Bank of Russia

FORECASTING REGIONAL INFLATION RATES USING MACHINE LEARNING METHODS: THE CASE OF SIBERIA MACROREGION

No. 91 / March 2022

WORKING PAPER SERIES

Oleg Semiturkin
Andrey Shevelev

Contents

Oleg Semiturkin
Siberian Main Branch of the Bank of Russia
E-mail: SemiturkinON@cbr.ru
Andrey Shevelev
Research and Forecasting Department of the Bank of Russia
E-mail: ShevelevAA@cbr.ru

Authors are grateful to Ivan Verbny and anonymous referees for useful comments and suggestions.

Bank of Russia Working Paper Series is anonymously refereed by members of the Bank of Russia Research Advisory Board and external reviewers.

Cover image: Shutterstock/FOTODOM

Address: 12 Neglinnaya street, Moscow, 107016
Tel.: +7 495 771-91-00, +7 495 621-64-65 (fax)
Website: www.cbr.ru

SUMMARY

This paper evaluates the quality of forecasting regional inflation rates using machine learning methods in the case of the Siberia Macroregion[1] and Siberian regions. At the first stage of our study, we forecast regional inflation rates for various periods using several machine learning and benchmarking methods. At the second stage, we combine the forecasts with machine learning methods and weight them based on the quality metrics obtained. In the final part of this paper, we compare the obtained quality metrics with our benchmarks and confirm the stability of the results achieved using the Diebold – Mariano test.

Based on the results of our study, we conclude that the quality of forecasting inflation rates in the Siberia Macroregion and Siberian regions using machine learning methods is comparable with traditional econometric methods. At the same time, it is necessary to assess the quality of forecasting through machine learning methods for each region in advance to determine whether it makes sense to use them over traditional econometric methods. For most Siberian regions and the Siberia Macroregion as a whole, machine learning methods work better than benchmarks for periods longer than a year, in contrast to forecasts for 1-3 quarters ahead. Forecasting with combining machine learning models is in most cases preferable to any one.

Key words: monetary policy, machine learning, gradient boosting, elastic net, time series forecasting.

JEL-classification: E31, E37, E52, E58.

[1] The Siberia Macroregion comprises the pre-2018 borders of the Siberian Federal District (i.e. it includes the Trans-Baikal Territory and the Republic of Buryatia).

1. INTRODUCTION

After the Bank of Russia switched to a policy of inflation targeting, forecast accuracy of the price growth rate became a prerequisite for an efficient monetary policy. The purpose of our study is to assess the quality of inflation forecasting in Siberian regions using machine learning methods.

Advanced machine learning methods are applied successfully in many data processing areas, but they are relatively new in predicting macroeconomic indicators (in particular, inflation). In our study, we apply popular machine learning methods, such as lasso regression, ridge regression, elastic net, random forest decision trees and gradient boosting decision trees to forecast regional inflation rates in Siberia. To assess the quality of forecasting, we compared the forecast quality metrics obtained by machine learning methods with the forecast quality using the ARIMA and ARIMAX models that are widely used to accomplish similar tasks. We then obtained a combined forecast based on weighted predictions using machine learning methods and performed a test to identify significant differences between the benchmark method and the combined forecast method. For each model, the quality of the forecasts obtained was compared with the quality of the indicator-based benchmark ($RMSE_x/RMSE_b$). The set of such relative assessments is a formal criterion for achieving the goal of our study.

2. LITERATURE REVIEW

The first works containing results of forecasting macroeconomic variables using machine learning methods have been relatively recently published. C. Chakraborty and A. Joseph (2017) describe various methods of machine learning with respect to accomplishing the tasks of central banks. In particular, they consider an example of forecasting UK inflation rates for a medium-term horizon – up to two years. As explanatory variables, the authors used 12 macroeconomic indicators for the period from 1988 Q1 to 2015 Q4 (112 observations): GDP, labour productivity, gross disposable household income, money supply, private sector debt, unemployment rate, employment rate, GBP real effective exchange rate, Bank of England policy rate, difference in yield of indexed and non-indexed 5-year securities and global commodity price index. As a result, the authors conclude that machine learning methods, such as neural networks, support vectors, random forest models, etc., produce better quality forecasts than the methods traditionally used for these purposes, such as the AR(1) autoregressive model.

J. Jung et al. (2018) used three different machine learning methods to forecast real GDP growth rates in 7 countries (US, UK, Germany, Mexico, Philippines, Vietnam and Spain). The sample was comprised of 29 macroeconomic indicators on a quarterly and annual basis from the 1970–90s to 2016. The work also concludes that the above machine learning methods consistently outperform the methods traditionally used in IMF forecasts.

I. Baybuza (2018) proves the consistency of ML methods in forecasting inflation rates in Russia compared with traditional alternatives. The author uses a CPI time series and a sample of 92 macroeconomic monthly indicators from February 2002 to June 2016 (173 observations). The selected data reflect the state of business activity, industrial production, the money market, the employment rate, the balance of payments, and prices for the Russian economy's main exports. The random forest and boosting methods for predicting monthly inflation rates showed results comparable to the basic AR(1) model. At the same time, they showed significantly better results when comparing average inflation forecasts on a horizon greater than two months. The author concludes that the mentioned methods are promising for forecasting inflation rates in Russia. In addition, the author proves that ensemble models based on untransformed data show better forecasting results than similar models that use transformed data.

A study by K. Yakovleva (2018) describes a methodology for calculating a high-frequency indicator that reflects the dynamics of economic activity in Russia. As inputs, the author uses news articles from an Internet resource covering economic events in Russia and abroad. Such news articles are analysed using text analysis and machine learning methods. First, the author, using the LDA probabilistic model, builds an indicator that reflects a topic's quantitative component, i.e. how often the topic is mentioned in the news. Then the author builds an indicator characterising the emotional component of the news and defines it using supervised machine learning methods. Based on the obtained indicators, the author predicts the business activity index. The calculated results showed that the use of news articles can be an important component in forecasting Russian economic activity, as testified by the strong quality of the resulting model.

An article by E. Pavlov (2020) studies the applicability of neural networks for forecasting inflation rates in Russia. The author includes 10 macroeconomic indicators from February 2002 to August 2018 (200 observations) in the sample: CPI, GDP physical volume index, labour productivity (ratio of real GDP to employed population), money supply M2, the volume of loans issued in real terms, the unemployment rate, exports in real terms, oil prices in US dollars, real disposable income and the money market interest rate. The models analysed have demonstrated positive results over a planning horizon of more than one month, thus making them a good tool for short- and medium-term forecasting. Machine learning methods, such as neural networks and support vectors, have generally proven more efficient than one-factor and simple linear models.

In an article by O. Barkan et al. (2021), which is a joint effort with the Bank of Israel, the authors assess the quality of forecasting the inflation components in the United States using neural networks. The study concludes that forecast quality using neural networks is better than AR(p) and VAR(p), especially at lower levels of the US consumer price index hierarchy.

In their study, O. Ozgur et al. (2021) compared the predictive power for inflation rates in Turkey using the regularisation methods and ARIMA and VAR methods. As a result, machine learning methods showed a better forecast in the medium term compared to benchmarks.

In the Bank of Russia report ‘Forecasting Russian CPI with Data Vintages and Machine Learning Techniques’, its authors M. Mamedli and D. Shibitov show how the predictive abilities of models change depending on the use of vintage data and seasonal adjustments. The authors use the elastic net method, ensemble machine learning methods, as well as neural networks and compare the results with the AR(p) forecasting method. They show that the actual forecast error can be significantly underestimated when conducting a pseudo-real time experiment. The authors emphasise the importance of taking these into account when making real-time forecasts.

There are important differences between ours and similar studies. In particular, we not only apply machine learning methods to forecast regional inflation rates in Russia, but we also employ a method for combining weighted forecasts with machine learning models and comparing them with the results obtained by individual machine learning methods as well as the ARIMA and ARIMAX benchmark methods. As factors, we use the exchange rate, the past dynamics of regional inflation rates and their derivatives (lagged and average values, standard deviations). Similar studies use arrays of many different factors for machine learning, which can significantly complicate a transition from pseudo-real-time to real-time forecasting. In addition, a broader data set compared to benchmarks makes it unclear where the source of the improved forecast lies – method or data. In our study, both benchmarks and machine learning use the same set of information, allowing us to draw better conclusions about the relative quality of forecasting methods.

3. DATA

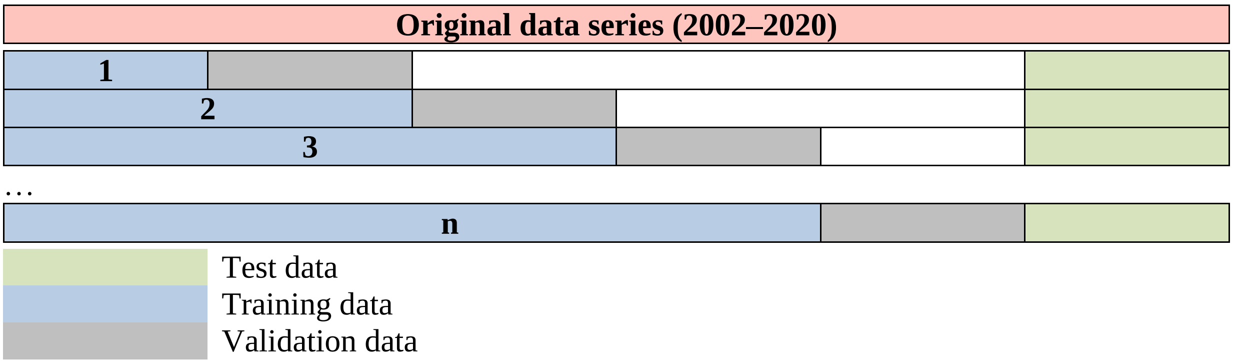
As inputs, we used the CPI[2] from February 2002 to December 2020 for the Siberia Macroregion and 12 Siberian regions. The daily exchange rate data were converted to monthly indicators by calculating the monthly average. For the L1- and L2-regularisation methods, the original series were pre-normalised. The literature review shows that ensemble methods use various approaches to data preprocessing in practice for similar tasks, however unadjusted data are preferable. We also obtained the best estimates of ensemble forecasts based on raw unadjusted regional inflation data.

As predictors for machine learning methods, we used the CPI and national currency rate series in the form of lagged values from 1 to 12 months as well as mean values and standard deviations for the previous 3, 6, 9 and 12 months. Accordingly, we obtained a set of factors based on two statistical series.

4. RESEARCH METHODS

The study used the principle of supervised learning. We divided the initial data set into training, validation and test samples. To solve the problem of transferring information from the test sample to the result and to select the model hyperparameters, we searched for optimal parameters through cross-validation (Chart 1).

Chart 1. Cross-validation principle

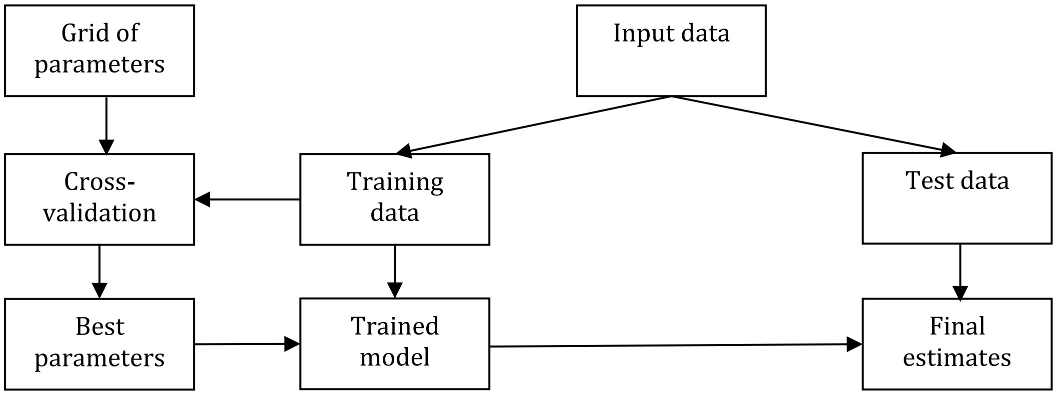


Source: compiled by the authors.

Accordingly, we avoided transferring information from the test set to the model and obtained more impartial estimates. Initially, the series was divided into training and test components, after which a grid of model hyperparameters was established using expert knowledge. Next, the training dataset was divided into a specified number of control (validation) datasets, and the model was trained on each of them. For each set of hyperparameters, the cross-validation correctness mean was calculated based on the validation samples. As a result, a combination with the best quality ratings was selected. The general process of dividing data and evaluating final parameters is shown in the chart (Chart 2).

[2] Federal State Statistical Service. URL: https://fedstat.ru/ (accessed on 12 February 2021).

Chart 2. Process of dividing data and evaluating final parameters



Source: https://scikit-learn.org[3].

CPI forecasts were calculated using three L1- and L2-regularisation methods (lasso regression, ridge regression, elastic net) and two ensemble methods (decision trees random forest and decision trees gradient boosting). The results were compared using two benchmarking methods (ARIMA and ARIMAX). Let us briefly describe the particular characteristics and application practice of each forecasting method.

For forecasting the CPI time series using ARIMA and ARIMAX methods, the best parameters were selected by choosing the optimal value on the out-of-sample (similarly to machine learning methods). This was done for optimal comparability between benchmark and machine learning forecasts. The augmented Dickey – Fuller test was used to determine the difference order to bring the data to a stationary form. The model's hyperparameter lag length was set in the range of 1- 24. In the ARIMAX model, the nominal ruble exchange rate is used as an augmentation; otherwise the principle of choosing the best model parameters is similar to that described above.

The L1- and L2-regularisation methods are a development of the least squares method. In addition to the square error restriction, they add another restriction, which in the case of ridge regression minimises the value of coefficients in regressors, and in lasso regression equates insignificant coefficients to zero. In fact, using lasso regression, we select the most important regressors for the model. In this case, the coefficients must retain the maximum predictive power of the resulting model. The model quality is assessed as follows:

$$Q = \sum_{i=1}^{l} L(y_i, a(x_i)) + \gamma V(w) \to \min_{w} \quad (1) ,$$

where Q is the model quality indicator, L is the loss function, $V(w)$ is the model regulariser, y_i are the real values of the function, γ is the parameter for adjusting the regulariser weight in the model, $a(x_i)$ are the predicted values of the function, w are the coefficients in the model regressors, l is the dimension of the model parameter vectors.

For ridge regression, the loss function is written as follows:

$$L(y_i, a(x_i)) = (y_i - a(x_i))^2 \quad (2) ,$$

L2 regulariser in the following form:

[3] This site provides a description of the library «scikit-learn» for the Python programming language.

$$V(w) = \|w\|_{l2}^2 = \sum_{n=1}^{d} w_n^2 \quad (3),$$

For lasso regression, the loss function is written as follows:

$$L(y_i, a(x_i)) = |y_i - a(x_i)| \quad (4),$$

L1-regulariser in the following form:

$$V(w) = \|w\|_{l1} = \sum_{n=1}^{d} |w_n| \quad (5),$$

Introducing a regulariser makes it possible to reduce the tendency to overfit and improve the generalising properties of the model compared to the linear one. In our task, we define the regulariser weight by searching from predetermined values, evaluating and choosing the best characteristics of the model.

The elastic net model combines the properties of lasso regression and ridge regression, i.e. it uses L1- and L2-regularisation with specific weights. The elastic net regulariser is written as follows:

$$V(w) = \partial \sum_{n=1}^{d} |w_n| + (1 - \partial) \sum_{n=1}^{d} (w_n)^2 \quad (6),$$

where $\partial \in [0; 1]$is the model hyperparameter defining the weight of L1- and L2-regularisation. In our study, it equals 0.5, similar to Chakraborty and Joseph (2017).

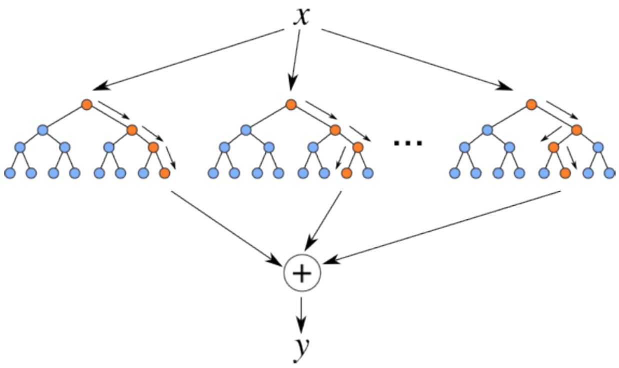
In general, all L1- and L2-regularisation methods are distinguished by their simplicity, speed of operation and ability to interpret the obtained results. The specifics of applying such models are the need to adjust the regularisation parameters and eliminate any possible model overfitting. In addition, attributes must be normalised before being used in the model.

For ensemble machine learning methods, the backbone lies in the decision tree method, which in general resembles a graph consisting of root and leaf vertices. When choosing the parameter of transition from one vertex to another, the criterion of information content is minimised. In the case of accomplishing a regression task, the goal is to minimise the mean square error (MSE) at each vertex of the built tree.

$$H(Q_m) = \frac{1}{N_m} \sum_{y \in Q_m} (y - \overline{y}_m)^2 \quad (7),$$

where $\mathrm{H(Q_m)}$ is the information content criterion (root-mean-square error), y are the real values of the function, $\overline{\mathrm{y}}_\mathrm{m}$ are the obtained estimates of the function, $\mathrm{N_m}$ is the number of values. The main disadvantage of decision trees is the possibility of uncontrolled overfitting and low generalisation capacity, so nowadays, instead of decision trees, a random forest decision tree is typically used. The essence of this method is that not one, but several decision trees are built on the sample. For each tree, a subsample of data is randomly allocated, upon which the model is built and the obtained parameters are then averaged. This significantly reduces the possibility of model overfitting when compared with a separate decision tree. Unlike the decision tree method, attributes are randomly selected at the nodes, and the optimal solution is searched for based on them (Chart 3).

Chart 3. Random forest model operation chart

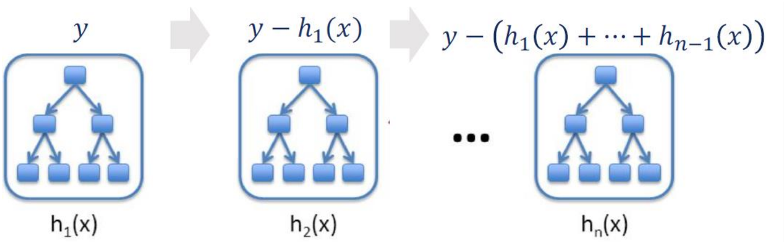


Source: https://scikit-learn.org.

In the case of regression, the algorithm provides a forecast for each tree, after which the obtained values are averaged. The advantages of this method include resistance to overfitting, distributed training of the model, and no need for preliminary preparation of the attribute series. Its main disadvantage lies in moderate interpretability of the obtained results compared with a decision tree.

The gradient boosting decision trees method uses many decision trees, but, unlike the random decision tree forest model, they are built not on data subsamples, but on a full sample. After building the first tree, we obtain the solutions h1(x), at the second step, we build a tree in which we use the errors of the first y-h1(x) as an input parameter. Thus, at the second step, we are looking for a solution h2(x) to minimise the errors of the first tree. We can take as many such steps as necessary to obtain the required model parameters (Chart 4).

Chart 4. Gradient boosting model operation chart



Source: https://scikit-learn.org.

The main advantages of this method are high forecasting results and no need to preprocess the data. Its disadvantages include high demands for the precise setting of parameters, poor work with sparse data and low results interpretability. Gradient boosting is one of the most powerful and efficient forecasting methods used in various fields of machine learning. In our task of forecasting regional inflation rates, it produced some of the best results.

Based on the obtained data, the Root Mean Square Error ($RMSE_x$) forecast quality metric was calculated for each model and correlated with the $RMSE_b$ of the benchmark.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(a_i - y_i)^2} \quad (8),$$

where m is the number of forecast points, a_i is the predicted CPI value, y_i is the real CPI value on the test out-of-sample.

Therefore, $RMSE_x/RMSE_b$ was obtained for each model. If the obtained relative indicator is less than 1, then the model's RMSE is better than the benchmark, and vice versa.

Having prepared forecasts with the above machine learning methods, we received a combined forecast. The method is based on averaging several forecasts, taking into account the quality of their forecast. The forecast weight for each model is calculated based on the inverse $RMSE_x$ estimate to the sum of similar estimates for all models:

$$W_x = \frac{1/RMSE_x}{\sum_{x=1}^{N} 1/RMSE_x} \quad (9),$$

$$\sum_{x=1}^{N} W_x = 1 \quad (10),$$

where W_x is the forecast weight of the model x in the combined forecast, N is the number of machine learning models included in the combined model.

This approach allows increasing the contribution of the best models with the minimum RMSE and minimising the forecast weight using the models, as the quality shown on the out-of-sample deteriorates.

To alternatively assess the stability of the results obtained, we conducted the Diebold – Mariano test (Diebold, Mariano, 1995). This allowed us to identify significant differences between the benchmark method and the combined forecasting method. This test is resistant to deviations from the assumptions of normality, autocorrelation and deviations from the zero mean error rate. The null hypothesis of the test is the absence of differences between predictive properties of the two models, i.e. the average difference between the loss functions of the compared models is equal to zero ($\bar{d}$):

$$\bar{d} = \frac{1}{T}\sum_{t=1}^{T}(g(e_{At}) - g(e_{Bt})) \quad (11),$$

where $g(e_{At})$ is the loss function of model A, $g(e_{Bt})$ is loss function of model B, T is the forecasting horizon of the time series.

In our case, the mean square error MSE acts as a loss function. The decisive statistic (S) of the test is as follows:

$$S = \frac{\bar{d}}{\sqrt{\hat{\sigma}^2/T}} \quad (12)$$

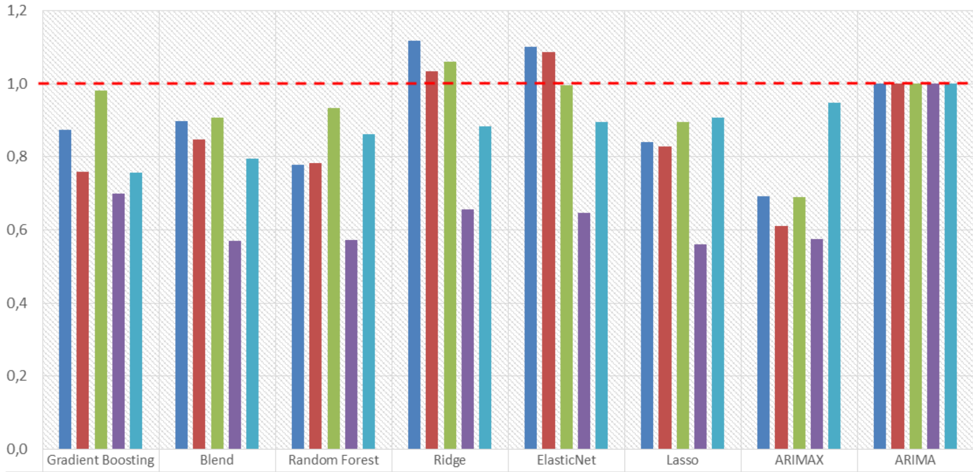
where $\hat{\sigma}^2$ is the estimated variance of a number of differences in the loss function values, T is the forecasting horizon of the time series.

In our study, we estimate the p-value as the probability that the null hypothesis is accepted and that the combined forecast based on machine learning models and the ARIMAX model do not differ significantly.

5. RESULTS

In the CPI case for the Siberia Macroregion (Chart 5), most machine learning models performed better than benchmarks for a 2-year forecast horizon. Moreover, the general trend suggests that machine learning shows comparable estimates with benchmark models for a period of one year or less. The combined forecast model and all machine learning methods employed showed better results in the 2-year medium-term forecast than the benchmarks.

Chart 5. CPI forecast quality estimates for the Siberia Macroregion ($RMSE_x/RMSE_b$)

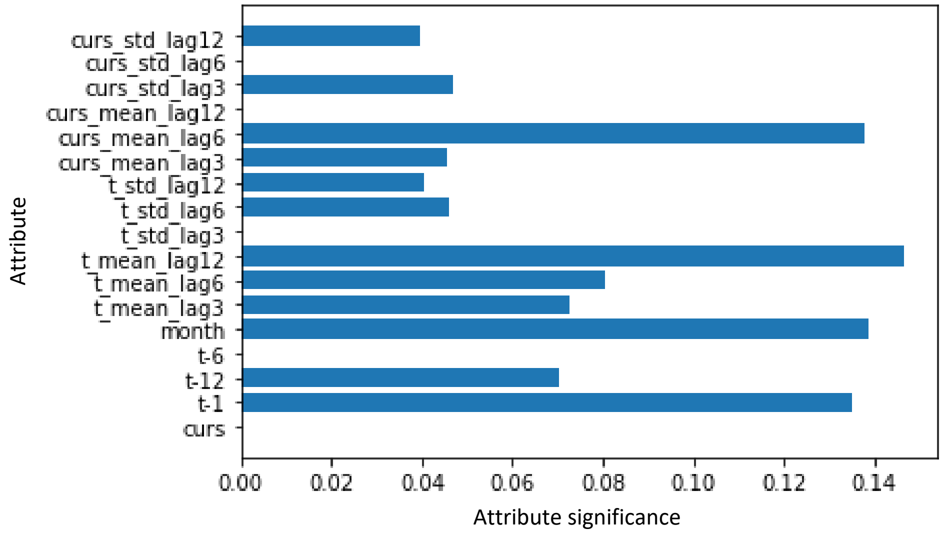


	Gradient Boosting	Blend	Random Forest	Ridge	ElasticNet	Lasso	ARIMAX	ARIMA
3	0,87	0,90	0,78	1,12	1,10	0,84	0,69	1,00
6	0,76	0,85	0,78	1,03	1,09	0,83	0,61	1,00
9	0,98	0,91	0,93	1,06	0,99	0,89	0,69	1,00
12	0,70	0,57	0,57	0,66	0,65	0,56	0,58	1,00
24	0,76	0,79	0,86	0,88	0,90	0,91	0,95	1,00

Source: authors' estimates.

A group of factors that have the greatest influence on the forecasted indicator in the gradient boosting model (Chart 6) includes such factors as the average CPI value over the last 12 months (t_mean_lag12) and the previous CPI value (t-1). The month factor is the ordinal number of the month in the year that is widely used in practice in this kind of models. The nominal exchange rate is best reflected in the model as the average over the last 6 months (curs_mean_lag6). The CPI and nominal exchange rate standard deviations have the least contribution to the model quality.

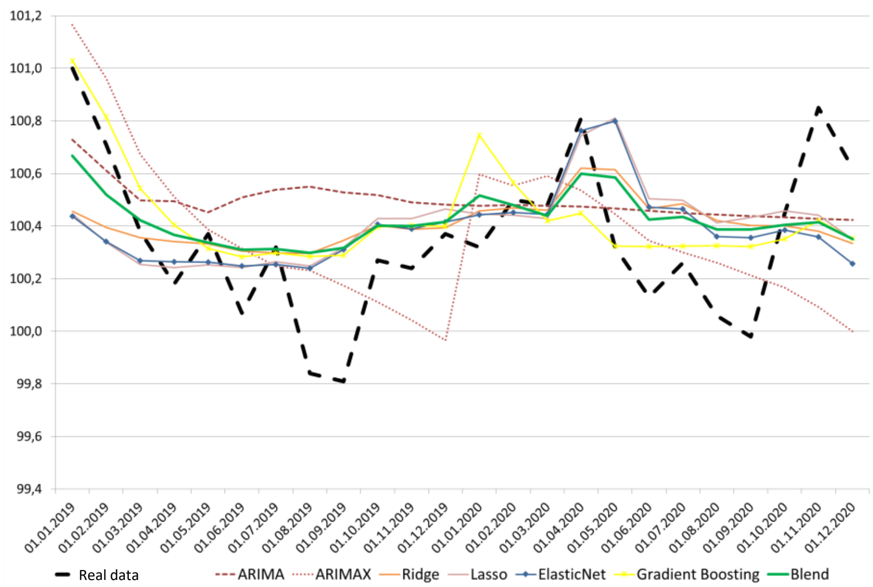
Chart 6. Significance of factors in the gradient boosting model for the Siberia Macroregion



Source: authors' estimates.

A graphical comparison of the real out-of-sample series and forecasts obtained from the study of data for the Siberia Macroregion for 24 months (Chart 7) also demonstrates that the forecasts by machine learning methods show the best predictive power. The acceleration of inflation rates in April 2020 was mainly predicted by the L1- and L2-regularisation methods. The benchmark models mainly show the general direction of inflation dynamics, but some local short-term accelerations or decelerations are not reflected in forecasts by these methods.

Chart 7. CPI forecasting for the Siberia Macroregion for 24 months



Source: authors' estimates.

The averaged results of the model quality assessment for 12 Siberian regions are presented in the Table 1. Assessed values are sorted by the column with forecasts for 24 months (Chart 7). The cells show relative assessment of the model quality by forecast period and modelling method relative to the ARIMA model which quality is taken as 1. The lower the cell value, the lower the $RMSE_x$ of the model relative to ARIMA is, which means it is a better forecast.

Table 1. Averaged results for regions of the Siberia Macroregion

Average $RMSE_x/RMSE_b$					
Forecast period (months)	**3**	**6**	**9**	**12**	**24**
Gradient Boosting	0.87	0.76	0.98	0.70	0.76
Blend	0.90	0.85	0.91	0.57	0.79
Random Forest	0.78	0.78	0.93	0.57	0.86
Ridge	1.12	1.03	1.06	0.66	0.88
Elastic Net	1.10	1.09	0.99	0.65	0.90
Lasso	0.84	0.83	0.89	0.56	0.91
ARIMAX	0.69	0.61	0.69	0.58	0.95
ARIMA	1.00	1.00	1.00	1.00	1.00

Source: authors' estimates.

Generalised data for the Siberian regions confirmed the authors' hypothesis that machine learning methods provide better forecasts for the periods of 12 and 24 months. The combined forecast (blend) method on the horizon of 12 and 24 months showed the best result almost on a par with lasso and random forest.

As an additional assessment, we tested the absence of significant differences between the benchmark forecasts (ARIMAX) and the combined forecast (blend) method to predict the CPI for the Siberia Macroregion. As a null hypothesis, we assumed that the forecasts of these models do not differ significantly. As a result, we obtained the following p_values for the null hypothesis (Table 2):

Table 2. Results of the Diebold – Mariano test for the Siberia Macroregion CPI forecast

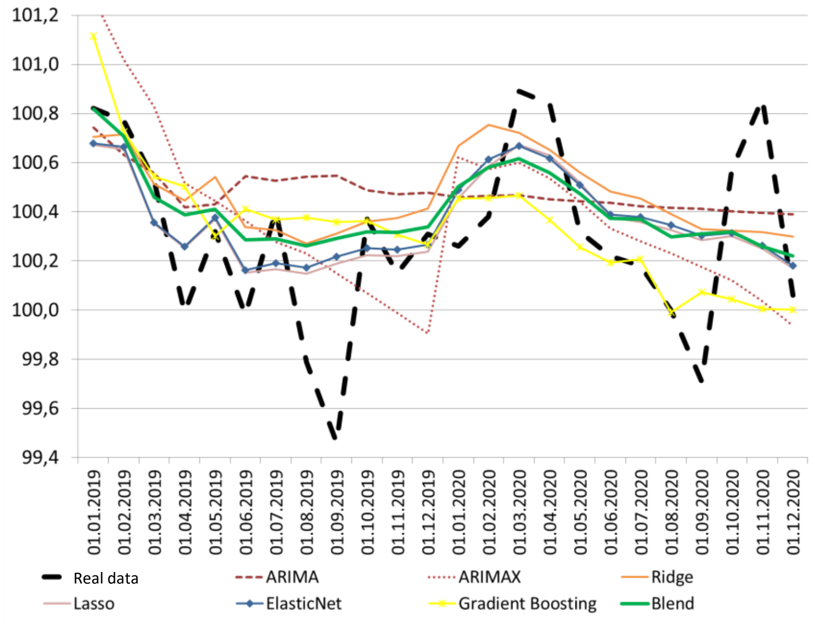
Forecast period (months)	3	6	9	12	24
P_value DM_test (blend vs ARIMAX)	0.04	0.1917	0.1856	0.15	0.017

Source: authors' estimates.

The results allow us to reject the null hypothesis at the 95% confidence level for forecasts using the combined forecast and ARIMAX methods for 3 and 24 months. Moreover, the ARIMAX forecast turns out to be more accurate on a 3-month horizon, and the combined forecast on a 24-month horizon. For other forecasting periods, we cannot accept the alternative hypothesis at the 95% confidence level. The achieved result is confirmed by the conclusions based on the $RMSE_x/RMSE_b$ ratios. Therefore, we can assume that there are significant differences between the forecasts using the benchmark (ARIMAX) method and the combined forecast (blend) method for the Siberia Macroregion CPI and it is appropriate to apply the combined forecasting method to accomplish such task.

Forecasting the CPI for twelve regions of the Siberia Macroregion, on average, also produced the best results using machine learning methods for forecast periods of a year or more. As an example, we present forecasts for the Novosibirsk Region for 24 months (Chart 8). The ARIMAX model showed a forecast similar in quality to machine learning methods. The best results were shown by the L1- and L2-regularisation methods and combining machine learning models due to a better forecast of the dynamics in 2020 Q1.

Chart 8. CPI forecasting for the Novosibirsk Region for 24 months



Source: authors' estimates.

For the Novosibirsk Region, machine learning models also perform better than benchmarks for periods of more than a year (Table 3).

Table 3. CPI forecast quality ratings for the Novosibirsk Region

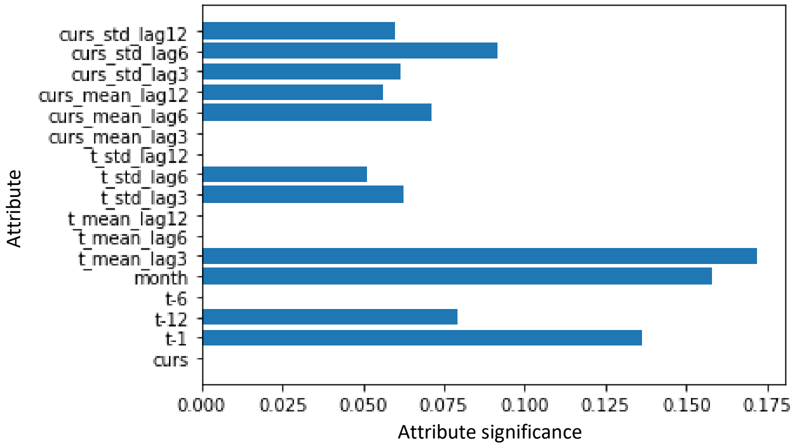
RMSE$_x$/RMSE$_b$					
Method / Forecast period (months)	**3**	**6**	**9**	**12**	**24**
Blend	0.94	1.05	0.91	0.73	0.87
Elastic Net	0.92	1.03	0.88	0.64	0.89
Gradient Boosting	0.86	1.10	0.96	0.88	0.89
Lasso	0.99	1.06	0.94	0.77	0.90
Ridge	1.07	1.11	1.03	0.80	0.91
ARIMAX	0.70	0.97	1.13	0.72	0.96
ARIMA	1.00	1.00	1.00	1.00	1.00
Random Forest	0.99	1.13	0.96	0.81	1.16

Source: authors' estimates.

For periods of 1 and 2 years, the forecasts obtained by machine learning methods are comparable to the quality of forecasts that use benchmark methods. We can conclude that it is possible to use machine learning methods to forecast CPI for the Novosibirsk Region for 24 months by combining forecasting methods, in particular the elastic net and gradient boosting together with the ARIMAX benchmark method.

The gradient boosting method mostly uses averaged data, the standard deviations of the exchange rate and the original CPI series for the forecast (Chart 9). The most important factors in this model were the average CPI values for 3 months (t_mean_lag3), the CPI value with a lag of 1 (t-1), the number of the month (month) and the standard deviation of the exchange rate for the last 6 months (curs_std_lag6).

Chart 9. Model factor significance for the Novosibirsk Region by the elastic net method



Source: authors' estimates.

Therefore, we can conclude that it is possible to use machine learning methods to forecast inflation rates in the Siberia Macroregion and Siberian regions together with traditional econometric methods. At the same time, it is necessary to assess the quality of forecasting using machine learning methods for each region in advance to determine whether it makes sense to use them over traditional econometric methods.

6. CONCLUSION

According to the results of this study, we can conclude that the quality of inflation forecasts in the Siberia Macroregion and Siberian regions using machine learning methods is comparable to the traditional econometric methods ARIMA and ARIMAX. However, it is necessary to assess the forecast quality in advance and consider whether it makes sense to use them for each region. For most Siberian regions and the Siberia Macroregion, machine learning methods show better quality than benchmarks for periods beyond a year unlike forecasts for up to one year. Forecasting with combining machine learning models is in most cases preferable to any one.

A possible area for the development of this study could be a transition from pseudo-real time forecasting to real time forecasting. In this case, we would be able to rely on scenario forecasts of the nominal ruble exchange rate and add scenario forecasts of other factors. Assessing the quality of forecasting inflation components and comparing the quality of forecasts before and after an economic crisis would also be a challenge to develop this study.

In addition, an expert adjustment can be added to the obtained forecast, which will help the model take into account one-time factors that have not manifested themselves in the past but that can affect regional inflation rates.

If necessary, it is possible to expand the outlined approach to other Russian regions, while maintaining an individual account of regional specifics and expert adjustment of the results obtained.

7. REFERENCES

1. Athey, S. & Imbens, G.W. (2019). Machine Learning Methods Economists Should Know About. Annual Review of Economics, 11(1), Pp. 685–725.
2. Barkan, O., Benchimol, J., Caspi, I., Hammer, A. & Koenigstein, N. (2021). Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks. Bank of Israel Working Paper.
3. Baybuza, I. (2018). Inflation Forecasting Using Machine Learning Methods. Russian Journal of Money and Finance, 77(4), Pp. 42–59. (In Russian)
4. Chakraborty, C. & Joseph, A. (2017). Machine Learning at Central Banks. Bank of England Working Paper, 674.
5. Diebold, F.X. & Mariano, R.S. (1995). Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13(3), Pp. 253–263.
6. Jung, J., Patnam, M. & Ter-Martirosyan, A. (2018). An Algorithmic Crystal Ball: Forecast-based on Machine Learning. IMF Working Paper, 18/230.
7. Mamedli, M. & Shibitov, D. (2021). Forecasting Russian CPI with Data Vintages and Machine Learning Techniques. The Bank of Russia Working Paper Series, 70.
8. Ozgur, O. & Akkoç, U. (2021). Inflation Forecasting in an Emerging Economy: Selecting Variables with Machine Learning Algorithms. International Journal of Emerging Markets.
9. Pavlov, E. (2020). Forecasting Inflation in Russia Using Neural Networks. Russian Journal of Money and Finance, 79(1), Pp. 57–73. (In Russian)
10. Yakovleva, K. (2018). Text Mining-based Economic Activity Estimation. Russian Journal of Money and Finance, 77(4), Pp. 26–41. (In Russian)