

FORECASTING THE INFLATION IN THE REPUBLIC OF NORTH MACEDONIA BY USING ARIMA MODEL

ANA JOSIMOVSKA NIKOLOV¹ AND SONJA GEGOVSKA-ZAJKOVA²

Abstract. Modeling and forecasting inflation play a critical role in the economic management of all countries, particularly developing ones. This paper is focused on predicting inflation in the Republic of North Macedonia. The selected method for econometric inflation modeling is the Box-Jenkins approach, utilizing an ARIMA model. The analyzed time series focuses on the consumer price index (CPI), based on 2010 as the reference year. It encompasses 249 observations, with a monthly frequency, spanning from January 2002 to September 2022. The analysis demonstrates that the model accurately forecasts inflation for the initial two months beyond the sample period, with a mere 0.1 and 0.5 percentage point disparity between the actual and predicted inflation rates, respectively. Consequently, it can be affirmed that the Box-Jenkins methodology provides highly precise inflation forecasts for the Republic of North Macedonia, particularly within the first two months beyond the sample period of the time series.

1. INTRODUCTION

Inflation refers to the sustained increase in the general level of prices for goods and services in an economy over a period of time. The inflation rate is an important economic indicator that affects the purchasing power of citizens, the value of savings and investments, the exchange rate of the country, as well as the overall economic stability.

Controlling inflation represents one of the most significant economic challenges faced by all countries, particularly developing nations in the process of restructuring their economies. This transformation involves promoting market mechanisms, enhancing trade opportunities, facilitating technology transfers, and attracting foreign direct investments. To achieve this, monetary authorities must employ various tools and policies to monitor and control undesirable high or volatile inflation when necessary. The negative consequences of inflation on economies are well-known. Inflation can erode the purchasing power of the national currency, resulting in deteriorating social conditions and living standards within the country. Elevated prices can also lead to uncertainty and discourage both

2010 *Mathematics Subject Classification.* Primary: 62M10, Secondary: 91B84.
Key words and phrases. Time series, Box-Jenkins methodology, ARIMA model.

domestic and foreign investors from investing in the economy, negatively impacting investment and economic growth while increasing poverty. Furthermore, excessive price increases worsen a country's conditions for international trade, as domestic products become more expensive and less competitive in regional and global markets.

Over the past couple of years, inflation has risen significantly in most countries worldwide, largely due to the CoVid crisis, and the Republic of North Macedonia is no exception.

The Consumer Price Index (CPI) represents a key and frequently used indicator of the price level in an economy. It measures the overall change in consumer prices based on a representative basket of goods and services over time.

Main topic of this paper is forecasting and modeling the inflation in the Republic of North Macedonia. These predictions are made by ARIMA modeling technique, more specifically the model relies on the approach by Box-Jenkins methodology. Econometric modeling for this study was carried out using the EViews 8 software package.

The insights derived from this study will contribute to a better understanding of inflation dynamics within the domestic economy, shedding light on potential future price movements. Moreover, a deeper comprehension of the country's inflation situation and expectations for future inflation can assist monetary policymakers in formulating appropriate policies and measures to address the challenges posed by adverse economic trends.

2. METHODOLOGY AND FINDINGS

The time series that is the subject of analysis in this paper represents the Consumer Price Index, with a base year of 2010. The time series has a monthly frequency and covers the period from January 2002 to September 2022, resulting in a total of 249 observations. At the time of data acquisition and model development, the latest available data point was for September 2022, making it the last data point in the time series. However, at the time of writing the paper, the latest available data on inflation was the data for February 2023. This allows a direct assessment of the forecasting capabilities of the model. The data was obtained from the State Statistical Office. Figure 1 provides a graphical representation of the time series. The data allows us to see how the living costs in North Macedonia has evolved from the beginning of 2002 to September 2022. As expected, throughout the entire period, an upward trend is observed, indicating a continuous increase in the consumer price index. However, it generally represents a gradual and stable rising in the consumer price index, particularly until the beginning of 2022. However, in the last few months of the analyzed period, there is a noticeable sharp increase, which was not only the case in the domestic economy but also everywhere in the world. For comparison, during 2020 and 2021, the average annual inflation rate in the domestic economy was 1.2% and 3.2%, respectively, while in the period from January to September 2022, the average annual inflation rate stood at 12.4%, indicating a significant acceleration in the inflation rate during this period.

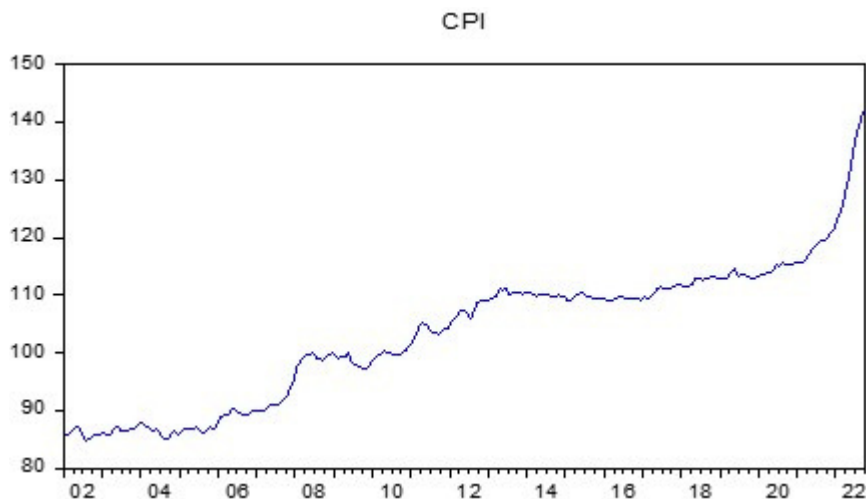


FIGURE 1. Consumer price index with base year 2010

High inflation rates in 2022 were observed almost worldwide, and the main reasons for such occurrences are attributed to economic shifts related to the CoVid-19 pandemic. Nearly all economies adopted stimulative fiscal and monetary policies during 2020 and 2021 to address the economic crisis caused by the pandemic. Another reason is the unexpectedly rapid recovery of demand in 2021, which led to supply shortages for certain products. Additionally, the conflict in Ukraine had a further impact, causing price shocks in energy and food. North Macedonia, as a relatively small and open economy, did not remain immune to such events. Therefore, during 2022, relatively high inflation rates were observed in the domestic economy. From the available data, looking solely within the framework of 2022, it can be concluded that inflation is consistently accelerating. In January, the annual inflation rate was 6.7%, while in September, it reached 18.7%. Within the scope of this paper, the goal is to conduct econometric modeling of the time series of the consumer price index. Subsequently, based on the model, predictions of the index in the future, or predictions of the inflation rate in the upcoming period in the domestic economy, will be generated. Considering that the latest available inflation data is from February 2023, a comparison can be made between the obtained predicted inflation rates for the period from October 2022 to February 2023 and the actual inflation values. This will serve as a check of the model's forecasting performance.

3. IDENTIFICATION AND SELECTION OF THE MODEL

The graphical representation of the time series (Figure 1) indicates that it is a non-stationary time series. A non-stationary time series exhibit trends, seasonal

patterns, or other structural changes that affect their statistical characteristics. Using non-stationary time series data in financial models generates unreliable and misleading outcomes, resulting in inadequate comprehension and forecasting. The series of the Consumer Price Index clearly indicates that there is an upward trend throughout the analyzed period. In order to confirm the non-stationarity of the series, an appropriate correlogram has been obtained. The null hypothesis being tested states that the series has a unit root and is non-stationary, as opposed to the alternative hypothesis which asserts that the series is stationary.

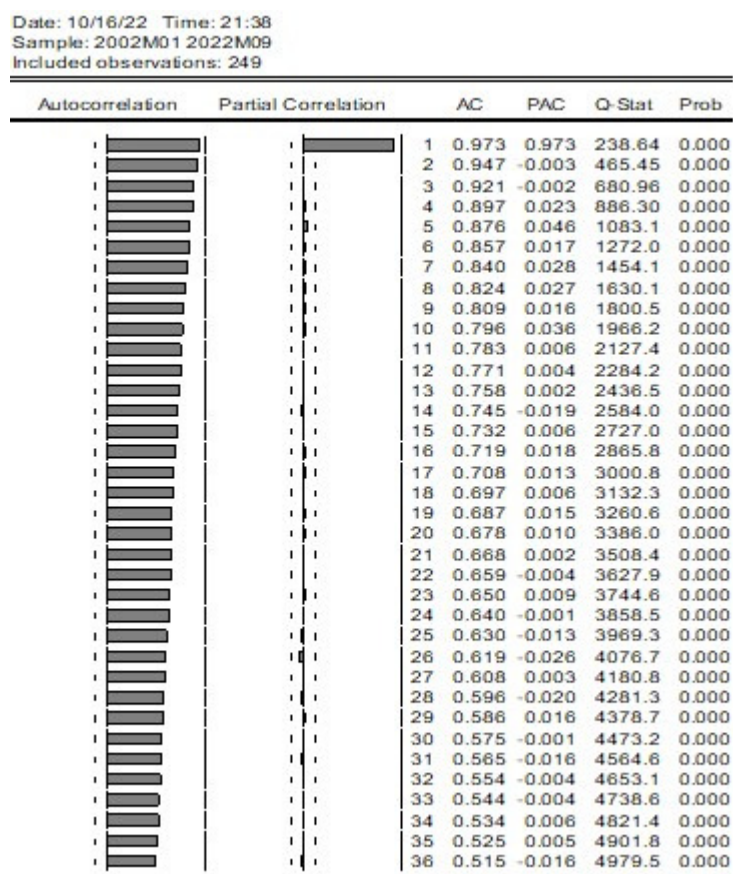


FIGURE 2. Correlogram of the time series consumer price index

Figure 2 displays the correlogram of the time series of the Consumer Price Index, depicting autocorrelation and partial autocorrelation functions with 36 lags. The null hypothesis being tested states that the correlation between observations at a certain lag is equal to zero, while the alternative hypothesis states that the correlation is different from zero. As one can observe from the correlogram in

Figure 2, the autocorrelation function is characterized by statistically significant coefficients at all 36 lags, whereas the partial autocorrelation function is characterized by a statistically significant coefficient only at the first lag. In other words, the correlation between observations at time points t and $t - 1$ is so strong that, when looking at the autocorrelation function, it extends through the other periods. Therefore, in the autocorrelation function, the coefficients for all 36 lags are statistically significant. This correlogram, where the autocorrelation function shows a slow decline until the very end, is a typical example of a non-stationary time series. However, to confirm the non-stationarity of the time series with certainty, it's best to conduct the Augmented Dickey-Fuller (ADF) unit root test ([5]).

Null Hypothesis: CPI has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 2 (Automatic - based on SIC, maxlag=15)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.255732	0.9983
Test critical values:		
1% level	-3.995800	
5% level	-3.428198	
10% level	-3.137485	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(CPI)
 Method: Least Squares
 Date: 10/16/22 Time: 21:44
 Sample (adjusted): 2002M04 2022M09
 Included observations: 246 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CPI(-1)	0.003364	0.013153	0.255732	0.7984
D(CPI(-1))	0.384491	0.065800	5.843347	0.0000
D(CPI(-2))	0.149883	0.067898	2.207481	0.0282
C	-0.311700	1.101809	-0.282899	0.7775
@TREND("2002M01")	0.000591	0.002088	0.283086	0.7774

R-squared	0.258649	Mean dependent var	0.225854
Adjusted R-squared	0.246344	S.D. dependent var	0.692915
S.E. of regression	0.601543	Akaike info criterion	1.841477
Sum squared resid	87.20673	Schwarz criterion	1.912724
Log likelihood	-221.5017	Hannan-Quinn criter.	1.870165
F-statistic	21.02055	Durbin-Watson stat	2.011964
Prob(F-statistic)	0.000000		

FIGURE 3. ADF unit root test with constant and time trend on the time series consumer price index

In Figure 3, the results of the ADF unit root test with a constant and trend are observed. In a similar way, an ADF unit root test with a time series constant on the time series consumer price index has been conducted, but without a trend, as well as an ADF test for a unit root without a constant and without a trend on the time series consumer price index. The results of the ADF unit root test with three different specifications indicate that the series of the Consumer Price Index has a unit root and is non-stationary. Specifically, the t-statistics in the three different

specifications are characterized by p-values that are very high (above 0.99, i.e., greater than the three levels of statistical significance of 0.01, 0.05, and 0.1). This suggests that we do not reject the null hypothesis, which states that the series has a unit root and is non-stationary. Considering that the time series is non-stationary, the next step is to transform the series into a stationary one, and for this purpose the finite differences ($y_t - y_{t-1}$) have been calculated. The graphical representation of the differenced series is provided in Figure 4.

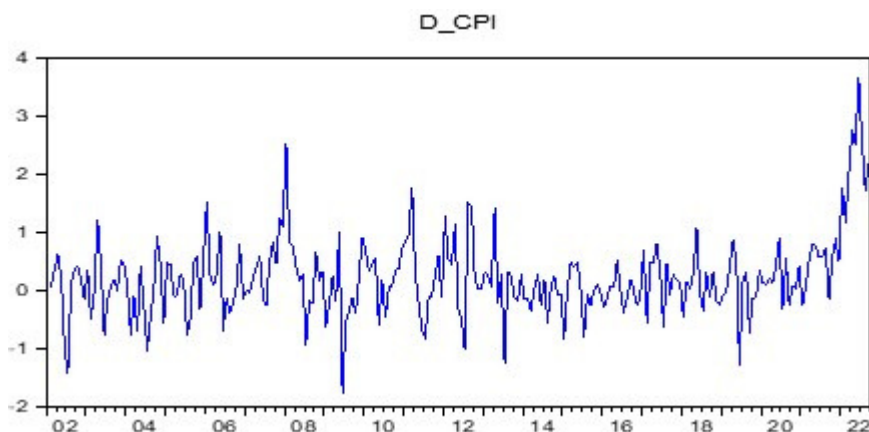


FIGURE 4. Finite differences of the consumer price index with base year 2010

As one can observe from the graphical representation in Figure 4, the differenced series is characterized by values oscillating around a mean value, resembling a stationary series. Towards the end of the series, slightly higher values are noticed, which deviate from the rest of the series.

The correlogram of the differenced series, as shown in Figure 5, is characterized by different autocorrelation and partial correlation functions compared to the original series. The autocorrelation function exhibits a much faster decline compared to the original series, which could indicate that the differenced series is stationary.

Figure 6 displays the ADF unit root test with a constant and trend on the differenced time series of the consumer price index. Furthermore, an ADF unit root test with a time series constant on the time series consumer price index has been conducted, but without a trend, as well as an ADF unit root test without a constant and without a trend on the time series consumer price index. The ADF unit root test conducted on the differenced series (Figure 6) suggests rejecting the null hypothesis, which states that the series has a unit root and is non-stationary (the p-value for the t-statistic in all three different specifications is 0, i.e., lower than the three levels of statistical significance of 0.01, 0.05, and 0.1). This means that the differenced series does not have a unit root and is stationary, characterized by

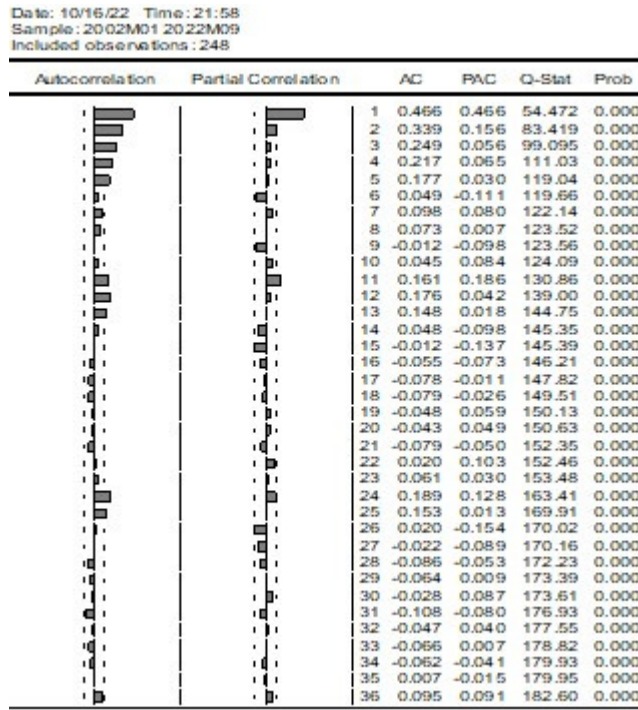


FIGURE 5. Correlogram of the differenced time series consumer price index

a constant mean and variance. Due to this, modeling will be conducted on the differenced series. Given that differencing the original time series once achieved stationarity, we can say that the series is integrated of the first order ([6]). According to the correlogram in Figure 5, statistically significant coefficients in the autocorrelation function are found at lags 1, 2, 3, 4, 5, 11, 12, 13, 24, and 25, while in the partial autocorrelation function, statistically significant coefficients are observed at lags 1, 2, 11, 15, 24, and 26. Based on this, it is necessary to create various combinations of AR and MA processes, resulting in different ARIMA models. Among all these models, the one with the best performance needs to be selected. Typically, the most suitable model is chosen based on criteria such as all coefficients being statistically significant, having the highest adjusted coefficient of determination (Adjusted R-squared), and having the lowest values for information criteria (Akaike, Schwarz, Hannan-Quinn). In this research, based on the statistically significant coefficients in the autocorrelation and partial autocorrelation functions of the differenced series, several different ARIMA models have been constructed. To maintain the focus on the main research question, this paper will not present all possible combinations of AR and MA processes. Instead, only a few ARIMA models will be presented, and one of them will be selected

Null Hypothesis: D(CPI) has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 1 (Automatic - based on SIC, maxlag=15)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.714450	0.0000
Test critical values:		
1% level	-3.995800	
5% level	-3.428198	
10% level	-3.137485	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(CPI,2)
 Method: Least Squares
 Date: 10/16/22 Time: 21:55
 Sample (adjusted): 2002M04 2022M09
 Included observations: 246 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(CPI(-1))	-0.456429	0.067977	-6.714450	0.0000
D(CPI(-1),2)	-0.155259	0.064438	-2.409439	0.0167
C	-0.030637	0.077738	-0.394104	0.6939
@TREND("2002M01")	0.001106	0.000553	1.997981	0.0468

R-squared	0.283199	Mean dependent var	0.007398
Adjusted R-squared	0.274313	S.D. dependent var	0.704777
S.E. of regression	0.600380	Akaike info criterion	1.833619
Sum squared resid	87.23039	Schwarz criterion	1.890616
Log likelihood	-221.5351	Hannan-Quinn criter.	1.856569
F-statistic	31.87030	Durbin-Watson stat	2.013234
Prob(F-statistic)	0.000000		

FIGURE 6. ADF unit root test with constant and trend on the differenced time series consumer price index

as the best. During the modeling process, other possible combinations of AR and MA processes were considered for additional validation, but they will not be presented in the paper.

4. ESTIMATION OF THE MODEL PARAMETERS

According to the statistically significant coefficients of the autocorrelation function and partial autocorrelation function in the correlogram of the differenced time series, five ARIMA models have been constructed, and they are as follows:

1. $ARIMA(1, 1, 1) : y_t = \varphi_0 + \varphi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t,$
2. $ARIMA(1, 1, 2) : y_t = \varphi_0 + \varphi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t,$
3. $ARIMA(1, 1, 3) : y_t = \varphi_0 + \varphi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \epsilon_t,$
4. $ARIMA(1, 1, 3) : y_t = \varphi_0 + \varphi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_3 \epsilon_{t-3} + \theta_4 \epsilon_{t-4} + \epsilon_t,$
5. $ARIMA(11, 1, 1) : y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_{11} y_{t-11} + \theta_1 \epsilon_{t-1} + \epsilon_t,$

Below is the result of conducting the regression ARIMA (11, 1, 1) (see Figure 7). This analysis has been applied to the other four models as well. The null

hypothesis is individually set for each of the parameters included in the model and states that the coefficient is equal to 0 and is statistically insignificant ($\beta = 0$). The alternative hypothesis states that the coefficient is different from 0 ($\beta \neq 0$) and is statistically significant.

Dependent Variable: D(CPI)
 Method: ARMA Conditional Least Squares (Marquardt - EVIEWS legacy)
 Date: 10/16/22 Time: 22:43
 Sample (adjusted): 2003M01 2022M09
 Included observations: 237 after adjustments
 Convergence achieved after 13 iterations
 MA Backcast 2002M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.641771	2.170256	-0.295712	0.7677
AR(1)	0.813596	0.066597	12.21667	0.0000
AR(11)	0.216564	0.055509	3.901424	0.0001
MA(1)	-0.474168	0.099981	-4.742579	0.0000
R-squared	0.312610	Mean dependent var		0.236624
Adjusted R-squared	0.303759	S.D. dependent var		0.692306
S.E. of regression	0.577667	Akaike info criterion		1.757097
Sum squared resid	77.75201	Schwarz criterion		1.815629
Log likelihood	-204.2159	Hannan-Quinn criter.		1.780689
F-statistic	35.32108	Durbin-Watson stat		2.026515
Prob(F-statistic)	0.000000			
Inverted AR Roots	1.01	.83-.43i	.83+.43i	.43-.76i
	.43+.76i	-.06+.84i	-.06-.84i	-.51-.65i
	-.51+.65i	-.78-.24i	-.78+.24i	
Inverted MA Roots		Estimated AR process is nonstationary		
	.47			

FIGURE 7. Results from the ARIMA (11, 1, 1) model

The regression results indicate that in the ARIMA (1, 1, 3) model, the coefficients θ_2 and θ_3 are statistically insignificant at all three levels of statistical significance. Similarly, in the ARIMA (1, 1, 4) model, the coefficients θ_2 , θ_3 , and θ_4 are statistically insignificant (the p-values for the respective t-statistics are greater than the three levels of statistical significance, 0.01, 0.05, and 0.1). Therefore, these two models can be eliminated. Three models remain from which the most appropriate model needs to be chosen, namely ARIMA (1, 1, 1), ARIMA (1, 1, 2), and ARIMA (11, 1, 1). In these three models, all coefficients included in the models are statistically significant, except for the constant, which is statistically insignificant in the ARIMA (1, 1, 2) and ARIMA (11, 1, 1) models, but we will disregard this. Moreover, in all three models, the coefficients are statistically significant at all three levels of statistical significance, except for the coefficient θ_2 in the ARIMA (1, 1, 2) model, which is statistically significant only at a level of 0.1 (the p-value for the corresponding t-statistic is greater than 0.05 but less than 0.1). From these three models, the best one would be the one with the highest adjusted coefficient

of determination and the lowest information criteria (Akaike, Schwarz, Hannan-Quinn). Among the ARIMA (1, 1, 1), ARIMA (1, 1, 2), and ARIMA (11, 1, 1) models, the ARIMA (11, 1, 1) model has the highest adjusted coefficient of determination and simultaneously has the lowest information criteria compared to the other two models. Therefore, it can be concluded that the best model to choose is the ARIMA (11, 1, 1) model, whose mathematical representation is:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_{11} y_{t-11} + \theta_1 \epsilon_{t-1} + \epsilon_t.$$

This means that the value of the differenced time series of the consumer price index in period t depends on its value in the previous period ($t - 1$), its value from 11 periods ago ($t - 11$), as well as the error from one period ago ($t - 1$).

5. TESTING THE IDENTIFIED MODEL

In the final stage of the Box-Jenkins methodology, a model verification is performed ([8]).

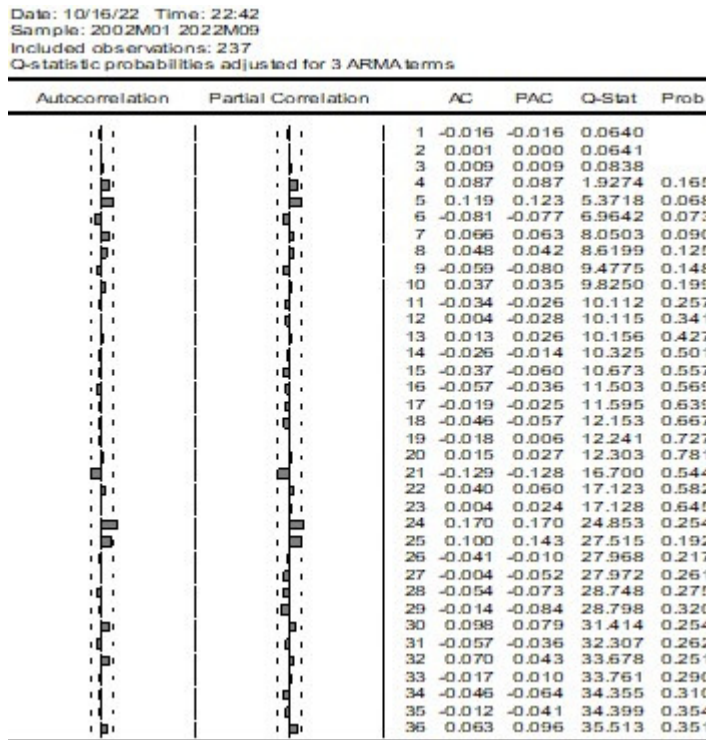


FIGURE 8. Correlogram of the residuals of the ARIMA (11, 1, 1) model

The residuals in the identified ARIMA model should follow a white noise process. To verify whether the residuals of the model follow a white noise process, it is necessary to analyze their correlogram.

In our case, the p-value for the Ljung-Box Q-statistic lags up to 36 periods is 0.351, which is greater than the three levels of statistical significance (see Figure 8). Therefore, in this case, the null hypothesis that there is no autocorrelation among the residuals up to 36 lags is accepted, meaning that the residuals of the model up to 36 lags follow a white noise process and are not serially correlated. Thus, it can be said that the residuals in the ARIMA (11, 1, 1) model follow a white noise process, confirming that the model is good, and predictions can be made using it.

6. PREDICTING INFLATION USING THE MODEL

In this section the forecasts of the consumer price index based on the ARIMA (11, 1, 1) model are presented, as shown in Figure 9. Since the latest available data was for September 2022, the forecasting is done for the next 12 months, i.e., for the period from October 2022 to September 2023. Based on the obtained forecasted values for the consumer price index, annual inflation rates are derived for each month in the period from October 2022 to September 2023.

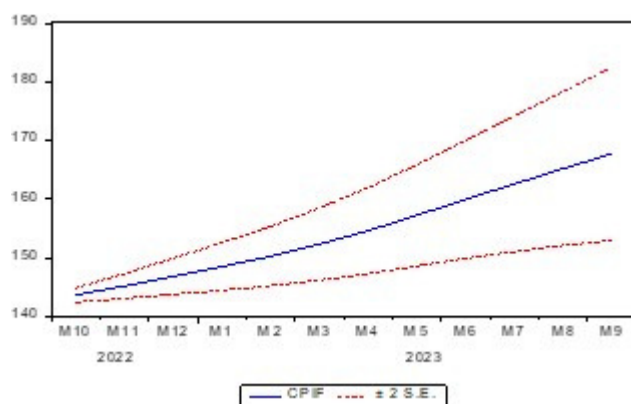


FIGURE 9. Forecast of the Cost of Living Index using the ARIMA (11, 1, 1) model

A direct comparison can be made between the realized values and the forecasted values of inflation for the period from October 2022 to February 2023, which is shown in Figure 10.

The results of forecasting the consumer price index shown in a Table 1 suggest continued growth in the index from October 2022 to September 2023.

According to the model's forecast, the annual inflation rate in October 2022 would be 19.6%, while the realized inflation rate for that month was 19.8%. Therefore, the difference between the two values is extremely small, just 0.1 percentage points. For the month of November, the model predicted an inflation rate

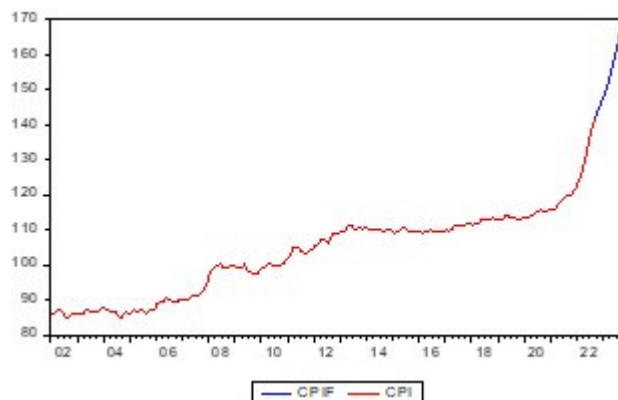


FIGURE 10. Realized and forecasted values of the consumer price index with the ARIMA (11, 1, 1) model

of 20%, while the realized inflation rate was 19.5%, representing a small deviation of only 0.5 percentage points. For the remaining three months, the model forecasts inflation rates of 20.9%, 20.5%, and 20.7%, while the realized inflation rates were 18.7%, 17.1%, and 16.7%, respectively. This means that the difference between the two series, in absolute terms, was 2.2 percentage points, 3.4 percentage points, and 4.1 percentage points, respectively. The results indicate that the model accurately predicted inflation in October and November 2022 with exceptional precision. However, in each subsequent month, its accuracy slightly declined.

Period	Realized inflation rate (%)	Predicted inflation rate (%)	Difference
10.2022	19.8	19.6	0.1
11.2022	19.5	20.0	-0.5
12.2022	18.7	20.9	-2.2
1.2023	17.1	20.5	-3.4
2.2023	16.7	20.7	-4.1
3.2023		20.5	
4.2023		19.6	
5.2023		19.3	
6.2023		18.1	
7.2023		17.8	
8.2023		18.2	
9.2023		18.2	

TABLE 1. Comparison of realized and forecasted inflation rates

7. CONCLUSION

As suggested by the theory of forecasting with ARIMA models, the chosen model for inflation in the Republic of North Macedonia demonstrates exceptionally accurate predictions, especially for the first two months, with discrepancies of 0.1 percentage points and 0.5 percentage points, respectively ([7]). The discrepancies between the actual inflation and the model's predictions increase over the next three months, reaching 2.2 percentage points, 3.4 percentage points, and 4.1 percentage points, respectively. From the results of the empirical research, it can be concluded that the Box-Jenkins methodology provides exceptionally good results in predicting inflation in North Macedonia in the first three months outside the sample range for the time series. This is in line with previous findings from the application of the same methodology for predicting inflation in other countries.

REFERENCES

- [1] F. Alvarez, R. E. Lucas Jr., W. E. Weber, *Interest rates and inflation*, American Economic Review, 91(2), (2001), pp.219-225.
- [2] A. A. Ariyo, A. O. Adewumi, C. K. Ayo, *Stock price prediction using the ARIMA model*, in: UKSim-AMSS 16th international conference on computer modelling and simulation, IEEE, (2014), 106-112.
- [3] J. Bai, P. Perron, *Estimating and testing linear models with multiple structural changes*, Econometrica, (1998), pp.47-78.
- [4] G. E. P. Box, G. M. Jenkins, *Time series analysis: forecasting and control*, 2 nd edn. Holden Day, San Francisco, (1976).
- [5] D. C. Montgomery, G. C. Runger, *Applied statistics and probability for engineers*, John Wiley & sons, (2010).
- [6] R. F. Engle, *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation*, Econometrica: Journal of the econometric society, (1982), pp.987-1007.
- [7] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, A. Lachhab, *Forecasting of demand using ARIMA model*, International Journal of Engineering Business Management, 10(2):184797901880867, (2018).
- [8] V. Karadzic, B. Pejovic, *Inflation forecasting in the Western Balkans and EU: a comparison of Holt-Winters, ARIMA and NNAR Models*, Amfiteatru Econ, 23, (2021), pp.517-532.
- [9] M. A. Salam, S. Salam, M. Feridun, *Modeling and forecasting Pakistan's inflation by using time series ARIMA models*, Economic Analysis Working Papers, No.1, (2007).

ANA JOSIMOVSKA NIKOLOV
HALKBANK AD SKOPJE,
SS CYRIL AND METHODIUS 54, 1000 SKOPJE, NORTH MACEDONIA
Email address: ana.josimovska@gmail.com

SONJA GEGOVSKA-ZAJKOVA
SS CYRIL AND METHODIUS UNIVERSITY,
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGIES,
RUDJER BOSHKOVIK 18, 1000 SKOPJE, NORTH MACEDONIA
Email address: szajkova@feit.ukim.edu.mk

Received 21.11.2023

Revised 9.7.2024

Accepted 19.7.2024