Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/51/20230675

Comparison of ARIMA and ARIMA Error Regression Models: Evidence from the Russian Consumer Price Index

Zeyi Cai[1,a,*]

[1]St.Michael's College,University of Toronto, 81 St Mary St, Toronto,Ontario, Canada
a. zcloris.cai@mail.utoronto.ca
**corresponding author*

Abstract: Consumer Price Index (CPI) is regarded as a common approach for measuring inflation. The present study examines the Russia CPI in the context of political and social upheavals, especially under the wars and the COVID-19. Due to the unstable political situation, the inflation rate in Russia sharply grow in 2014 and 2022 which creates two of the largest increase over the past decade after the Crimean war and the Ukrainian war were announced, thus it is crutial to make both long-term and short-term trends prediction. The paper aims to choose the best fitting model to estimate the future value of Russia monthly CPI data, and ultimately provides a suggestion and reference for monetary and fiscal authorities in Russia when making the policies to reduce inflation risk. The study applies the ARIMA (Autoregressive Integrated Moving Average) model to forecast Russia CPI, and uses regression with ARIMA errors to analyze the Russia CPI according to the ARIMA fitted values of the crude oil price in Europe. According to the result, ARIMA (2,3,1) is the best fitting model that can make comparatively sensible prediction for the future values.

Keywords: CPI, crude oil, 2014 Crimean crisis, COVID-19, Russia-Ukraine war, ARIMA

1. Introduction

1.1. Research Background

The Russian Federation attacked and took over Crimea in early 2014 in response to the culmination of the Euromaidan revolution in Ukraine. On the peninsula, Russian forces captured Ukrainian military structures while acting without marks and backed by Kremlin denials. As a result of the sanctions imposed, tensions between the West and Russia increased [1]. The dramatic decline of the Russian ruble versus the USD and Euro at the end of 2014 is a result of the crisis that arose in 2014 as a result of rapid declines in crude oil prices and political upheaval [2]. In 2014, Russia's inflation rate was 11.36%. That is 4.91 percent greater than the previous year (2013) and 1.55% less than the following year (2015) [3]. Many fiscal problems have a detrimental impact on the Russian economy as a whole, and specifically on the banking industry and its capital [1].

Furthermore, the embargo on agricultural imports from the US and other Western nations, which came into effect in 2014, had an impact on Russian consumers by restricting Russia's purchase of agricultural and food items, significantly rising food costs, and decreasing consumption [4].

The worldwide economy was significantly impacted by the COVID-19 epidemic, which resulted in lockdowns and restricted economic activity. Furthermore, the epidemic caused a significant drop in global oil demand, resulting in a steep drop in oil prices, including Brent-Europe [5].

The dynamics of income and buying power of the residents were represented in a study. It stated that the consumer price index in Russia as a whole was 104.9%, from the end of 2019 to the end of 2020, and it rose sharply for this year in all federal districts [5].

New economic sanctions against Russia were imposed by the US, Europe, and many other Western nations as a result of Russia's invasion of Ukraine on February 24, 2022. The empirical data reveal that Russia's invasion of Ukraine, together with the COVID epidemic, resulted in a large spike in global food and crude oil prices. The spike in the world food price index following the invasion was mostly due to an increase in the price of dairy and oils. Following the invasion, there was an increase in inflation in nations that placed strong sanctions on Russia, as well as countries that were not participating in the conflict in any way [4].

1.2. Research Purpose and Significance

Inflation, which is considered to be the most important indicator that ensure the internal balance in the economy, is extremely important in the development and growth of Russia's economy, especially under the turbulent political situation due to military operations and the global pandemic. A common approach for measuring inflation is thought to be the CPI. Applying the CPI values is the cornerstone of the creation of long-term plans around the globe. However, given the high level of uncertainty, predicting CPI levels provides a forecasting challenge because of the various forecasting techniques which is widely used [6]. Therefore, after narrowing the range of the choice for a forecasting model, ARIMA modeling and regression with ARIMA errors will be explored in this research to estimate the future value of Russia monthly CPI data, and the comparison will propose the best fitting model.

It is obvious that the currently major purpose of the Bank of Russia's inflation targeting regime is to preserve price stability. The study also aims to provide a data analysis to help policy makers consider the options available to the central bank for implementing its monetary policy aimed at reducing inflation to sustainable low levels [7].

2. Methodology

2.1. Data Description

The secondary data utilized in this article was gathered from the Federal Reserve Economic Data(https://fred.stlouisfed.org). To determine the best-fitting model, this article employed R program and used Russia monthly CPI data for a total of ten years, from 2012 to 2022, to undertake the first stage of the time series analysis of Russia's consumer price index. To guarantee data availability and consistency, data from the same period were utilized for the second portion as a sample information for spotting patterns, formulating forecasts, and gaining more unbiased analysis for the time series analysis of the European crude oil prices linked with Russia CPI.

2.2. ARIMA Models

Among statistical learning techniques, ARIMA modelling is known as the most sophisticated approach to time series forecast. The three components of ARIMA (p, d, q) modeling are, respectively, the autoregressive (AR), the moving average (MA), and the stationarity of the time series [8].

The third part is addressed first since the time series can be simulated only when it is stationary. A time series' stationarity states that for a given amount of time, the mean and variance stay constant regardless of where the time series is located. Additionally, as the value of 'h' increases, the covariance

DOI: 10.54254/2754-1169/51/20230675

of any two terms with a 'h'-second difference in time decreases until it is zero. At the level, time series seldom ever remain stationary. They often become stationary when temporal differencing is used. The order of integration (or stationarity) is '1' if the temporal differencing is only performed once. If stationarity is seen after two differentiating, the integration is considered to be '2'. A time series is referred to as integrated of order '0' if stationarity exists within the level itself. The third term in the ARIMA model denotes the integration of time series. If the data is not converted into a stationary time series, ARIMA modelling cannot be performed [8].

The following stage is to go for the AR portion. AR is an abbreviation for autoregressive models. The future values are determined by the lagged values of the time-series. The AR model examines prior data values and makes assumptions about them. The order "p" implies that historical data from period "p" are utilized to forecast current values. The following is a pth-order AR process:

$$y_t = \alpha + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \ldots + \emptyset_p y_{t-p} + \varepsilon_t \tag{1}$$

Where y_t is the stationary response variable at time t, $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ is the dependent variable at different time lags, $\alpha = \mu\left(1 - \emptyset_1 - \emptyset_2 - \ldots - \emptyset_p\right)$, with $\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ as the parameter estimation part, and ε_t represents the error term.

The MA term is the next phase in the ARIMA modeling process. MA is an abbreviation for moving average. It is the lag of data from a random term generated by a random process (white noise process or term).

The MA (q), can be written as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{2}$$

Where q is the moving average's lag count and $\theta_1, \theta_2, \ldots, \theta_q$ are the coefficients for parameter estimation.

Combining the ARIMA (p,d,q) model with the AR, the MA, and the difference method yields the ARIMA model, where d is the order in which the data must be differenced. As a result, the final step in ARIMA modeling is to determine the lag length using information criteria.

As a result, the ARIMA (p,1,q) model is written as:

$$y_t = \beta_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \ldots + \emptyset_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \tag{3}$$

Where y_t is the first order differenced series, and $\emptyset, \beta$ and θ are the coefficients for parameter estimation [8].

2.2.1. Augmented Dickey-Fuller Test

The ADF Test is used in this research to determine whether or not the time series is stationary. The following is the definition of a random walk with trend and drift:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p+1} + \varepsilon_t \tag{4}$$

Where the number of included p which is the AR process's lag order, should be large enough to make the residuals serially uncorrelated [9]. If the ADF test result's p-value exceeds 0.05, the time series is not stationary as what the null hypothesis states and should be differenced.

2.2.2. ACF & PACF

In addition to the ADF test, this research employs the autocorrelation function (ACF) and partial autocorrelation function (PACF) to assess the time series' stationarity, and both of them are plotted on the correlograms for the time series analysis.

The ACF is a mathematical expression that represents the degree of continuity over the relevant variable lags:

$$r_k = \frac{\sum_{t=1}^{T-k}(y_t-\bar{y})(y_{t+k}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2} \tag{5}$$

The PACF is a mathematical expression that calculates the degree of connection of two variables:

$$\begin{cases} r_1 & if\ k = 1 \\ \frac{r_k-\sum_{j=1}^{k-1}(P_{k-1,j}\cdot r_{k-j})}{1-\sum_{j=1}^{k-1}(P_{k-1,j}\cdot r_j)} & if\ k = 2,3,\dots \end{cases} \tag{6}$$

Where $P_{k,j} = P_{k-1,j} - P_{k,k}P_{k-1,k-j}$ for $j = 1,2,\dots,k-1$. If the ACF and PACF plots show that the majority of the coefficients are within critical levels, the time series can be termed stationary.

2.2.3. The Ljung-Box Test

After an ARMA(p,d,q) model has been fitted to the data, the residuals of a time series are subjected to the Ljung-Box test. The test looks at the residuals' m autocorrelations. If the autocorrelations are really minute, then we get to the conclusion that the model does not have a significant lack of fit [10]. The Box Ljung Test's null hypothesis, H0, indicates that our model does not show lack of fit. The alternative hypothesis, Ha, is that the model exhibit a lack of fit. A great p-value indicates that the time series is not autocorrelated [10].

3. Results and Discussion

3.1. Russia CPI Time Series Observation

A time series plot of the Russia CPI from March 2012 to March 2022 is demonstrated in Figure 1.

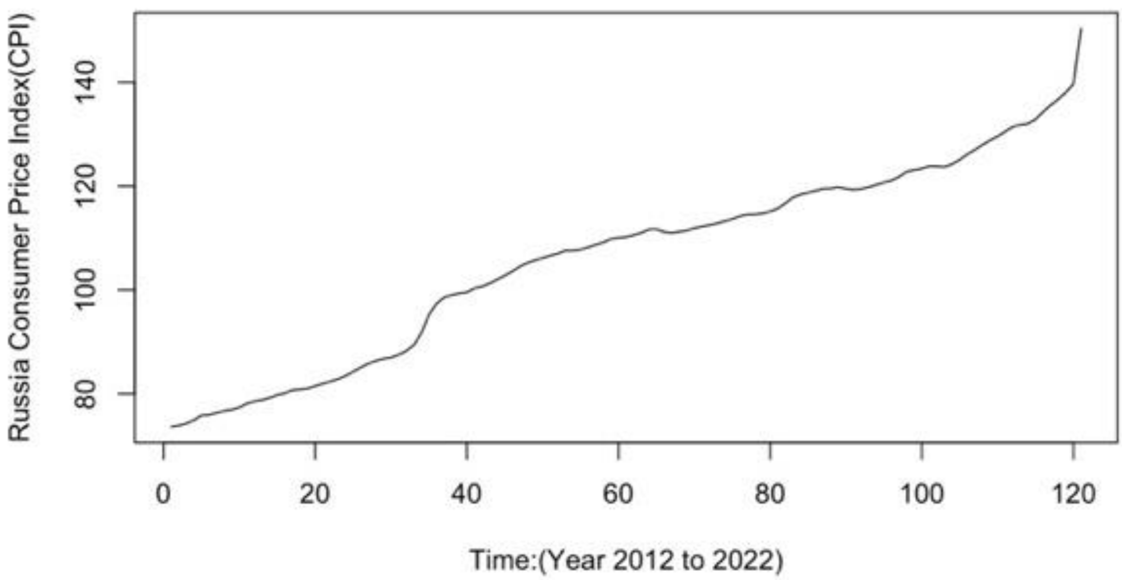

Figure 1: Time series plot of Russia CPI.

The descriptive data for the CPI time series are displayed in Table 1. There were 121 total observations; the reported CPI peaked at 150.35 in March 2022, and the lowest documented data was

73.68 in March 2012. The figure exhibits an overall upward trend. Additionally, there is no apparent seasonal variation in the series.

Table 1: Descriptive statistics of Russia CPI.

Statistic	CPI (2012-2022)
No. of observations	121
Minimum	73.68
1st Quartile	87.56
Median	110.22
Mean	106.30
3rd Quartile	119.56
Maximum	150.35
Range	(73.68, 150.35)
Standard deviation	18.55

Table 2 shows that the Dickey-Fuller statistic is 6.976 and the resulting large p-value of 0.99 indicates that the time series is not stationary.

Table 2: ADF Test for CPI.

Dickey-Fuller	Lag order	p-value
6.976	0	0.99

One way to enhance the stationarity of the series is by applying a transformation through the process of taking differences. Table 3 displays the ADF test results for the series following the third difference. The outcome indicates that the p-value is ultimately below 0.05. Hence, it is possible to reject the null hypothesis.

Table 3: ADF Test for CPI third difference.

Dickey-Fuller	Lag order	p-value
-7.57	0	0.01

However, it is important to first analyze the ACF and PACF before making any conclusions about stationarity.

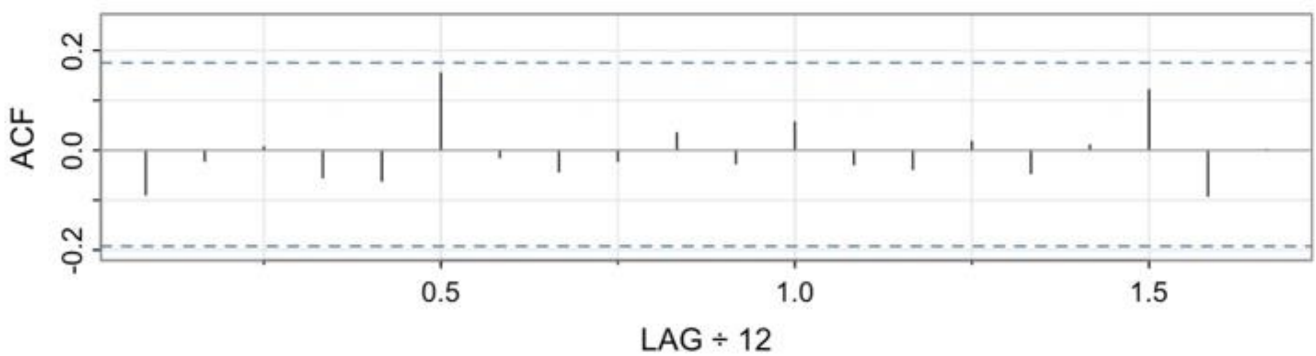

Figure 2: ACF for the third difference.

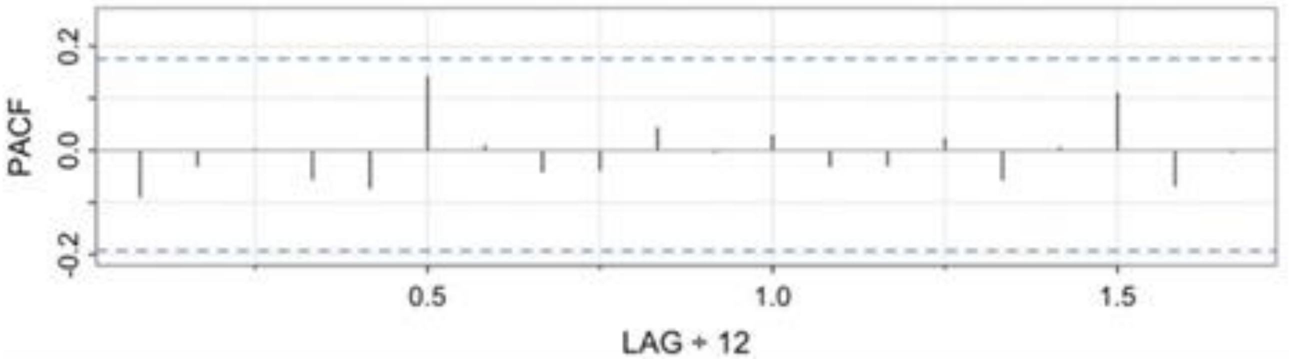

Figure 3: PACF for the third difference.

For the third difference in the Russia CPI series, Figures 2 and 3 show the ACF and PACF. The majority of coefficients fall within a critical range. The third difference time series is stationary which can finally be fitted into the ARIMA model and estimated its parameters.

3.2. ARIMA Model for CPI

Table 4 displays the results for automatically fitted ARIMA models with log-likelihood, AIC, ME, MAE, RMSE, and MPE following log-transformation of the data. The autocorrelation of the series can be reduced by using log-transformation. ARIMA (2,3,1) is the fit model before transformation, while ARIMA (1,3,1) (1,0,1) [12] is the fit model after the transformation.

Table 4: Results from different ARIMA models.

Model	Log-likelihood	AIC	ME	RMSE	MAE	MPE
ARIMA(1,3,1)(1,0,1) [12]	-185.73	381.47	0.0353	1.0819	0.6984	320.328
ARIMA(2,3,1)	-183.68	375.37	0.0688	1.1379	0.7314	320.0436

Also, a model that has the lowest AIC and the highest log-likelihood surpasses the others. The log-likelihood, AIC, ME, RMSE, MAE and MPE will be considered as the important standards to measure forecasting performance in the study.

In Table 5, both the ARIMA (1,3,1)(1,0,1) [12] and ARIMA (2,3,1) Ljung-Box test show that the p-value is less than 0.05, which imply significant autocorrelation for residuals.

Table 5: Ljung-Box test of ARIMA (1,3,1)(1,0,1) [12] and ARIMA (2,3,1).

data: Residuals from ARIMA(1,3,1)(1,0,1)[12]	data: Residuals from ARIMA(2,3,1)
Q* = 43.045, df = 20, p-value = 0.002017	Q* = 34.717 df = 21, p-value = 0.03032
Model df: 4. Total lags used: 24	Model df: 3. Total lags used: 24

3.3. Comparative Analysis

By comparing two best fit models using different methods, this study finds out that the ideal model is ARIMA (2,3,1) because it best reflects the stochastic variance in the data and satisfies the aforementioned requirements for forecasting accuracy. The results of ARIMA (2,3,1) parameters are shown in Table 6. The results of residuals from ARIMA (2,3,1) are shown in Figure 4.

DOI: 10.54254/2754-1169/51/20230675

Table 6: ARIMA (2,3,1) parameters.

Variable	Coefficient	Standard Error
AR(1)	-1.0983	0.0995
AR(2)	-0.6227	0.1001
MA(1)	-1.0000	0.1314

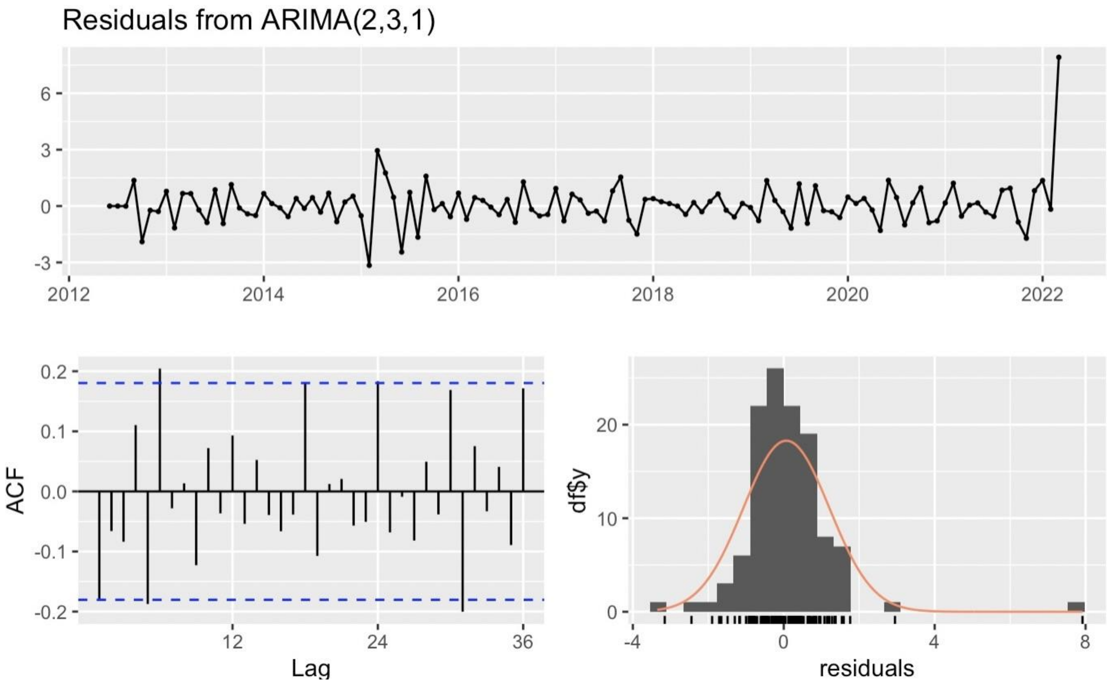

Figure 4: Residual result for ARIMIA (2,3,1).

3.4. Regression with ARIMA Errors for CPI Using Crude Oil Price Observation

Table 7 shows the descriptive statistics of both Russia CPI and crude oil price: Brent-Europe from March 2012 to March 2022.

Table 7: Descriptive statistics of Russia CPI.

Statistic	CPI (2012-2022)	Europe-Brent
No. of observations	121	121
Minimum	73.68	18.38
1st Quartile	87.56	51.59
Median	110.22	64.75
Mean	106.30	71.72
3rd Quartile	119.56	97.13
Maximum	150.35	125.45
Range	(73.68, 150.35)	(18.38,125.45)
Standard deviation	18.55	25.9218

Figure 5 shows a time series plot of both the Russia CPI and European oil price for the same period. CPI time series has an overall increasing trend while the oil price time series has dramatic fluctuations over the period.

DOI: 10.54254/2754-1169/51/20230675

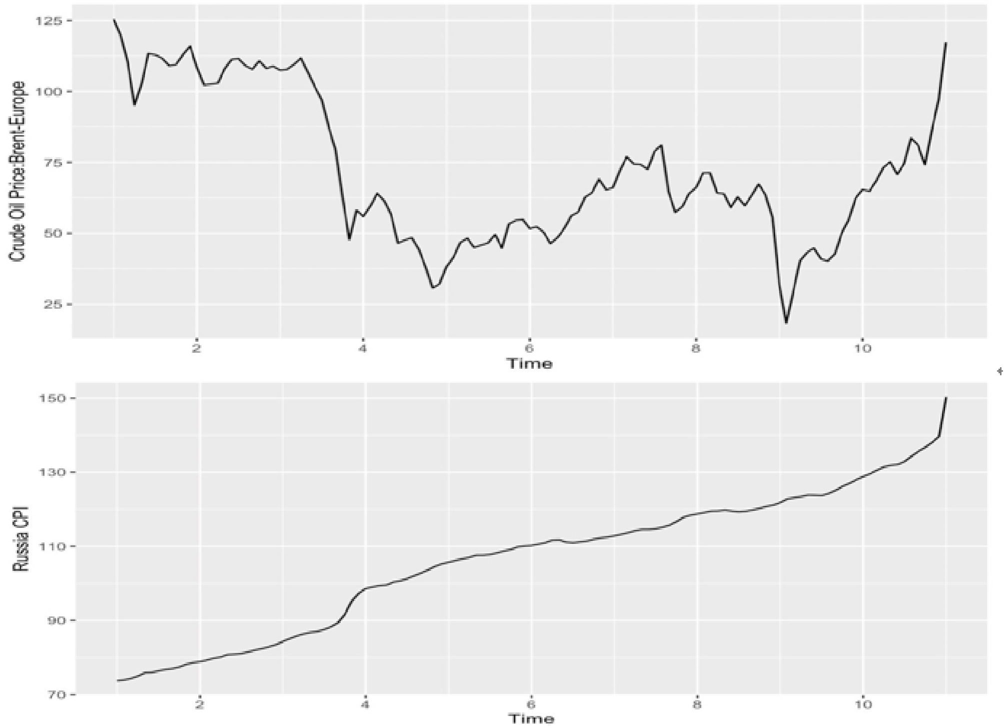

Figure 5: Time series plot of Russia CPI and European crude oil price.

In this section, we set the oil price as the independent variable, and fit the CPI data by regression with ARIMA errors model which is suitable for time series data that are autocorrelated.

The outcomes of two alternative regression with ARIMA errors models are shown in Table 8. As previously stated, a model with the largest log-likelihood and other indexes that are the lowest will be selected in advance.

Table 8: Results from different ARIMA models.

Model	Log-likelihood	AIC	ME	RMSE	MAE	MPE
ARIMA(0,0,0)errors	-498.07	1002.14	-2.28e-14	14.8402	11.8419	-1.9478
ARIMA(1,3,0) errors	-159.23	324.46	0.0740	0.9209	0.3968	0.0495

In Table 9, the ARIMA (0,0,0) errors Ljung-Box test shows that the p-value is smaller than 0.05, which implies significant autocorrelation for residuals. However, the ARIMA (1,3,0) errors Ljung-Box test indicates little autocorrelation for residuals.

Table 9: Ljung-Box test of residuals from ARIMA (0,0,0) errors and ARIMA (1,3,0) errors.

data: Residuals from ARIMA(0,0,0)errors	data: Residuals from ARIMA(1,3,0) errors
Q* = 812.32, df = 24, p-value < 2.2e-16	Q* = 10.556 df = 23, p-value = 0.9873
Model df: 0. Total lags used: 24	Model df: 1. Total lags used: 24

3.5. Comparative Analysis

The ACF plot in Figure 6 and a Ljung-Box test shown in Table 9, which describe the residuals from the ARIMA (1, 3, 0) errors model, both indicate that the residuals are white noise, since all autocorrelations are below the threshold level and the output of the p-value is greater than 0.05.

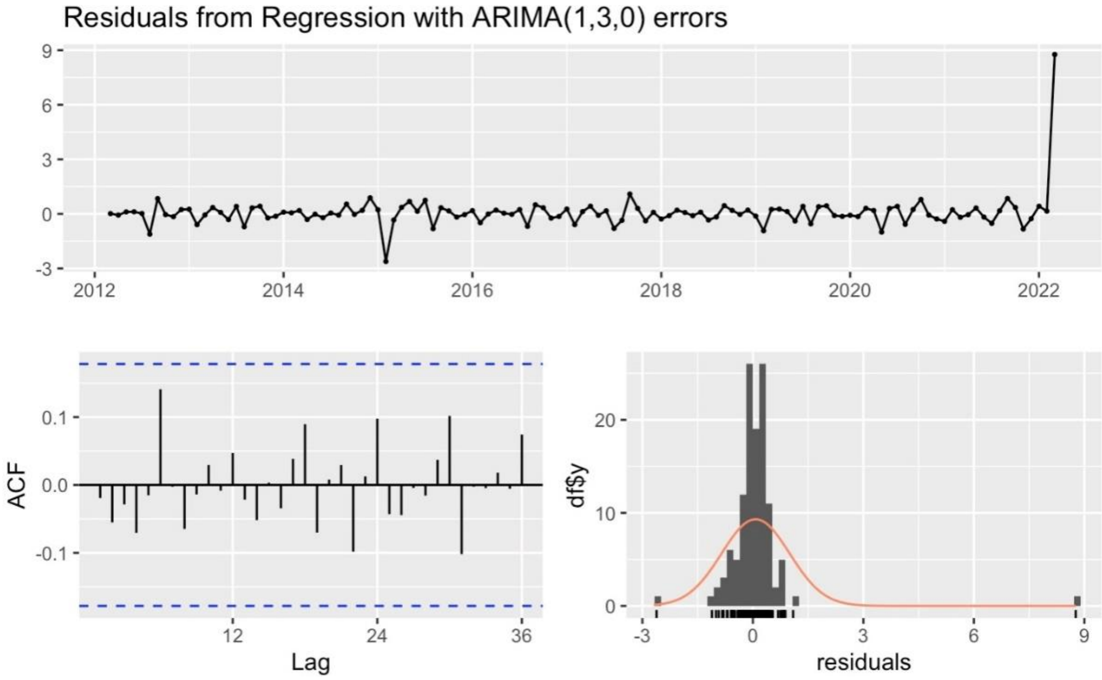

Figure 6: Residual result from ARIMA (1,3,0) errors.

By comparing two models using different methods, this study finds out that ARIMA (1,3,0) errors is the best model since it reflects the stochastic variance comparatively well and satisfies the aforementioned requirements for forecasting accuracy. The results of ARIMA (1,3,0) errors parameters are shown in Table 10.

Table 10: ARIMA (1,3,0) errors parameters.

Variable	Coefficient	Standard Error
AR(1)	-0.3134	0.1787
xreg	0.0000	0.0093

3.6. Comparative Analysis for From ARIMA and Regression with ARIMA Errors

After spotting and analyzing the model, it is crucial to apply the model to predict the future CPI data as the aim of the study. Figure 7 indicates that the CPI in Russia is anticipated to continue increasing over the following two years, with a predicted range from January 2023 to January 2025. In contrast, Figure 8 can only predict the values for the next two months, with a fault between March 2022 and May 2022. Overall, it is clear that ARIMA (2,3,1) is the model that fits the data the best.

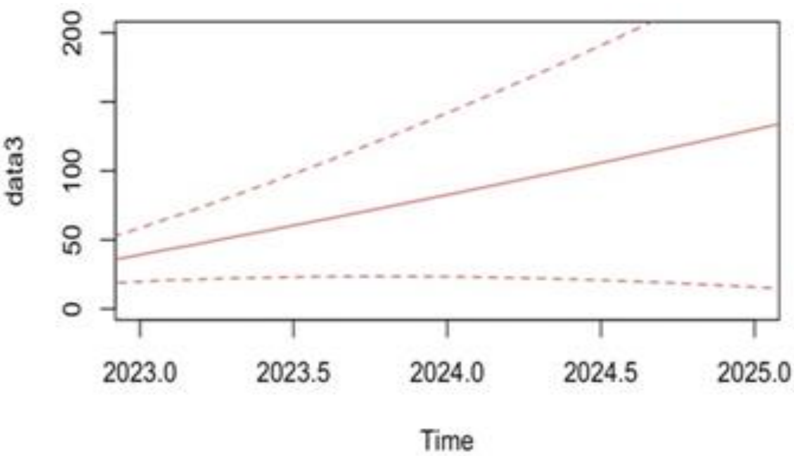

Figure 7: Forecast result from ARIMA (2,3,1).

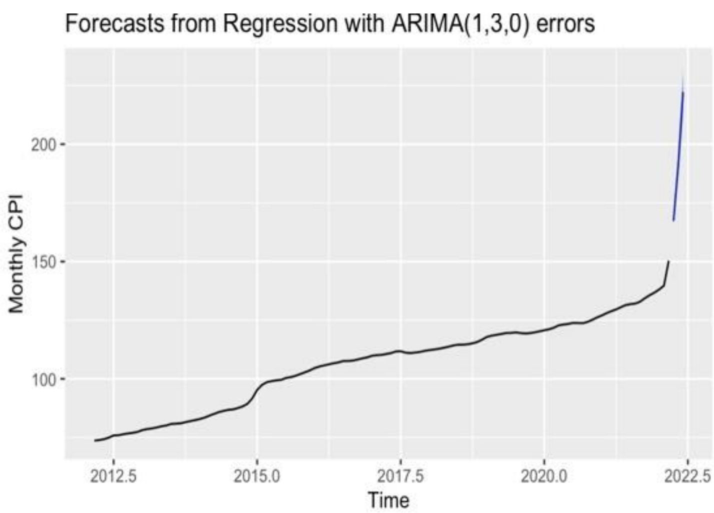

Figure 8: Forecast result from ARIMA (1,3,0) errors.

4. Conclusion

The analysis of the ARIMA and regression with ARIMA errors models was conducted to explore the monthly CPI for Russia over the past decade from March 2012 to March 2022. The study's main goal was to forecast the monthly CPI for Russia, and the best-fitting model was chosen based on how effectively it captured stochastic variance. The ARIMA (2, 3, 1) model was found to be a reliable model for predicting Russia's CPI over the following two years. However, as previously mentioned, when the study came to forecast the future CPI values by ARIMA(1,3,0) errors, the result is quite abnormal and not even give an output with a required period. That is probably because the currently available dataset only includes the data before March 2022 which is a special endpoint for statistics since it is the date that Russia launched the war against Ukriane, and it created a sharp increasing trend in CPI data that is comparatively difficult to forecast. Generally, CPI in Russia had been trending upwards over the predicted period. The result of this study suggests that monetary and fiscal authorities in Russia should adopt more prudent monetary policies to address the inflation risk in Russia. It is critical to substantially enhance the quality of identifying current short-term significant trends in consumer prices that the Bank of Russia considers as it makes decisions regarding monetary policy.

DOI: 10.54254/2754-1169/51/20230675

References

[1] *Cosgrove, J. (2020). The Russian invasion of the Crimean peninsula, 2014-2015: A post-cold war nuclear crisis case study. Applied Physics Lab, Johns Hopkins University Laurel Md Laurel.*

[2] *Ara Stepanyan, Agustin Roitman, Gohar Minasyan, Dragana Ostojic, and Natan Epstein.(2015). Spillover. THE SPILLOVER EFFECTS of Russia's Economic Slowdown on Neighboring Countries. Retrieved from https://www.imf.org/-/media/Files/Publications/SpilloverNotes/SpilloverNote3.ashx.*

[3] *Russia inflation rate in 2014. Russian Federation Inflation Rate in 2014. (n.d.). Retrieved from https://www.statbureau.org/en/russia/inflation/2014.*

[4] *Ozili, P. K. (2022). Global economic consequence of Russian invasion of Ukraine.SSRN, 4064770.*

[5] *Yurina, E. A., Tishchenko, E. S., Fedotova, E. V., Markov, B. E., & Kuzmenko, N. I. (2022). New Approaches to the Economic Security of Russian Regions in Terms of the COVID-19 Pandemic. In Geo-Economy of the Future: Sustainable Agriculture and Alternative Energy. Cham: Springer International Publishing, 699-715.*

[6] *Kasyan, E. A., & Nemirovich-Danchenko, M. M. (2022). Selection of a forecasting model for Russia consumer price index based on the application of Analytic Hierarchy Process. In Journal of Physics: Conference Series, 2291, 1, 012006.*

[7] *Benkraiem, R., Garfatta, R., Lakhal, F., & Zorgati, I. (2022). Financial contagion intensity during the COVID-19 outbreak: A copula approach. International Review of Financial Analysis, 102136.*

[8] *GAUTAM, R. S., & KANOUJIYA, J. (2022). Inflation Targeting: An Application of ARIMA Modelling Using Forecasting of CPI and WPI. Iconic Research and Engineering Journals, 5(11), 195-198.*

[9] *Jalil, A., & Rao, N. H. (2019). Time series analysis (stationarity, cointegration, and causality). In Environmental Kuznets Curve (EKC).Academic Press, 85-99.*

[10] *Stephanie. (2023). Ljung box test: Definition. Statistics How To. Retrieved from https://www.statisticshowto.com/ljung-box-test/.*