Predicting Failures of Russian Banks Using Parametric and Non-parametric Techniques

Maximilian Negele
University College Roosevelt
maximilian.negele@gmail.com

ABSTRACT

In this paper I compare three methods to predict bank failures in the Russian banking sector. Based on data ranging from 1997 to 2004 I test the predictive performance of Random Forests and Rotation Forests and compare it to logistic regression using four different time horizons for failure. The sample size ranges between 1,960 and 10,500 and includes 9 different financial ratios as predictors. I conclude that Random Forests outperform both Rotation Forests and logistic regression. Rotation Forests slightly outperform logistic regression in smaller failure time horizons. Overall, it can be concluded that all three models perform well in comparison to similar models in the literature.

Keywords
Banks, Russia, prediction, machine learning, logit

INTRODUCTION

Banks have a crucial role in the economy of a country. They allocate financial capital to its optimal use and keep the savings of the public. When banks fail, the consequences can be devastating. In 1998, Russia was hit by a severe financial crisis, during which the ruble was devalued by more than 50% and the government unilaterally put a moratorium on its treasury bills. As an immediate consequence, many Russian banks went into default. Between August 1998 and August 1999, the total number of banks decreased from 1600 to 1390, despite efforts of the Central Bank of Russia (CBR) to act as a lender of last resort. During the course of the crisis, the liabilities of Russian banks at the CBR increased from 10.5 billion rubles to 71.7 billion rubles. The result of the crisis was economic desolation. In 1998, GDP shrunk by 4.6%, investments declined by 6.7%., and inflation hiked to 85.5%. The general population suffered greatly, which is reflected in poverty statistics - over the course of the crisis, the rate of people living below the poverty line of 394 rubles per month increased from 20% to 30%. (Herr, 2016). While the consequences of financial crises are not entirely attributable to failing banks, a more stable financial system can ameliorate at least some of the repercussions of financial crises. As a result, policy makers and regulatory authorities have a great interest in determining the reasons for bank failure, and in developing early warning systems for failure events. Such early warning systems have been developed using *parametric methods* from statistics (e.g. Martin (1977), Karminsky and Kostrov (2014)) as well as *non-parametric methods* from machine learning (e.g. Frydman et. al. (1985), Tanaka et. al (2016)).

While considerable progress has been made in the last decades, I have identified some gaps in the literature that I would like to address in this paper.

(1) While many authors have applied versions of neural networks to bankruptcy prediction, there is only one article in which a tree-based ensemble learning technique is employed. (Tanaka et. al. 2016). In my paper, I apply Tanaka et. al.'s "Random Forest EWS" to the Russian bank sector, a large sector that is not covered in their paper. Generally speaking, I expand the literature on tree-based methods – methods that have not been as thoroughly explored in the context of bankruptcy prediction as other machine learning techniques.

(2) The usefulness of principal component analysis (PCA) has been acknowledged by some authors (Canbas et.al. 2003), but the method has not been widely applied by researchers. I address this point by including rotation forests in my analysis, an ensemble learning technique based on PCA. To the best of my knowledge, this is the first analysis to employ rotation forests for bankruptcy prediction.

(3) Tree-based models have been successful in bankruptcy prediction in the banking sectors of different countries (Tam, 1990). However, there is no paper on the Russian banking sector that utilizes their potential. My paper is thus unique in its application of tree-based methods to the Russian banking sector.

(4) The sample size of many analyses is small, due to lack of available data on bank failures. The Russian bank sector has seen many failures in the last two decades, especially in the 1990. Therefore, my analysis features a larger sample size than the related literature.

(5) There has been no assessment of the volatility of prediction techniques in the literature. In my analysis, I show how the predictive accuracy of techniques can vary between different bootstrap samples.

My paper is structured as follows: First, I will elaborate shortly on my chosen methodology, followed by an introduction of the data and variables of choice. After that, I will present the results of my analysis, followed by a short conclusion.

METHODOLOGY

The format of this paper does not allow for an in-depth introduction of logistic regression, Random Forests and Rotation Forests. However, I will give a short overview of the most important facts. All three methods are used for classification and only the first is a parametric method, while the other two are non-parametric methods. Logistic regression is frequently used and well-known statistical technique, in which the dependent variable is categorical or binary. Random Forests are based on simple decision trees and were developed in 2001 by Leo Breiman (2001). They exhibit a high degree of classification accuracy in comparison to decision trees and correct for overfitting. Rotation Forests were developed by Rodriguez and Kuncheva (2006) in an attempt to increase the predictive performance of Random Forests by including PCA in the algorithm. The choice of Rotation Forests and Random Forests for this paper is motivated by the peculiar lack of literature using these techniques for bankruptcy prediction. Logistic regression is chosen as a reference technique, as it has been frequently applied in the bankruptcy prediction literature.

DATASET AND VARIABLES

The data used in this analysis stems from a dataset constructed by Karas and Schoors (2005) and contains a quarterly time series of balance sheet indicators and legal data of all Russian banks for the period of 1995 until 2010.

Dependent Variable

The dependent variable of interest is a binary variable indicating the occurrence of a bank failure within a specified time span. The event of failure is operationalized as the date of license revocation by the Central Bank of Russia (CBR).

I construct a dummy variable for failure called revdum which is either 1 or 0, depending on its time distance from a failure event. This variable serves as the dependent variable to be classified. I focus on four different time spans: 3 months, 6 months, 12 months, and 24 months. As a result, there are four different dependent variables: revdum1, revdum2, revdum4, and revdum8. In the context of prediction, the dependent variable can thus be interpreted in the following way: if the dependent variable is 1 and covers a time span of k quarters, then it means that a bank will fail at least within k quarters. For example, if a bank from the test set is predicted to be 1 based on a time span of 8 quarters, then it means that the bank is predicted to fail within at least 8 quarters. In this case, it could be that the bank already fails within 1 quarter, but this is not obvious from the prediction.
From a policy perspective, knowing that some bank will fail at least within a certain time span is more useful than knowing that some banks will fail within exactly a certain time span, while some others will not be detected.

Sample

Following the construction of the dependent variable, I extract eight random sub-samples from the original dataset. Half of the samples are used as training samples, while the other half are used as holdout samples. Each training sample is used for one of the four different failure time spans. I construct the samples from the perspective of a policy maker. In order to predict bank failures, a regulatory authority will use past data in order to classify future events. Consequently, the training samples and holdout samples are drawn from different time periods: the training sample is drawn from a time period of 3 years, from the first quarter of 1997 to the first quarter of 2000. The advantage of this sample selection is that the training period covers the Russian financial crisis of 1998, which means that there are a high number of bank failures available to train the model. The holdout sample covers a time period of 4 years, spanning from the second quarter of 2000 to the second quarter of 2004. The test sample is thus chosen such that no additional risk-taking incentives were put in place during the specified time period. Since tree-based models are sensitive to class imbalance, the dataset is cropped such that exactly 20% of the banks included in each sample are banks that are about to fail. This is in line with Lanine et. al.'s (2006) estimate that approximately 20% of Russian banks failed between 1988 and 2004.

Independent Variables

Nine variables measuring different determinants for failure risk are included in the three models, each of which measures a different aspect of a bank's financial structure. The choice of variables is motivated by previous studies of bankruptcy prediction, both in Russia and in other countries, as well as common concepts of banking theory.

Financial characteristic	**Variable (name)**	**Expected effect on failure probability**
Capital risk	Capital adequacy ratio (sk_ta)	negative
Liquidity risk	Liquidity ratio (liq_ta), non-gov't securities/total assets (ndo_ta)	Negative, positive
Default risk	Loans/total assets (lo_ta), non-performing loans/total assets (pzs_ta)	Positive, positive
Earnings	Return on assets (bp_ta)	negative
Size	Log of total assets (ln_ta)	negative
Deposits of firms	Deposits of firms/total assets (Vdul_ta)	practically, not theoretically motivated
Deposits of individuals	Deposits of individuals/total assets (vdfl_ta)	practically, not theoretically motivated

RESULTS

The metric for comparison between the results of the three models is the area under the receiver operating characteristic curve (ROC curve). ROC analysis is an established and commonly used statistic within bankruptcy prediction and machine learning in general. (e.g. Kolari, 2002) The ROC curve is a plot of the true positive rate (TRP) of a classifier against its false positive rate (FPR), depending on the choice of discrimination threshold. The area under the ROC curve measures the level of discrimination a classifier can achieve. It ranges from 0 to 1. An area of 0.5 indicates that a classifier is making predictions no better than chance, while an area of 1 indicates perfect prediction. For practical purposes, an area under the curve of 0.8 is considered to be good (Hosmer et. al., 1989).

Logit Model

The regression equation describing the logistic model is specified as:

$$revdum_i = \beta_1 sk\_ta + \beta_2 bp\_ta + \beta_3 liq\_ta + \beta_4 pzs\_ta + \beta_5 ndo\_ta + \beta_6 vdul\_ta + \beta_7 vdfl\_ta + \beta_8 lo\_ta + \beta_9 \ln\_ta + \beta_0$$

for each i in {1,2,4,8}.

The coefficients are mostly in line with theory, but for the sake of brevity, I will refrain from an extensive interpretation. In any case, the aim of this study is a comparison of predictive power, not an analysis of the causes of bank failure. The results of the regression for the training set and the areas under the ROC curve are shown in the following table:

Variables	One quarter before failure	Two quarters before failure	Four quarters before failure	Eight quarters before failure
sk_ta (Capital/Total Assets)	-3.027***	-2.735***	-2.608***	-2.144***
	(0.354)	(0.250)	(0.186)	(0.147)
bp_ta (Net Income/Total Assets)	-5.658***	-5.282***	-4.785***	-4.189***
	(1.130)	(0.795)	(0.610)	(0.485)
liq_ta (Liquid Assets/Total Assets)	-6.094***	-6.537***	-6.100***	-5.288***
	(0.699)	(0.487)	(0.330)	(0.230)
pzs_ta (Non-performing Loans/Total Assets)	3.298***	2.283***	1.445***	0.00451
	(1.044)	(0.715)	(0.543)	(0.435)
ndo_ta (Non-government Securities/Total Assets)	1.306***	1.282***	1.008***	0.503***
	(0.404)	(0.283)	(0.208)	(0.167)
vdul_ta (Deposits of Firms/Total Assets)	-1.613	-1.722*	-0.583	-0.748
	(1.300)	(0.895)	(0.661)	(0.463)
vdfl_ta (Deposits of Individuals/Total Assets)	-7.627***	-5.229***	-4.801***	-4.435***
	(1.384)	(0.864)	(0.584)	(0.444)
lo_ta (Loans/Total Assets)	0.405	0.220	-0.00538	0.204
	(0.377)	(0.262)	(0.186)	(0.144)
ln_ta (Log of Total Assets)	-0.236***	-0.265***	-0.267***	-0.199***
	(0.0400)	(0.0293)	(0.0212)	(0.0161)
Constant	0.714***	0.936***	1.162***	0.843***
	(0.235)	(0.170)	(0.129)	(0.102)
ROC area (training set)	0.8737	0.8572	0.8269	0.7827
ROC area (test set)	0.8085	0.7614	0.7028	0.6517
Observations	1,960	3,785	6,860	10,500

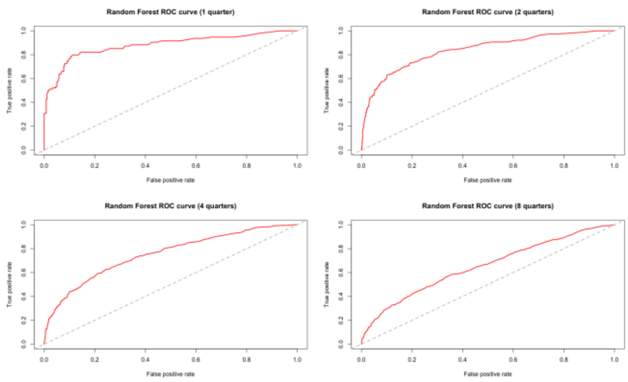
The predictive performance of the logit model decreases from 0.8025 to 0.6517 as the time period until failure becomes large. As expected, performance in the test set is worse than in the training set. In comparison to other studies, my logit model performs better and worse in different contexts. In the training set, the logit model underperforms in comparison to benchmarks set by other papers. For example, for 1 quarter Lanine et. al (2006) achieve an area under the curve (AUC) of 0.9482 in their logit model and an area of 0.9683 using their modified trait recognition approach. However, in terms of test sample performance, my logit model performs better than Lanine et. al.'s logit model by 0.0505 and even outperforms the trait recognition approach employed by Kolari (2002) by 0.0575. This pattern persists for all failure time periods and the difference between the models increases as failure time periods become longer. For instance, my logit model outperforms Lanine et. al. by 0.1412 and Kolari et. al. by 0.1129.

There may be different reasons why the performance is better than in the literature. The most likely explanation is the buildup of my samples. The samples with longer failure time periods also include observations from shorter time periods. The predictive edge of my logit model might thus stem from observations from shorter time periods that are easier to classify. Nevertheless, due to its good performance, I conclude that my logit model is a suitable benchmark for comparison of Random Forests and Rotation Forests.

Random Forest Model

The performance of the Random Forest algorithm depends on a number of parameters. I specify the parameters such that predictive performance is maximized. One important parameter choice is the number of single decision trees to grow. A decrease-of-error diagrams shows that 500 trees are sufficient for this analysis, as the error converges to a stable level at this number. Another important parameter is the number of variables per subset used for classification. Depending on the number of variables chosen there is a trade-off between diversity and accuracy. If more variables are included in the decision tree, then classification accuracy increases. If less variables are included, then the ensemble of trees is more diverse. Both diversity and accuracy are desirable in ensembles, but there is no general rule which is more important. In this case I chose 3 variables per subset which is equal to the square root of the total number of variables. Moreover, I did not find that pruning the trees increases predictive power. Hence, I allowed the algorithm to grow the maximum number of terminal nodes.

The ROC curve results from the four different quarters are displayed in the following figure:

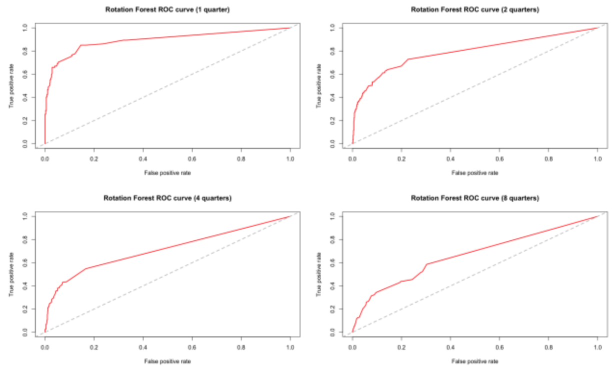


The results show that Random Forests outperform logistic regression by a solid margin. In the four different time periods, starting from 1 quarter, Random Forests outperform logistic regression by differences in ROC curve area of 0.0793753, 0.0801819, 0.0477478, and 0.0026845. Therefore, in accordance with Tanaka (2016), I conclude that the Random Forest model is better at predicting bank failures than at least one conventional statistical approach. The Random Forest model also outperforms the existing trait recognition models by Kolari et.al. (2002) and Lanine et. al. (2006), given that one takes ROC curve area as the sole criterion. It is unclear, however, how comparable my model is to other models in the literature in terms of sample selection and interpretability. Many authors do not elaborate whether their failure predictions for longer time periods also involve observations with shorter time periods. It should be noted that this caveat does not play in for a failure period of one quarter, where only observations from one quarter before are included. This is evidence that the superior performance of the Random Forest model is inherent to the technique itself, not the sample selection.

Rotation Forest Model

The current implementation for the Rotation Forest classifier allows to manipulate the value of two parameters: number of variable subsets (K) and number of trees per ensemble (L). Rodriguez and Kuncheva (2007) state that 3 features per subset and 10 trees per ensemble worked best for their analysis. Accordingly, I select the parameter K such that 3 features per subset emerge and the parameter L such that 10 trees are grown per forest. I find that this leads to superior performance in comparison to other specifications. As of 2016, the R-implementation of Rotation Forests does not allow for measures of variable importance or feature importance. Neither does it allow for internal diagnostic measures, like decrease of error rate in Random Forests.

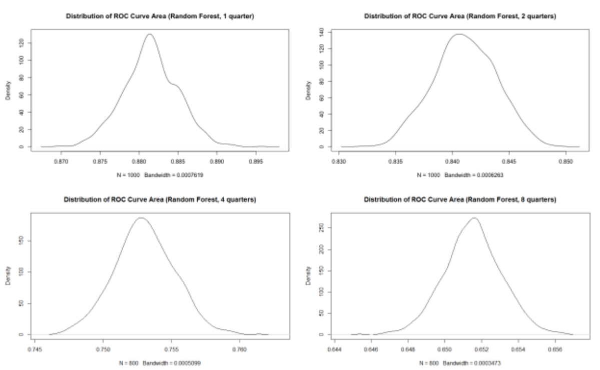
The ROC curve results for the four different results are displayed in the following figure:

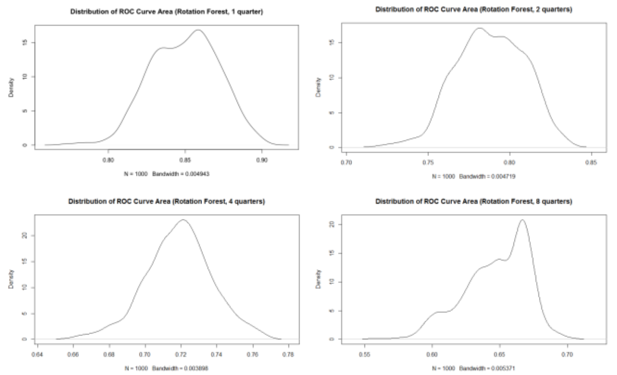


Rotation Forests outperform the logit model in 1, 2, and 4 quarters by 0.0218408, 0.0307726, and 0.0522938, respectively. For 8 quarters, Rotation Forests fall behind the logit model by 0.0511. In contrast, Rotation Forests perform worse than Random Forests in all time periods. This result is surprising, given that Rodriguez and Kuncheva (2006) claim that Rotation Forests outperform Random Forests and other machine learning techniques on a broad variety of benchmark datasets. One possible explanation is that Rotation Forests work best only under certain conditions. The defining characteristic of Rotation Forests is that they apply PCA to subsets of the data. The extraction of principal components is especially useful if the set of variables is large and includes many variables that are related to each other. If all variables are already quite independent of each other, then the extraction of principal components might not be as effective. It might be the case that the selected variables are already so different from each other that PCA did not make a great difference.

Volatility of Random Forests and Rotation Forests

When comparing prediction techniques, it is interesting to know whether a prediction model is consistent if it makes many predictions in a row. A model that sometimes delivers good predictions but is far off at other times is not particularly useful for policy makers, since they will never know when they can rely on its predictions. In order to test whether Random Forests and Rotation Forests are consistent, I let each model conduct 1000 predictions. For Random Forests with 4 quarters and 8 quarters, I conduct 800 predictions, due to lack of computational resources. I come to the conclusion that both Random Forests and Rotation Forests are very consistent if they make large numbers of predictions. The density plots shown in the following figure show that the volatility of areas under the ROC curve is low for both techniques.





CONCLUSION

My conclusion is that Random Forests outperform both logistic regression and Rotation Forests. Rotation Forests slightly outperform logistic regression in three out of four failure time periods. This result is in line with prior research on Random Forests in bankruptcy prediction (Tanaka, 2016), but contradicts the research on the performance of Rotation Forests (Rodriguez and Kuncheva, 2006). Assuming that the approaches are comparable, I also find that my Random Forest model outperforms models by Kolari et. al. (2002) and Lanine et. al. (2006). Based on my results, I can give the following suggestions for further research:

(1) How can the predictive performance of Random Forests and Rotation Forests be increased in the context of bankruptcy prediction?
(2) In which kind of countries and under which circumstances are tree-based ensemble learning techniques especially useful for prediction?
(3) A meta-suggestion: how can already existing knowledge about bank failure prediction better be utilized?

ROLE OF THE STUDENT

Maximilian Negele was an undergraduate student working under the supervision of Dr. Alexei Karas when the research in this report was performed. The topic was proposed by the supervisor. The theoretical work, data analysis, interpretation, formulation of the conclusions and the writing were done by the student.

REFERENCES

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Frydman, H., Altman, E. I., & KAO, D. L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. The Journal of Finance, 40(1), 269-291.

Herr, H. (2002, May 26). Finanzkrise in Russland im Gefolge der Asienkrise. Retrieved from: http://www.bpb.de/apuz/25451/die-finanzkrise-in-russlandim-gefolge-der-asienkrise?p=all

Karas, A., & Schoors, K. J. (2005). Heracles or Sisyfus? Finding, Cleaning and Reconstructing a Database of Russian Banks. SSRN Electronic Journal.

Karminsky, A. M., & Kostrov, A. (2014). The probability of default in Russian banking. Eurasian Economic Review, 4(1), 81-98.

Kolari, J., Glennon, D., Shin, H., & Caputo, M. (2002). Predicting large US commercial bank failures. Journal of Economics and Business, 54(4), 361-387.

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

Martin, D. (1977). Early warning of bank failure: A logit regression approach. Journal of Banking & Finance, 1(3), 249-276.

Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(10), 1619-1630.

Tam, K. Y., & Kiang, M. (1990). Predicting bank failures: a neural network approach. Applied Artificial Intelligence an International Journal, 4(4), 265-282.

Tanaka, K., Kinkyo, T., & Hamori, S. (2016). Random Forests-based Early Warning System for Bank Failures. Economics Letters, 148, 118-121