Comparative Analysis of Methods of Forecasting the Consumer Price Index for Food Products (on the Example of the Altai Territory)

Stepan Mezhov and Maxim Krayushkin
Altai State University, 61 Lenin avenue, Barnaul, Russia
megoff@mail.ru, kramaks-97@mail.ru

Keywords: Consumer Price Index, Time Series, Neural Network, Decision Tree, Error Back Propagation, Gradient Boosting, Forecast.

Abstract: At the moment, there are no uniform universal methods for forecasting regional indicators of economic development in general and the consumer price index in particular. But depending on how accurate and reasonable the forecasts of the consumer price index will be, the budget of the region will be drawn up so correctly and the parameters of the forecast of socio-economic development, in the calculation of which this indicator is used, will be accurately predicted. The article presents a comparative analysis of methods for forecasting the consumer price index for food products. First, the most popular methods of forecasting the consumer price index were identified. Then models of time series, neural networks and decision trees were built, as well as retro-forecasts of the consumer price index for food products based on them. It is revealed that neural networks provide higher accuracy of forecasts compared to other models. The result of the work was the forecast of the consumer price index for food products in the Altai Territory for 2021 based on the constructed neural network model. The constructed neural network models can be used in relevant organizations to increase the accuracy of the forecast of this indicator. In addition, such an approach can be used as a basis for forecasting other indicators that characterize the socio-economic development of regions.

1 INTRODUCTION

Forecasts of the socio-economic development of the region, including the forecast of the consumer price index (hereinafter also – CPI), are sent to the Ministry of Finance of the region and to the Ministry of Economic Development of the Russian Federation. Using these data, the Ministry of Finance of the region develops the main parameters of the regional budget, and the Ministry of Economic Development of the Russian Federation clarifies the forecast of socio-economic development of the Russian Federation and monitors the socio-economic development of the region.

At the moment, there are no uniform universal methods for forecasting regional indicators of economic development in general and the consumer price index in particular.

The following are identified as the most popular methods used to predict the consumer price index:

- Based on time series analysis.
- Based on artificial neural networks.

This conclusion was made based on a meaningful analysis of 30 scientific papers.

Also, the construction of forecasts of the consumer price index based on decision trees is currently gaining popularity. In 4 out of 30 cases they were used [1, 2, 3, 4], and in 6 cases other methods: [5, 6, 7, 8, 9, 10].

With regard to other methods, it should be noted that such as regression analysis, factor analysis, and a method based on the construction of a system of balanced indicators were used to predict this indicator.

As factors for the construction of CPI models, the authors usually used such as: the price index of manufacturers of industrial products, retail trade turnover, the volume of paid services to the population, the dollar exchange rate, the price index for agricultural products and others.

As specific information technologies for building models, the authors used both the programming languages: R, Python and the like, as well as modern software packages for statistical data analysis, for example, the Statistica software package.

2 CONSTRUCTION AND COMPARATIVE ANALYSIS OF FORECAST CPI MODELS FOR FOOD PRODUCTS IN THE ALTAI TERRITORY

2.1 Construction of a Forecast Model of the CPI for Food Products in the Altai Territory Based on Time Series Analysis

As it has already been revealed, the method based on time series analysis is very popular in forecasting the consumer price index. To build such models, a representative sample was first determined. It is believed that it is advisable to train models with data from 2017, since similar economic conditions have been formed since that time.

That is, first, data on the CPI for food products in the Altai Territory from January 2017 to December 2020 were taken. Then the analysis of the time series was carried out.

The analysis was carried out based on the expert analysis of a number of data, and then using the construction and analysis of the correlogram.

Based on this analysis and the specifics of this group of products, it was determined that a number of data have all deterministic components (trend and seasonality). It is believed that if they are available, it is advisable to build a Holt-Winters model [11].

It should be noted that to build a forecast using a time series model for one year, it usually takes at least 2 years to train it [11].

We consider it appropriate to train the model first with data from 2017 to 2018 and test it on 2019 data, and then train it with data from 2018 to 2019 and test it on 2020 data.

Schematically, the Holt-Winters model can be shown as follows (1):

$$IPC_t = A + B * t + S_i, \quad (1)$$

where: IPC_t – forecast value of CPI for food products, A – constant, B – angular coefficient, t – forecast period, S_i – seasonality coefficient.

It should be noted that when constructing the Holt-Winters model, special coefficients characterizing the sensitivity of a data series to components are usually also estimated.

Initially, the Holt-Winters model of the CPI for food products was built according to data from 2017 to 2018. Her work on training and test data is presented below (Figure 1).

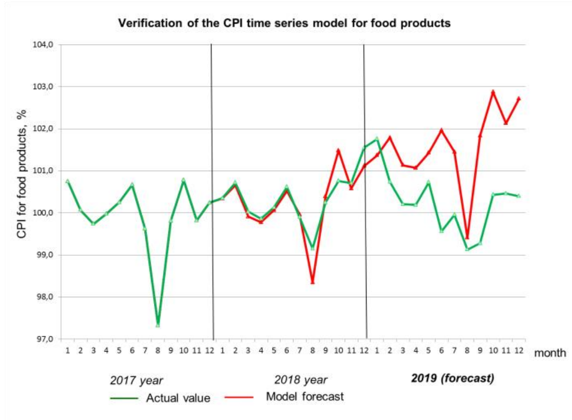

Figure 1: Operation of the CPI time series model for food products trained from 2017 to 2018.

The value of the middle absolute error (hereinafter also referred to as MAE) of the model on the training data is 0,23 percentage points (hereinafter also referred to as p. p.), which is unacceptable. It should be noted that for 8 months of 2019, the forecast for the MAY model was 1,73 p. p.

The choice in favor of the MAE indicator was made due to the fact that in order to solve our task, it is important that the deviation of real data from the predicted values obtained by the model on the training data was on average no more than 0,05 percentage points, and on the forecast average no more than 0,35 percentage points. Otherwise, in our case, it is impractical to use this model in the future for forecasting.

After that, a model of time series of CPI for food products was built according to data from 2018 to 2019. The work on training and test data is shown in Figure 2.

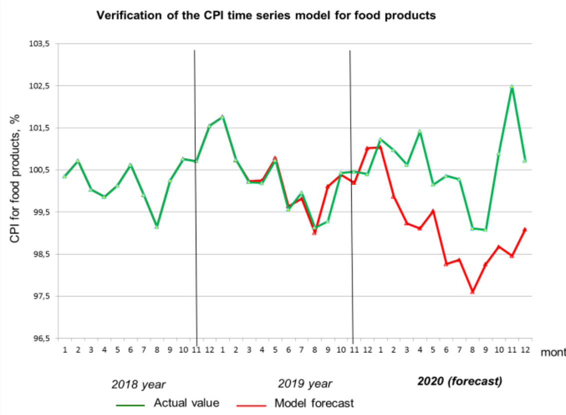

Figure 2: Operation of the CPI time series model for food products trained from 2018 to 2019.

The value of the average absolute error (MAE) of the model on the training data is 0,19 percentage points, which is unacceptable. Note that for 6 months of 2020, the forecast for the model MAY was 2,02 percentage points.

Thus, in our case, this model is impractical to use in forecasting.

2.2 Construction of a Forecast Model of the CPI for Food Products in the Altai Territory Based on Neural Networks

As already noted, it is advisable to train a neural network model on a data set starting in 2017.

Based on the analysis of Pearson pair correlation coefficients and expert analysis of possible factors affecting the CPI for food products, it was found that it would be advisable to take the CPI for food products as factors for training the neural network model: The CPI of the Russian Federation for food products, the producer price index for agricultural products sold in the Altai Territory, the producer price index for industrial goods according to the type of economic activity «Food Production» in the Russian Federation. This is due to the consideration of real economic processes that provide the necessary correlation.

The training of neural networks with a teacher was chosen as a paradigm, the training rule was error correction, the architecture was a multilayer neural network, the learning algorithm was reverse propagation. The choice was made during the analysis of various learning algorithms for solving the forecasting problem.

The Deductor Studio software environment was used to build neural networks. This is due to a number of factors: the simplicity of building a model, the possibility of additional training and a user-friendly interface for a user who does not have high qualifications [12].

When using the learning algorithm with a teacher, the weighting coefficients of the neural network are adjusted in such a way as to minimize deviations of the predicted values from the values of the test sample [13].

Of course, within the framework of solving our task, it is important that the modal value of the CPI be as close as possible to statistical, that is, the model error should be small enough.

Numerous publications on industrial applications of multilayer networks with a learning algorithm by the method of back propagation of errors have confirmed its fundamental operability in practice [14].

It should be noted that it is important when building models of neural networks that it is necessary to achieve the absence of the effect of retraining. After all, then the model works well on training data, but the forecast when using this model turns out to be inaccurate. Therefore, it is important to correctly approach the selection of network parameters. That is, it is necessary to find the optimal number of hidden layers, the activation function, the optimal number of training epochs so that the neural network is not retrained.

The scheme of operation of this algorithm is presented below (Figure 3).

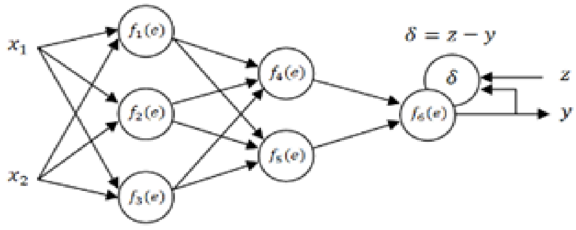

Figure 3: Diagram of the operation of the error back propagation algorithm.

When using this algorithm, the weighting coefficients are found at the beginning of the first epoch of training. Then they are adjusted from the output to the inputs, and not vice versa, in order to reduce the error of training and testing.

Neural networks have been trained for epochs. At first, 10,000 epochs were taken. Then it is determined on which of the epochs the error on the test sample is minimal. This served as a criterion for stopping learning.

A two-year sample for training a neural network with a prediction period of up to 1 year is quite enough.

Note that the constructed neural networks have 3 hidden layers of 8 neurons in each of them, and the activation function of neurons is sigmoidal.

The graph of neural networks can be represented as follows (Figure 4).

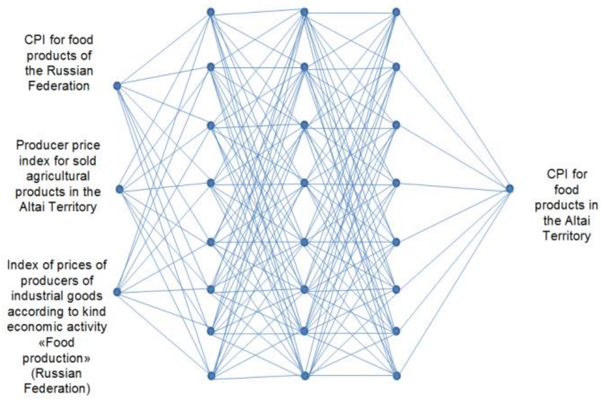

Figure 4: Graph of constructed neural networks of CPI for food products.

It should be noted that in our case, three hidden layers of 8 neurons in each of them are enough. This conclusion is made because it is believed that the number of hidden layers should not exceed the number of network inputs, and the number of neurons should not exceed the number of observations.

First, a neural network of CPI for food products was built according to data from 2017 to 2018. The results of her work on training and test data are shown in Figure 5.

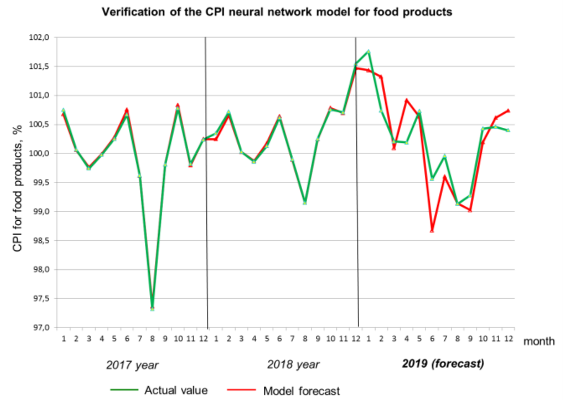

Figure 5: Verification of a neural network for predicting CPI for food products, trained on 2017-2018 data.

The value of the average absolute error of the model (MAE) was 0,03 percentage points, which is generally acceptable. Note that the forecast error obtained using the neural network model is 0,29 p. p.

Then the CPI neural network for food products was built according to data from 2018 to 2019. The results of her work on training and test data are shown in Figure 6.

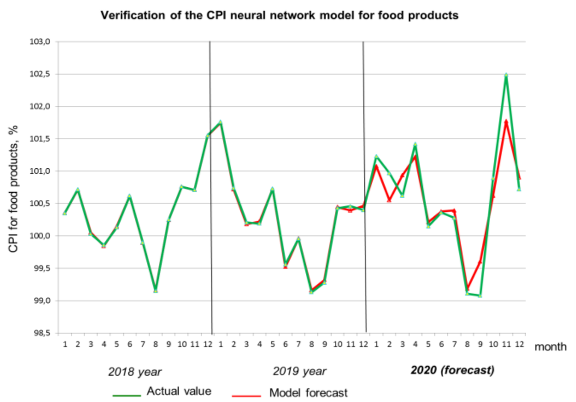

Figure 6: Verification of a neural network for predicting CPI for food products, trained on 2018-2019 data.

The value of the average absolute error of the model (MAE) in this case was 0,02 percentage points, which is also acceptable. The forecast error for 6 months of 2020 based on the neural network model is 0,31 percentage points.

These two examples confirm the applicability of neural network models in predicting CPI, which significantly increase the accuracy of the forecast.

2.3 Forecasting the Consumer Price Index for Food Products in the Altai Territory Based on Decision Trees

The next fairly popular approach to constructing forecasts of the consumer price index is the approach based on the construction of decision trees.

The process of building a decision tree can be represented as the following steps:

- Definition of input and output parameters.
- Building a tree.
- Evaluation of the quality of the decision tree.

We will use the same input and output parameters as for training neural networks.

At the second stage, we will build the tree itself. The ported version of the Statistica software package is used, because of the possibility of a fairly fast and convenient construction of a decision tree in it by the «Gradient Boosting» method, which is widely used in forecasting.

The idea of «Gradient Boosting» is an iterative process of sequential tree construction. The new tree is trained using information about errors made at the previous stage. This technique uses the idea that the next model will learn from the mistakes of the previous one [15].

Models have an unequal probability of appearing in subsequent models, and those that give the greatest error will appear more often. That is, many trees are being built, each of which can be more accurate. The process ends when the required accuracy is reached or the accuracy begins to fall due to retraining [15].

In the Statistica software package, it is possible to build different decision trees using different methods. But it is for forecasting that it is better to use the «Gradient Boosting» method when constructing them.

A lot of trees are being built in this package. As a result of «Gradient Boosting», the most accurate tree is determined.

In addition, the second most popular method for predicting using decision trees is the «Random Forest».

The essence of this method is to use an ensemble of decision trees. By itself, the decision tree provides an extremely low quality of classification, but due to the large number of them, the result is significantly improved.

Schematically, a fragment of the decision tree model can be represented as follows (Figure 7).

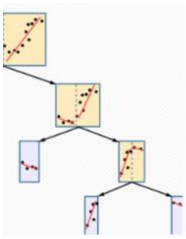

Figure 7: A fragment of the decision tree.

Initially, a decision tree of the CPI for food products was built according to data from 2017 to 2018 (Figure 8).

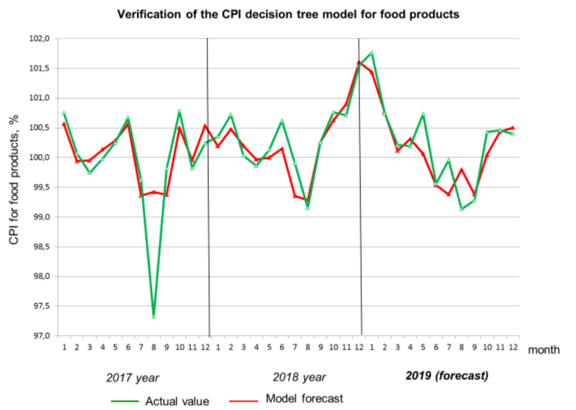

Figure 8: The work of the CPI decision tree for food products trained from 2017 to 2018.

The value of the average absolute error (MAE) of the model on the training data is 0,28 p. p. Note that for 8 months of 2019, the MAE forecast for the model was 0,32 p. p.

After that, a decision tree of the CPI for food products was built according to data from 2018 to 2019 (Figure 9).

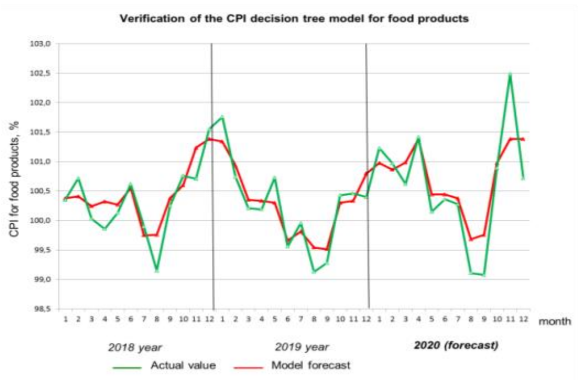

Figure 9: The work of the CPI decision tree for food products trained from 2018 to 2019.

The value of the average absolute error (MAE) of the model on the training data is 0,24 p. p. Note that for 6 months of 2020, the value of the average absolute error of the forecast for the model is 0,53 p. p.

2.4 Comparative Analysis of Forecast Models and the Construction of a Forecast of the CPI for Food Products in the Altai Territory for 2021

In this section, the forecast of the CPI for food products in the Altai Territory was built on the basis of an optimal model. Initially, the optimal model was selected based on the absolute verification of retro-forecasts (Table 1).

Table 1: The fact of absolute verification of retro-forecasts.

Training period	Forecast period	MAE retro-forecast, p. p.		
		Time Series model	Neural network model	Decision Tree model
from 2017 to 2018	May-December 2019	1,73	0,29	0,32
from 2018 to 2019	July-December 2020	2,02	0,31	0,53

Table 1 shows that neural network models are the most accurate.

Then it was on their basis that the forecast of the CPI for food products in the Altai Territory was built. The forecast value of December 2021 by December 2020 was 109,3 percent.

3 CONCLUSION

Forecasting the consumer price index is of great importance for regional development. The correctness of the formation of the budget of the region will depend on how accurate and reasonable the forecast of this indicator is.

During the analysis of scientific papers devoted to the prediction of the CPI, it was revealed that the most common methods of forecasting this indicator are an approach based on time series analysis, forecasting based on the construction of artificial neural networks and decision trees.

Then models of time series, neural networks, and CPI decision trees for food products were built. It should be noted that neural network models were chosen as the best, on the basis of which the forecast of the CPI for food products for 2021 was built. When constructing econometric models, programs such as R-Studio, Deductor Studio, Statistica were used.

REFERENCES

[1] V. A. Baykov and A.Y. Tarasova, “An integrated approach to forecasting the inflation rate”, Bulletin of the Moscow University of Finance and Law, 2019, no. 7, p. 49-57.

[2] N.N. Karabutov, “Interrelation of indices determining the level of inflation”, Economy, Taxes, Law, 2020, no. 10, pp. 40-47.

[3] A. A. Skrobotov and A.V. Tsarev, “Forecasting of macroeconomic indicators of the Russian economy”, Economic development of Russia, 2020, no. 18, pp. 45-55.

[4] S. G. Shulgin, “Selection of variables for instability analysis and forecasting using gradient boosting models”, System monitoring of global and regional risks, 2018, no.8, pp. 115-153.

[5] E. V. Balatsky, N. A. Ekimova, and M. A. Yurevich, “Short-term forecasting of inflation based on marker models”, Forecasting problems, 2019, no. 5, p. 28-40.

[6] I. A. Vakhrushev, “Forecasting the trend dynamics of the stock market based on macroeconomic factors using a diffuse index”, Scientific Journal of ITMO Research University, The series «Economics and Environmental Management», 2020, no. 5, pp. 42-48.

[7] F. E. Huseynova, “Analysis of medium-term trends in the dynamics of inflationary processes of the Russian economy”, Skif. Questions of student science, 2019, no. 7, p. 12-20.

[8] I. N. Dementieva, “Application of the index method in studies of consumer sentiment of the population”, Economic and social changes, 2019, no. 1, pp. 153-173.

[9] E. B. Mitsek and S. A. Mitsek, “Analysis of factors of dynamics of the main macroeconomic variables of the Russian Federation”, Questions of management, 2020, no. 1, pp. 47-61.

[10] A. Yu. Yakimchuk, A. I. Teplenko, and M. N. Konyagina, “The influence of the key rate on inflation rates in modern Russia”, Bulletin of the Academy of Knowledge, 2020, no. 37, pp. 490-497.

[11] I. N. Dubina, Mathematical and statistical methods in empirical socio-economic research: textbook, Moscow, Finance and Statistics, 2010, 415 p.

[12] Deductor - platform capabilities. Official website of BaseGroup Labs, 2021, [Online]. Available: https://basegroup.ru/deductor/description.

[13] F. Wasserman, “Neurocomputer technology”, Theory and practice: a textbook, Moscow, 1992.

[14] S. A. Terekhov, Lectures on the theory and applications of artificial neural networks: textbook, - Snezhinsk, 1998.

[15] Database of examples of solving specific management tasks in the statistica system. Official website of Statsoft, 2021, [Online]. Available: http://statsoft.ru/solutions/ExamplesBase/tasks.