Do Higher Interest Rates on Loans and Deposits and Advertising Spending Cuts Forecast Bank Failures? Evidence from Russia

Lev Fomin, PJSC Bank Otkritie Financial Corporation
lfomin@nes.ru

This study builds a probabilistic model of Russian bank defaults. Microdata from the monthly financial and regulatory statements that Russian banks submit to the Bank of Russia are analysed, covering the period from July 2010 to December 2017. A model incorporating a standard set of reliable predictors of bank defaults is augmented by three novel predictors: the excess of deposit and loan rates over the respective cross-section averages, and the ratio of spending on advertising to the bank’s assets. These predictors are statistically significant in logit regressions that forecast bank defaults and improve the forecasting power of the model, although relatively moderately. The too-big-to-fail premise is not supported by the data.

Keywords: bank closure, bankruptcy, CAMEL indicators, Russia’s banking sector

JEL Codes: G2, G21, G33

Citation: Fomin, L. (2018). Do Higher Interest Rates on Loans and Deposits and Advertising Spending Cuts Forecast Bank Failures? Evidence from Russia. Russian Journal of Money and Finance, 78(2), pp. 94–112. doi: 10.31477/rjmf.201902.94

1. Introduction

In the past ten years, the number of banks operating in Russia has halved, a phenomenon for which the literature provides several historical explanations (see e.g. Mamonov, 2017, 2018). First, during this period, the Russian economy faced two economic crises, in 2008–2009 and 2014–2016, which had a negative impact on the banking sector. Second, in 2013, when Elvira Nabiullina became Governor of the Bank of Russia, the regulator adopted a stricter stance on rehabilitation of the banking system and began cleansing the financial sector. As Figure 1 shows, in 2014–2017, the number of banks in operation fell by 40%, from 923 to 561, which is three times more than the 13% drop in the previous period, 2010–2013.
The risk of bank defaults remains the principal risk in the banking system. Bank failures have an impact on almost every sector of the economy. They affect not only individual and corporate customers, who could lose their deposits or access to loans, but also counterparty banks and professional securities and foreign exchange market participants. In small Russian towns, where there are only a few credit institutions, the failure of any one of them is likely to cause local economic activity to slow substantially, if not come to a total standstill.

Although the growth in the number of revoked banking licences observed since 2013 had slowed by 2017, that year was important for other reasons. The Bank of Russia revoked the licence of Yugra Bank, one of the top 30 Russian banks by assets, and, by appointing a provisional administration, began the resolution procedure for Otkrytie and B&N Bank, which were among the top 15. Although these developments do not overturn completely the too-big-to-fail premise supported in the abundant research on bank defaults (Hwang et al., 1997), they seem to be more consistent with the too-many-to-fail notion (Brown and Dinç, 2011).

Forecasting Russian bank defaults therefore remains an important task. Early warning systems that can spot troubled banks well before their licences are revoked are in great demand. For this purpose, most studies typically employ regularly published (quarterly or monthly) balance sheet data from banks available in the public domain.

In this paper, I will consider the following potential causes of bank failures: (i) misalignment of management/owners’ incentives and the interests of depositors, which may result in deliberate bankruptcy, and (ii) unsustainable business models, which make credit institutions vulnerable to adverse market forces. In the former case, the task of forecasting bank failures is likely to be further complicated by the fact that, if the management’s incentives are distorted, they may be motivated to forge the credit institution’s financial statements, and this will adversely affect the prediction quality. In this paper, I tackle this problem by introducing new predictors.
that, to the best of my knowledge, have not been analysed before. These predictors are constructed from balance sheet items that are not subject to the Bank of Russia’s regulatory requirements, which makes them less likely to be forged. These are (i) the excess of the deposit rate offered by the institution over the industry average, (ii) the excess of the loan rate charged by the institution over the industry average, and (iii) the monthly change in the ratio of spending on advertising to assets.

The objective of this paper is to build up a probabilistic default model for Russian banks which is capable of predicting their bankruptcy with a high level of precision by augmenting a conventional set of predictors with the three abovementioned predictors. To the best of my knowledge, these predictors have never been used before in the literature. I will explore whether adding these new predictors to what is an otherwise standard bank failure forecasting model increases the model’s predictive power. For the purpose of this study, I gathered monthly balance sheet data from Russian banks covering the time period from July 2010 to December 2017. The sample included 1,063 banks, of which 470 (or over 40%) went bankrupt during the sample period. Although most studies of bank failure models employ quarterly data, we use monthly data, as do e.g. Mamonov (2017, 2018) and Mäkinen and Solanko (2018).

This paper contributes to the literature by showing that the three novel predictors, the excess of deposit and loan rates over the respective cross-section averages along with the rate of change in the advertising spending to assets ratio, improve the quality of forecasts by the otherwise standard state-of-the-art model of bank failures, as comparison of the respective ROC curves demonstrates.

A more rigorous (and challenging) way to establish the marginal contribution of the three novel predictors would be to compare the out-of-sample performance of forecasts based on such best-practice models as Mamonov (2017, 2018), Peresetsky and Zhivaykina (2017), and Mäkinen and Solanko (2018) both with and without the three predictors on top of the original predictors. I leave this worthy exercise for future research.

The paper is structured as follows. Section 2 briefly discusses the related literature. Section 3 describes the methodology and data. Section 4 presents the results, and Section 5 concludes.

## 2. Analysis of previous experience

Bank default risk has long been studied both in Russia and abroad. Over time, a methodology for conducting this line of investigation has evolved. The majority of modern studies use financial, macroeconomic and bank activity variables as predictors. In this section, I do not intend to survey the vast literature on predicting bank defaults. Its purpose is rather narrow: I only cite those studies that document the forecasting value of predictors that belong to my set of conventional or best-practice predictors.
Most studies use a binary choice model, the logit or probit regression. The first paper to apply this method to study bank defaults was Martin (1977), which suggested breaking financial variables down into five groups. This methodology is now known as CAMEL, and it employs the following indicators: capital adequacy, asset quality, management, earnings, and liquidity management.

A recent study, Mäkinen and Solanko (2018), shows that the risk of default is affected not only by CAMEL factor values, but also by their monthly variation. One interesting feature of this article is that the authors' model was estimated on monthly (Russian) data, whereas most studies on bank failures work with quarterly data, as is standard practice. Two other papers on bank failures which employ monthly data on Russian banks are Mamonov (2017, 2018). Mäkinen and Solanko (2018) argue that monthly data are better suited to explaining bank defaults.

I base my choice of potentially informative predictors on the empirical literature on forecasting bank failures in Russia:

1. The size of the bank negatively influences the probability of default (Zubarev, 2012; Chernykh and Theodossiou, 2011; Peresetsky et al., 2011), which could be interpreted as the too-big-to-fail premise.
2. The bank being in state or foreign ownership decreases the risk of default (Micco et al., 2007; Hanafi and Santi, 2013).
3. There is a non-linear relationship between the size of the bank and the risk of default (Golovan et al., 2003; Karminsky and Kostrov, 2013). In the class of polynomials, the quadratic polynomial was shown to explain this relationship best.
4. Participation of the bank in the state-run deposit insurance programme negatively influences the probability of default (Demirgüç-Kunt and Detragiache, 1998) since the risk of a bank run is reduced.
5. Adverse developments in the macroeconomy affect the probability of a bank default (Golovan et al., 2004; Männasoo and Mayes, 2009; Shen and Hsieh, 2011).
6. The riskiness of assets on a bank's balance sheet affects the probability of bank default (Cole and White, 2012).

3. Data and model

I decide which variables to use as predictors to build the probabilistic model of bank defaults on the basis of my analysis of empirical research on bank failures in Section 2.

The set of explanatory variables will include:
- Financial indicators;
- Macroeconomic indicators;
- Bank activity indicators.
I use the abovementioned CAMEL classification to select financial indicators. The data was taken from monthly reporting forms published on the Bank of Russia website:

- Balance sheet statement, Form 101;
- Income statement, Form 102;
- Capital statement, Forms 123 and 134;
- Required regulatory ratios, Form 135.

**Capital_a**: the ratio of capital to net assets. This variable is one of a bank's capital adequacy indicators. The expected sign of its influence on the bank's default probability is negative as banks with more capital tend to be more stable (Keeley, 1990; Boyd and De Nicolo, 2005; Berger and Bouwman, 2013).

**Arrears_l**: the ratio of arrears on the bank's balance sheet to the total amount of loans extended to individuals and firms. This variable is an asset quality indicator. The expected sign of its influence on the bank's default probability is positive: the more toxic assets the bank has, the higher its default probability is.

**Loans_a**: the ratio of loans to the economy to the bank's net assets. It reflects the bank's credit policy and its appetite for risk. The expected sign of its influence on the probability of default is positive.

**Profit_a**: the ratio of the bank's net profit to assets. This variable is a profitability indicator. The expected sign of its influence on the probability of default is negative: higher profits mean that the bank's business model is sustainable and effective.

**N3**: the current liquidity ratio set by the Bank of Russia; a liquidity indicator. The expected sign of its influence on the probability of default is negative: banks with insufficient liquidity are less likely to survive a bank run.

**Pledge_l**: the fraction of collateralised loans in the bank's loan portfolio; a riskiness indicator. The expected sign of its influence on the probability of default is negative: banks with a higher fraction of collateralised loans are less exposed to the risk of counterparty default.

I use the real output of goods and services from key business activities as the macroeconomic indicator. The monthly time series, seasonally adjusted, in constant prices, labelled $Y_{growth}$, is taken from the Federal State Statistics Service website www.gks.ru. The expected sign of its influence on the probability of default is negative: the risk of a bank default increases during a crisis or recession.

I use the following bank activity indicators:

**Ln_assets, Ln_assets2**: the log and the squared log of the bank's net assets. The squared indicator is required because its influence on the default probability is non-linear. The expected signs are negative for the squared log and positive for the log of assets. This is explained by the fact that medium-sized banks crash more often than their larger or smaller peers do.

**Insurance**: a dummy variable that equals 1 if the bank participates in the deposit insurance programme and 0 otherwise. The expected sign of its influence on the probability of default is negative: the participation of a bank in the programme
disincentivises depositors to run it, as they are guaranteed to be repaid even if the bank fails.

_Gov_fi eign_: a dummy variable that equals 1 if the bank is state-owned or a subsidiary of a foreign credit institution and 0 otherwise. In case of any problems, such banks can rely on government or parent company support, which greatly decreases the probability of default.

I also use the following predictors, which, to the best of my knowledge, have not been used before and are likely to be informative about bank failures:

_Deposits_r_: the ratio of the interest rate that a bank pays on deposits to the cross-section weighted average:

$$Deposits_r = r_{it}^D / \bar{r}_t^D,$$

where $r_{it}^D$ is the average deposit rate paid by bank $i$ in month $t$ and $\bar{r}_t^D$ the cross-section average deposit rate in month $t$ defined as

$$\bar{r}_t^D = \frac{\sum_{j=1}^{n} D_{jt} \times r_{jt}^D}{\sum_{j=1}^{n} D_{jt}},$$

where $D_{jt}$ is the amount of household deposits on the balance sheet of bank $i$ in month $t$. It is conceivable that a bank experiencing flight of its core deposits might start to raise household deposits more actively by increasing interest rates. One practical difficulty with this variable is that it is not readily available from a bank’s financial or regulatory statements, and nor is historical data always published on the bank’s website. (Even if it were, it would take a great deal of effort to collect this data.) I therefore use a proxy for the interest rates offered by a bank on deposits. I take the information on monthly accrued interest to be paid to individual customers from banks’ financial statements published by the Bank of Russia and divide it by the average amount of household deposits in the month in question. The result is multiplied by 12 to obtain the annualised rate. The expected sign of its influence on the probability of default is positive.

_Loans_r_: the ratio of the average interest rate a bank is paid on loans to the cross-section weighted average:

$$Loans_r = r_{it}^L / \bar{r}_t^L,$$

where $r_{it}^L$ is the average loan rate that bank $i$ is paid in month $t$ and $\bar{r}_t^L$ the cross-section average loan rate in month $t$ defined as

$$\bar{r}_t^L = \frac{\sum_{j=1}^{n} L_{jt} \times r_{jt}^L}{\sum_{j=1}^{n} L_{jt}},$$

where $L_{jt}$ is the amount of loans on the balance sheet of bank $i$ in month $t$. Like the previous indicator, this is a proxy variable, and it is obtained in a similar way.
This predictor is meant to capture the incentive of a bank on the verge of bankruptcy to “gamble for resurrection” by shifting its lending to more profitable, but also riskier, borrowers. A loan rate that is well above the cross-section average is thus likely to signal that the bank is willing to bear excessive risk and is therefore not particularly picky as far as the quality of its assets is concerned.

Adv_a: the ratio of advertising spending to the bank’s assets. Attracting new customers is an objective of every commercial bank, and banks use advertising to do this. However, if the bank’s financial standing is poor and the management is aware of this, advertising spending is likely to be cut. The reason why this indicator is defined as the ratio of ad spending to the bank’s assets is that advertising budgets in large and small banks differ substantially. The expected sign of its influence on the probability of default is negative: only banks that expect to stay in business spend money on advertising.

Figure 2 shows that cross-section asset-weighted average proxy rates in accordance with the key interest rate. In 2013Q3 and 2013Q4, the average deposit rate crossed the key rate. This can be explained by the fact that the key rate was introduced only on 13 September 2013 with an initial value of 5.5%, replacing the refinancing rate. This comparison permits the assumption that the constructed proxy rates are close to the actual rates offered or charged by banks on average.

Table 1 contains descriptive statistics for the variables involved in this study, showing the sample mean of each variable, its standard deviation, minimum and maximum values, and the expected sign of its influence on bank default probability.

In this study, I use the logit regression to build a probabilistic model of bank defaults. This specification has been widely employed in empirical research on forecasting bank defaults.
Table 1. Descriptive statistics

| Variable      | Mean   | Std. dev. | Minimum | Maximum | Expected sign |
|---------------|--------|-----------|---------|---------|---------------|
| Capital_a     | 0.252  | 0.239     | -0.875  | 19.0    | -             |
| Arrears_l     | 0.0573 | 0.0936    | 0       | 0.988   | +             |
| Loans_a       | 0.529  | 0.203     | 0       | 0.988   | +             |
| Ln_assets     | 22.1   | 1.97      | 13.3    | 30.8    | -             |
| Ln_assets2    | 493.9  | 88.9      | 176.1   | 949.7   | +             |
| Profit_a      | 5.24×10⁻³ | 40.4×10⁻³ | -2.26   | 1.50    | -             |
| H3            | 1.22   | 0.695     | 0.0471  | 4.67    | -             |
| Y_growth      | 1.02   | 0.0269    | 0.959   | 1.08    | -             |
| Pledge_1      | 1.10   | 0.674     | 1.35×10⁻⁴ | 7.98   | -             |
| Deposits_r    | 1.28   | 0.393     | 5.44×10⁻⁴ | 8.67   | +             |
| Loans_a       | 0.977  | 0.566     | 7.16×10⁻³ | 11.3   | +             |
| Adv_a         | 8.31×10⁻⁵ | 5.43×10⁻⁴ | 0       | 7.43×10⁻⁴ | -             |
| Number of 0’s | 15 633 |           | 57 912  |         | -             |
| Number of 1’s | 62 563 |           | 10 982  |         | -             |

Mäkinen and Solanko (2018) use the linear probability model, i.e. the OLS regression, with fixed effects instead of a binary choice model. The idea of introducing bank fixed effects is quite appealing in that banks may potentially have many time-invariant latent variables which could affect the probability of default and which are also difficult to account for explicitly. In their study, the authors build an explanatory regression rather than a predictive model. However, this framework does not seem suitable for my purposes, as the linear probability model is notorious for being able to produce predicted probabilities below 0 and above 1, which are difficult, if not impossible, to interpret. Mäkinen and Solanko (2018) avoid interpreting predicted probabilities, instead focusing on the statistical and economic significance of explanatory variables.

\[
P(\text{Default}_{i,t+h} = 1| x_{1,i,t}, \ldots, x_{K,i,t}) = \left(1 + \exp(\beta_0 + \beta_1 x_{1,i,t} + \ldots + \beta_K x_{K,i,t})\right)^{-1},
\]

where \( P(\text{Default}_{i,t+h} = 1| x_{1,i,t}, \ldots, x_{K,i,t}) \) is the probability of default of bank \( i \) on date \( t + h \) conditional on the values of predictors \( x_{1,i,t}, \ldots, x_{K,i,t} \), and \( \beta_0, \beta_1, \ldots, \beta_K \) are unknown parameters to be estimated.

In this paper, a bank default is defined as one of the following events:
1. The bank’s licence has been revoked.
2. The bank is under resolution.
3. The bank has been taken over by another bank but its financial standing was poor before the transaction.
Most studies define bank default using only a single criterion, revocation of its licence. This is not necessarily sufficient since, for instance, the launch of a resolution procedure means that the bank keeps its licence but is nevertheless experiencing serious problems and likely to be on the verge of bankruptcy. Definitions of bank defaults similar to mine have already been used in other studies that produced meaningful results (Logan, 2001; Arena, 2008; Karminsky and Kostrov, 2013, etc.).

I consider four specifications of the proposed model which vary in the data collected. The model that uses a standard set of predictors (excluding Deposits_r, Loans_r, and Adv_a) is labelled the standard model. The standard model augmented with the three novel predictors is labelled the modified model. For each of the models, I define the default dummy in two alternative ways: (i) the bank will default in \( h \) months, and (ii) the bank will default within \( h \) months. The signal that the bank is likely to go bankrupt within a certain period of time is valuable. I forecast bank defaults at horizons of one to six months, as it has been documented that the forecasting power of default models based on bank credit ratings deteriorates sharply beyond the six-month horizon (Hirtle and Lopez, 1999). In order to establish whether this lack of forecastability at longer than the six-month horizon applies to Russian data or not, I also include the forecast horizon of 12 months in the analysis. From the practical point of view, earlier indicators of heightened risk of default are undoubtedly more valuable. The performance of the model that forecasts defaults at the 12-month horizon will give us a sense of the extent to which this task is implementable.

In what follows I compare the forecast performance of the two rival specifications at different forecast horizons.

4. Results

Table 2 in the Appendix shows that, in the standard model, all predictors except the current liquidity ratio are statistically significant at the one-month horizon. Many related studies find that predictors belonging to the class of liquidity indicators are not statistically significant (Zubarev, 2012; Martin, 1977). This can be explained by the peculiarities of the Russian system: in case of liquidity shortages, banks can apply for a loan from the Bank of Russia.

The signs of the estimated coefficients are as anticipated. It is interesting to note that, at the two-month and longer forecast horizons, predictors related to the size of the bank cease to be significant. At the same time, as is clear from Table 3 (see Appendix), which shows logit regressions forecasting defaults within \( h \) months, the size of the bank remains significant at all forecast horizons.

We now turn to Tables 4 and 5 in the Appendix, which show estimated logit regressions for the modified model. It is clear that variables capturing
the bank’s size are statistically insignificant. It therefore appears that the too-big-to-fail premise is not supported by this analysis. The current liquidity ratio remains insignificant in the modified model as well. It is worth noting that the signs of the estimated coefficients in the modified model are well in line with the anticipated signs, and the majority of predictors keep their statistical significance (or insignificance) at all forecast horizons. The coefficients of the three novel predictors, the excess of loan and deposit rates over respective cross-section averages and the rate of change in the ad-spending-to-assets ratio, are also significant and are consistent with theoretical predictions with regard to the coefficient signs.

It is interesting to note that the coefficient of the variable $Adv_a$ in Table 4 (see Appendix) drops sharply when the forecast horizon increases from one to two months, as well as from two to three months. This confirms that banks that are on the verge of bankruptcy cut their ad spending just before they default.

A meaningful comparison of alternative models in terms of the quality of forecast they produce requires a criterion that is simple and clear enough to be interpreted intuitively. The ROC analysis seems an appropriate candidate for such a criterion. A specific feature of the logistic regression is that it does not predict strictly 1 or 0, but rather gives a probability in the range of 0 to 1. The user of the model can therefore set an arbitrary threshold which may be other than 0.5, depending on his/her goals, after which the model will be deemed to have predicted a certain result (see the Appendix for details). In ROC analysis, the AUC statistic serves as the model’s quality metric.

Inspecting the AUC values shown in the last line of Tables 2 to 5 in the Appendix, we see that the value of the AUC is highest for the one-month-ahead forecast, while the quality of the forecast is worse at longer horizons. Notably, the forecast quality for models predicting default in $h$ months deteriorates faster as $h$ grows compared with models predicting default within $h$ months.

Bank defaults are thus best forecast at the one-month horizon. The study by Mäkinen and Solanko (2018), which also employs monthly data, arrives at a similar conclusion. For practical purposes, however, this horizon seems unsatisfactorily short.

That said, if we use only one-month-ahead forecasts, the advanced model will be more precise (AUC = 0.846 vs. 0.79), although the gain is quite moderate. This means that prediction precision increases by about 5%, which is not a very impressive result.

The ROC curves shown in Figures 3 and 4 suggest that the modified model (Figure 4) performs somewhat better than the baseline model (Figure 3). It is also notable that, as the forecast horizon lengthens, the forecasting power of both models declines.

To check my results for robustness, I estimate the standard and modified models using the data with the 2017 observations excluded and evaluate
the forecast performance of both models on data covering 2017 only. As Table 6 in the Appendix demonstrates, the forecasting power of both models is still high (AUC = 0.766 and 0.738), although the model is trained on a sample different from that used for evaluating forecast performance. It is worth noting that the forecasting power of the modified model out of sample is even higher than that of the standard model in sample. It is also noteworthy that, although its forecasting power declines substantially at the 12-month horizon, the modified model is nevertheless able to signal a deterioration in bank balance sheets one year ahead.

**Figure 3.** ROC-curve for the standard model, different forecast horizons

![ROC-curve for the standard model](image)

**Figure 4.** ROC-curve for the modified model, different forecast horizons

![ROC-curve for the modified model](image)

This therefore permits the conclusion that our main findings are robust. The modified model has a higher forecasting power than the standard model. This implies that such predictors as the excess of the deposit and loan rates over respective cross-section averages and the ratio of advertising spending to the bank's assets help predict bank defaults, especially within the next month, and thus improve the model's overall forecasting power.
5. Conclusion

In this study, I have built a probabilistic model for bank default based on microdata from Russian banks covering the period from July 2010 to December 2017. The best forecast performance was achieved when the default of a bank was forecast one month ahead. As I have shown in this paper, the forecasting power of models at the three-month forecast horizon is 5–10% weaker than at the one-month horizon.

The main finding of this study is that three novel predictors improve the quality of the forecast by 5% in terms of the ROC-AUC criterion. I conclude, therefore, that the excess of deposit and loan rates over the respective cross-section along with a decline in the ad-spending-to-assets ratio signal a bank’s poor financial standing, which is likely to result in a default.

It is remarkable that the liquidity ratio turns out to be statistically insignificant. My study is not the first study based on Russian data to arrive at this conclusion. The lack of significance of the liquidity indicator might be related to the peculiarities of interaction between credit institutions and the Bank of Russia.

My models also show that the too-big-to-fail premise (i.e. the notion that large banks are less exposed to the risk of default since their bankruptcy would mean serious losses for the economy, so the regulator has a strong incentive to bail them out) is not supported by the data.

References

Arena, M. (2008). Bank Failures and Bank Fundamentals: A Comparative Analysis of Latin America and East Asia During the Nineties Using Bank-level Data. Journal of Banking and Finance, 32(2), pp. 299–310.

Berger, A. N. and Bouwman, C. H. (2013). How Does Capital Affect Bank Performance During Financial Crises? Journal of Financial Economics, 109(1), pp. 146–176.

Boyd, J. H. and De Nicolo, G. (2005). The Theory of Bank Risk Taking and Competition Revisited. Journal of Finance, 60(3), pp. 1329–1343.

Brown, C. O. and Dinç, I. S. (2011). Too Many to Fail? Evidence of Regulatory Forbearance When the Banking Sector Is Weak. The Review of Financial Studies, 24(4), pp. 1378–1405.

Chernykh, L. and Theodossiou, A. K. (2011). Determinants of Bank Long-term Lending Behavior: Evidence from Russia. Multinational Finance Journal, 15(3–4), pp. 193–216.

Cole, R. A. and White, L. J. (2012). Déjà Vu All Over Again: The Causes of US Commercial Bank Failures This Time Around. Journal of Financial Services Research, 42(1–2), pp. 5–29.

Demirgüç-Kunt, A. and Detragiache, E. (1998). The Determinants of Banking Crises in Developing and Developed Countries. IMF Staff Papers, 45(1), pp. 81–109.

Fawcett, T. (2006). An Introduction to ROC Analysis. Pattern Recognition Letters, 27(8), pp. 861–874.
Golovan S., Karminsky M., Kopylov, A. and Peresetsky, A. (2003). Probability of Default Models of Russian Banks. I. Models with Clustering. NES Working Paper, N 39. [In Russian].

Golovan, S., Evdokimov, M., Karminsky, A. and Peresetsky, A. (2004). Probability of Default Models of Russian Banks. II. Models and Macroeconomic Environment. NES Working Paper, N 43, 2004. [In Russian].

Hanafi, M. M. and Santi, F. (2013). The Impact of Ownership Concentration, Commissioners on Bank Risk and Profitability: Evidence from Indonesia. Eurasian Economic Review, 3(2), pp. 183–202.

Hirtle, B. and Lopez, J. (1999). Supervisory Information and the Frequency of Bank Examinations. FRBNY Economic Policy Review, 4, pp. 1–20.

Hwang, D.-Y., Lee, C. F. and Liaw, K. T. (1997). Forecasting Bank Failures and Deposit Insurance Premium. International Review of Economics and Finance, 6(3), pp. 317–334.

Keeley, M. C. (1990). Deposit Insurance, Risk, and Market Power in Banking. American Economic Review, 80(5), pp. 1183–1200.

Keeley, M. C. (2011). Probability of Default Models of Russian Banks. Economic Change and Restructuring, 44(4), pp. 297–334.

Männasoo, K. and Mayes, D. G. (2009). Explaining Bank Distress in Eastern European Transition Economies. Journal of Banking and Finance, 33(2), pp. 244–253.

Martin, D. (1977). Early Warning of Bank Failure: A Logit Regression Approach. Journal of Banking and Finance, 1(3), pp. 249–276.

Micco, A., Panizza, U. and Yanez-Pagans, M. (2007). Bank Ownership and Performance: Does Politics Matter? Journal of Banking and Finance, 31(1), pp. 219–241.

Peresetsky, A., Karminsky, A. and Golovan, S. (2011). Probability of Default Models of Russian Banks. Economic Change and Restructuring, 44(4), pp. 297–334.

Peresetsky, A. and Zhivaykina, A. (2017). Russian Bank Credit Ratings and Bank License Withdrawal 2012–2016. Journal of New Economic Association, 4 (36), pp. 49–80. doi: 10.31737/rja.2017.4.3

Shen, C. H. and Hsieh, M.-F. (2011). Prediction of Bank Failures Using Combined Micro and Macro Data. International Review of Accounting. Banking and Finance, 3(2), pp. 1–40.

Zubarev, A. (2012). The Factors of Stability of Russian Banks during the Crisis of 2008–2009. Economic Policy, 4, pp. 126–142.
APPENDIX

**Applying ROC curves and AUC to assess and compare logistic models (Fawcett, 2006)**

A specific feature of logistic regression is that it does not predict strictly 1 or 0 but rather the probability that lies in the range of 0 to 1. Therefore, the user of model can set the threshold, which may be other than 0.5, depending on his/her goals, after which the model will be deemed to have predicted a certain result. After setting the threshold, model results can be represented as shown in Table 7 below, where

- TP (True Positives) are correctly predicted positive outcomes;
- TN (True Negatives) are correctly predicted negative outcomes;
- FN (False Negatives) positive outcomes predicted as negative (Type I error);
- FP (False Positives) negative outcomes predicted as positive (Type II error).

True Positives Rate: $TPR = TP + FN$.
False Positives Rate: $FPR = TN + FP$.
Sensitivity equals the true positives rate: $Se = TPR$.
Specificity is the share of true negative outcomes that are correctly identified by the model: $Sp = TN + FP$.

To draw a ROC curve, sensitivity is plotted on the $x$-axis and specificity on the $y$-axis. AUC is the area under the curve that takes values from 0.5 (random guess; the ROC curve looks like a diagonal line $y = x$) to 1 (the ideal model).

The higher AUC is, the more predictive power the model has.
### Table 2. Baseline model, default in $h$ months

|          | $h = 1$ (1) | $h = 2$ (2) | $h = 3$ (3) | $h = 4$ (4) | $h = 5$ (5) | $h = 6$ (6) | $h = 7$ (7) |
|----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Intercept | 33.5*** (8.3) | 25.9*** (8.2) | 30.2*** (8.1) | 33.2*** (8.1) | 30.9*** (8.2) | 35.0*** (8.2) | 38.2*** (8.7) |
| Capital_a | -2.269*** (0.505) | -1.156* (0.484) | -1.250** (0.481) | -1.435** (0.488) | -1.405** (0.492) | -1.385** (0.491) | -2.461*** (0.571) |
| Arrears_l | 2.439*** (0.505) | 2.103*** (0.515) | 2.200*** (0.510) | 2.001*** (0.533) | 1.716** (0.566) | 1.761** (0.555) | 0.374 (0.812) |
| Loans_a  | 1.135*** (0.300) | 1.085*** (0.289) | 1.082*** (0.289) | 0.929** (0.287) | 0.955*** (0.289) | 0.880** (0.290) | 1.082*** (0.311) |
| Ln_assets | -1.669* (0.659) | -1.014 (0.659) | -1.191 (0.648) | -1.316* (0.646) | -1.168 (0.649) | -1.189 (0.652) | -1.508* (0.693) |
| Ln_assets2 | 0.033* (0.014) | 0.020 (0.014) | 0.023 (0.014) | 0.026 (0.014) | 0.023 (0.014) | 0.023 (0.014) | 0.029 (0.015) |
| Profit_a | -17.0*** (0.9) | -11.7*** (1.2) | -6.27*** (1.68) | -4.53* (1.88) | -6.89*** (1.75) | -1.10 (2.10) | -3.53 (2.54) |
| H3       | 0.001 (0.003) | -0.004 (0.010) | 0.001 (0.003) | -0.000 (0.005) | -0.003 (0.009) | 0.002 (0.002) | 0.002 (0.011) |
| Insurance | -0.563*** (0.143) | -0.526*** (0.137) | -0.539*** (0.136) | -0.568*** (0.135) | -0.636*** (0.135) | -0.630*** (0.136) | -0.590*** (0.146) |
| Gov_foreign | -1.053*** (0.264) | -1.036*** (0.249) | -1.052*** (0.250) | -1.053*** (0.250) | -1.009*** (0.244) | -1.005*** (0.244) | -0.975*** (0.258) |
| Y_growth | -16.7*** (2.0) | -17.4*** (1.9) | -19.3*** (1.9) | -20.5*** (1.9) | -20.1*** (1.9) | -23.9*** (1.9) | -22.5*** (2.0) |
| Pledge_l | -0.622*** (0.098) | -0.456*** (0.093) | -0.527*** (0.095) | -0.543*** (0.095) | -0.409*** (0.093) | -0.453*** (0.095) | -0.549*** (0.102) |
| N        | 59 308 | 58 920 | 58 532 | 58 143 | 57 754 | 57 364 | 55 204 |
| AUC      | 0.789 | 0.737 | 0.710 | 0.719 | 0.709 | 0.717 | 0.715 |

Note: *** – $p < 0.001$; ** – $p < 0.01$; * – $p < 0.05$. 

### Table 3. Baseline model, default within $h$ months

|         | $h = 1$ (1) | $h = 2$ (2) | $h = 3$ (3) | $h = 4$ (4) | $h = 5$ (5) | $h = 6$ (6) | $h = 12$ (7) |
|---------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|
| Intercept | 33.5***     | 30.7***     | 31.2***     | 32.4***     | 32.8***     | 33.8***     | 37.1***      |
|          | (8.3)       | (5.9)       | (4.8)       | (4.2)       | (3.7)       | (3.4)       | (2.6)        |
| Capital_a | -2.269***   | -1.693***   | -1.534***   | -1.499***   | -1.475***   | -1.447***   | -1.777***    |
|          | (0.505)     | (0.353)     | (0.286)     | (0.249)     | (0.224)     | (0.205)     | (0.157)      |
| Arrears_l | 2.439***    | 2.375***    | 2.372***    | 2.323***    | 2.253***    | 2.208***    | 1.683***     |
|          | (0.505)     | (0.365)     | (0.302)     | (0.267)     | (0.245)     | (0.228)     | (0.190)      |
| Loans_a  | 1.135***    | 1.129***    | 1.117***    | 1.071***    | 1.054***    | 1.029***    | 0.971***     |
|          | (0.300)     | (0.211)     | (0.171)     | (0.148)     | (0.133)     | (0.122)     | (0.090)      |
| Ln_assets | -1.669*     | -1.360**    | -1.298***   | -1.298***   | -1.270***   | -1.251***   | -1.251***    |
|          | (0.659)     | (0.470)     | (0.383)     | (0.332)     | (0.298)     | (0.274)     | (0.205)      |
| Ln_assets2 | 0.033*      | 0.027**     | 0.025**     | 0.025***    | 0.025***    | 0.024***    | 0.024***     |
|          | (0.014)     | (0.010)     | (0.008)     | (0.007)     | (0.006)     | (0.006)     | (0.004)      |
| Profit_a | -17.0***    | -16.2***    | -14.5***    | -13.1***    | -12.5***    | -11.4***    | -8.86***     |
|          | (0.9)       | (0.8)       | (0.8)       | (0.7)       | (0.7)       | (0.7)       | (0.6)        |
| H3       | 0.001       | 0.000       | 0.001       | 0.001       | 0.000       | 0.001       | 0.008***     |
|          | (0.003)     | (0.003)     | (0.002)     | (0.002)     | (0.002)     | (0.001)     | (0.002)      |
| Insurance| -0.563***   | -0.565***   | -0.570***   | -0.581***   | -0.604***   | -0.620***   | -0.664***    |
|          | (0.143)     | (0.081)     | (0.070)     | (0.063)     | (0.058)     | (0.084)     | (0.044)      |
| Gov_foreign | -1.053***   | -1.055***   | -1.059***   | -1.063***   | -1.058***   | -1.054***   | -1.056***    |
|           | (0.264)     | (0.183)     | (0.148)     | (0.128)     | (0.114)     | (0.103)     | (0.075)      |
| Y_growth | -16.7***    | -17.1***    | -17.9***    | -18.7***    | -19.2***    | -20.2***    | -22.3***     |
|           | (2.0)       | (1.4)       | (1.1)       | (1.0)       | (0.9)       | (0.8)       | (0.6)        |
| Pledge_l | -0.622***   | -0.542***   | -0.539***   | -0.543***   | -0.519***   | -0.511***   | -0.523***    |
|           | (0.098)     | (0.068)     | (0.056)     | (0.048)     | (0.043)     | (0.040)     | (0.030)      |
| N        | 59 308      | 59 308      | 59 308      | 59 308      | 59 308      | 59 308      | 59 308       |
| AUC      | 0.789       | 0.766       | 0.747       | 0.741       | 0.736       | 0.733       | 0.728        |

Note: *** – $p < 0.001$; ** – $p < 0.01$; * – $p < 0.05$. 
Table 4. Modified model, default in $h$ months

| $h$ | Intercept | Capital_a | Arrears_l | Loans_a | Ln_assets | Ln_assets2 | Profit_a | H3 | Insurance | Gov_foreign | Y_growth | Pledge_l | Deposits_r | Loans_r | Adv_a | N | AUC |
|-----|-----------|-----------|-----------|---------|---------|-----------|----------|-----|-----------|------------|---------|---------|-----------|---------|-------|----|-----|
|     | 29.3**    | -2.887*** | 2.963***  | 1.273***| -1.354  | 0.026     | -17.7*** | -0.020| -0.835*** | -1.128***  | -16.4***| -0.861***| 0.778***  | 0.074***| -6493.4***| 52.818 | 0.846 |
| 1   | (9.9)     | (0.650)   | (0.640)   | (0.349) | (0.794) | (0.017)   | (1.1)    | (0.047)| (0.188)   | (0.295)    | (2.2)   | (0.113) | (0.093)   | (0.017) | (1170.5)| 52.818 | 0.773 |
| 2   | 18.7      | -1.111    | 2.622***  | 1.053** | -0.431  | 0.007     | -12.5***  | 0.001 | -0.704*** | -0.994***  | -17.6***| -0.609***| 0.739***  | 0.058** | -1625.3***| 52.488 | 0.730 |
| 3   | 22.6*     | -1.290*   | 2.631***  | 1.005** | -0.587  | 0.010     | -7.26***  | 0.034**| -0.749*** | -0.998***  | -19.2***| -0.640***| 0.634***  | 0.046** | -534.4  | 52.156 | 0.730 |
| 4   | 25.2**    | -1.531*   | 2.341***  | 0.784*  | -0.616  | 0.011     | -5.37*    | 0.008 | -0.767*** | -0.975***  | -21.2***| -0.641***| 0.665***  | 0.043*  | -460.7  | 51.821 | 0.743 |
| 5   | 25.9**    | -1.626*   | 1.804*    | 0.813*  | -0.652  | 0.011     | -8.03***  | 0.008 | -0.793*** | -1.003***  | -21.4***| -0.578***| 0.622***  | 0.064***| -591.0  | 51.487 | 0.729 |
| 6   | 31.4**    | -1.644*   | 2.112**   | 0.676*  | -0.730  | 0.013     | 2.06      | -0.034| -0.798*** | -0.971***  | -25.8***| -0.617***| 0.653***  | 0.049***| -638.5  | 51.152 | 0.738 |
| 7   | 34.9***   | -2.806*** | 0.750     | 0.722*  | -1.021  | 0.018     | -1.78     | -0.901 | -0.829*** | -0.903***  | -24.8***| -0.677***| 0.657***  | 0.047** | -142.3  | 49.292 |       |

Note: *** – p < 0.001; ** – p < 0.01; * – p < 0.05.
Table 5. Modified model, default within $h$ months

|          | $h = 1$ (1) | $h = 2$ (2) | $h = 3$ (3) | $h = 4$ (4) | $h = 5$ (5) | $h = 6$ (6) | $h = 12$ (7) |
|----------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|
| Intercept| 29.3***     | 25.3***     | 25.3***     | 26.1***     | 27.0***     | 28.3***     | 32.5***      |
|          | (9.9)       | (7.0)       | (5.7)       | (5.0)       | (4.5)       | (4.1)       | (3.1)        |
| Capital_a| -2.887***   | -1.925***   | -1.674***   | -1.587***   | -1.569***   | -1.529***   | -1.833***    |
|          | (0.650)     | (0.454)     | (0.369)     | (0.323)     | (0.292)     | (0.269)     | (0.209)      |
| Arrears_l| 2.963***    | 2.943***    | 2.858***    | 2.765***    | 2.607***    | 2.536***    | 1.809***     |
|          | (0.640)     | (0.453)     | (0.372)     | (0.331)     | (0.307)     | (0.287)     | (0.243)      |
| Loans_a  | 1.273***    | 1.144***    | 1.097***    | 1.005***    | 0.972***    | 0.931***    | 0.716***     |
|          | (0.349)     | (0.244)     | (0.198)     | (0.171)     | (0.154)     | (0.141)     | (0.105)      |
| Ln_assets| -1.354      | -0.930      | -0.828      | -0.788      | -0.778*     | -0.780*     | -0.796**     |
|          | (0.794)     | (0.568)     | (0.465)     | (0.404)     | (0.363)     | (0.333)     | (0.249)      |
| Ln_assets2| 0.026       | 0.017       | 0.015       | 0.014       | 0.014       | 0.014*      | 0.014**      |
|          | (0.017)     | (0.012)     | (0.010)     | (0.009)     | (0.008)     | (0.007)     | (0.005)      |
| Profit_a | -17.7***    | -17.0***    | -15.4***    | -14.0***    | -13.4***    | -12.0***    | -9.06***     |
|          | (1.1)       | (1.0)       | (0.9)       | (0.8)       | (0.8)       | (0.7)       | (0.64)       |
| H3       | -0.020      | -0.010      | 0.020       | 0.017       | 0.013       | 0.006       | -0.020       |
|          | (0.047)     | (0.028)     | (0.012)     | (0.012)     | (0.012)     | (0.012)     | (0.014)      |
| Insurance| -0.835***   | -0.784***   | -0.791***   | -0.803***   | -0.822***   | -0.840***   | -0.941***    |
|          | (0.188)     | (0.132)     | (0.108)     | (0.094)     | (0.084)     | (0.077)     | (0.058)      |
| Gov_foreign| -1.128***  | -1.071***   | -1.052***   | -1.038***   | -1.037***   | -1.030***   | -1.002***    |
|          | (0.295)     | (0.201)     | (0.161)     | (0.139)     | (0.125)     | (0.114)     | (0.083)      |
| Y_growth | -16.4***    | -17.3***    | -18.1***    | -19.1***    | -19.8***    | -20.9***    | -23.9***     |
|          | (2.3)       | (1.6)       | (1.3)       | (1.1)       | (1.0)       | (0.9)       | (0.7)        |
| Pledge_l | -0.861***   | -0.747***   | -0.717***   | -0.703***   | -0.684***   | -0.677***   | -0.647***    |
|          | (0.113)     | (0.079)     | (0.065)     | (0.056)     | (0.050)     | (0.046)     | (0.034)      |
| Deposits_r| 0.778***   | 0.805***    | 0.790***    | 0.807***    | 0.815***    | 0.836***    | 0.959***     |
|          | (0.093)     | (0.072)     | (0.064)     | (0.060)     | (0.057)     | (0.055)     | (0.048)      |
| Loans_r  | 0.074***    | 0.073***    | 0.071***    | 0.070***    | 0.080***    | 0.084***    | 0.082***     |
|          | (0.017)     | (0.014)     | (0.013)     | (0.011)     | (0.011)     | (0.011)     | (0.011)      |
| Adv_a    | -6493.4***  | -3216.1***  | -1911.3***  | -1420.1***  | -1228.2***  | -1119.4***  | -400.2***    |
|          | (1170.5)    | (575.9)     | (374.6)     | (289.8)     | (247.0)     | (219.4)     | (126.1)      |
| N        | 52 818      | 52 818      | 52 818      | 52 818      | 52 818      | 52 818      | 52 818       |
| AUC      | 0.846       | 0.809       | 0.782       | 0.772       | 0.766       | 0.762       | 0.756        |

Note: *** – p < 0.001; ** – p < 0.01; * – p < 0.05.
### Table 6. Analysis of robustness, default within $h$ months

|                  | In-sample (with 2017) |                        | Out-of-sample (without 2017) |                        |
|------------------|-----------------------|-------------------------|-----------------------------|-------------------------|
|                  | Baseline              | Modified                | Baseline                    | Modified                |
|                  | $h = 1$               | $h = 12$                | $h = 1$                     | $h = 12$                |
|                  | $(1)$                 | $(2)$                   | $(3)$                       | $(4)$                   |
| Intercept        | 33.5***               | 29.3**                  | 32.5***                     | 30.5**                  |
|                  | (8.3)                 | (9.8)                   | (3.1)                       | (10.0)                  |
| Capital_a        | -2.269***             | -2.887***               | -1.833***                   | -2.453***               |
|                  | (0.505)               | (0.650)                 | (0.209)                     | (0.557)                 |
| Arrears_l        | 2.439***              | 2.963***                | 1.809***                    | 2.444***                |
|                  | (0.505)               | (0.640)                 | (0.243)                     | (0.566)                 |
| Loans_a          | 1.135***              | 1.273***                | 0.716***                    | 1.084***                |
|                  | (0.300)               | (0.349)                 | (0.105)                     | (0.318)                 |
| Ln_assets        | -1.669*               | -1.354                  | -0.796**                    | -1.222                  |
|                  | (0.659)               | (0.794)                 | (0.249)                     | (0.811)                 |
| Ln_assets2       | 0.033*                | 0.026                   | 0.014**                     | 0.023                   |
|                  | (0.014)               | (0.017)                 | (0.005)                     | (0.017)                 |
| Profit_a         | -17.0***              | -17.7***                | -9.06***                    | -16.9***                |
|                  | (0.9)                 | (1.1)                   | (0.64)                      | (1.0)                   |
| H3               | 0.001                 | -0.020                  | -0.020                      | 0.001                   |
|                  | (0.003)               | (0.047)                 | (0.014)                     | (0.003)                 |
| Insurance        | -0.563***             | -0.835***               | -0.941***                   | -0.603***               |
|                  | (0.143)               | (0.188)                 | (0.058)                     | (0.150)                 |
| Gov_foreign      | -1.053***             | -1.128***               | -1.002***                   | -1.249***               |
|                  | (0.264)               | (0.295)                 | (0.083)                     | (0.308)                 |
| Y_growth         | -16.7***              | -16.4***                | -23.9***                    | -18.4***                |
|                  | (2.0)                 | (2.2)                   | (0.7)                       | (2.1)                   |
| Pledge_l         | -0.622***             | -0.861***               | -0.647***                   | -0.637***               |
|                  | (0.098)               | (0.113)                 | (0.034)                     | (0.105)                 |
| Deposits_r       | 0.778***              | 0.959***                | 0.769***                    | 1.023***                |
|                  | (0.093)               | (0.048)                 | (0.097)                     | (0.051)                 |
| Loans_r          | 0.074***              | 0.082***                | 0.072***                    | 0.080***                |
|                  | (0.017)               | (0.011)                 | (0.018)                     | (0.011)                 |
| Adv_a            | -6493.4***            | -400.2***               | -5531.6***                  | -252.7*                 |
|                  | (1170.5)              | (126.1)                 | (1142.1)                    | (125.3)                 |
| N                | 59 308                | 52 818                  | 53 808                      | 47 907                  |
| AUC_2017         | 0.743                 | 0.791                   | 0.644                       | 0.732                   |

Note: *** – $p < 0.001$; ** – $p < 0.01$; * – $p < 0.05$.

### Table 7. Actual outcomes vs. model predictions

| Model prediction | Actual data |
|------------------|-------------|
|                  | 1           | 0           |
| 1                | TP          | FP          |
| 0                | FN          | TN          |