FORECASTING THE VOLUME OF INVESTMENTS IN INDUSTRIAL ENTERPRISES OF SURKHANDARYA REGION USING THE ARIMA MODEL

Hojiqulova Feruza Dona qizi
Отдел экономики
Teacher
Termez State University
orcid: 0009-0000-0120-4011

Abstract

This article presents an analysis and forecasts of the volume of investments in industrial enterprises of Surkhandarya region, the number of industrial enterprises and the volume of products obtained from industrial enterprises for 2010-2023. The article is devoted to studying the growth dynamics and development directions of the industrial sector of the region, highlighting the importance of industry in the economic growth of the region. The article determines the prospects for growth and development in the industrial sector of Surkhandarya region, and also offers recommendations for the effective implementation of investment policy, modernization of industry and expansion of production.

Key words: ARIMA, investment volume, industrial products, Economic Analysis, investment flows, industrial development, investment trends, industrial investments, investment forecast.

INTRODUCTION

Surkhandarya region is located in the southwestern part of Uzbekistan, its industrial potential is important for the country's economy. In recent years, special attention has been paid to the development of industry, opening new enterprises and attracting investments in the region. Investments in industrial enterprises can lead to many positive changes, especially in ensuring the economic growth of the territory, the creation of new jobs, the introduction of modern technologies and Environmental Protection. It is important in forecasting the size of investments, Economic Planning and strategic decision-making. One effective method of statistical modeling in this process is the ARIMA (AutoRegressive Integrated Moving Average) model. The ARIMA model is widely used in data forecasting through time series analysis and is known as an effective tool in the study of economic systems, including the industrial sector. This study examines the application of the ARIMA model in forecasting the volume of investments in industrial enterprises of Surkhandarya region. Forecasts make it possible to determine the factors affecting the development of the industrial sector in the region and anticipate future investment flows. This, in turn, contributes to the formation of territorial economic policies and the effective implementation of investment strategies. This study aims to analyze the capabilities of the ARIMA model and its practical importance in forecasting industrial investments.

Literature review

The impact of foreign investment on the development of the economy of the region, the study of econometric models of attracting foreign investment in the development of the economy of the region Khatamov N. made by [1]. Scientific proposal and practical recommendations for making economic-statistical analysis and forecast of competitiveness of enterprises in the housing and construction boozer Ibragimov Q. from development by [2].Multi-option forecast parameters of the targeted development of the regional industry until 2030 due to the correlation-regression correlation between privatization of industry transformation processes in the region, investments in fixed capital and demographic factors Murodkhojayeva F. developed by[3]. In contrast to the above, statistical analysis and predictive state of industrial enterprises are given.

Research Methodology

This study used techniques such as economic, statistical and structural analysis. The article also analyzes the articles of several economists as a basis. Analysis and discussion of the results this article gives an economic-statistical analysis of active industrial enterprises in Surkhandarya region. It is known that statistical observation is analyzed in two forms and is based on statistical reporting and specially organized [4] and levels stability, vibration coefficient, trend equation, additional growth, stability index [5] s.

Analysis and discussion of results

A number of econometrics have been used to forecast the indicators of the volume of investment in industrial enterprises(billion soums)to 2024-2029. Typical time series modeling uses multi-factor trend models, distributed lag models, autoregression models, ARIMA(Autoregressive integrated moving average) models, and other models. It is known that from these modeling methods, the ARIMA model is distinguished by a high accuracy of forecast values. For this reason, we used Arima models in our scientific research work. As far as ARIMA models are concerned, Jenkins and Boks proposed to distinguish this model separately from a group of non-stationary series, which together took sequential differences and proved that it could be brought into the Arima-type stationary form.Also, the boxing and Jenkins methodology consists of three parts, which are: Stage 1, determination Stage 2, evaluation Stage 3, forecasting.

The above processes are stages of ARIMA forecasting. The predictive ARIMA (Autoregressive integrated moving average) i.e.,the "integral autoregression to the sliding mean" model has (p,d, q) order, where P indicates the autoregression parameter, d indicates the integration part, q indicates the sliding mean parameter. The general outline of the ARIMA model would be as follows:

$$\Delta^d y_t = c + \sum_{i=1}^{p} a_i \Delta^d y_{t-1} + \sum_{j=1}^{q} a_i e_{t-j} + e_t \quad [1]$$

or,

$$y_t = \alpha + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t + \mu_1 \varepsilon_{t-1} + \ldots + \mu_q \varepsilon_{t-q} \quad [3]$$

Where, $\varepsilon_t -¿$ is the stationary time series,, $c, \alpha_i, \beta_i -¿$ is the model parameter, $\Delta^d - d$ s the time series subtraction of the order.

We used the data provided by the Surkhandarya Region Department of Investment, Industry and trade to model the indicators of investment volume included in the industrial enterprises of Surkhandarya region (table 1). During the study, the volume of investment in industrial enterprises of Surkhandarya region (Bln. sum) we forecast using ARIMA models.

table 1

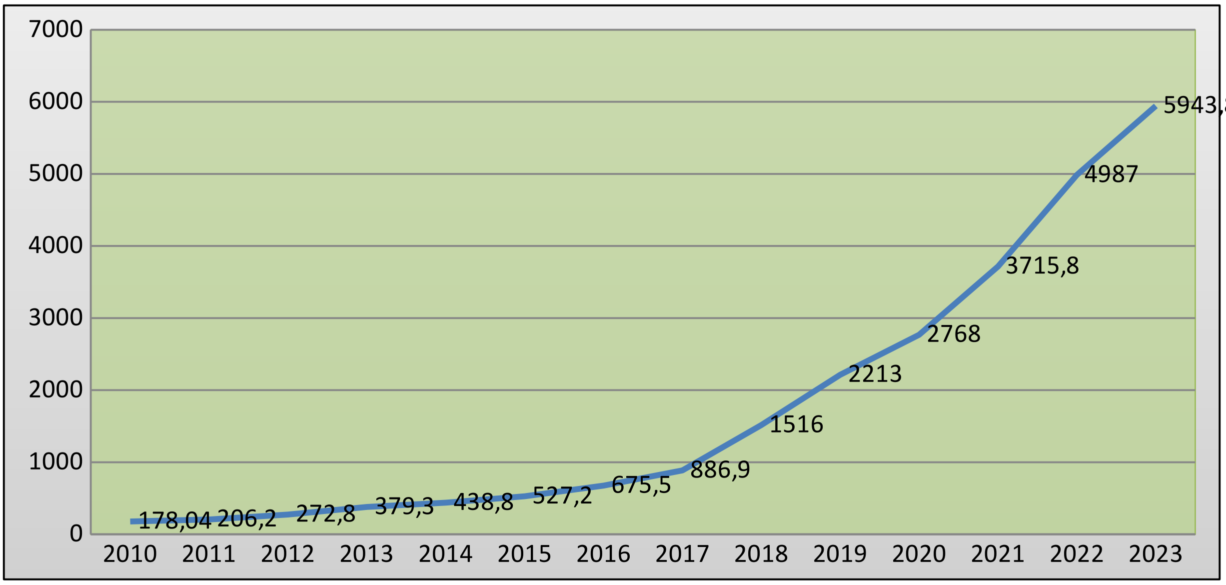
The volume of investment in industrial enterprises of Surkhandarya region(Bln.sum)[1]

Years	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Specification	178	206,2	272,8	379,3	438,8	527,2	675,5	886,9	1516	2213	2768	3715,8	4987	5943,8

A gretl 2023c interpretation of Gretl's software was used for modeling. Typically, ARMA models are used to model stationary time series. In the ARIMA model, however, it also has the potential to model non-stationary time series. It is known that the stationary or constant represents the nature of the process of not changing its properties over time. Initially, the indicators in Table 1 are the volume of investment in industrial enterprises of Surkhandarya region (billion. sum)when modeling, we check the time series stationary. To do this, we look at

[1] Information of surkhandarya Region investment, industrial and Trade Department.

the drawing of the timed row (Figure 1).



3.3.1-rasm. The volume of investment in industrial enterprises of Surkhandarya region(Bln.sum)[2]

According to the visual analysis of Figure 1, we can see that the time series is not stationary. Therefore, we check the constancy of its first differences. To do this, we will use the ADF-extended Dickey-Fuller test, which is considered one of the most common methods of checking for stability(Table 2). The Dickey-Fuller test (ADF test, Dickey - Fuller test) is a method used in Applied Statistics and econometrics to perform a stationary test to analyze timed rows , representing one of these single root tests (the unit Root Test). It was proposed by David Dickey and Wayne Fuller in 1979 .

2-jadval

Results of the extended Dickey-Fuller test [3]

[2] Author development.
[3] Author development

```
Augmented Dickey-Fuller test for d_n
testing down from 4 lags, criterion AIC
sample size 9
unit-root null hypothesis: a = 1

  test without constant
  including 3 lags of (1-L)d_n
  model: (1-L)y = (a-1)*y(-1) + ... + e
  estimated value of (a - 1): 0.0860391
  test statistic: tau_nc(1) = 0.26011
  asymptotic p-value 0.7616
  1st-order autocorrelation coeff. for e: 0.079
  lagged differences: F(3, 5) = 3.432 [0.1088]

  test with constant
  including 3 lags of (1-L)d_n
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0060559
  test statistic: tau_c(1) = -0.0187554
  asymptotic p-value 0.9558
  1st-order autocorrelation coeff. for e: -0.057
  lagged differences: F(3, 4) = 2.907 [0.1646]
```

The first difference in the time series according to the results of the unchanged test according to Table 2 is also not stationary. For this reason, we will check the second difference of the ARIMA model(Table 3).

Table 3.

Results of the extended Dickey-Fuller test [4]

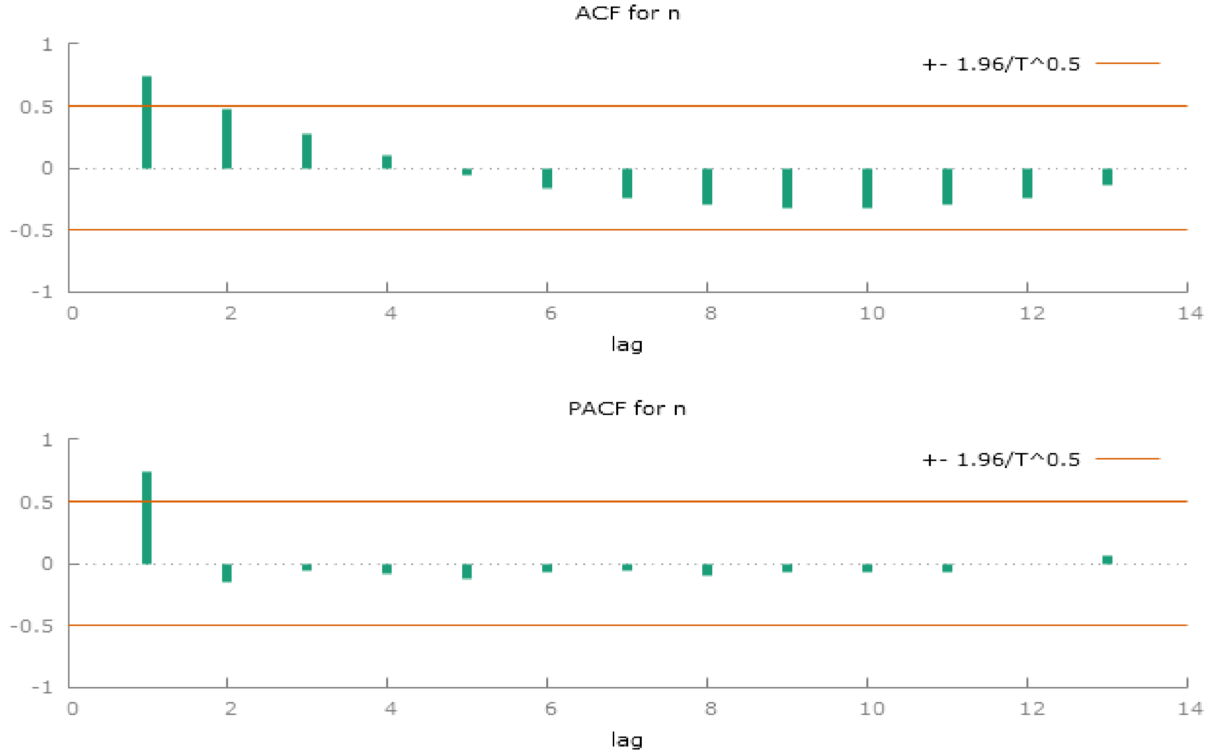
```
Augmented Dickey-Fuller test for d_d_n
testing down from 4 lags, criterion AIC
sample size 9
unit-root null hypothesis: a = 1

  test without constant
  including 2 lags of (1-L)d_d_n
  model: (1-L)y = (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0673956
  test statistic: tau_nc(1) = -0.124097
  asymptotic p-value 0.641
  1st-order autocorrelation coeff. for e: 0.020
  lagged differences: F(2, 6) = 6.172 [0.0350]

  test with constant
  including one lag of (1-L)d_d_n
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -2.1232
  test statistic: tau_c(1) = -4.42483
  asymptotic p-value 0.0002631
  1st-order autocorrelation coeff. for e: 0.307
```

According to Table 3, the invariant test results have a p-value of 0.641 and the invariant test results have a p-value of 0.0002631. These values are smaller than the desired significance levels, i.e. α=0.1,α=0.05,α=0.01. So the second difference in the time series is stationary. This means that the order d of the ARIMA(p,d,q)model is 2. So far, our model looks like this:

$$ARIMA(p,2,q) \qquad [3]$$

[4] Author development

At the next stage, we determine the remaining P and q orders of the ARIMA model. To do this, we will examine the ACF and PACF correlograms. It is shown in Figure 2.



2-rasm. Temporal series correlogram

Figure 2 shows that ACF is decreasing exponentially, and in turn a discontinuity after the first lag is observed in PACF. This was thus found to be an autoregression specific case, and the order of p and q was determined to be p=1 and q=0. Therefore, we experiment with the ARIMA(1,2,0) model order (Table 4).

Table 4

Regression analysis results [5]

[5] Author development

Model 14: ARIMA, using observations 2012-2023 (T = 12)
Dependent variable: (1-L)^2 n
Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
const	94.9568	17.7216	5.358	<0.0001	***
phi_1	0.351331	0.315030	1.115	0.2648	
theta_1	−0.999999	0.260017	−3.846	0.0001	***

Mean dependent var	77.38667	S.D. dependent var	212.9434
Mean of innovations	−25.68829	S.D. of innovations	166.0018
R-squared	0.992769	Adjusted R-squared	0.992046
Log-likelihood	−79.32677	Akaike criterion	166.6535
Schwarz criterion	168.5932	Hannan-Quinn	165.9354

		Real	*Imaginary*	*Modulus*	*Frequency*
AR					
	Root 1	2.8463	0.0000	2.8463	0.0000
MA					
	Root 1	1.0000	0.0000	1.0000	0.0000

According to table 3.3.4, we will use the capabilities of the Gretl program due to the fact that the results were not obtained from the procedures of the ARIMA model found on the correllogram. In Gretl, we experience the ARIMA(2,1,3)model layout from 'ARIMA lag selection'(table 5).

Table 5

ARIMA lag tanlash natijalari [6]

```
Estimated using AS 197 (exact ML)
Dependent variable Surxondaryoviloyatisanoatkor, T = 12
Criteria for ARIMA(p, 2, q) specifications
-------------------------------------------------------------
 p, q          AIC           BIC           HQC        loglik
-------------------------------------------------------------
 0, 0     165.6750      166.6448      165.3160      -80.8375
 0, 1     165.8342      167.2889      165.2956      -79.9171
 0, 2           NA            NA            NA            NA
 0, 3           NA            NA            NA            NA
 1, 0     167.2421      168.6968      166.7035      -80.6210
 1, 1     166.6535      168.5932      165.9354      -79.3268
 1, 2     164.9957      167.4203      164.0981      -77.4979
 1, 3     166.4573      169.3667      165.3801      -77.2286
 2, 0     163.4107      165.3504      162.6926      -77.7054
 2, 1     162.8761*     165.3007*     161.9785*     -76.4381
 2, 2     164.7169      167.6263      163.6397      -76.3584
 2, 3     165.8986      169.2929      164.6419      -75.9493
 3, 0     163.2576      165.6821      162.3599      -76.6288
 3, 1     164.7202      167.6296      163.6430      -76.3601
 3, 2     166.3134      169.7078      165.0567      -76.1567
 3, 3     168.0967      171.9759      166.6604      -76.0483

'*' indicates best, per criterion
'NA' indicates that a specification could not be estimated
Log-likelihood ('loglik') is provided for reference
```

The order p=2 and q=1 according to Table 5 are significant by the Akank criterion. For this reason we check the ARIMA(2,2,1)values for the experiment based on the above case(table6).

Table 6

Regression analysis results [7]

[6] Author development
[7] Author development

Model 18: ARIMA, using observations 2012-2023 (T = 12)
Dependent variable: (1-L)^2 n
Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
const	87.5862	28.5755	3.065	0.0022	***
phi_1	−0.491809	0.245565	−2.003	0.0452	**
phi_2	−0.755171	0.179305	−4.212	<0.0001	***
theta_1	0.636258	0.264590	2.405	0.0162	**

Mean dependent var	77.38667	S.D. dependent var	212.9434
Mean of innovations	0.145362	S.D. of innovations	128.5208
R-squared	0.995313	Adjusted R-squared	0.994271
Log-likelihood	−76.43807	Akaike criterion	162.8761
Schwarz criterion	165.3007	Hannan-Quinn	161.9785

		Real	*Imaginary*	*Modulus*	*Frequency*
AR					
	Root 1	-0.3256	-1.1037	1.1507	-0.2957
	Root 2	-0.3256	1.1037	1.1507	0.2957
MA					
	Root 1	-1.5717	0.0000	1.5717	0.5000

From table 6 we can see that all parameters of the model have statistical significance, that is, the coefficients φ_1, φ_2, and 〖theta〗_1 have statistical significance. Also, according to the table, all parameters of the model are of statistical significance. Also the model's approximation error is MARE = 10.107%.(Figure3).

```
Model estimation range: 2012 - 2023
Standard error of the regression = 128.521

       Surxondaryoviloy~         fitted      residual

2012                272.8         303.0         -30.2
2013                379.3         413.3         -34.0
2014                438.8         601.4        -162.6
2015                527.2         592.9         -65.7
2016                675.5         791.7        -116.2
2017                886.9         896.4          -9.5
2018               1516.0        1213.5         302.5
2019               2213.0        2280.7         -67.7
2020               2768.0        2714.9          53.1
2021               3715.8        3572.1         143.7
2022               4987.0        4865.9         121.1
2023               5943.8        6076.4        -132.6

   Forecast evaluation statistics using 12 observations

   Mean Error                                  0.14536
   Root Mean Squared Error                     128.52
   Mean Absolute Error                         103.24
   Mean Percentage Error                      -5.4135
   Mean Absolute Percentage Error              10.107
   Theil's U2                                  0.52243
```

Figure 3. Model approximation error . Figure 3 shows that the model's approximation error fully satisfied the required 15% threshold . Further experiments found that there was no autocorrelation in the remains. Also the model residues obey the law of normal distribution. Figure 4 examined the autocorrelation of the remains and observed no autocorrelation state in the

remains (Figure 4).

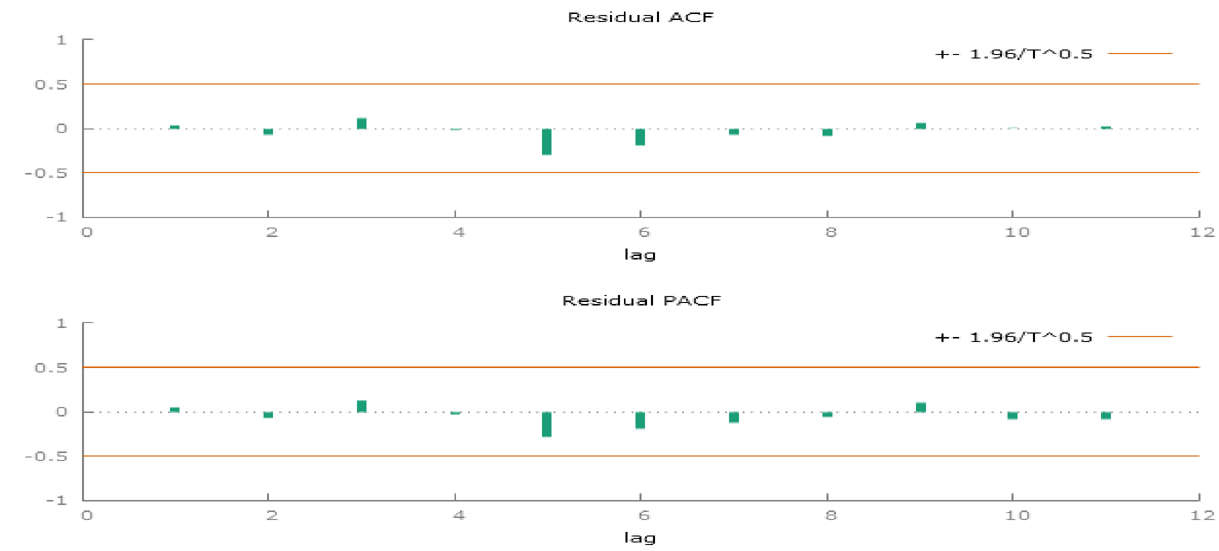
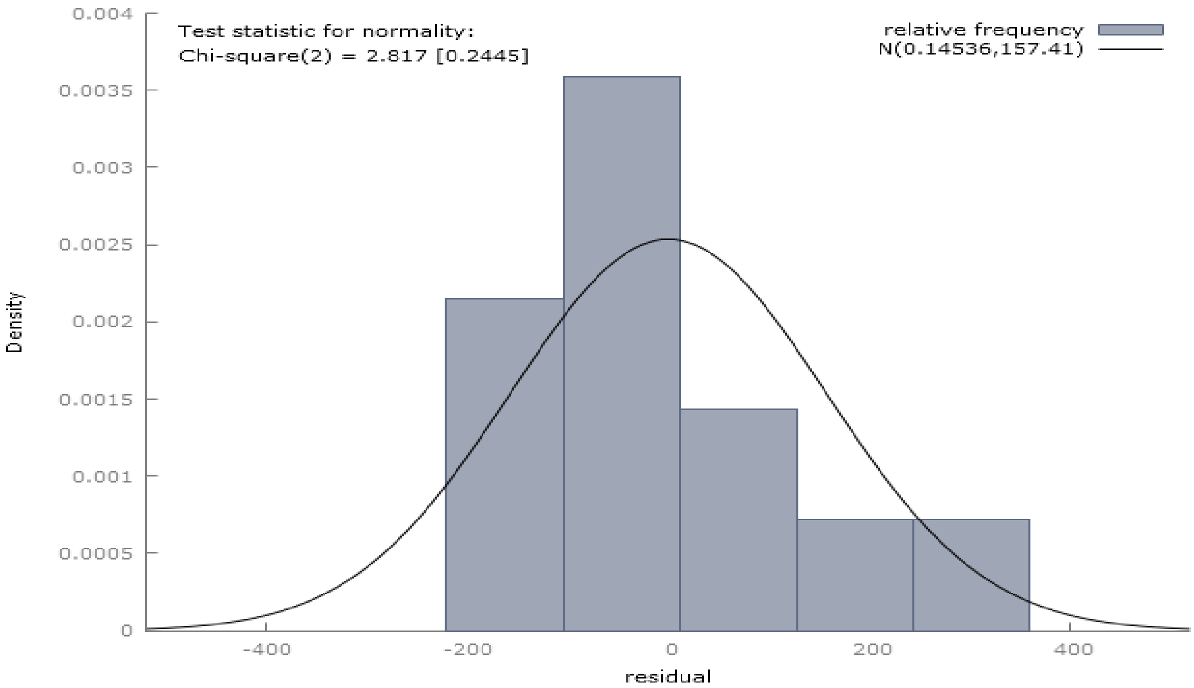


Figure 4. Correlagram of residues

Based on Figure 4, the correlagrammation of residues is subject to the criterion requirements. That is, the lags that pierce the system were not detected. Next we take advantage of Gretl's capabilities to examine the law of normal distribution(Figure5).



5. Residue normality In experiments according to figure 5, the remnants of the model obey the normal tasimot law. That is, Chi-square(2) = 2.817 [0.2445]. This value is greater than 0.1. Next we will place the normal feed indicators (table 7).

Table 7

Indicators of the law of Normal distribution [8]

Frequency distribution for residual, obs 3-14
number of bins = 5, mean = 0.145362, sd = 157.405

interval	midpt	frequency	rel.	cum.	
< -104.49	-162.62	3	25.00%	25.00%	*********
-104.49 - 11.782	-46.353	5	41.67%	66.67%	**************
11.782 - 128.05	69.917	2	16.67%	83.33%	******
128.05 - 244.32	186.19	1	8.33%	91.67%	***
>= 244.32	302.46	1	8.33%	100.00%	***

Test for null hypothesis of normal distribution:
Chi-square(2) = 2.817 with p-value 0.24448

From Table 7 we can also see that Chi-square(2) = 2.817 [0.2445]. That this value is greater than 0.1. Thus the general outline of the model was as follows:

ARIMA(2,2,1) [4];

or, $\Delta^2 y_t = ¿87.586 - 0.492\Delta^2 y_{t-1} - 0.755\Delta^2 y_{t-2} + 0.636\varepsilon_{t-1}$ [5¿].

The value of this (4) modeling AKAIKE criterion is 162.8761. It is known that the best model is chosen in which the value of the Akaike criterion is the smallest. Thus, the optimal module was found to be the Arima(2,2,1) model. Also, using (4)and(5), the volume of investment in the industrial sector of Surkhandarya region (mlrd.so ' m) we can predict. We can see that this process has the following appearance if we see the predictive state of the ARIMA model in gretl capabilities(Figure 6).

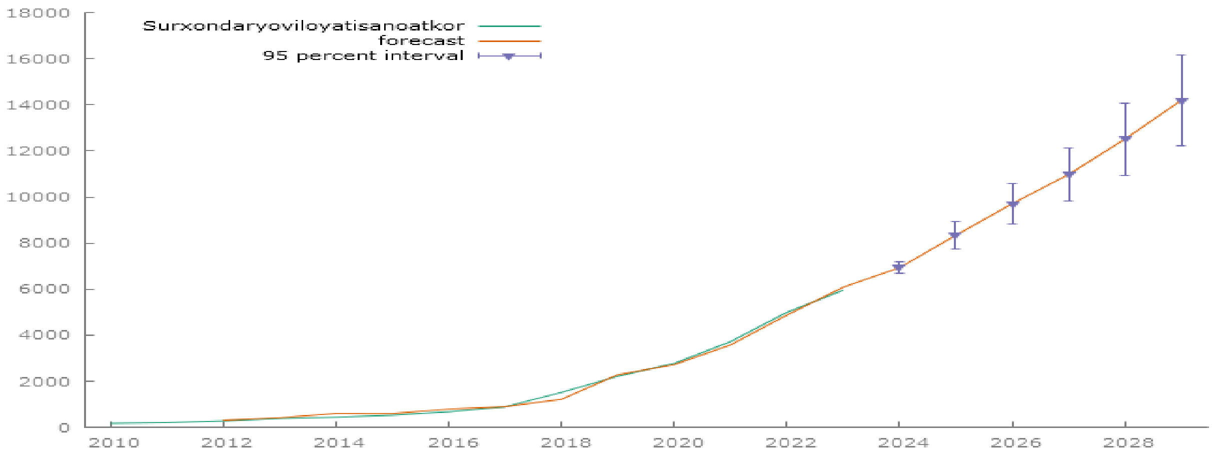


Figure 7. The volume of investment in industrial enterprises of Surkhandarya region(Bln.su m) of predictive value[9]

[8] Author development
[9] Author development

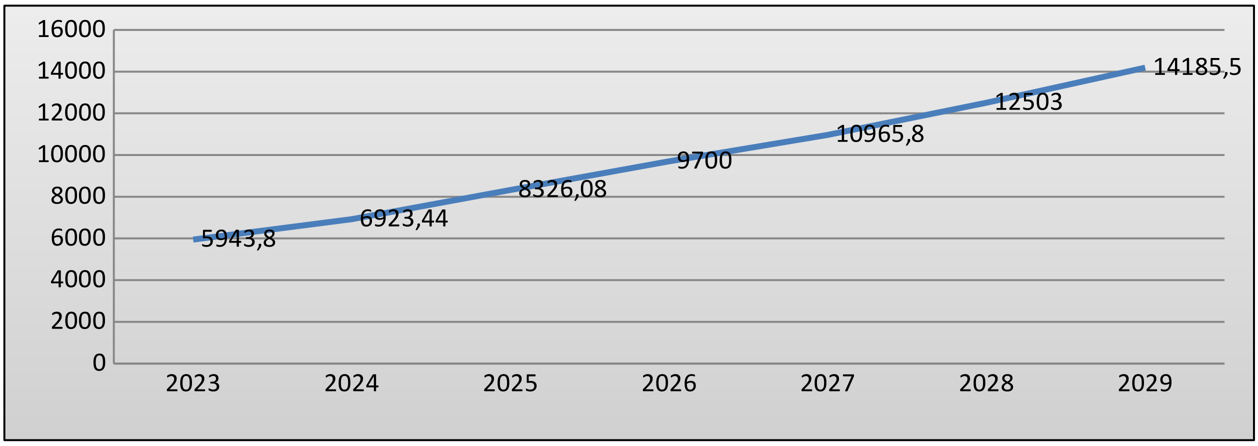


Figure 7. The volume of investment in industrial enterprises of Surkhandarya region(mlrd.so 'm)' s forecast value for 2024-2029 [10]

As can be seen from Figure 6 and Figure 7, the volume of investment in the industrial sector of Surkhandarya region(Bln.sum) it has been found that there is a tendency to grow and that a state of growth can be observed when the factors affecting this state are not affected. If we determine the general view of the indicators of the forecast and confidence intervals of our model, we will make the following view(Table 8)

Table 8

Indicators of forecast and confidence intervals [11]

Years	Indicator real values	Indicator theoretical values	Standard error	95% confidence interval lower limit	95% confidence interval upper limit
2010	178.040				
2011	206.200				
2012	272.800	303.006			
2013	379.300	413.287			
2014	438.800	601.423			
2015	527.200	592.896			
2016	675.500	791.732			
2017	886.900	896.403			
2018	1516.00	1213.54			
2019	2213.00	2280.75			
2020	2768.00	2714.93			
2021	3715.80	3572.14			
2022	4987.00	4865.86			
2023	5943.80	6076.39			
2024		6923.44	128.521	6671.55	7175.34
2025		8326.08	304.100	7730.06	8922.11
2026		9700.24	438.922	8839.97	10560.5

[10] Author development
[11] Author development

2027		10965.8	590.898	9807.64	12123.9
2028		12503.0	797.858	10939.3	14066.8
2029		14185.5	1008.94	12208.0	16163.0

According to Table 8, the volume of investment in industrial enterprises of Surkhandarya region (mlrd.so ' m) 2010-2023-year data was given, and the state of stationary was not observed in investments in the annual industry network collected using statistical observations. Using the Dickey-Füller test, it was found that there is a stationary presence when the second order integrative is I (2).As the most optimal model for the ARIMA model, we calculated that ARIMA is(2,2,1). The ARIMA(2, 2, 1) model has determined the value indicators for the volume of investment in industrial enterprises of Surkhandarya region up to 2029. The volume of investment in industrial enterprises in 2023 is 5943,80 Bln.sum, this figure is 14,185. 5 as of 2029 Bln.sum. The volume of investment in industrial enterprises over the past 6 years according to the forecast results is 8241.7 Bln.sum M will increase, and in 2025-2029 this figure is expected to be 115.5%. The highest increase in the cross-sectional of years will occur in 2029, the volume of investment in industrial enterprises will be 14185.5 billion soums. Descriptive statistics conducted between the volume of investment in industrial enterprises and the forecast of the volume of investment in industrial enterprises calculated using the ARIMA model showed positive results. As a result of the formation of the base of investment volume invested in industrial enterprises, their correct formation, as well as the identification of factors affecting these processes, it will be possible to achieve forecast indicators. We forecast the volume of investment in industrial enterprises of Surkhandarya region based on the data from 2010-2023 above, as well as the value of products produced by industrial enterprises of Surkhandarya region through the data provided by the Department of industry, investment and trade of Surkhandarya region using the gretl program through the ARIMA model, based on the data provided (Bln.sum), and the number of industrial enterprises was projected in the above order, and the models produced the following outline(Table 9).

Table 9

General formula views of ARIMA models [12]

Name	Model formula
The volume of investment in the industrial sector of Surkhandarya region(billion sum).	$\Delta^2 y_t = ¿87.586 - 0.492\Delta^2 y_{t-1} - 0.755\Delta^2 y_{t-2} + 0.636\varepsilon_{t-1}$ yoki, ARIMA(2,2,1)
The value of the product produced in the industrial network of Surkhandarya region(billion sum).	$\Delta^2 y_t = ¿50.444 - 0.919\Delta^2 y_{t-1} - 0.738\Delta^2 y_{t-2} - 0.999\,\varepsilon_{t-1}$ yoki, ARIMA(2,2,1)

[12] Author development

The number of industries established in Surkhandarya region, in the cross section of years (PCs)	$\Delta^2 y_t = ¿65.020 + 0.894\Delta^2 y_{t-1} - 0.596\Delta^2 y_{t-2} + 1.0006\varepsilon_{t-1}$ yoki, ARIMA(2,1,1)

On the basis of the data available in Table 9, formulas were generated that were specific to the criteria requirements in the gretl program. Also, the overall forecast results of ARIMA models are expressed as follows(table 10).

Table 10

General forecast results of ARIMA models [13]

Years	The volume of investment in the industrial sector of Surkhandarya region(Bln.sum).	The value of the product produced in the industrial network of Surkhandarya region(Bln.sum).	The number of industries established in Surkhandarya region, in the cross section of years (PCs)
2024	6923.44	4978.18	603
2025	8326.08	5142.43	918
2026	9700.24	6233.20	1125
2027	10965.8	7106.13	1167
2028	12503.0	7628.84	1128
2029	14185.5	8768.64	1112

Conclusion: The volume of investment in the industrial sector of Surkhandarya region in recent years(billion sum), the value of the product produced by industrial enterprises(billion sum) and the number of industrial enterprises (PCs) 5943, respectively 80mlrd.so 'm, 4137,30 billion soums. the investment in the industrial sector is 14,185 in 2029, if it is m and 400 5 billion soums., the value of the product produced by industrial enterprises is 8768,64 billion soums. and the number of industrial enterprises in 2027 is projected to reach 1167. In 2025-2029, these figures are expected to be 115.5%, 112.2% and 114.8% respectively. As we know from the table, the volume of investments in the industrial sector(billion sum), the value of products produced in the industrial sector(billion sum) and the number of enterprises being established in the industrial sector represents positive results. The products created at the enterprises created by the inclusions introduced into the industrial network will not only improve the way of living the population, but also contribute to the creation of new jobs, the improvement of population income, the growth of the country's gross domestic product.

List of used literature

1. Khatamov N. Mintaqa iqtisodiy tizimini rivojlantirishda xorijiy investitsiyalarni jalb qilishning ekonometrik modellari (Surxondaryo viloyati misolida). Urganch – 2024.

[13] Author development

2. Ibragimov Q. Uy-joy qurilish bozorida korxonalarning raqobatbardoshligining iqtisodiy statistik tahlili va uni prognozlash. Urganch – 2024.
3. Murodkhojayeva F. Sanoatda transformatsiyalash jarayonining mintaqaviy xususiyatlari (Namangan viloyati misolida). Tashkent-2022.
4. Ibragimov Q. economic-statistical analysis of active construction enterprises in Surkhandarya region. 2022.6p.
5. Тураев Б.Оценка устойчивости тенденций развития промышленности строительных материалов сурхандарьинской области 2023.8р
6. Information of the Surkhandarya Region investment, industrial and Trade Department.