

ADVANTAGES OF FORECASTING ECONOMIC PROCESSES USING THE ARIMA MODEL IN PYTHON

Normamatova Yulduz Ravshan qizi

Abstract. This thesis highlights the advantages of forecasting economic processes using the AutoRegressive Integrated Moving Average (ARIMA) model implemented in the Python programming language. The empirical part of the study uses official statistical data on wheat production in the Surkhandarya Region for 2010–2024. Using the pandas, NumPy, Matplotlib, and statsmodels libraries in Python, the stationarity of the time series was tested using the Augmented Dickey–Fuller (ADF) test, the optimal ARIMA model was selected, and forecast values for 2027–2028 were calculated. The results of the study demonstrate that Python is an open-source, fast, and effective platform for economic forecasting.

Keywords: Python, ARIMA model, econometric modeling, forecasting, time series, wheat, agriculture, statsmodels.

Introduction

In the context of the digital economy, modeling and forecasting economic processes have become important components of managerial decision-making. In particular, forecasting agricultural production volumes in advance is essential for ensuring food security, using resources rationally, and developing regional development programs.

In recent years, the use of open-source software tools has become increasingly widespread in scientific research. One such tool is the Python programming language, which integrates economic data processing, statistical analysis, econometric modeling, and forecasting capabilities within a single platform. Python libraries such as pandas, NumPy, Matplotlib, and statsmodels are widely used to analyze economic indicators.

One of the most widely used classical methods for time-series forecasting is the ARIMA model. This model makes it possible to identify the underlying patterns

of a time series based on historical observations and forecast future values. The ARIMA model is considered one of the reliable methods for short-term forecasting in economics, finance, energy, and agriculture.

The purpose of this thesis is to highlight the advantages of forecasting economic processes using the ARIMA model in Python and to present empirical results using wheat production in the Surkhandarya Region as an example.

Research Methodology. Official statistical data on wheat production in the Surkhandarya Region for the period 2010–2024 were used for the study. The data were imported into Python using the pandas library, followed by an initial descriptive statistical analysis.

The Augmented Dickey–Fuller (ADF) test was applied to determine the stationarity of the time series. After obtaining a stationary series, ACF and PACF plots were used to identify the parameters of the ARIMA model. The model was evaluated using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Log-Likelihood criteria. Based on the optimal model, forecast values for 2027–2028 were calculated.

The general form of the ARIMA model is expressed as follows:

$$Y_t = c + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_s \varepsilon_{t-s} + \varepsilon_t$$

where Y_t denotes the current value of the time series, φ denotes the autoregressive coefficients, θ denotes the moving-average coefficients, and ε_t denotes the random error term.

Research Results and Their Analysis. In this study, official statistical data on wheat production in the Surkhandarya Region for the period 2010–2024 were analyzed using the Python programming environment. Initially, the data were

imported from an Excel file using the pandas library, and a time series was constructed. Subsequently, the dynamics of the time series were analyzed graphically.

According to the analysis, certain fluctuations in wheat production were observed during 2010–2024. In particular, production declined sharply in 2018, while a stable upward trend was recorded in 2021–2023. Although production decreased slightly in 2024, the overall trend remained at a high level.

Table 1

Wheat Production Volume in the Surkhandarya Region (2010–2024)¹

Year	Wheat production volume (thousand tons)
2010	595,2
2011	576,1
2012	594,8
2013	600,6
2014	606,4
2015	607,9
2016	604,2
2017	577,5
2018	476,8
2019	554,2

¹ O'zbekiston Respublikasi Prezidenti huzuridagi Statistika agentligi ma'lumotlari asosida muallif ishlanmasi.



2020	598,7
2021	631,5
2022	645,4
2023	659,2
2024	653,6

In the next stage, the stationarity of the time series was tested using the Augmented Dickey–Fuller (ADF) test. The ADF test results for the original series indicated that the series was non-stationary. Therefore, first-order differencing was performed, and the differenced series was tested again. The results confirmed that the series had become stationary, thereby satisfying the necessary condition for constructing the ARIMA model.

Several ARIMA models were estimated in order to select the optimal model. The models were compared based on the Akaike Information Criterion (AIC).

Table 2

Comparison of ARIMA Model Results ²

Model	AIC
ARIMA(1,1,0)	146,189
ARIMA(0,1,1)	146,188
ARIMA(1,1,1)	148,173
ARIMA(2,1,0)	147,980

² Python dasturida muallifning hisob-kitoblari



ARIMA(2,1,1)	148,465
--------------	---------

As shown in Table 2, the lowest AIC value is 146.188, corresponding to the ARIMA(0,1,1) model. Therefore, this model was selected as the optimal model for forecasting.

Based on the selected model, forecast values for 2025–2028 were calculated using Python's forecast() function.

```
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA
```

```
data = pd.read_excel("bugdoy.xlsx")
```

```
series = data["Bugdoy"]
```

```
model = ARIMA(series, order=(0,1,1))
result = model.fit()
```

```
forecast = result.forecast(steps=4)
```

```
print(forecast)
```

The following forecast values were obtained from the calculations.

Table 3

Forecast Values of Wheat Production in the Surkhandarya Region³

³ Python dasturida muallifning hisob-kitoblari



Year	Forecast (thousand tons)
2025	653,52
2026	653,52
2027	653,52
2028	653,52

The forecast results indicate that no substantial changes in wheat production are expected in the coming years. Taking into account the dynamics of the existing time series, the ARIMA(0,1,1) model forecasts that production will stabilize at around 653 thousand tons. These results have practical significance for regional agricultural planning, resource allocation, and management decision-making aimed at ensuring food security.

Conclusion. This study analyzed the possibilities of forecasting economic processes using the ARIMA model in the Python programming language and conducted an empirical study based on wheat production in the Surkhandarya Region during 2010–2024. Official statistical data from the Statistics Agency under the President of the Republic of Uzbekistan were used for the study.

The time series was initially processed using Python's pandas library, and its stationarity was tested using the Augmented Dickey–Fuller (ADF) test. After obtaining a stationary time series, several ARIMA models were estimated, and the optimal model was selected based on the Akaike Information Criterion (AIC). The calculations showed that the ARIMA(0,1,1) model was the most appropriate.

Based on the selected model, forecast values for 2025–2028 were calculated. The results indicate that wheat production is expected to stabilize at approximately 653.52 thousand tons in the coming years. This result suggests that wheat production

in the Surkhandarya Region is not expected to undergo substantial changes in the near future.

The following advantages of Python in econometric modeling were identified during the study:

1. being an open-source and free software platform;
2. rapid processing of large volumes of statistical data;
3. the ability to estimate ARIMA models simply and reliably using the statsmodels library;
4. convenient graphical visualization of forecasting results;
5. ensuring the reproducibility of scientific research;
6. the ability to rapidly update the model when new economic data are added.

Thus, the Python programming language was confirmed to be a modern, convenient, and effective tool for modeling and forecasting economic processes. The ARIMA model, in turn, can be widely applied in empirical research in economics and agriculture because it provides reliable results for short-term forecasting.

References

1. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. *Time Series Analysis: Forecasting and Control*. 5th Edition. John Wiley & Sons, 2015.
2. Hyndman R. J., Athanasopoulos G. *Forecasting: Principles and Practice*. 3rd Edition. OTexts, 2021.
3. Hamilton J. D. *Time Series Analysis*. Princeton University Press, 1994.
4. McKinney W. *Python for Data Analysis*. 3rd Edition. O'Reilly Media, 2022.
5. Seabold S., Perktold J. *Statsmodels: Econometric and Statistical Modeling with Python*. Proceedings of the 9th Python in Science Conference, 2010.
6. Harris C. R., Millman K. J., van der Walt S. J. et al. *Array Programming with NumPy*. Nature, 2020.



7. Hunter J. D. *Matplotlib: A 2D Graphics Environment*. Computing in Science & Engineering, 2007.
8. O‘zbekiston Respublikasi Prezidenti huzuridagi Statistika agentligi. *Qishloq xo‘jaligi statistikasi. 2010–2024-yillar*.