Non-performing loans	47611	369878.6	5511744	0	3.26e+08
PZS_F	Non-performing loans to individuals and small business	22912	194123.4	1609185	0	5.70e+07
RA	Working assets	58448	9228285	1.27e+08	0	8.36e+09
RES	Loss reserves	58342	597177.8	9699533	0	7.39e+08
SDB	Funds on vostro accounts	15076	348963.5	2569998	0	1.04e+08
SK	Capital	59629	1596060	2.29e+07	-5.72e+08	1.46e+09
SO_LONG	Over 1 year liabilities	50757	5458812	8.32e+07	0	4.65e+09
SRTS	Funds on settlement accounts	41798	2813746	3.14e+07	0	1.84e+09
VBCB	Issued securities	40401	763762.7	5467567	0	2.02e+08
VDFL	Over 30 days deposits of individuals	51810	2813877	6.47e+07	0	4.31e+09
VDFL_30	Less than 30 days deposits of individuals	54222	551249.4	9284596	0	7.60e+08
VDUL	Deposits of legal entities	45212	2392271	2.64e+07	0	1.33e+09
NORM_H3	Current liquidity ratio	42792	102571.1	3170775	-1000	1.00e+08
ODB_q*	Operational revenues	40005	1092664	1.88e+07	-1.34e+08	1.95e+09
ORB_q*	Operational costs	39753	1121457	2.08e+07	-1.34e+08	2.22e+09
PDFL_q*	Interest income from loans to individuals	39671	42527.78	610626.1	-3725184	4.63e+07
PDK_q*	Interest income from loans to industry	40344	97565.66	1424873	-2.56e+07	1.31e+08
PDMBK_q*	Interest income from loans to credit institutions	30771	8649.174	116576.5	-976662	9345929
RSA_q*	Personnel expenses	40004	42522.98	552660.7	-471395	4.86e+07
RUB2_q*	Expenses from revaluation of foreign funds	36357	832045.2	1.69e+07	-1.39e+08	1.89e+09

Notes: in thousands of Russian rubles; in accordance with Russian Accounting Standards.

Alexander M. Karminsky
National Research University Higher School of Economics, Professor, Doc. (Ec., Tech.),
E-mail: karminsky@mail.ru