

5/2022,
sentyabr-
oktyabr
(№ 00061)



COMPARISON OF ARIMA MODEL AND ANN MEASURES IN FORECASTING REGIONAL MACROECONOMIC INDICATORS

Namozov Gafur Shokulovich

Doctoral student of Termez State University. Termez, Uzbekistan.

tofa82@mail.ru

DOI: https://doi.org/10.55439/EIT/vol10_iss5/a29

Abstract

In this article, the process of modeling and forecasting the volume of the gross regional product of the Surkhandarya region was carried out. The forecasting efficiency of the ARIMA and ANN models is also compared. As a result, the best ARIMA model in forecasting is ARIMA (1, 0, 1)(0, 1, 0)₄, and according to ANN, the model of the output neuron MLP (4-7-1) turned out to be the best. In addition, the volume of the gross regional product is predicted until 2025.

Keywords: GRP volume, ARIMA model, ANN model, Box-Jenkins method, Akaike information criterion, Schwartz information criterion, forecasting.

Introduction

Forecasting has been one of the most popular methods for planning the state's socioeconomic development and future strategic goals, as well as ensuring the population's well-being, in recent years. As a result, one of the pillars of forecasting is the use of time series in predicting the state of the global market. However, constructing such a forecasting model is difficult due to the great variability of the rules that underpin the time series. Many scholars have proposed and developed various approaches for analyzing and projecting socio-sustainable development efforts throughout the last few decades. One of these is that the ARIMA model has been used as a primary tool for evaluating many new modeling approaches (Gosasang V., Chan., W. and Kiattisin, S. 2011.)[2]. However, ARIMA is a common variable model, developed based on the assumption that predicted time series are linear and stationary, and a well-known function in artificial neural networks, forecasting and system modeling, has played a major role in time series analysis and forecasting in recent years. (J. Yao and C.L. Tan, S. Yaser and A. Atiya)[3-4], 1996. Although many studies have been conducted to study economic and social development forecasting in neuronal network forecasting, none of the many studies conducted in Uzbekistan so far have used the ANN and ARIMA models to predict GRP in the region. In particular, we raise three research questions. Initially, this study developed models to predict the approximate value of GNP volume in subsequent years using ARIMA and ANN indicators. Second, we reviewed and compared the results and trends of the projected values of gross regional product volume. Thirdly, we used the ANN and ARIMA approaches to compare outcomes and trends in

forecasting regional GVA volumes to assess the significance of average absolute percentage error (MAPE) models and obtain new conclusions.

Literature review

As we already know, quantitative forecasting models can actually be divided into two categories: time series-based models and causal methods. Time series analysis seeks to determine the appearance of a model by requiring previous data, and extrapolation to the future, in other words, allows the prediction required to provide a prediction that demonstrates a repetitive experience over time [8]. These categories include simple methods, sliding averages, trend curve analysis, exponential alignment, and sliding averages (ARIMA) models integrated into autoregression.

Mitrea, C.A., Lee, C. K. M., WuZ. compared different prediction methods such as the sliding mean (MA) and the sliding mean (ARIMA) model integrated into the autoregression as a forward (NN) and nonlinear autoregressive network with neural network (NN) models. Inputs a nonlinear autoregressive exogenous model (NARX) is a nonlinear autoregressive model with exogenous inputs. (PRICE), the results showed that forecasting using (NN) offers better prognostic indicators [5].

Nowrouz Kohzadi, and labeling Kaastra(1996), compared neural networks with ARIMA using monthly data on live cattle and wheat prices in 1950-1990. Neuron network models achieved a much lower mean square error than the ARIMA model [6]. On the other hand, the results show that the application of these indicators to traditional methods based on previous data output has several advantages. In this study, an attempt was made to select the best model by using ARIMA and ANN models and comparing the results in forecasting the GVA volume of the region.

Research methodology

(ARIMA)- Autoregressive integrated moving average model: The economic time series obtained using observations are generally not stationary, except in rare cases. Such time series will contain time-dependent non-random components. If the random balances of the time series form a stationary row, then the time series forms a non-stationary row. The **Autoregressive Integrated Moving Average** (ARIMA) model is used to describe such series [10].

ARIMA models allow the modeling of integrated time series. If the time series is not stationary and its -order differences are stationary, then this time series is considered to be integrated-order [11]. In other words, if the time series is not stationary, ARIMA switches to its -order (until stationary) differences to create a model.

The ARIMA model is usually expressed as ARIMA (p, d, q). p is the order of auto regression, -d is the order of the sliding average, and -q is the order of the differences. The model has the following general appearance:

$$\Delta^d Y_t = c + \sum_{i=1}^p \alpha_i \Delta^d Y_{t-1} + \sum_{j=1}^q \beta_j \varepsilon_{t-j} + \varepsilon_t$$

Here c , α_i , β_i are model parameters; $\Delta^d - d$ are operator of ordinal differences. (e.g. $\Delta^1 Y_t = Y_t - Y_{t-1}$ - first-order differences) [16].

ANN (**artificial neural networks**) model: Artificial neural networks are mathematical models, as well as their biological neural networks - software or hardware based on the principle of organization and operation of nerve cells networks of a living organism. Each

biological neuron receives signals from neighboring neurons through special nerve fibers. Individual artificial neurons also communicate with each other in different ways. This allows us to create different neural networks with different architectures, study rules, and capabilities [7].

The philosophy of the ANN approach is to develop an architecture inspired by the biological nervous system. The main advantage of neural networks is their flexibility in nonlinear modeling [14].

We used a forward-facing backward propagating neural network (in other words, (MLP) **-Multilayer perceptron-** also known as multilayer perceptron network) neural network model. The neural network model consists of three parts, the input layer, the latent layer, and the output layer are shown in Figure 1, and each input is measured by the corresponding w-weight. The sum of the weights with the input and the bias f (x) constitute the input to the activation function (Figure 1).

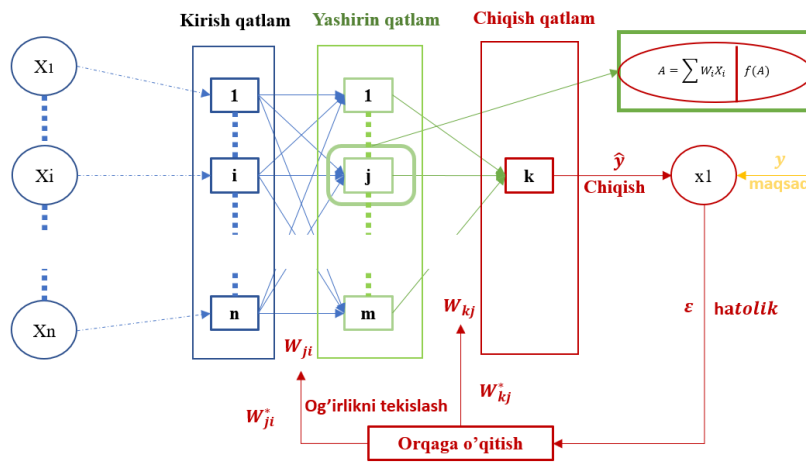


Figure 1. A normal backward propagating neural network

In an MLP network, two steps are considered: the start-up phase, in which the input sample is presented to the trained network and trained through successive layers of neurons until the result is achieved. Network weights or parameters are known as training and are needed to minimize error. Therefore, the study of the MLP network is controlled. In this study, we use the following three-layer feedback nets:

$$F = F \left\{ \beta_0 + \sum_{j=1}^j \beta_j G \left[\sum_{k=1}^k \gamma k_t X_j \right] \right\} \quad (2)$$

Here F is the output function of the output layer unit, β_0 -bias unit (equal to 1), G is the output function of the hidden layer units j, γk_t - k denotes the weight for the connection connecting the β_j input to the hidden j, - the weight of the output from the hidden layers in the output layer unit, X -input vector. MLP nets are typically layered nets trained with static back training and forward orientation.

Problem statement: Data processing was carried out using the statistical analysis software "Statistica10". Typically, statistical testing is performed to evaluate the performance of the ARIMA and ANN models. This test has an average absolute percentage error (MAPE). This criterion The following expression shows the process of calculating MAPE.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \cdot 100\% \quad (3)$$

Here:

n - number of observations, Y_t observed -real indicators, \hat{Y}_t -predicted indicators

Analysis and results

Today, GRP is one of the most important indicators of the economy of Surkhandarya region. According to statistics, in 2021 the volume of GRP in Surkhandarya region in current prices will reach 30090.5 billion soums. soums and increased by 7.8% compared to 2020. The GDP deflator index was 112.7% of prices in 2020.

Table 1 shows the volume of GRP (Y_t) (billion soums) in Surkhandarya region for the period from the 3rd quarter of 2012 to the 4th quarter of 2021.

Volume of GRP in Surkhandarya region from the 3rd quarter of 2012 to the 4th quarter of 2021 (billion soums)

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
1- quarte r	-	649,3 3	832,2 1	1013, 91	1407, 35	1746, 2	2245, 1	3092, 1	4078	4390, 9
2- quarte r	-	1130, 37	1414, 82	1673, 36	2214, 73	2954, 9	3971, 8	7009, 2	7303, 3	9139, 6
3- quarte r	944,5	1561, 17	1894, 64	2237, 47	2743, 34	2784, 3	3426, 7	5940, 6	5741, 3	7904, 3
4- quarte r	1316, 68	1534, 49	1935, 3	2238, 71	2807, 12	6216, 8	7418, 3	6307, 4	6881, 0	8655, 7

The data in Table 1 are taken from the official website of the Statistics Department of Surkhandarya region [17]. Quarterly values were calculated from the values published with increasing results on the site. Its dynamics are shown in Figure 2.

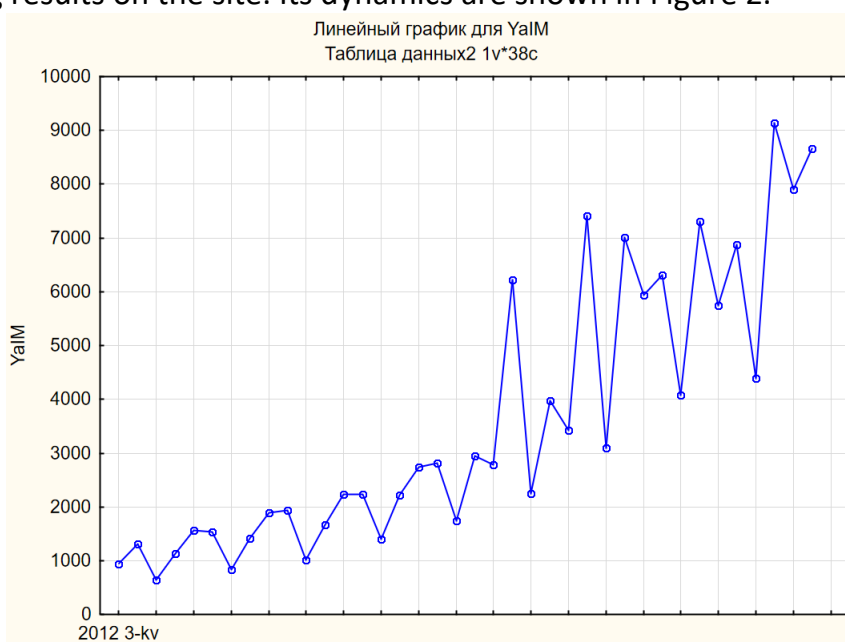


Figure 2. Growth dynamics of GRP in Surkhandarya region in 2012-2021

ARIMA model

Model identification in ARIMA is performed by determining the p, d, and q values for the series before evaluating the model using ACF and PACF diagrams.

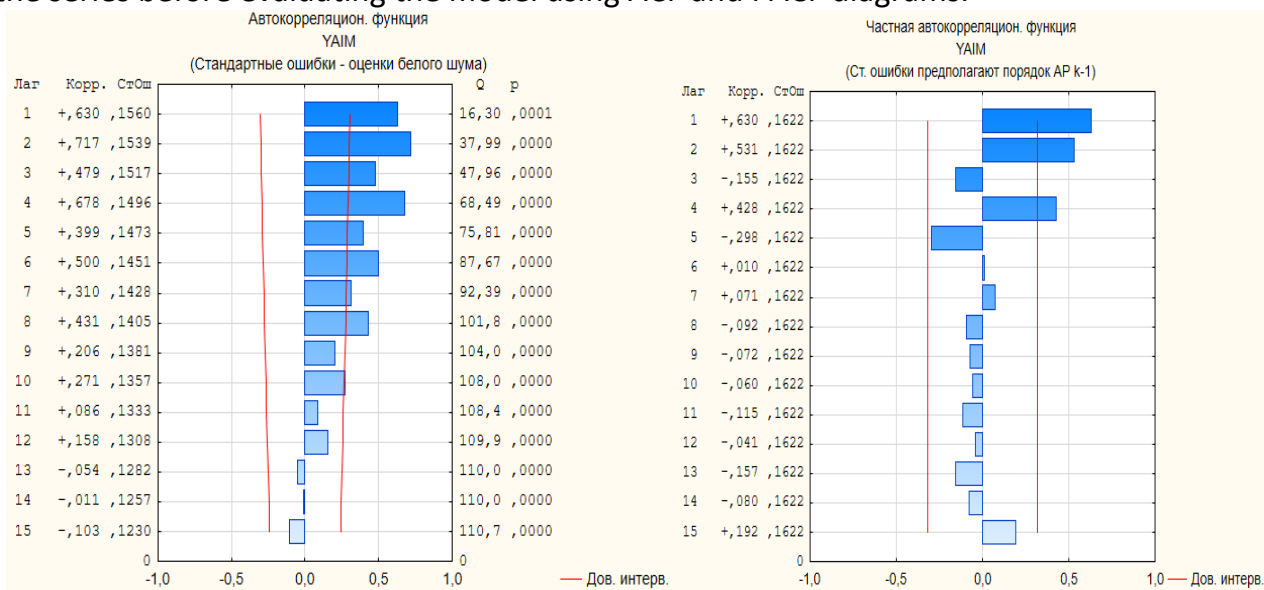


Figure 3. Correlogram of autocorrelation and special autocorrelation functions on the first differences of the time series

The figure above (Figure 3) shows the autocorrelation with columns and the specific autocorrelation coefficients and the 95% confidence interval limits with horizontal lines. It can be seen that the coefficients of the autocorrelation function are gradually approaching zero in the form of a sinusoid, and in the diagram of a special autocorrelation function there is a sudden insignificant lag after the first significant coefficient, i.e. there is a break. This indicates the order p = 1 and q = 1 in the ARIMA model. It can also be seen from the time series chart that it has a 4-quarter seasonality. Hence, the correlogram is specific to the autoregression process and it is necessary to take into account the presence of seasonality in the time series. Furthermore, the fact that the first-order differences of the time series are stationary means that it is first-order integrated, i.e., d = 1.

Based on the above, it is advisable to test the following models:

ARIMA (1,0,1); ARIMA (1,0,1)(1,0,0)4; ARIMA (1,0,1)(0,1,0)4;

ARIMA (1,0,1)(2,0,0)4; ARIMA (1,0,1)(2,1,0)4; ARIMA (1,0,1)(3,0,0)4

Evaluate the model parameters and check its adequacy. Typically, the following information criteria are used when comparing several models.

1. Akaike information criterion, AIC:

$$AIC = \frac{k + m}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right) \quad (4)$$

2. Swarz information criterion, SIC:

$$SIC = \frac{(p + q) \ln n}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right) \quad (5)$$

Here k is the number of parameters to be evaluated.

The AIC and SIC information criteria are used to determine the compatibility quality of the Box and Jenkins model. The model with the smallest AIC and SIC value is selected.

The AIC and SIC criteria (AIC = 644.95 SIC = 654.77) of the ARIMA (1,0,1) (3,0,0) 4 model and the parameters of the determination coefficient R2 = 0.88 are important. From this calculation, the most optimal model form could be selected. However, the determination coefficient of the ARIMA (1,0,1) (0,1,0) 4 model is R2 = 0.86, although the AIC and SIC criteria are the smallest (AIC = 568.11 SIC = 572.69).

One of the indicators characterizing the deviation from the actual values is the average absolute percentage error (approximation error coefficient), calculated by the above formula (3).

While the average absolute percentage error is up to 10%, the model is considered bad for high accuracy from 10% to 20% on average from 20% to 100%.

The results of the assessment show that MAPE for model ARIMA (1,0,1) (3,0,0) 4 = 14.22%, MAPE for model ARIMA (1,0,1) (0,1,0) 4 11.42%, Hence, the smallest indicator belongs to ARIMA (1,0,1) (0,1,0) 4 models (MAPE = 11,42%), indicating the most optimal model form for the time series. The results of the analysis show that ARIMA (1,0,1) (0,1,0) 4 model form is adequate for the time series, which means that this model form can be used in forecasting.

$$\Delta Y_{aIM} = -0.33 * Y_{aIM}_{t-1} - 0.84 * \varepsilon_{t-1}$$

Here: GRP - the volume of gross regional product of Surkhandarya region. Forecasting the value of GRP at constant prices (in 2012 prices). The forecast results are shown in Figure 4.

Прогнозы; Модель:(1,0,1)(0,1,0) Сезонный лаг: 4 (Таблица данных1)				
Исход.:YAIM				
Начало исходных: 1 Конец исходн.: 38				
Набл. N	Прогноз	Нижний 95,0000%	Верхний 95,0000%	Ст.ошиб.
39	5888,57	3884,826	7892,32	983,707
40	10637,27	8513,590	12760,95	1042,587
41	9401,97	7164,775	11639,17	1098,315
42	10153,37	7808,148	12498,59	1151,349
43	7386,24	3804,389	10968,09	1758,453
44	12134,94	8286,596	15983,28	1889,283
45	10899,64	6802,098	14997,18	2011,622
46	11651,04	7318,611	15983,46	2126,936
47	8883,91	3369,911	14397,90	2707,009
48	13632,61	7728,425	19536,79	2898,564
49	12397,31	6127,174	18667,44	3078,221
50	13148,70	6532,834	19764,57	3247,956

Figure 4. Forecast table and schedule

ANN model

In the development of the model using the initial data for the analysis, the autoregression section for the time series of the neural network was used. When we set the training to 80% and the control to 20%, the latent neurons of the MLP network were assumed to have a minimum value of 2, a maximum value of 8, and a number of iterations of 100.

Table 3

Top 5 Model Architecture Table

N	Архитектура	Производительность обуч.	Контр. производительность.	Ошибка обучения	Контрольная ошибка	Алгоритм обучения	Функция ошибки	Ф-я актив. скрытых нейр.	Ф-я актив. выходных нейр.
1	MLP 4-6-1	0,977392	0,996533	52485,71	36690,08	BFGS 98	Сум. квадр.	Логистическая	Тождественная
2	MLP 4-6-1	0,972248	0,996910	78688,65	43521,87	BFGS 79	Сум. квадр.	Логистическая	Тождественная
3	MLP 4-7-1	0,970641	0,998841	85667,30	6629,36	BFGS 63	Сум. квадр.	Логистическая	Тождественная
4	MLP 4-8-1	0,977224	0,997976	53561,15	11953,09	BFGS 96	Сум. квадр.	Логистическая	Тождественная
5	MLP 4-2-1	0,969353	0,996621	91986,82	28288,50	BFGS 54	Сум. квадр.	Гиперболическая	Тождественная

Activation functions for latent neurons: exactly the same (coronary), sigmodal and hyperbolic, the same (coronary) option was selected for the output neurons. The program selects the best 5 out of 50 network results from this model in the learning process (Table 3). When analyzing the data obtained to select the best model, each model is compared on parameters such as performance and error, network projection graph, residual distribution histogram, residual distribution [13].

For example, in model 3, the training efficiency is 0.970, the control error is 6629.36, the latent layer activation function is sigmoid, the output layer activation function is the same (crown), the MLP architecture has 4 inputs, 7 latent neurons and 1 output. Layer is present [13]. According to the ANN (4-7-1) model architecture, the weights and bias values connected from each input layer to the latent layer neuron node and from the latent layer to the output layer are shown in the following figure (Figure 5).

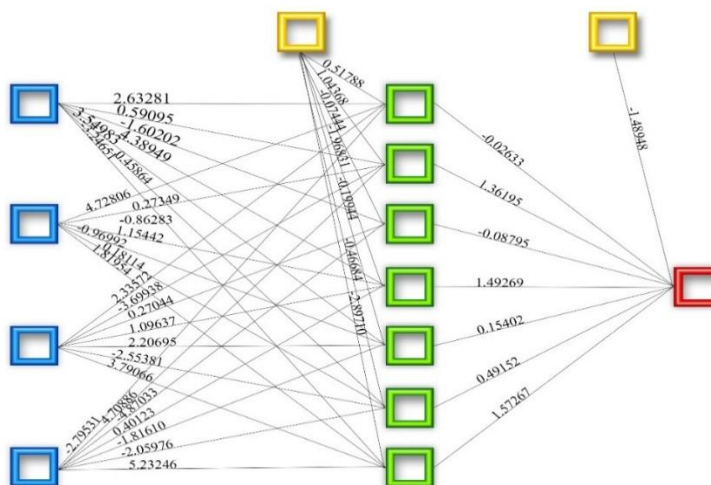


Figure 5. The architecture of the ANN (4-7-1) model is the associated weight and bias values of neuron nodes

It is also possible to add that. MLP (4-7-1) model mathematical view, hidden layer activation function sigmoid

$$\delta(z) = \frac{1}{1+e^{-z}}$$

$$z_1 = \delta(2.63281 \cdot x_1 + 4.72806 \cdot x_2 + 2.33572 \cdot x_3 - 2.79531 \cdot x_4 + 0.51788)$$

$$z_2 = \delta(0.59095 \cdot x_1 + 0.27349 \cdot x_2 - 3.69938 \cdot x_3 + 4.70886 \cdot x_4 + 1.04368)$$

$$z_3 = \delta(-1.60202 \cdot x_1 - 0.86283 \cdot x_2 + 0.27044 \cdot x_3 - 4.87033 \cdot x_4 - 0.07444)$$

$$z_4 = \delta(4.38949 \cdot x_1 + 1.15442 \cdot x_2 + 1.09637 \cdot x_3 + 0.40123 \cdot x_4 - 1.96831)$$

$$z_5 = \delta(3.54983 \cdot x_1 - 0.96992 \cdot x_2 + 2.20695 \cdot x_3 - 1.8161 \cdot x_4 - 0.19944)$$

$$z_6 = \delta(0.45864 \cdot x_1 + 0.18114 \cdot x_2 - 2.55381 \cdot x_3 - 2.05976 \cdot x_4 - 0.46684)$$

$$z_7 = \delta(-3.24651 \cdot x_1 + 1.81954 \cdot x_2 + 3.79066 \cdot x_3 + 5.23246 \cdot x_4 - 2.8971)$$

Activation function in the output layer Identical

$$\vartheta(x) = x \quad (7)$$

$$\hat{y} = (-0.02633 \cdot z_1 + 1.36195 \cdot z_2 - 0.08795 \cdot z_3 + 1.49269 \cdot z_4 + 0.15402 \cdot z_5 + 0.49152 \cdot z_6 + 1.57267 \cdot z_7 - 1.48948)$$

That indicates this model form can be used in forecasting.

The distribution histogram of the residues of each model is then compared. Figure 6 shows a distribution graph of 1 mesh residue.

A good distribution index of the model is an approximate view of the normal distribution of residues. As can be seen from Figure 6, this model has a normal distribution. Compared to all five nets, the third network is of higher quality.

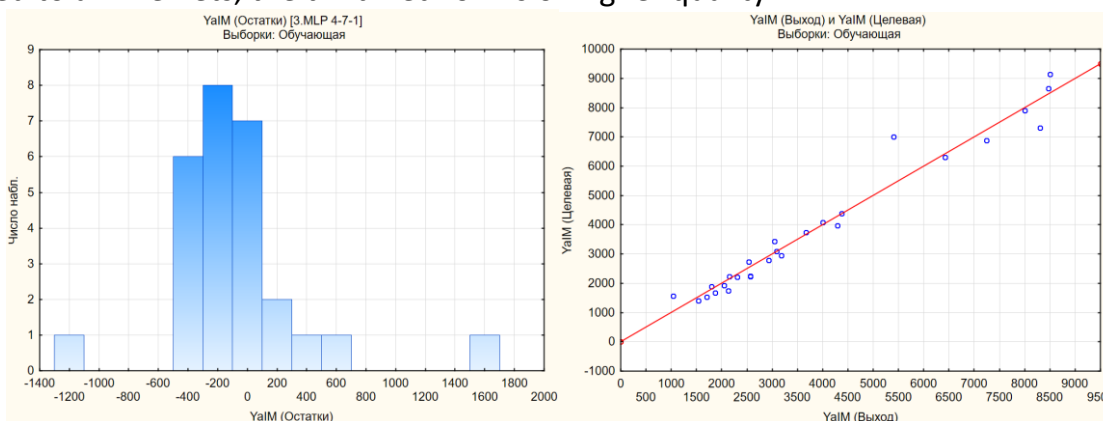


Fig. 6 Distribution histogram diagram of MLP 4-7-1 network residues

Figure 6 shows the remnants of the MLP 4-7-1 network scattered around a straight line. When all the dots are in a straight line, they give information about good model quality.

By all parameters, the MLP 4-7-1 model can be said to be the best for the next forecast. For this model, the determination coefficient is $R^2 = 0.97$, MAPE - 7.93% error is a very acceptable result.

Comparison of ARIMA and ANN forecast indicators:

The comparison of the ARIMA and ANN models was evaluated by calculating the errors between the actual figure and the forecasts obtained using these models. Numerous studies have reported that ANN models perform better in forecasting than ARIMA models, and the above idea was confirmed in this paper (Table 4).

Comparison table between the best ARIMA and ANN model

Table 4

No	Type of model	MAPE
1	Arima (1, 0, 1)(0, 1, 0)4	11,42
2	ANN (4-7-1)	7,93

Alternatively, if we look at the results of the ARIMA and ANN models, we can see the following view (Table 5).

Table 5

Table Arima and ANN medels

period	Fact	ANN	ARIMA	period	Fact	ANN	ARIMA
2019 1-quarter	3092,100	3084,97	3005,71	2022 1-quarter	-	5126,51	5888,57
2019 2-quarter	7009,200	5407,92	4734,77	2022 2-quarter	-	10628,54	10637,27
2019 3-quarter	5940,600	5875,34	4464,74	2022 3-quarter	-	9544,70	9401,97
2019 4-quarter	6307,400	6421,63	8628,38	2022 4-quarter	-	10247,11	10153,37
2020 1-quarter	4078,000	4002,53	3999,06	2023 1-quarter	-	6096,69	7386,24
2020 2-quarter	7303,300	8308,37	7915,64	2023 2-quarter	-	12196,35	12134,94
2020 3-quarter	5741,300	5795,10	6760,37	2023 3-quarter	-	10650,71	10899,64
2020 4-quarter	6881,000	7246,62	6990,82	2023 4-quarter	-	11434,76	11651,04
2021 1-quarter	4390,900	4379,05	4739,94	2024 1-quarter	-	6890,53	8883,91
2021 2-quarter	9139,600	8499,08	7914,21	2024 2-quarter	-	13080,80	13632,61
2021 3-quarter	7904,300	8006,78	6497,97	2024 3-quarter	-	11087,56	12397,31
2021 4-quarter	8655,700	8475,79	7804,31	2024 4-quarter	-	11832,21	13148,70

As can be seen from the table above, the growth trend in GRP over 2022-2024 can be seen. In general, according to the forecast results of the ANN (4-7-1) model, the forecast result is that by 2024 the volume of GRP will reach 42891.09 billion. soums, and according to ARIMA (1, 0, 1) (0, 1, 0) 4 model, 48062.52 bln. soums. This can be seen as a 1.43% increase in the 2024 ANN model and a 1.60% increase in the ARIMA model compared to 2021.

Conclusions and suggestions

This study compared the forecasting effectiveness of the ARIMA and ANN models. The most suitable model for prognosis according to ANN was MLP (4-7-1) output neuron, and the best model for ARIMA was ARIMA (1, 0, 1) (0, 1, 0) 4. According to the MLP (4-7-1) model, by 2024, the volume of GRP will reach 42891.09 billion. soums, and 48062.52 bln. soums on ARIMA (1, 0, 1) (0, 1, 0) 4 models. soums.

In other words, the region's GRP will increase by 1.43% in the ANN model and 1.60% in the ARIMA model from 2024 to 2021. This, in turn, corresponds to the task of increasing the regional economy by 1.4-1.6 times through the balanced development of the regions identified in Goal 33 of the Development Strategy of New Uzbekistan for 2022-2026 [15].

Thus, the study concluded that the ANN model is more suitable for predicting GVA volume.

References

1. Балдин К.В. Эконометрика: Учебное пособие / Балдин К.В., Быстров
2. Gosasang, V., Chan., W. and KIATTISIN, S. 2011. A Comparison of Traditional and Neural Networks Forecasting Techniques for Container Throughput at Bangkok Port. The Asian Journal of Shipping and Logistics, Vol. 27, N° 3, pp. 463-482.
3. J. Yao and C.L. Tan, (2000) “A case study on using neural networks to perform technical forecasting of forex,” Neurocomputing, vol.34, pp. 79-98,
4. S. Yaser and A. Atiya, (1996) “Introduction to Financial Forecasting”, Applied Intelligence, vol. 6, pp 205-213.
5. Mitrea, C.A., Lee, C. K. M., WuZ. 2009. A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study”. International Journal of Engineering Business Management, Vol. 1, No. 2, p 1924.
6. NowrouzKohzadi, Milton S. Boyd, BahmanKermanshahi, and lebelingKaastra(1996), “A comparison of Artificial neural network and time series models for forecasting commodity prices”, Neurocomputing,pp. 10(2):169 181.
7. Боровиков В.П. Нейронные сети Statistica Neural Networks Методология и технологии современного анализа данных: “горячая линия –Телеком”, 2008 г. С. 392.
8. Елисеева И.И. Эконометрика: учебник / Под ред. Елисеевой И.И. - М. : Проспект, 2010. - 345с.
9. Григорьева Д.Р., Файзуллина А.Г. Анализ и прогнозирование экономического показателя объем продаж предприятия в системе Statistica (англ.яз) / Международный научный конгресс “Фундаментальные и прикладные научные исследования в странах Тихоокеанского и Атлантического бассейнов”.- Токио: Tokyo University Press, 2014.- №2 - С.284287.
10. Е.И.Кулинич. Эконометрия / Е. И. Кулинич. — М.: Финансы и статистика, 2010. — 304 с
11. О.А.Заяц (2009). Прогнозирование объемов производства молока на основе сезонной ARIMA-модели. – М.: Фундаментальные исследования, №6 - 61-66 с.
12. Namazov G.Sh. “Analysis of approaches for modeling and forecasting the socio-economic development of the region”. Journal of innovations in economy. 2021. Vol. 4, Issue 6. pp. 12-17 <http://dx.doi.org/10.26739/2181-9491-2021-6-2>
13. Хотамов О.Қ., Намазов Г.Ш. “Прогноз муаммоларини хал қилишда нейрон тўр моделларини қўллаш” Иқтисодиёт ва таълим журнали №5. 2020. – 4-12.
14. Отажанов У.А. Замоनावий иқтисодий масалаларни ҳал қилишда сунъий интеллектни қўллаш//“Ўзбекистонда инновацион менежмент стратегияларини амалга ошириш тажрибаси” мавзусидаги республика илмий-амалий конференцияси илмий мақола ва тезислар тўплами (2018 йил 9 октябрь). – Т.: ТДИУ, 2018.– 275-277 бб
15. O‘zbekiston Respublikasi Prezidentining 2022 yil 28 yanvardagi PF–60-sonli Farmoni bilan tasdiqlangan “2022–2026 yillarga mo‘ljallangan Yangi O‘zbekistonning taraqqiyot strategiyasi”
16. ru.wikipedia.org (2021) - свободная энциклопедия. - <https://u.to/JkNiGw>
17. www.surxonstat.uz (2021) - Surxondaryo viloyati Statistika boshqatmasi sayti ma’lumotlari.