IFC-Bank of Italy Workshop on "Machine learning in central banking"

19-22 October 2021, Rome / virtual event

Probability of default model with transactional data of Russian companies [1]

Gleb Buzanov and Andrey Shevelev,
Bank of Russia

[1] This presentation was prepared for the Workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the central banks and other institutions represented at the event.

Bank of Russia

PROBABILITY OF DEFAULT MODEL WITH TRANSACTIONAL DATA OF RUSSIAN COMPANIES

Andrey Shevelev[1]

Gleb Buzanov[2]

(1) Research and Forecasting Department
(2) Financial Stability Department

September 29, 2021

Overview of standard PD models & Our Goals

accounting data + macro indicators + Loan information → Probability of Default

Cons:

- Rare publication of data
- Data is published with lags
- There is no connection between agents

The global goal **of this project is to** ***improve*** **the existing models of the Bank of Russia for predicting the Probability of Default of Russian companies in terms of** ***quality*** **and decision** ***frequency*** **using transactional data of the Bank of Russia Payment System (BRPS).**

The purpose of this work **is to study the** ***usefulness*** **of the data of the Bank of Russia Payment System (BRPS) for improving the existing probability of default models of Russian companies.**

Bank of Russia Payment System (BRPS) data

A firm's Payment Data can be used as information about a change in its state:

- VAT as a proxy for revenue
- Income tax as a proxy for profit
- Personal income tax and insurance payments as a proxy for payroll
- Payment graph for counterparty risk assessment
- and so on

Key points:

- Cover most of firms' payments
- Payment Graph
- Daily data

Literature reviews (with transactional data)

A. Khandani, A. Kim, A. Lo (2010): Consumer Credit Risk Models via Machine-Learning Algorithms

- Monthly aggregated transactions for credit risk assessment

H. Kvammea, N. Sellereiteb, K. Aasb, S. Sjursen (2018): Predicting Mortgage Default using Convolutional Neural Networks

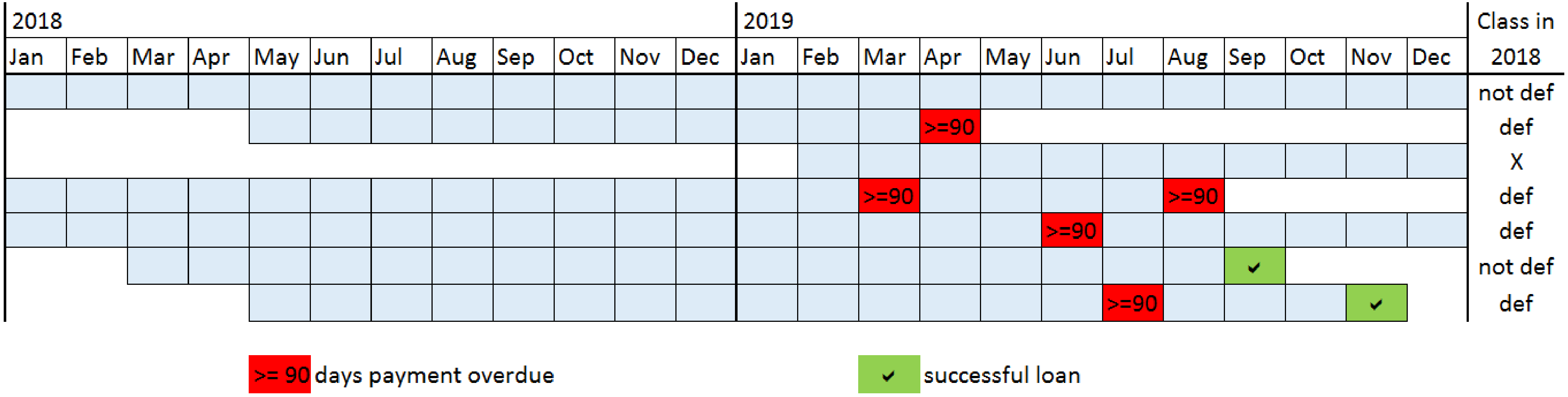
- Daily transactions for mortgage defaults prediction

D. Babaev, M. Savchenko, A. Tuzhilin, and D. Umerenkov (2019): E.T.-RNN: Applying Deep Learning to Credit Loan Applications

V. Shumovskaia, K. Fedyanin, I. Sukharev, D. Berestnev, and M. Panov (2020): Linking Bank Clients using Graph Neural Networks Powered by Rich Transactional Data

- Transaction based model for fraud and scoring SBER clients

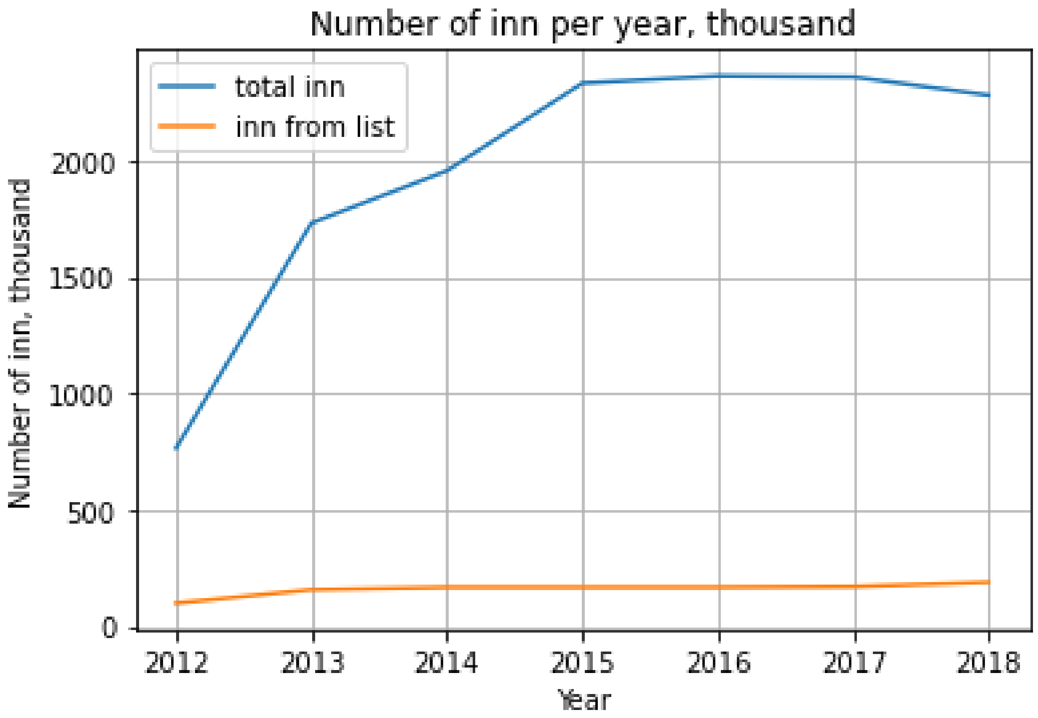
Definition of Default

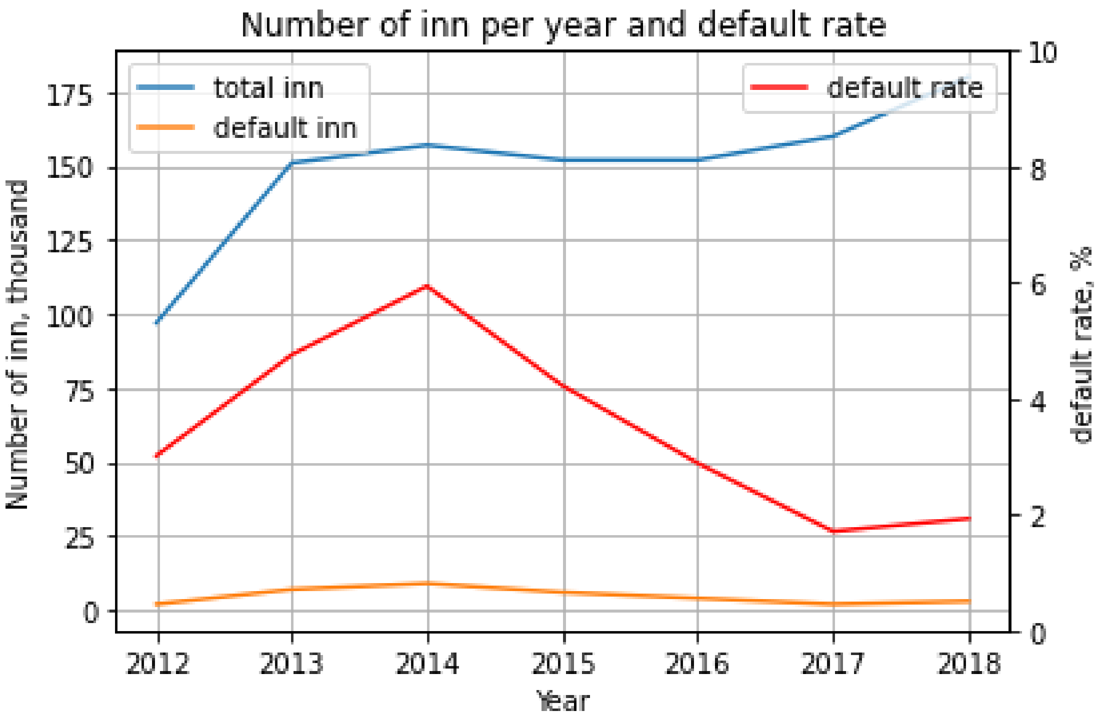


- Payment overdue >= 90 days [Basel II (III) IRB methodology]

Data & Default Rate

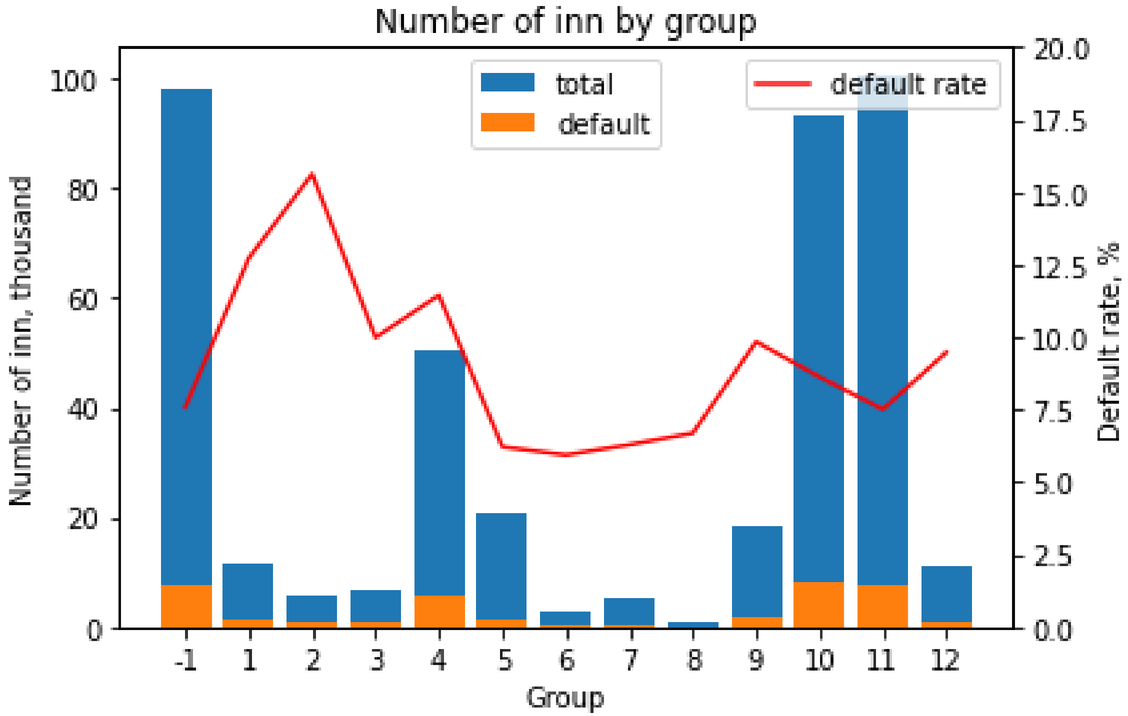
- Accounting data 2012-2018
- Default date 2013-2019
- Over 2 million unique companies per year
- About 200 thousand companies has a loan in next year after reported year
- 100-170k total unique companies
- 2-6% default rate





Default rate by industry, 2012-2018 year

Group	Industry name
1	Agro-industrial complex plant growing
2	Agro-industrial complex animal husbandry
3	Oil and gas industry
4	Construction
5	Real estate operations
6	Electricity and utilities sector
7	Chemical industry
8	Automotive
9	Consumer sector
10	Services sector
11	Wholesale trade (excluding fuel and minerals)
12	Metallurgy and mining
-1	Other

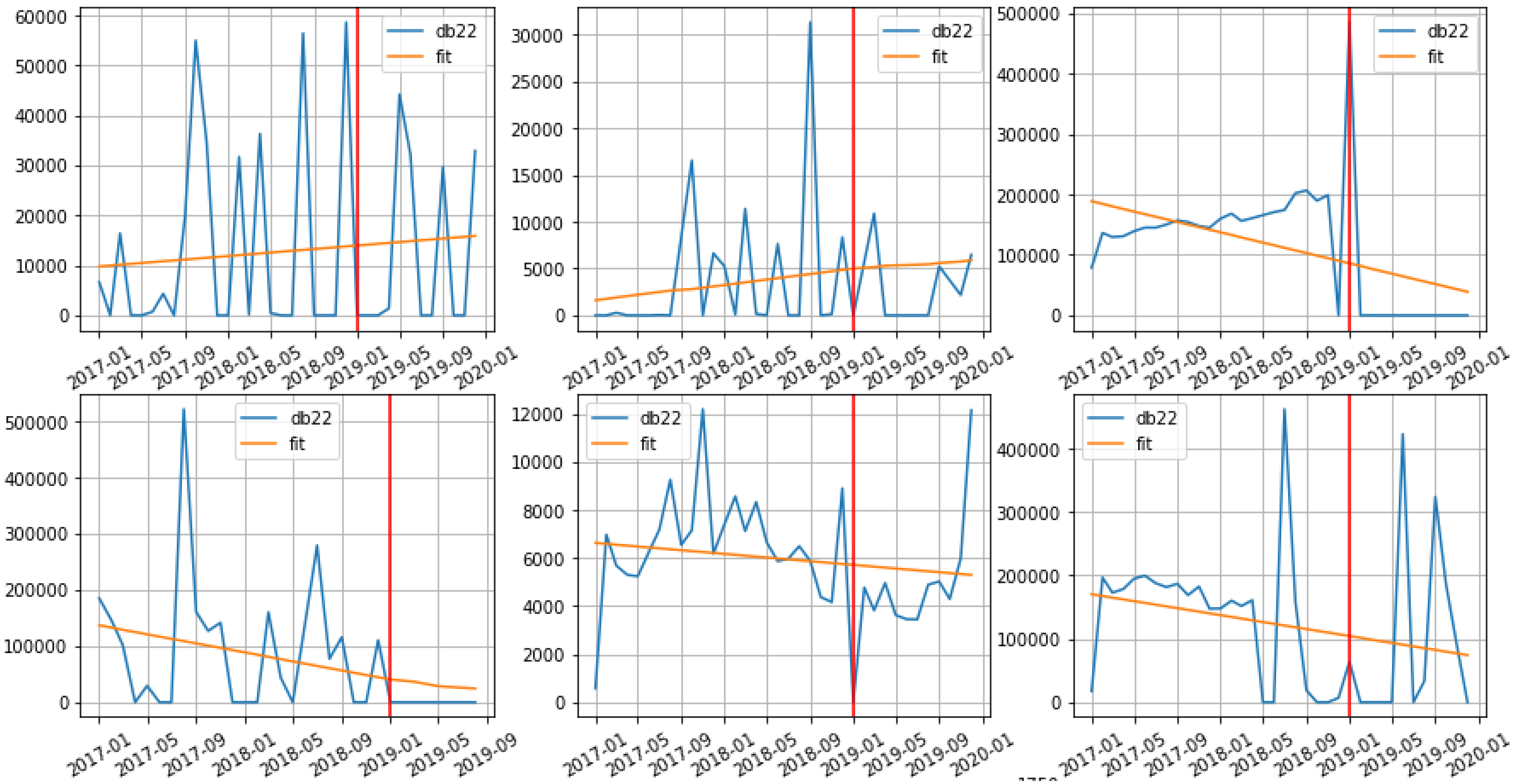


The Bank of Russia Payment System data

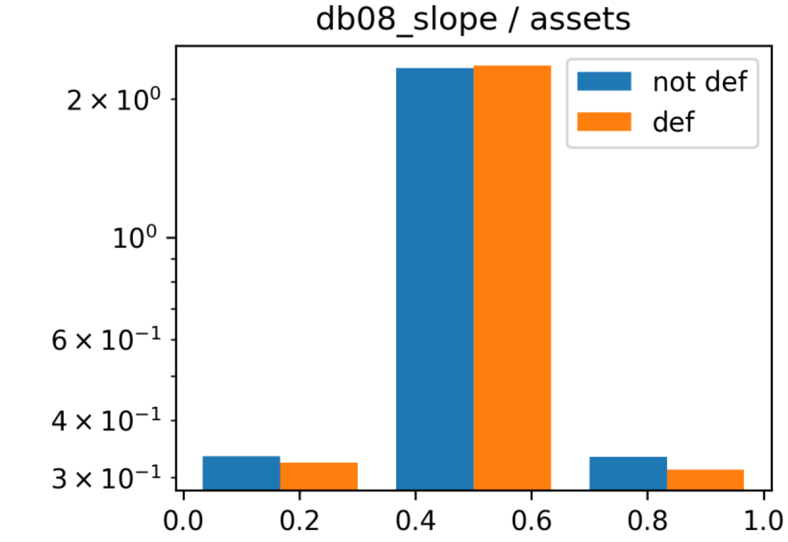
- From 2015 year
- ~ 10 million transaction per day

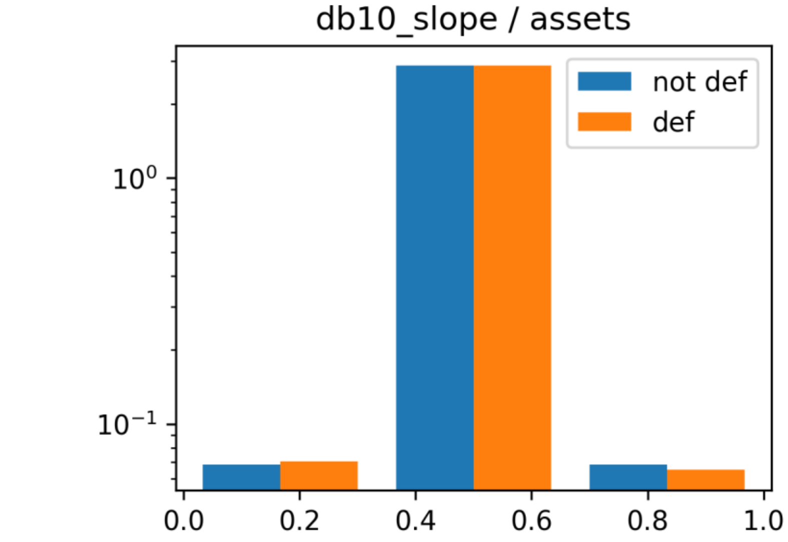
pd_main	pd_reestr	pd_status
• oper_dt • pmt_payment_doc_hk • inn_in • inn_out • amt • acc_in • acc_out • kbk_cd • pmt_type_kor_cd	• oper_dt • reestr_sqn • pmt_payment_doc_hk • inn_in • inn_out • trx_amt • acc_in • acc_out • trx_nm	• oper_dt • pmt_payment_doc_hk • pd_status_cd

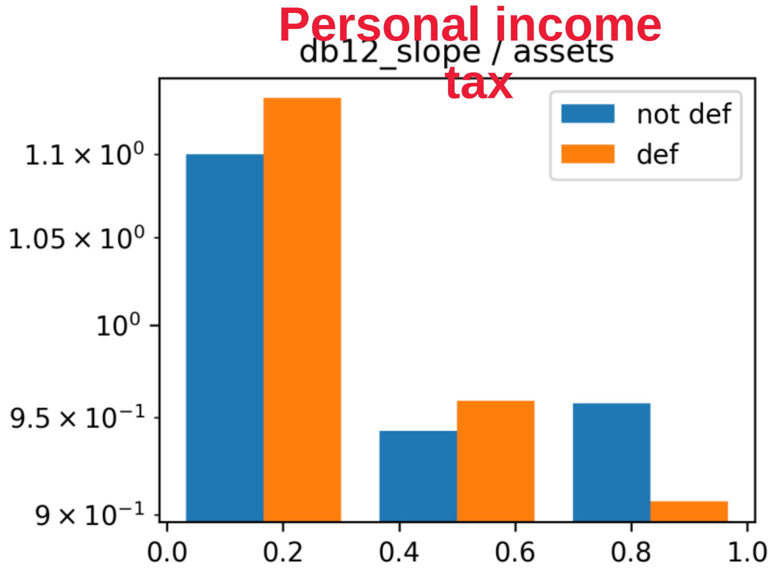
Example of BRPS data: social security payments

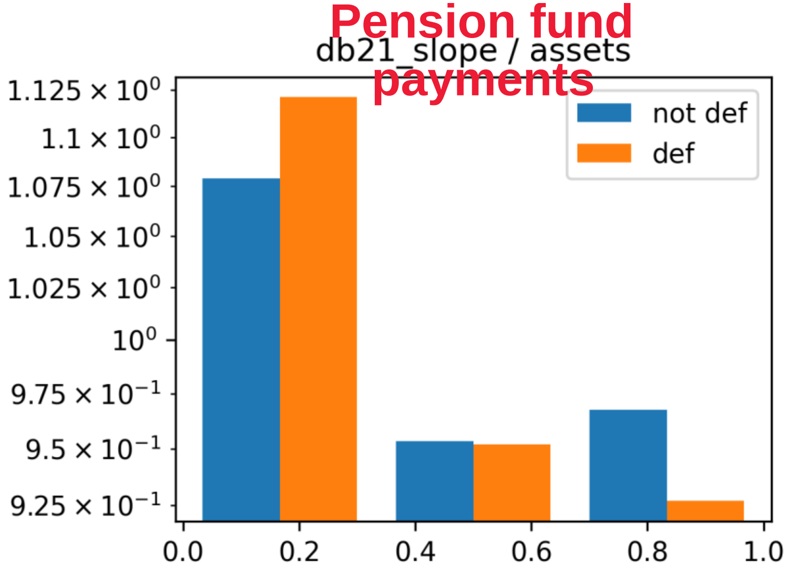


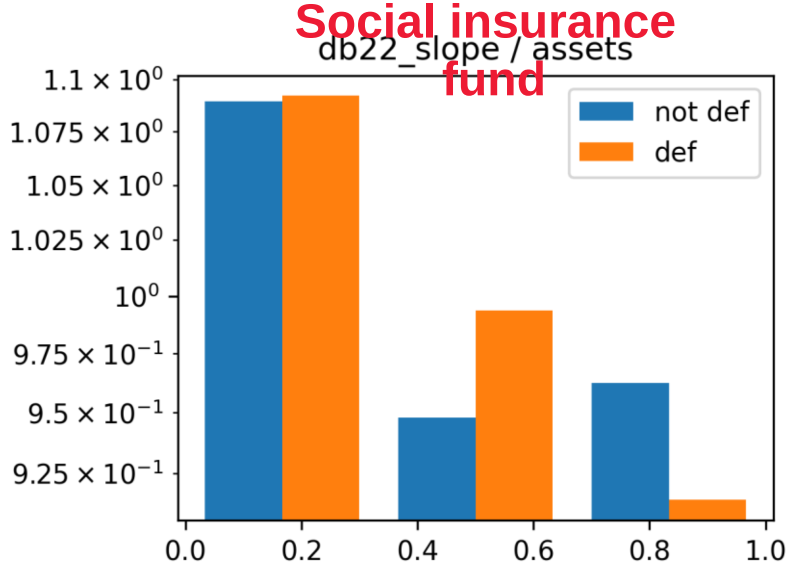
Histograms of some BRPS data slopes

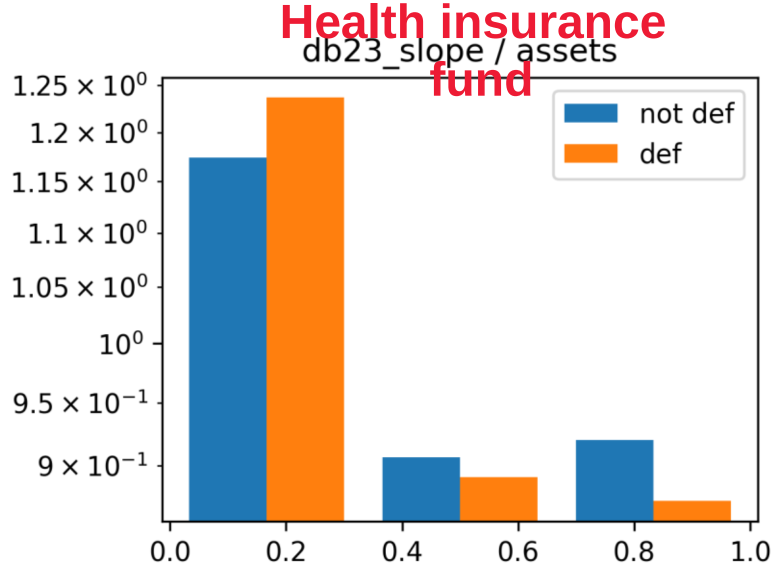








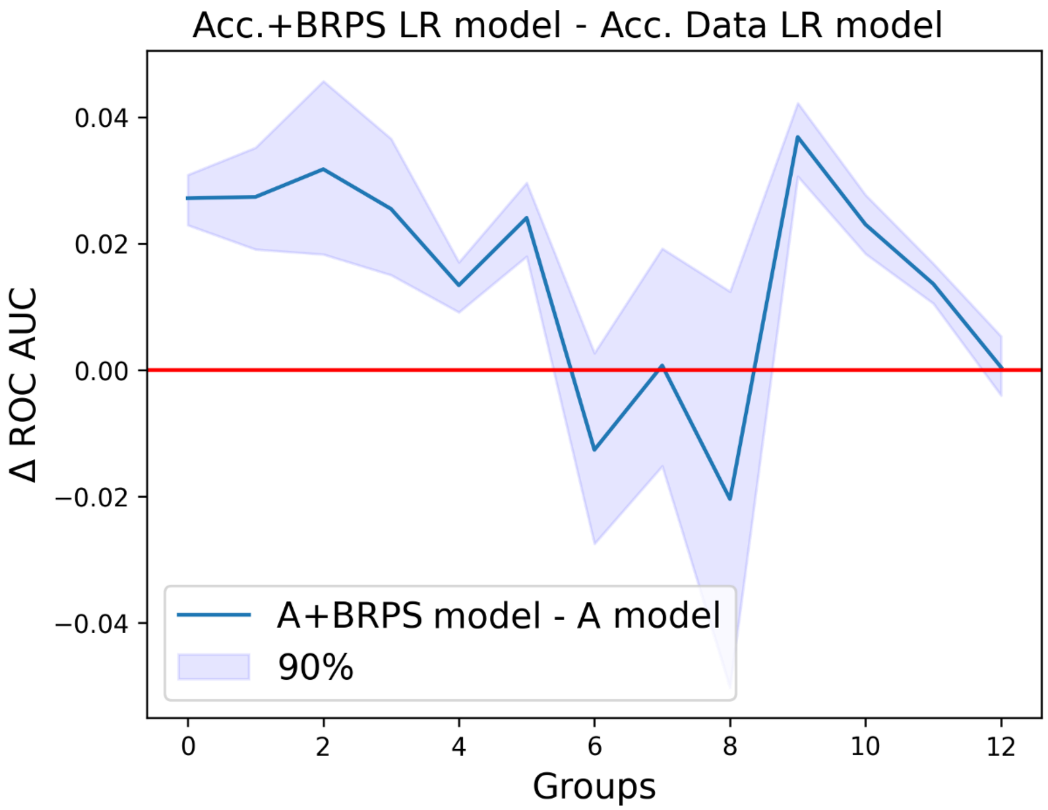




*All slopes are ranked, normalized, $\sum \text{def} = 1, \sum \text{not def} = 1$, logscale

Logistic Regression on AD vs A+BRPS data

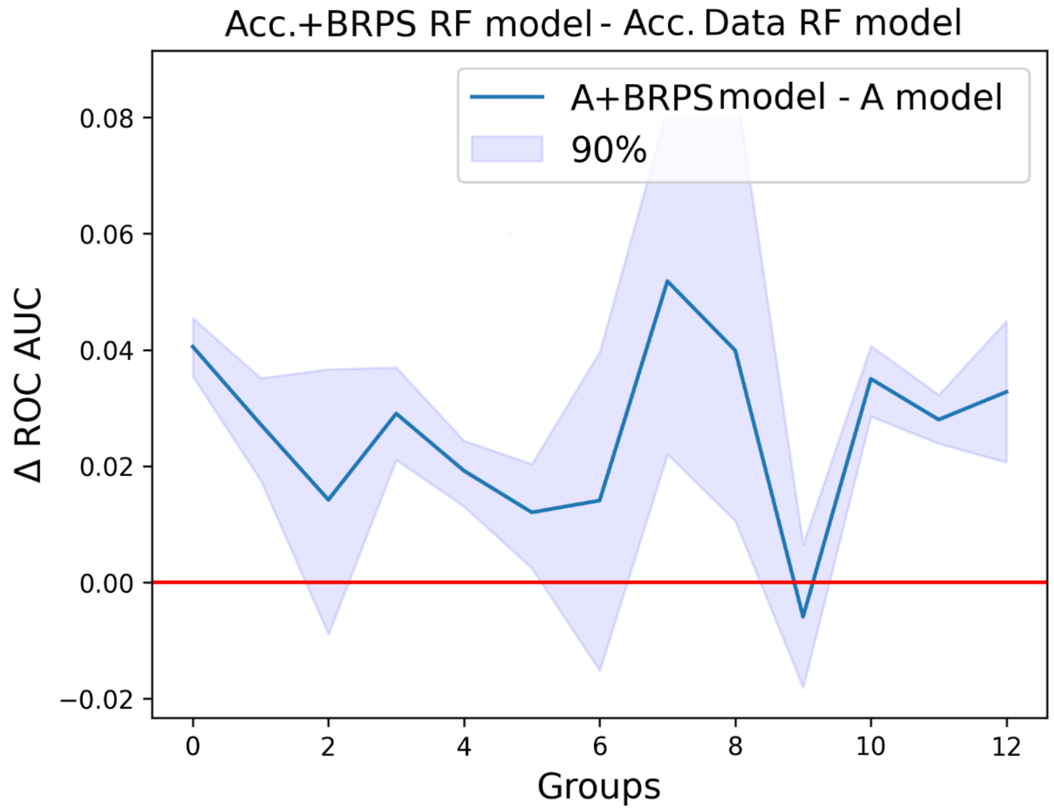
- L2 regularization
- Weighted likelihood



Combined score:
0.74 vs 0.76

Random Forest on AD vs A+BRPS data

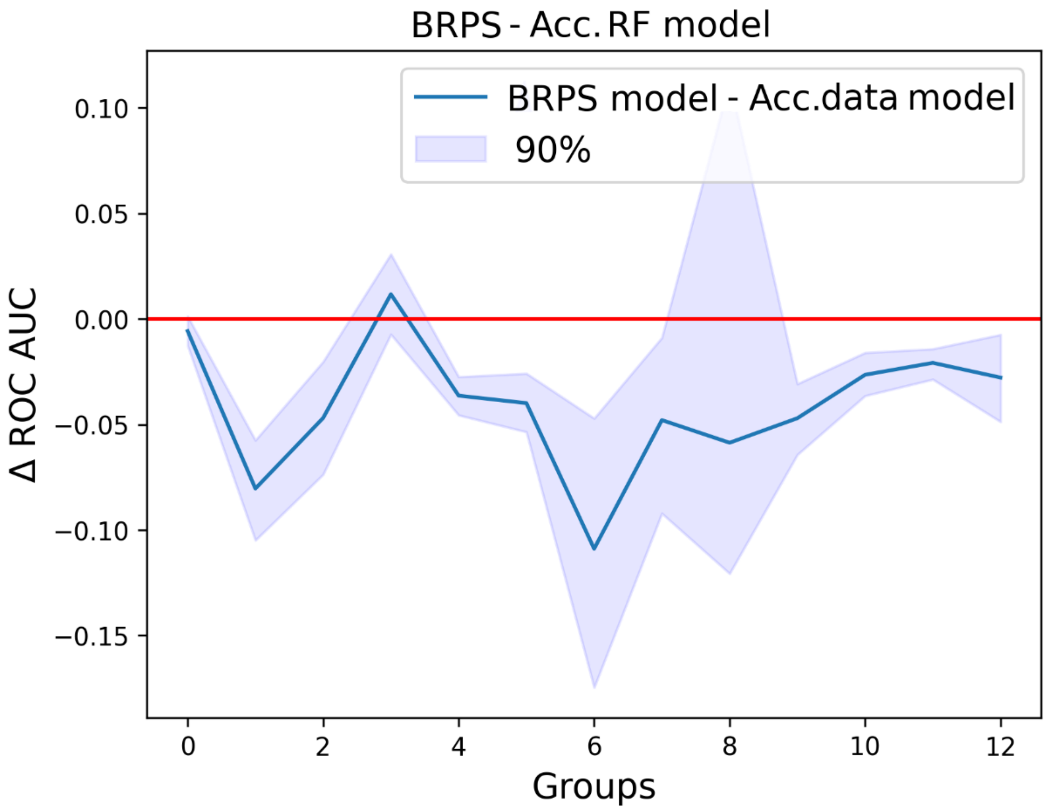
- Hyper parameters optimization
- Weighted Gini criterion for finding splits



Combined score:
0.72 vs 0.76

Random Forest on Accounting data vs BRPS data

- BRPS data only model is a useful for obtaining early estimates



Combined score:
0.72 vs 0.70

Conclusion

Results:

- BRPS data improve forecast quality
- Model with BRPS data only is a useful for getting early estimates

Next steps:

- Extending dataset up to date
- Neural Network model
- Graph Neural Network model
- Higher frequency
- Time series based

Bank of Russia

shevelevaa@mail.cbr.ru

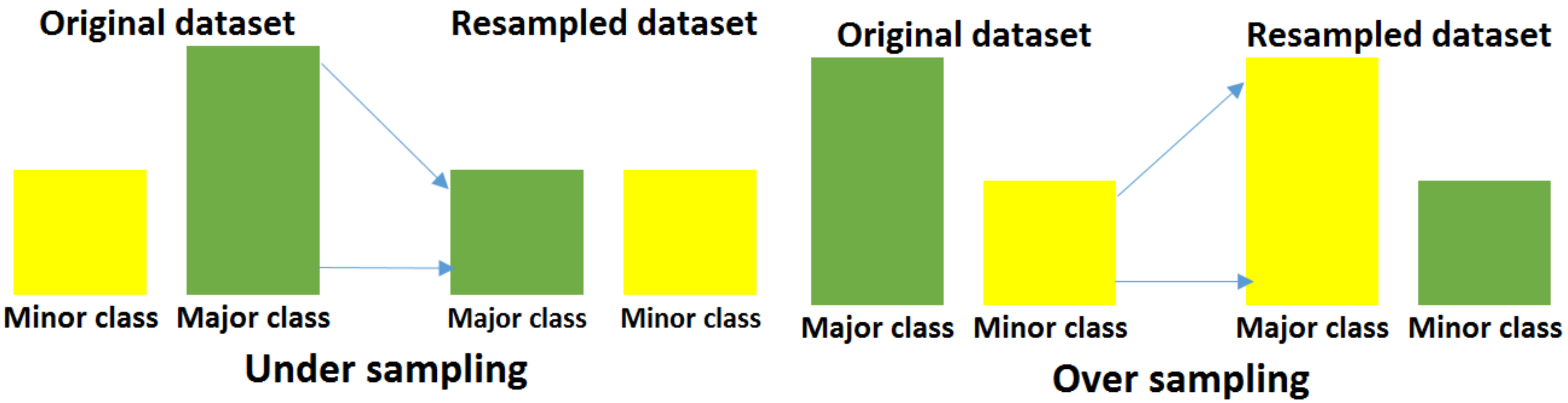
BACKUP SLIDES

Accounting data features

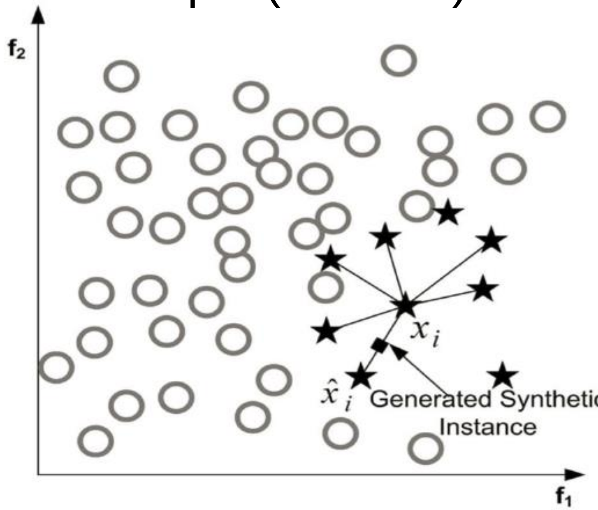
Name	Formula	Label
K1	'12003' / '15003'	Current liquidity ratio
K2	'22003' / '16003'	Return on assets
K3	'21003' / '21103'	Gross margin
K4	'24003' / '21103'	Net profit margin
K5	'22003' / '21103'	Operating margin
K6	'13003' / '16003'	Equity-to-asset ratio
K7	('14103'+'15103') / ('22003')	Debt/earnings from sales
K8	'22003' / '23303'	Interest coverage ratio
K9	'22003' / '23003'	Interest burden ratio
K10	('13003') / ('14003'+'15003')	Borrowed funds/Equity
K11	('13003'-'11003') / '12003'	Working capital to current assets ratio
K12	'24003' / '23003'	Tax burden ratio
K13	('21103' / '15003')	Short-term debt/Revenue
K14	'21103' / '12303'	Receivables turnover ratio
K15	'21203' / '15203'	Accounts payable turnover ratio

Imbalanced Data

- Random Under Sampling
- Random Over Sampling
- SMOTE
- Weighted likelihood estimation

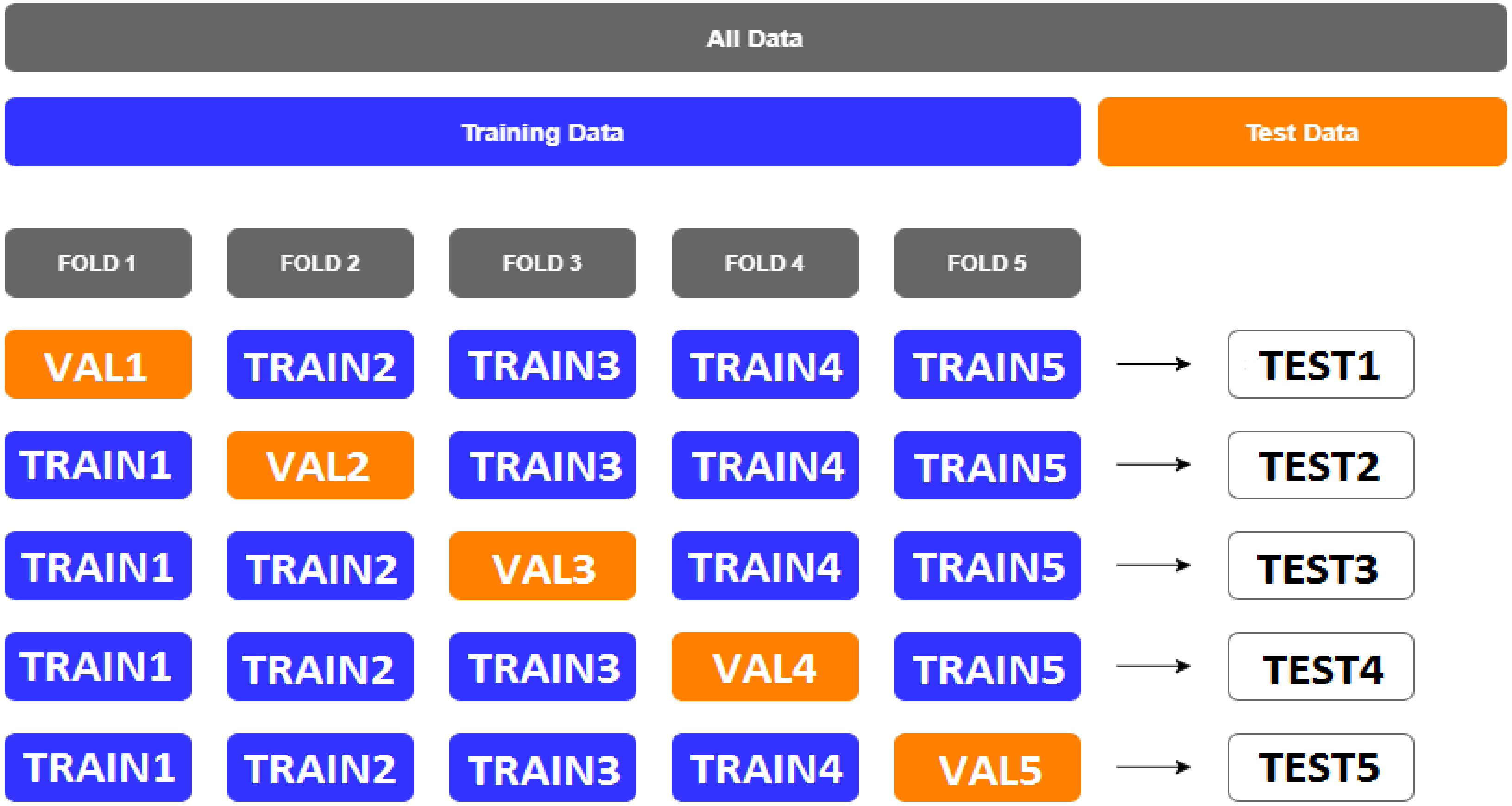


Synthetic Minority Oversampling Technique (SMOTE)



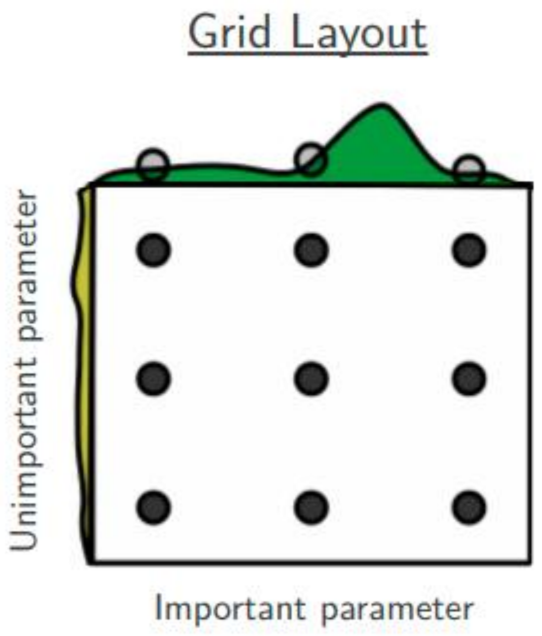
Cross validation

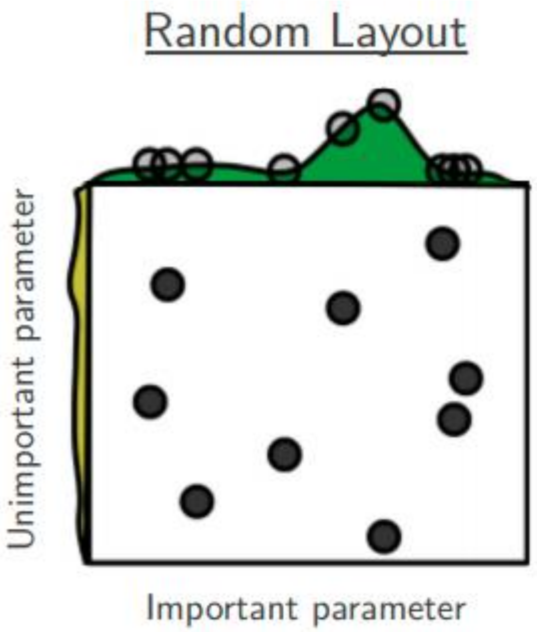
- K-Fold cross-validation
- Repeated k-Fold cross-validation

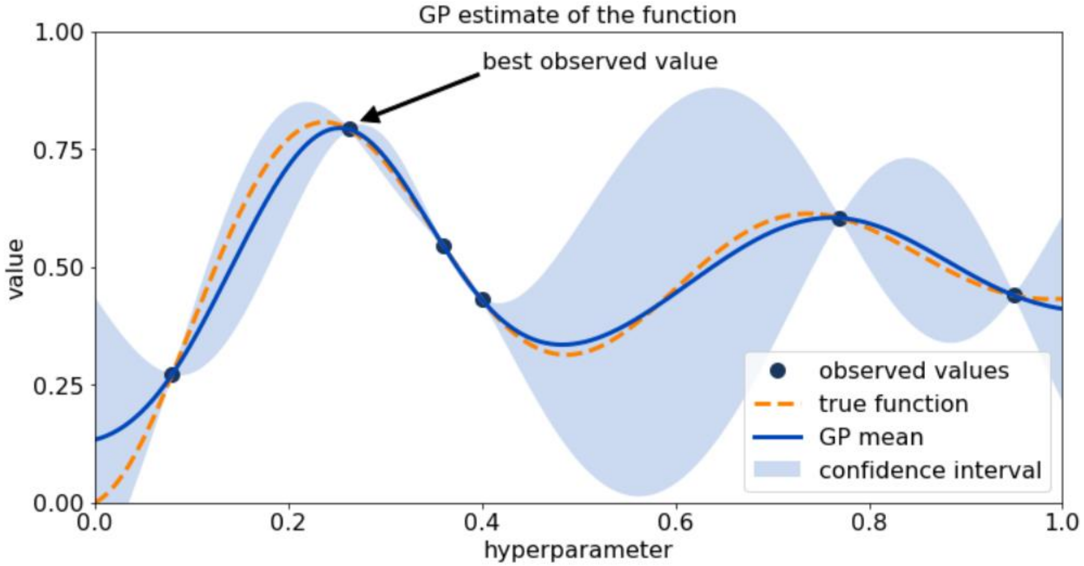


Optimization of Hyper Parameters

- Grid Search
- Random Search
- Bayesian optimization







Monthly aggregate BRPS data (transcript)

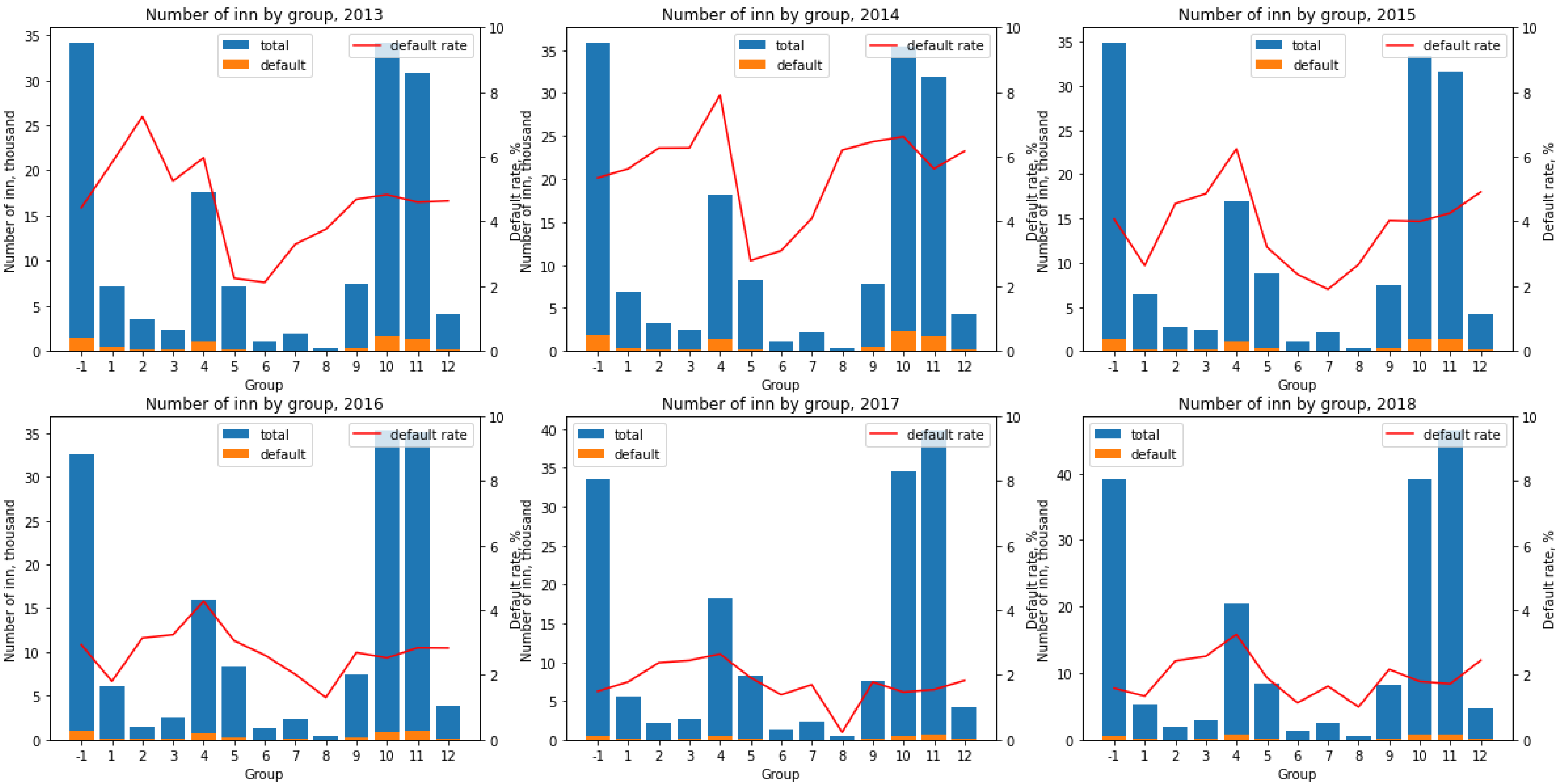
N	Name	Transcript
1	DT	Date of report
2	INN	Taxpayer Identification Numbers
3	OKVED_CODE	Russian Economic Activities Classification System Code
4	CNT_FI_ID_DB	Number of outgoing transactions
5	CNT_FI_ID_CR	Number of incoming transactions
6	PRC_MAX_DB	The share of the maximum outgoing turnover
7	PRC_MAX_CR	The share of the maximum ingoing turnover
8	REGNUM_MAX_DB	Number of maximum outgoing turnovers
9	REGNUM_MAX_CR	Number of maximum ingoing turnovers
10	TOT_DB	Outgoing payments value
11	TOT_CR	Incoming payments value
12	CNT_DB	Number of outgoing payments
13	CNT_CR	Number of incoming payments
14	DB06	Land tax
15	DB07	Gambling tax
16	DB08	Property tax
17	DB09	Transport tax

N	Name	Transcript
18	DB10	Other taxes
19	DB11	Corporate income tax
20	DB12	Personal income tax
21	DB13	Value added tax on goods imported to Russia
22	DB14	Value added tax on goods sold in Russia
23	DB15	Simplified tax
24	DB16	Single tax on imputed income for certain types of activities
25	DB17	Agricultural tax
26	DB18	Tax levied in connection with the application of the patent taxation system
27	DB20	Other payments for social needs
28	DB21	Pension fund payments
29	DB22	Social insurance fund payments
30	DB23	Health insurance fund payments
31	DB24	Federal customs service payments
32	DB30	Other budget payments
33	DB31	Client payments

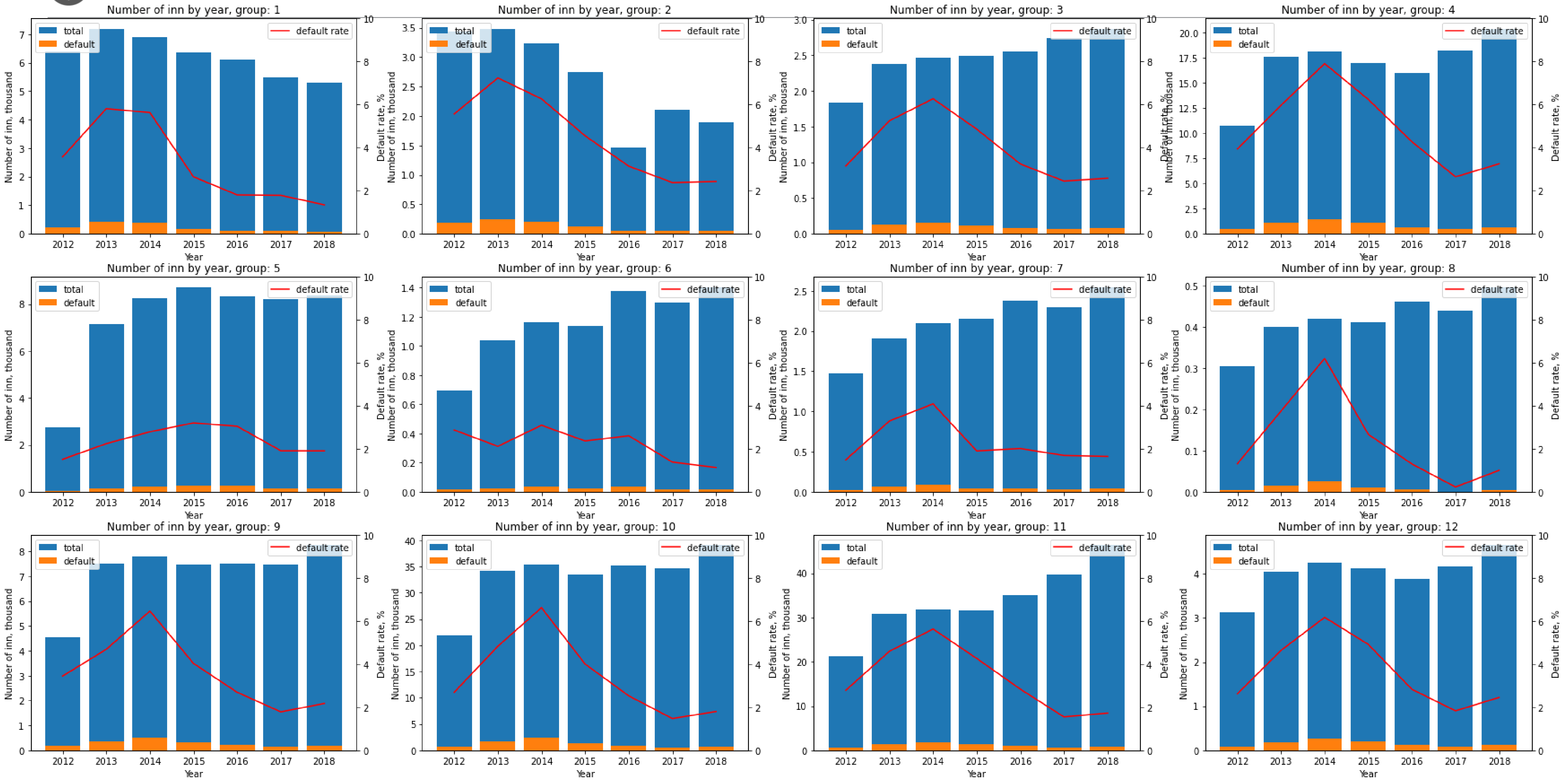
N	Name	Transcript
34	DB32	Payments to non-residents
35	DB33	Settlements with the exchange
36	DB35	Write-off from deposits
37	DB36	Payments to non-resident banks
38	DB37	Foreign currency purchase
39	DB38	Loan repayment
40	DB00	Other payments to be debited
41	CR39	VAT refund
42	CR40	Payment from the budget
43	CR41	Clients payments
44	CR42	Payments from non-residents
45	CR43	Settlements with the exchange
46	CR45	Deposit credits
47	CR46	Payments from non-resident banks
48	CR47	Foreign currency sale
49	CR48	Getting a loan
50	CR00	Other credit operations

Bank of Russia

BRPS data: default rate by industry, 2012-2018 year



BRPS data: default rate by industry



Model

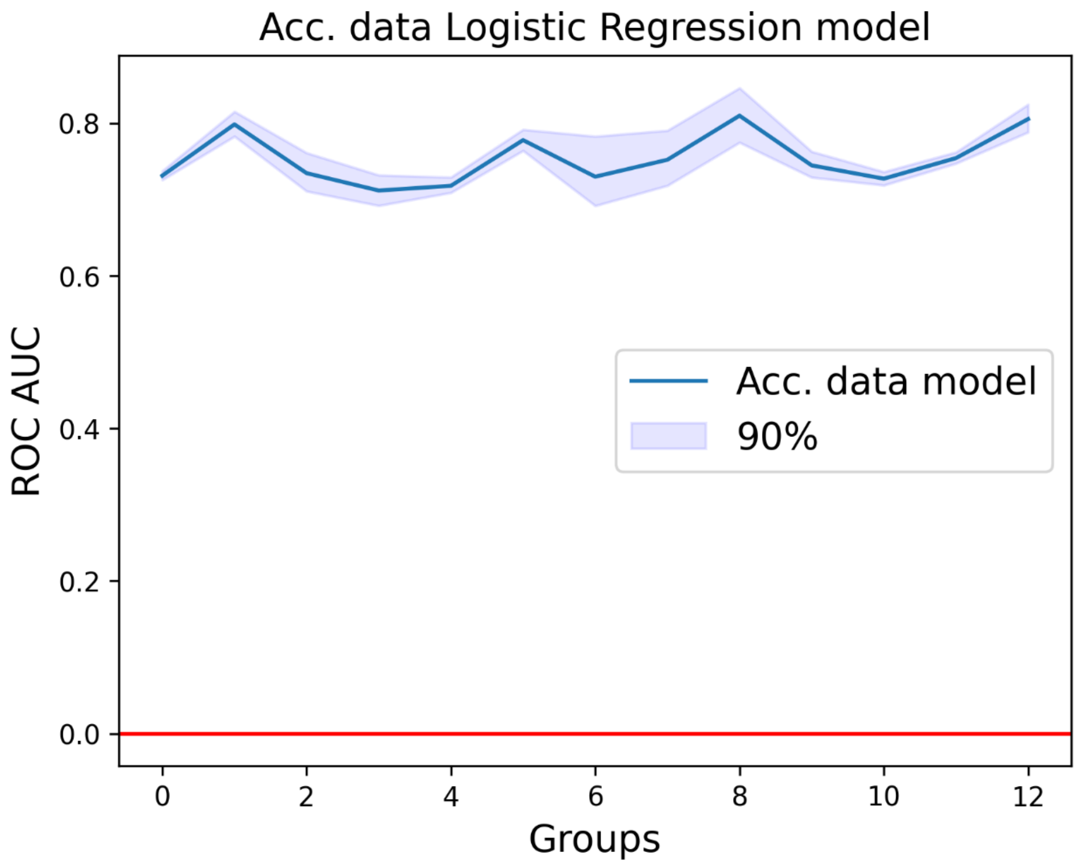
Baseline model includes:

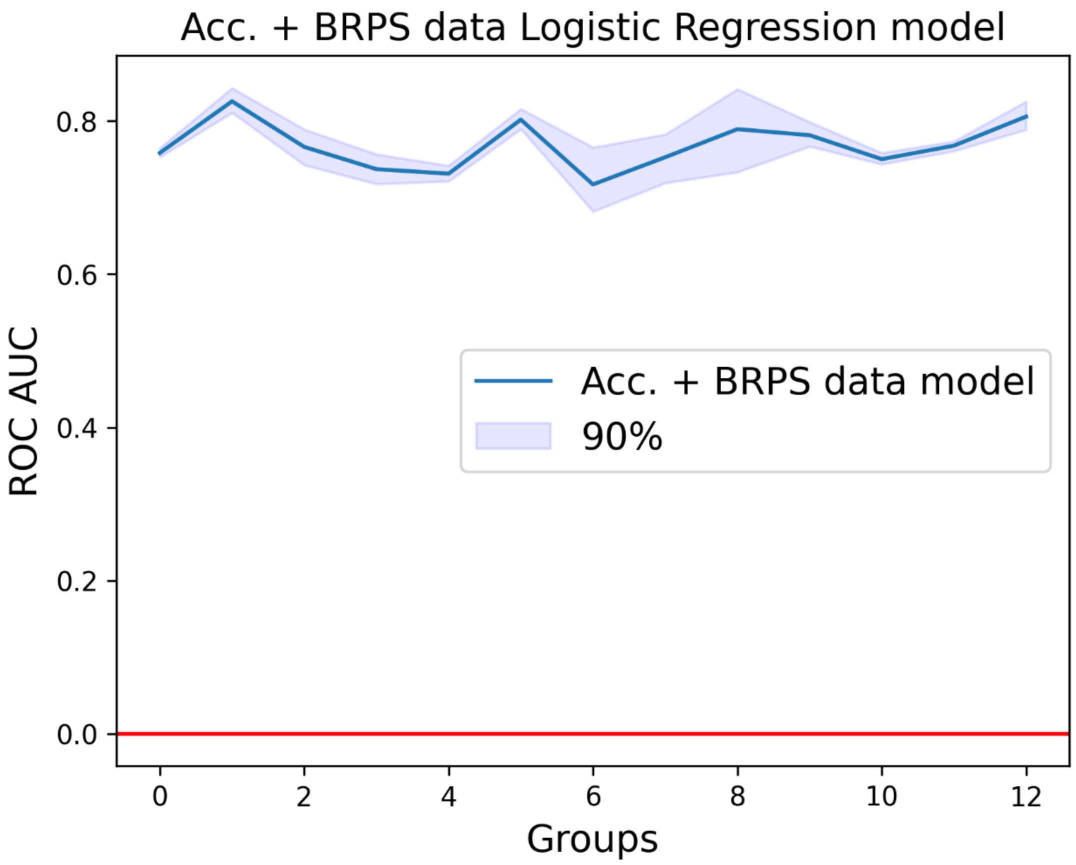
- Accountant data

Extended model includes:

- Accountant data
- Annual aggregated BRPS data normed by assets
- Slopes of monthly aggregated BRPS data
- Slopes of monthly aggregated BRPS data normed by assets

Logistic Regression results





Literature reviews

Babaev D., Savchenko M., Tuzhilin A., and Umerenkov D. (2019): E.T.-RNN: Applying Deep Learning to Credit Loan Applications

Bergstra J., Bengio Y. Random Search for Hyper-Parameter Optimization // Journal of Machine Learning Research 13 (2012) 281-305

Breiman L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324

Demeshev B., Tikhonova A. (2014). Default prediction for Russian companies: Intersectoral comparison // HSE Economic Journal, Vol. 18, No. 3, pp. 359—386. (In Russian)

Jackson R., Wood A. The performance of insolvency prediction and credit risk models in the UK: A comparative study // The British Accounting Review Volume 45, Issue 3, September 2013, Pages 183-202

Khandani A., Kim A, Lo A.(2010): Consumer Credit Risk Models via Machine-Learning Algorithms

Kvammea H., Sellereiteb N., Aasb K., Sjursen S. (2018): Predicting Mortgage Default using Convolutional Neural Networks

Louppe G., “Understanding Random Forests: From Theory to Practice”, PhD Thesis, U. of Liege, 2014.

Mogilat A. Bankruptcy in Russian real sector: Basic tendencies and financial indicators of a typical bankrupt // Nauchnye Trudy INP RAN, No. 13. 2015. pp. 156—186. (In Russian)

Mogilat A. Modelling financial distress of Russian industrial companies, or What bankruptcy analysis can tell // Voprosy Ekonomiki. 2019;(3):101-118. (In Russ.) https://doi.org/10.32609/0042-8736-2019-3-101-118

Shibitov, D., & Mamedli, M. (2019, August). The finer points of model comparison in machine learning: forecasting based on Russian banks' data. Retrieved from Bank of Russia Working Paper Series No. 43: http://www.cbr.ru/content/document/file/87572/wp43_e.pdf

Shirata C. Predictors of Bankruptcy after Bubble Economy in Japan: What can we learn from Japan Case? // the 15th Asia-Pacific Conference on International Accounting Issues, November 24, 2003

Shumovskaia V., Fedyanin K., Sukharev I., Berestnev D., and Panov M. (2020): Linking Bank Clients using Graph Neural Networks Powered by Rich Transactional Data