CHAPTER ELEVEN

INCREASE OF BANKS' CREDIT RISK FORECASTING POWER USING THE SET OF CREDIT RATINGS AND PROBABILITY OF DEFAULT MODELS

ELLA KHROMOVA

Abstract

The aim of this paper is twofold: first, to compare the divergence of credit ratings (CR) and the probability of default (PD) models of Russian banks; and second, to create a synergic reliable model. The research showed that there is a significant divergence in the predictions of CR and PD models: CR models tend to overestimate the probability of financial disease of a bank, whereas PD models provide underestimated results. Moreover, the paper introduces the process of derivation of single scale CR and PD econometric models for Russian banks, based on a 2008-2018 database. The usage of the synergic model of CR and PD has improved banks' credit risk forecasting power, comparing the separate CR and PD models. As a result, the percentage of predictions which fall into the one-point interval near the actual value increased by more than 15%, while the percentage of forecasts with less than three rating grades' deviation in a 21-grade rating scale reached 88%.

Keywords: banks; credit ratings; probability of default; ordered logit and probit models; synergic models

1. Introduction

The banking system is one of the main parts of the economy, as banks are key players in many financial markets. Numerous crises have proven that defaults by several banks can have a dramatic effect on the global banking system, and, afterwards, on the global economy. Furthermore, those crises have shown that the existing measures of banks' financial stability are not very reliable or useful. Therefore, a lot of research nowadays is aimed at improving instruments for analyzing banks' stability.

Currently, the most popular and reliable instruments for measurement of banks' financial stability are credit ratings (CR)and estimation of probability of default (PD). Rating agencies (RAs) act as independent experts, and assign CRs in symbolic form, where each grade represents a different level of a bank's financial stability, while the PD model gives a precise estimation of the probability of a bank becoming insolvent. Despite the fact that these measures are the most frequently used, there are still disadvantages to both. The main disadvantage is connected with over- and underfitting issues. Ratings given after the credit crunch of 2007 are considered too conservative, as rating agencies value their reputation. At the same time, probabilities of default are, on the contrary, too optimistic, due to the low number of observed defaults (unbalanced sample).

This aim of this paper is to construct models for estimating the two most popular measures of financial stability (CR and PD), based on publicly available information, and, secondly, to try to combine those two measures in order to achieve higher accuracy of estimation and higher forecasting power. The final model of this paper will be constructed on the basis of the CR model, combined with transformed ratings received from the PD model. Afterwards, this model will be compared with separate models.

The paper is based on the database of Central Bank of Russia, which consists of RAS statements of all Russian licensed banks for the period of 2008-2018.In addition, data on macroeconomic factors is taken from Rosstat and World Bank databases.

This paper is structured as follows: The first section contains a literature review of articles connected with CR and PD models, and states the main hypotheses. In the next section, the dataset will be described, and attention paid to the mapping of rating scales. In addition, the problem with an unbalanced sample for the probability of default model is emphasized. In the third section, binomial and multinomial ordered logit models are estimated separately (for PD and CR models, respectively) with the use of step-wise procedure and PCA analysis, and their predicted values are compared with actual figures, in order to measure their separate forecasting power. Finally, in section four, the combined synergic model is estimated and its out-of-sample forecasting power is measured in order to prove the initial hypothesis. Basing on these findings, conclusions are made, and further perspectives of the analysis are highlighted.

2. Literature review

The paper considers two seemingly separate areas of economic literature: underestimation of credit risk by default models, and overcautious assignments of credit ratings.

A lot of recent research pays much attention to the presence of the class imbalance problem in data on defaults and its influence on the estimation procedure. (Esarey and Pierce 2012, Karminsky and Khromova 2018, Karminsky and Kostrov 2017, Lanine and Vennet 2006). The main consequence of this problem is the underestimation of the 'rare' ('default') class, which leads to a deterioration of forecasting power for bank failures (Florez-Lopez and Ramon-Jeronimo 2014, Rösch and Scheule 2014). Garcia et al. (2012) discuss the class imbalance problem and methods to overcome it. Among the mostly used methods are random omission of non-defaults, random inclusion of defaults, and increase in weights of the rare class observations in a log-likelihood function. However, even the balancing data methods provided do not fully solve this problem, and a model gives underestimated results.

On the other hand, rating models are not fully reliable either. There exists a long-time tendency to estimate the differences between the ratings assessments of different RAs. Despite the fact that many rating agencies use similar letter designations, the approaches to financial analysis differ among them. It was observed that the rating agency Standard &Poor's approach is more cautious and conservative when evaluating the financial stability of banks, compared with its two largest competitors Fitch, and Moody's. Also, it was revealed that Moody's approach to the assessment of banking risks is the most liberal (Karminsky and Peresetsky 2011, Karminky and Khromova 2016). Many authors studied a consistent difference between the scores of the various rating agencies and the financial stability of corresponding banks (Morgan 2002). It was found that, previously, the activity of rating agencies has had little regulation, allowing them to avoid responsibility for inaccuracies (overestimation) in assigned ratings, while investors were suffering huge losses (Solovjova 2016). Santoni and Arbia (2013) noted that the reputation of RAs has steadily deteriorated due to some notable failures (Enron, WorldCom, Parmalat) and the subprime crisis (2007-2009). However, in the most recent times, it was shown that RAs are very cautious in their estimation of banks' financial stability, as their reputation fully depends on it. The reputation of a RA suffers more when it predicts a higher rating grade than it should. Therefore, nowadays, RAs tend to react sharply to any bad news from a well-performing bank, trying to predict the worst scenario of its performance, because for the RA, it is better to reassign the rating to a higher grade a bit later, than not to capture the worsening of financial performance and lose its reputation. It was shown thatafter the credit crunch of 2007, the ratings models have the property of overestimation of financial instability of a bank, and this gets worse if we apply the model out-of-sample (Karminsy and Khromova 2016).

Therefore, observing the divergence of ratings and PD modelling, an idea has occurred to different researchers, to combine these two forecasts in order to increase the power to predict the financial instability of a bank. Note that these two approaches give exactly opposite skewness of their predictors, which makes their combination even more reliable. For example, in 2007, Godlewski provided comparison of banks' credit ratings in emerging countries, and their corresponding probabilities of default. The research showed that the rating tends to aggregate banks' default risk information into intermediate-low rating grades, and thus caused ratings' partial divergence with the results of a PD scoring model. Following that, in 2016, Pompella and Dicanio introduced a new approach (the PC-Mahalanobis Method), which takes parts from PD and credit rating modelling, to test the validity of bank ratings assigned by rating agencies. However, the PC-Mahalanobis Method doesn't provide numerical interpretation of results and allows only the determination of whether an observed subject belongs to any of the two binary groups: healthy, or likely-to-fail, banks. In contrast, this research provides a method of forecasting the exact rating of a bank, with a 21-dimensional accuracy. Therefore, following a new literature stream, the main hypothesis of this paper is stated:

> *A combination of the credit rating model with the probability of default model would give higher forecasting power than those models separately.*

In order to check this hypothesis, this paper uses an algorithm of creation of a synergic model that was applied to the rating model and probability of default model of Russian banks by Karminsky and Khromova, in 2018.

The algorithm of this paper includes several steps. The first step was to study and compare different methodologies of banks' credit risk measurement, which are subdivided into Basel estimates (probability of default, loss given default, earnings at default, maturity), and ratings' estimates (IRB-approach, national RA, external RA). PD and CR models were chosen as a set of alternative models that would be considered in this research. In order to make a joint rating environment model, all national RA and external RA assessments were normalized into the base scale. The next step was to construct the PD and CR models separately, on the same dataset, using the basic rating scale adjustment provided by Karminsky and Sosurko in 2011. This part of the research is based on the review of factors of potential influence on credit risk of a bank which was summarized in the previous paper of the authors (Karminsky and Khromova2016). After the predicted values of both models were generated, calibration of ratings and PD by the methodology of Pomasanov and Vlasov (2008) was realized, in order to bring ratings and PD into a single scale. Then the forecasting errors of each model were compared by the descriptive statistics parameters of their distributions (mode, median, skewness). The divergence of both models from the perfect forecast was realized, and the optimal weight coefficients and monotonic transformations for these two models, which bring the forecasting errors' distribution closer to a normal distribution, were calculated. The obtained synergic model that consists of the set of alternative models was further checked for its out-of- sample fit.

3.Data

Firstly, it is important to understand the time period for the whole dataset. The last large change in the banking sector was after the credit crunch in 2007, when rating agencies raised the criteria for banks to receive high ratings and became more conservative in all their actions. Banks themselves started to suffer from stricter regulation, and there were an abnormal number of defaults in those years due to systematic factors. Because of that, it is reasonable to study data which was collected after 2007.

The models in this paper are constructed on the basis of data collected from the beginning of 2008 to the first quarter of 2018. In total, there are 41 quarterly periods. This allows us to achieve a most reasonable balance between using large amounts of data and focusing on most recent observations.

Now we proceed to the collection of data needed for the CR model. A dependent variable is a rating given by one of the top-three international agencies (Moody's, S&P, and Fitch) and the largest Russian agency –the Expert Rating Agency. Here, it should be noted that most of them have several rating types and scales (national, international, etc.). This paper will take into account all long-term ratings, on international and national scales. Data from other Russian rating agencies was omitted, as the number of ratings given in the period of our interest is too small (fewer than 100 ratings during the whole period), while process of scales mapping needs large amount of data in order to be accurate. Most of the data on ratings was taken from Thomson Reuters and CBonds. In the first selection, there were about 700 Russian banks which had at least one rating from the rating agencies of our interest in the last 41 quarters.

It is important to note that existing rating grades were extended for periods with no rating, in order to receive correct and full panel data. This can be reasonably done, as rating agencies do not need to reassign ratings every quarter, and, if a rating is given in a previous time period, we can assume that it is unchanged, given that there are no newer ratings for this bank.

After downloading data, mapping scales, and switching of letter-scale into numerical were completed. The mapping of rating scales in this paper is based mainly on the findings in the article by Karminsky and Sosurko (2011). The same modelling principles are applied to the newer data, in order to receive mapping which can be used for recent observations. After applying these principles, Table2-1 was created.

Table 2-1 Ratings mapping to a base scale

	Fitch Ratings – Long-term international rating in foreign currency	Fitch Ratings – Long-term international rating in national currency	Fitch Ratings – National scale (Russia)	Moody's Investors Service – Long-term international rating in foreign currency	Moody's Investors Service – Long-term international rating in national currency	Moody's Interfax Rating Agency – National scale (Russia)	S & P Global ratings – Long-term international rating in foreign currency	S & P Global ratings – Long-term international rating in national currency	S & P Global ratings – Long-term rating, national scale (Russia)	Expert Rating Agency – National scale (Russia)
8.0					Baa1	Baa1	BBB+	BBB+		
8.5	BBB+	BBB+	AAA(rus)				BBB	BBB	ruAAA	
9.0				Aaa.ru	Baa2	Baa2				
9.5	BBB	BBB					BBB-	BBB-		
10.0	BBB-	BBB-	AA+(rus)		Baa3	Baa3			ruAA+	A++
10.5							BB+	BB+		
11.0	BB+	BB+	AA(rus)	Aa1.ru	Ba1	Ba1	BB	BB	ruAA	
11.5	BB	BB	AA-(rus)							
12.0				Aa2.ru	Ba2	Ba2	BB-	BB-	ruAA-	
12.5	BB-	BB-	A+(rus)						ruA+	
12.5			A(rus)							
13.0	B+	B+	A-(rus)	Aa3.ru	Ba3	Ba3	B+	B+		
13.5			BBB+(rus)	A1.ru					ruA	A+
14.0	B	B	BBB(rus)	A2.ru	B1	B1	B	B	ruA-	
14.5	B-	B-	BBB-(rus)				B-	B-	ruBBB+	
15.0	CCC+	CCC+	BB+(rus)	A3.ru	B2	B2			ruBBB	
15.5			BB(rus)	Baa1.ru			CCC+	CCC+	ruBBB-	A
15.5			BB-(rus)						ruBB+	
16.0	CCC	CCC	B+(rus)	Baa2.ru	B3	B3			ruBB	
16.0				Baa3.ru						
16.5			B(rus)	Ba1.ru			CCC	CCC	ruBB-	
17.0				Ba2.ru					ruB+	
17.0			B-(rus)		Caa1	Caa1	CCC-	CCC-	ruB	
17.5				Ba3.ru						B++
17.5				B1.ru					ruB-	
18.0				B2.ru						
18.0	CC	CC		B3.ru	Caa2	Caa2				
18.5				Caa1.ru						B+
19.0	C	C	CCC(rus)	Caa2.ru	Caa3	Caa3				
19.5				Caa3.ru						B
20.0				Ca.ru	Ca	Ca				C++
20.5										C
21.0	D	D	D(rus)	C.ru	C	C	D	D	ruD	E

In this table, a value of8.0corresponds to the highest possible rating in Russia, that is, equal to the maximum sovereign rating for the period of our interest, while 21.0corresponds to the worst rating, which is assigned to insolvent banks.

Now, we switch to collecting data for the dependent variable of the probability of default model. As the default is the discrete event, the observable dependent variable can take only two values: 0 when there is no default, and the bank is still operating; and 1, when default has occurred.

This data can be easily found in open sources on the internet, such as the Banki.ru website. All in all, there were about 400 banks which defaulted in this period, while only 190 had at least one rating given by agencies of our interest. So, we only had those 190 banks left, because, for our final model (a combination of the CR and PD models) it is important that the set of banks should be the same.

There is an imbalanced data problem connected with probability of default models, arising due to the comparatively low number of defaults which occurred. In order to improve the situation, a random addition of defaults and extraction of non-defaults was applied, following the steps of He and Garcia (2009).

Observations of defaults were also used for constructing a ratings model. There is a need for this, as rating agencies rarely assign the worst-case rating (in our setting this is 21) after bank has defaulted. This can be explained by the fact that all ratings assigned need to be paid for, but the defaulting bank has no money for that. This causes the following problem: banks, which have had some positive rating in the past, and have defaulted sometime after that rating assignment, were given extensions to have the same initial rating in all periods, even after default. This, of course, causes a large bias in our models, and it was decided to assign a rating equal to 21, if we know that the bank has defaulted in that (or the previous) period. However, a lack of transition to high default group ratings still leads to low forecasting power for PD and CR models.

Then the data for explanatory variables was collected. For calculating those factors, data from banks' financial statements (from CBR website, forms 101-102) are used. Using the following data, explanatory variables were calculated. Explanatory variables for both models can be calculated as follows, in Table 2-2 (the right-hand table represents accounts from financial statements numbered in the left-hand table).

Table 2-2 Data for explanatory variables, Formulas for explanatory variables

#	Data
N1	Impaired loans
N2	% of capital held by investors
N3	% of impaired loans
N4	% of reserves for impaired loans
N5	Reserves for impaired loans
N6	Total assets
N7	Loans
N8	Net assets (w/o reserves)
N9	Cash and equivalents
N10	Deposits
N11	Share capital
N12	Net interest income
N13	Interest income
N14	Interest expenses
N15	Operational income
N16	Operational expenses
N17	Net profit/loss
N18	Net interest margin
N19	Inflation
N20	GDP growth rate
N21	GDP per capita
N22	Corruption Perception Index
N23	Sovereign rating

Var	Factor	Formula
X1	Return on assets	N17/N6
X2	Return on equity	N17/N11
X3	Net interest margin	N18
X4	Net profit with reserves / Assets	(N17+N5)/N8
X5	Interest income / Assets	N13/N6
X6	Interest expenses / Interest income	N14/ N13
X7	Operating expenses / Operating income	N16/N15
X8	Current ratio	(N7+N9)/N10
X9	Deposits / Equity	N10/N11
X10	Net assets / Deposits	N8/N10
X11	Liquid assets / Deposits	N9/ N10
X12	Loans / Deposits	N7/N10
X13	Tier I ratio	(N11+N17)/N8
X14	Equity / Assets	N2
X15	Impaired loans / Loans	N3
X16	Loan loss reserves / Loans	N4
X17	Loan loss reserves / Assets	N5/N6
X18	Impaired loans / Equity with LLR	N1/(N11+N5)
X19	Unreserved impaired loans / Equity	(N1-N5)/N11
X20	Loans net of reserves / NII	(N7-N5)/N12
X21	Ln (Assets)	ln(N6)
X22	Inflation	N19
X23	GDP growth rate	N20
X24	GDP per capita	N21
X25	Corruption Perception Index	N22
X26	Sovereign rating	N23

Data was aggregated from the site of the Russian Central Bank and its mobile database. Those databases include accounts for more than 1200 banks, but still, not all banks from our previous sample (banks with ratings in at least one period) were found. After matching those two samples, a pre-final sample was made, which included about 400 Russian banks. For those banks we have both dependent variable (rating and default), and explanatory variables (financial data).

Later, banks which had low amounts of financial data (less than 50% of the data needed in periods prior to default) and which were nationalized, were excluded. Nationalized banks were excluded, because the CR model in this paper aims to estimate stand-alone ratings, where government support is not included in the rating calculation. After all the modifications, we received the final sample, which included 338 banks.

All in all, the final sample includes 41 time periods and 338 banks (13,858 observations in total), each with ratings extended to all 41 periods, and data from financial statements on 26 factors.

4.Modelling

In order to achieve the goal of this paper, and to construct the final combined model, it is firstly necessary to construct each model (CR and PD) separately. This is needed to achieve the highest significance of each model, which, in turn, would increase the significance and the forecasting power of the final model.

4.1. Modelling Credit Ratings Separately

CRs are usually constructed using an ordered logit/probit model, due to the distribution of the dependent variable. Probit and logit models differ in their assumptions about the distribution of target variables. Probit assumes normal distribution of the target variable, while logit is based on the usage of the natural logarithm. Both models have certain drawbacks: logit models are very sensitive to multicollinearity of variables, whereas probit models are sensitive to the normality of distribution. However, logistic modelling is inherently more reliable for stable periods of economic development (Karminsky and Kostrov 2017). Based on these facts, logistic models are chosen as the basis for models in this paper.

Three different sets of explanatory variables were considered. Model 1 contained only financial variables received after the step-wise procedure, and was represented as follows:

$$Rat_{it} = b_1 X2_{it} + b_2 X3_{it} + b_3 X7_{it} + b_4\, X8_{it} + b_5\, X13_{it} + b_6\, X15_{it} + b_7\, X17_{it} + b_8\, X21_{it}$$

In Model 2, the macroeconomic coefficients for explaining systematic factors, and quadratic terms for considering the decreasing marginal effect of some variables, were added.

In Model 3, Principal Component Analysis (PCA)was applied to macroeconomic variables (Hotelling 1933). This was needed largely due to the high correlation between all the macroeconomic variables, and the large number of variables in total. All three models, after the step-wise procedure, are summarized in the Table 3-1below.

Table 3-1 Models for estimation of CR

	Model 1	Model 2 (with macro & squared)	Model 3 (with PCA)
X2	0.047**	0.045**	0.098**
X3	-2.960***	-2.624***	-2.351***
X7	-0.001		
X8	0.004***	0.004**	0.001***
X13	0.603***	0.552***	0.753****
X15	-1.359***	-1.392***	-1.458***
X17	1.580***	0.987*	1.197**
X21	-0.064*	1.167***	1.481***
$X7^2$		0.000	
$X21^2$		-0.040***	-0.067***
X22		-1.636	
X26		0.080***	
X24		-0.0001	
PC1			0.798***
PC2			-1.359***
LogL	-15,486.21	-15,469.67	-15,375.95
AIC	31,040.43	31,015.36	30,968.68
BIC	31,291.43	31,295.88	30,752.37

** – significant at 10%, ** – significant at 5%, *** – significant at 1%.*

From the results obtained, we can see that Model 3 performs better, according to the smallest AIC and BIC criteria. Most of the coefficients of Model 3 are significant and have low correlation.

4.2. Modelling Probabilities of Default separately

Same modelling techniques could be applied to the probability of default model, however, a binomial logit model was applied instead of a multinomial one. The same three types of model were considered. For the PD model, macroeconomic variables are even more important, as the occurrence of default strongly depends on the systematic risks and operational environment. The results are summarized in Table 3-2.

Table 3-2 Models for estimation of PD

	Model 1	Model 2 (with macro & squared)	Model 3 (with PCA)
X3	-10.353**	-11.515**	-12.358***
X5	12.130***	7.577**	10.349***
X6	-3.713***	-1.557***	-2.049***
X8	-0.001*	-0.001**	-0.002**
X10	0.0003**	0.0003**	0.0002*
X15	-17.867***	-18.738***	-19.354***
X17	34.144***	22.924***	22.142***
X21	0.739***	3.783*	
$X21^2$		-0.133*	
X22		-17.186***	
X26		2.127***	
X24		-0.001**	
X23		12.443***	
X25		10.097**	
PC1			0.972***
PC2			-2.358***
CONST	-25.742	-36.027*	-24.781***
LogL	-802.43	-682.52	-628.98
AIC	1,624.86	1,397.04	1,321.25
BIC	1,698.55	1,514.95	1,479.11
** – significant at 10%, ** – significant at 5%, *** – significant at 1%*			

Moreover, all three models were applied to different datasets with unbalanced data correction methods introduced by He and Garcia in 2009. In the first data sample (balanced-)10% of non-defaulting banks was randomly reduced. The second data sample, called balanced +, was generated by randomly adding observations with a dependent variable equal to 1 into our initial sample. The estimation results are systemized in Table 3-3.

Table 3-3 Comparison of PD models

		Model 1	Model 2 (with macro & squared)	Model 3 (with PCA)
Initial	LogL	-802.43	-682.52	-628.98
	AIC	1,624.86	1,397.04	1,321.25
	BIC	1,698.55	1,514.95	1,479.11
Balanced -	LogL	-796.24	-675.70	-642.31
	AIC	1,612.48	1,383.41	1,301.74
	BIC	1,685.40	1,500.08	1,419.38
Balanced +	LogL	-672.48	-603.98	-638.74
	AIC	1,598.34	1,299.36	1,270.77
	BIC	1,587.67	1,359.27	1,319.74

From AIC/BIC and information about the significance and correlation of variables, we can conclude that Model 3, estimated on the Balanced (-) sample, has the highest significance among all three models. Therefore, this model was chosen for further computations.

4.3. Forecasting and Measuring the Forecast Power of Separate Models

In order to construct a synergic model, we need to convert the estimated probabilities of default into numerical ratings. The transition matrix of PDs to CRs was taken from the paper by Pomasanov et al. (2008). This table was extrapolated, and resulted in Figure 3-1, which shows the correspondence of PDs to our base scale of CRs.

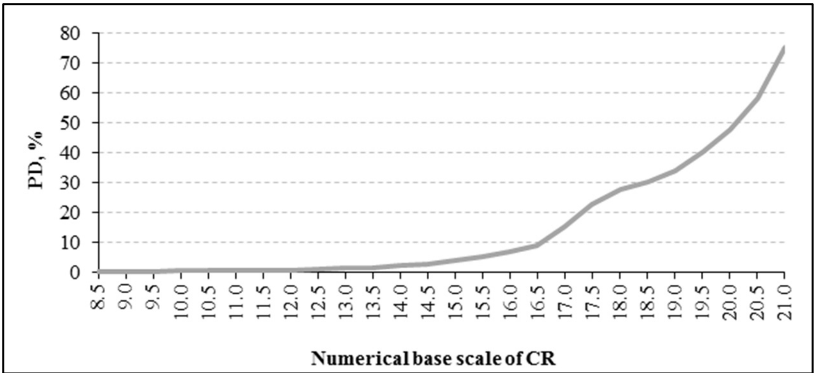

Figure 3-1 PDs corresponding to numerical ratings

Using this information, we are now able to calculate deviations of ratings received from predicted default probabilities, and actual average ratings for each bank in each period. Therefore, we can compare forecast errors of PD and CR models in the same scale. This can be seen in Figure 3-2, where black bars correspond to the CR model, while grey ones are from the PD model.

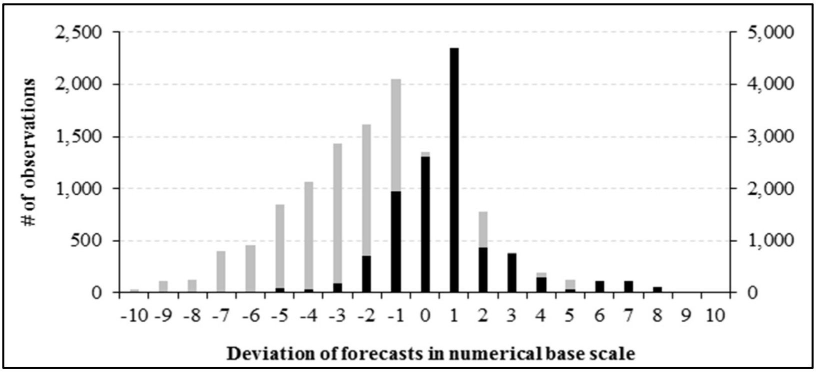

Figure 3-2 Forecast error for separate CR (black) and PD (grey) models

The histogram shows that the PD model has lower forecasting power, as there is a large number of serious deviations from actual values, and the range of them is significantly higher than in the model for ratings. This is consistent with our expectations and can be explained by problems of imbalanced data. Therefore, we can expect lower significance of the PD explanatory variable in the final combined model. Moreover, the distribution is skewed to the left (showing PD underestimation), which again proves the practical applicability of a combination of two models.

Another thing to mention here is the skewed nature of the CR histogram of deviations to the right. That means that the assumption about overfitting issues of rating models is proven. It is so, because predicted ratings tend to be one or more points greater than actual ones, while a larger rating means poorer financial stability of a bank. These are the main reasons why this paper is aimed at combing the CR, with the PD model.

5. Combined Model Estimation and Forecasting

The combined model can be a linear or a non-linear combination of ratings and PDs estimated previously. The general form of the synergic model is as follows:

$$Y_{it} = c + a_1 \text{ x } Rat_{it} + a_2 \text{ x } Rat\_PD_{it},$$

Where Y_{it} is the actual rating, Rat_{it} is a fitted value of rating estimated from the credit rating model, and Rat_PD_{it} is a fitted value of PD model transformed to the base rating scale. As was noted earlier, ratings received by transforming PDs are expected to have lower influence on the dependent variable (smaller coefficient).

To choose the best combined model, two types of model were estimated, and compared by information criteria:

- ordered logistic estimation of actual ratings on ratings predicted by both models;
- ordered logistic estimation of actual ratings on ratings predicted by both models and squared ratings estimated by CR model.

Table 4-1 Synergic models of credit risk

	Ordered logistic	Ordered logistic with Rat^2
Rat_PD	0.0116	0.0148**
Rat	3.2574***	57.9420***
Rat^2		-1.7423***
AIC	5,090.55	5,086.15
BIC	5,110.50	5,106.46
** – significant at 10%, ** – significant at 5%, *** – significant at 1%.*		

From Table 4-1, we can see that the significance of coefficients increases after adding squared ratings. This can be explained by the presence of the decreasing marginal effect of ratings. Therefore, the final combined model is chosen as an ordered logistic estimation of actual ratings on fitted values from the CR and PD models and squared fitted values of the CR model. All coefficients are significant, and both estimated ratings have coefficients with positive signs, which is consistent with expectations.

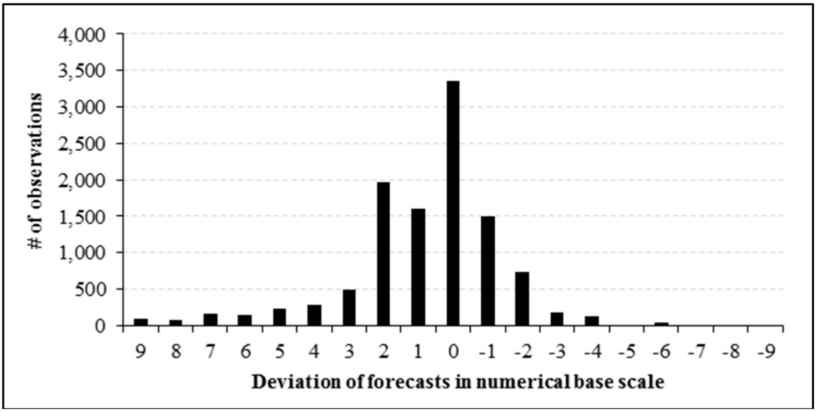

Figure 4-1 Forecast error for the combined model

From Figure 4-1, it can be noticed that the mode is 0 and the overall distribution is similar to normal, as it has less skewness and flatter tails. That shows that the combined model gives better predictions than two separate models, which is consistent with the main hypothesis of this paper.

Next, we proceed to testing the out-of-sample forecasting power of this model. For doing this, the model should be estimated on the whole dataset, but for time periods up to 29 only, so that three years (12 quarters) are left 'out-of-sample'. Next, predictions can be made for those time periods. Differences between these predicted values and actual ratings are summarized in the table below and showed in comparison with those figures received from the out-of-sample fit of the original credit rating model.

Table 4-2 Comparison of the forecasting power

Original CR model		Combined model	
Difference	% of observations	Difference	% of observations
<1	21.93%	<1	37.53%
<2	66.39%	<2	73.45%
<3	79.81%	<3	88.00%
3+	20.19%	3+	12.00%

From the table above, it can be concluded that the main goal of the paper – the increase in the forecasting power (out-of-sample) – was achieved. Comparing it to the separate CR model, the percentage of predictions which fall into the one-point interval near the actual value is increased by more than 15%, while percentage of observations which fall into the tree points interval near the actual value has reached 88%.

6. Conclusion

This paper focused on constructing models for banks' credit ratings and default probabilities, and, afterwards, combining them. From the results achieved, it can be argued that the main hypothesis was proven, and the combined model, which increases the forecasting power, was introduced. The combined model predicted more than 37% out-of-sample rating grades, with a forecasting error lower than 1 rating grade, out of a 21-dimensional base rating scale.

It is important to mention that this paper aimed to introduce a methodology, rather than a single final model. Further research may improve the model using more complex techniques of CR modelling, such as the artificial intelligence model, and correction of imbalances in the datasets of PD models. Moreover, qualitative explanatory variables can be added, in order to increase the forecasting power of both separate and combined models.

All of this means that there is a place for improvements in estimating the financial stability of a bank. But the methodology introduced in this paper will have a strong influence on future researches on this topic, as it solves natural biases of CR and PD models and can be successfully applied with other sets of data, models and variables.

References

Esarey, J., & Pierce, A. (2012). Assessing fit quality and testing for misspecification in binary-dependent variable models. *Political Analysis*, 20(4), 480–500.

Florez-Lopez, R., Ramon-Jeronimo, J.M. (2014). *The Journal of the Operational Research Society*, 65(3), 416-434.

Godlewski, C.J. (2007). Are Ratings Consistent with Default Probabilities? Empirical Evidence on Banks in Emerging Market Economies. *Emerging Markets Finance & Trade*, 43(4), 5-23.

He, H., Garcia, E. (2009), Learning from Imbalanced Data, *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441.

Karminsky A, Khromova E. (2018). Increase of banks' credit risks forecasting power by the usage of the set of alternative models. *Russian Journal of Economics,* 4(2), 155-174.

—. (2016). Extended Modelling of Banks' Credit Ratings, *Procedia Computer Science*, (91), 201-210.

Karminsky, A.M., Kostrov, A.V. (2017). The back side of banking in Russia: forecasting bank failures with negative capital. *International Journal of Computational Economics and Econometrics*, 7(1/2), 170-209.

Karminsky, A.M., Sosurko, V.V. (2011). The unified rating mapping: a step from the myth to reality. *Bankovskoe delo (Banking)*, 6, 58-63.

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

Peresetsky, A.A., Karminsky, A.M. (2011). Models for Moody's Bank Ratings. *Frontiers in Finance and Economics*, 8(1), 88-110.

Pomasanov, M., Vlasov, A. (2008). Calibration of national ratingsystems.*Rynokcennichbumag (Security market)*, 74-79 (in Russian).

Pompella, M., Dicanio, A. (2016). Ratings based Inference and Credit Risk: Detecting likely-to-fail Banks with the PC-Mahalanobis Method. *Economic Modelling*, 67, 34–44.

Rösch, D., Scheule, H. (2014). Forecasting probabilities of default and loss rates given default in the presence of selection. *The Journal of the Operational Research Society*, 65(3), 393-407.

Santoni, A., Arbia, B. (2013). Assessing Rating Agencies' Ability to Predict Bank Bankruptcy – The Lace Financial Case. *Economic Notes*, 42(2).

Solovjova, I. (2016). New Approaches to Regulating the Activities of Rating Agencies: a Comparative Analysis. *Procedia, Social and Behavioral Sciences*, 229, 115 – 125.

Authors note

Ella Khromova
National Research University Higher School of Economics, Shabolovka 26, Moscow, 119049, Russia
e-mail: ehromova@hse.ru
tel: +7 (916) 803 92 08
The author is grateful to acknowledge the help with data collection as well as overall support and guidance provided by Alexander M. Karminsky and Yaroslav Bulatov.