American Journal of Theoretical and Applied Statistics
2024; 13(4): 85-91
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20241304.14
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Modelling Petrol Prices in Kenya from 2014 to 2023 Using Sarimax Model: A Case Study of Nairobi County

Fidelis Nyamai[1], Joseph Esekon[1], Edwine Atitwa[2, *]

[1]Pure and Applied Sciences, Kirinyaga University, Kerugoya, Kenya
[2]Pure and Applied Sciences, University of Embu, Embu, Kenya

Email address:

fnyam99@gmail.com (Fidelis Nyamai), jesekon@kyu.ac.ke (Joseph Esekon), eatitwa@gmail.com (Edwine Atitwa)
[*]Corresponding author

To cite this article:

Fidelis Nyamai, Joseph Esekon, Edwine Atitwa. (2024). Modelling Petrol Prices in Kenya from 2014 to 2023 Using Sarimax Model: A Case Study of Nairobi County. *American Journal of Theoretical and Applied Statistics, 13*(4), 85-91.
https://doi.org/10.11648/j.ajtas.20241304.14

Received: 1 August 2024; **Accepted:** 16 August 2024; **Published:** 26 August 2024

Abstract: The requirement for petrol price information is crucial for majority of enterprises. This is because fluctuations in petrol prices impact inflation hence affecting daily lives of citizens. In analyzing the prices of petrol, researchers have employed several models but encountered various limitations. These limitations include; the Error Correction Model can examine only one co-integrating association. The Vector Autoregression (VAR) model does not account for the structural changes in the data. Additionally, the AutoRegressive Integrated Moving Average (ARIMA) model does not take into consideration the seasonal component in the data. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model assumes that over time the volatility is constant. Moreover, the Seasonal Autoregressive Integrated Moving Average (SARIMA) model does not integrate the external factors. Hence in this study Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model was employed since it captures seasonality in data and incorporates the exogenous variables. The research's aim was to model prices of petrol in Kenya for the period between 2014 to 2023 with exchange rates as an external factor. Secondary data was obtained from Energy and Petroleum Regulatory Authority (EPRA), Kenya National Bureau of Statistics (KNBS) and Central Bank of Kenya (CBK) websites. R software was used to analyze the data. By the use of historical data of petrol prices and exchange rates, the study sought to fit the best Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model, validate the model and predict the petrol prices. The petrol price data was found to be non-stationary using Augmented Dickey Fuller test (ADF). Regular differencing was conducted to make the data stationary. Seasonal differencing due to seasonality component available in the data was also performed. Best SARIMAX model was chosen from various SARIMAX models according to Box-Jenkins methodology which uses least Akaike Information Criterion (AIC) value. SARIMAX $(0, 1, 1)(2, 1, 2)_{12}$ model was selected since it had least Akaike Information Criterion (AIC) value of 656.3733 and the model validated using the hold out technique. The forecasts errors from the training set were; Mean Squared Error (MSE)=10.4970, Root Mean Square Error (RMSE)=3.239911, Mean Absolute Percentage Error (MAPE)=2.309268% while those from the testing set were; Mean Squared Error (MSE)=3271.1012, Root Mean Square Error (RMSE)=57.193542, Mean Absolute Percentage Error (MAPE)=26.695390%. There was less error in the training set than in the testing set as it was expected hence the model suited the data well and could be used for future predictions. The model was then used for five year forecast into the future. This study's findings will offer sound suggestions to policymakers, businesses and consumers. This study recommends a model to be fitted using other factors affecting petrol prices and fitting Fourier terms, Behavioral Assessment Tools (BATS) and Trigonometric Box-Cox ARMA Trend Seasonal (TBATS) models.

Keywords: Exogenous Variable, Petrol Price, Exchange Rate, Differencing, Forecasting

1. Introduction

Global economy is highly dependent on energy. Energy is the key factor propelling economic activities [1]. The bigger the nation's energy consumption, the greater its economic activity hence the superiority of its economy. Therefore, energy is vital for economic development.

Academicians and energy specialists have analyzed the past, present, and future price levels and fluctuations of petrol, considering its importance to the Kenyan economy. Globally, the prices of petrol have been increasing. For instance, over the past 20 years, the average petrol prices have doubled [2]. In 2001, the average petrol price globally was at 0.60 dollars per litre but to date it is at 1.32 dollars per litre.

Petrol prices globally and locally vary from time to time and are influenced by various factors [3]. To start with is the crude oil prices. When the prices of crude oil change they cause the petrol prices to change since petrol is produced from the crude oil. Secondly is the exchange rates since the crude oil is imported in US dollars. When a country's currency is weak, the imported crude oil becomes expensive leading to increase in the prices of petrol. Additionally are the taxes and subsidies. When taxes on petrol prices increase, the prices also increase and vice versa. Moreover governments intervene by providing subsidies to the petrol and this makes the price lower for consumers. The demand and supply dynamics also cause variation in petrol prices. When demand of petrol increases and the supply decreases, the prices also increase and vice versa. Distribution and marketing costs as well affect the prices of petrol. Changes in these costs cause variation in the petrol prices.

These petrol price variations have various effects. To start with, it affects inflation rates. Increase in petrol prices causes increase in inflation rates hence leading to increased cost of living. Secondly, the variations affect the cost of transportation. The increase in the price of petrol causes increase in transportation cost hence affecting transport businesses and travellers [4]. Additionally, they affect Government Budget. As petrol prices increase, government imposes price controls and also subsidies to stabilize fuel prices hence causing pressure on the government budget. Lastly, they affect the spending of consumers. When petrol prices increase, the spending on fuels decrease and this is common among individuals with low income.

The prices of petrol in Kenya are determined by the EPRA [5]. The following formula is used to calculate the landed cost of petroleum products:

$$C_l = \frac{\sum(V_{irp} \times C_{irp})}{\sum V_{irp}}$$

where

1. C_l is the weighted average price, expressed in Kenyan shillings per liter of petroleum products imported via an On-Time Scheduling (OTS) primary storage depot that has been gazetted,
2. V_{irp} is the amount in litres of a petroleum product merchandise that was imported through the OTS and released at the Mombasa port between the 10th day of the previous month and the 9th day of the month pricing is done, and
3. C_{irp} is the unit cost of a petroleum product cargo imported via the OTS that is released at the port of Mombasa in Kenya between the 10th day of the preceding month and the 9th day of the pricing month.

The wholesale prices of petrol are determined as

$$P_w = C_l + J_{hc} + C_p + L_{ps} + P_t - f - L_{pt} + C_{ss} + L_{ss} + C_f + M_w + Y + T + VAT$$

where

1. P_w: is the highest possible wholesale cost,
2. C_l: the cost at which imported petroleum products are landed,
3. J_{hc}: jetty handling costs,
4. C_p: primary cost of storage,
5. L_{ps}: Secondary storage losses allowed,
6. C_{ss}: costs on secondary storage,
7. P_t: costs on primary transportation,
8. L_{pt}: primary transport allowed losses,
9. $L_s s$: allowable losses for secondary storage,
10. C_f: financing costs inventory,
11. f: factor for adjusting price calculation,
12. M_w: margin on wholesale,
13. Y: any additional considered expense that the Authority has approved,
14. T: taxes applicable not including VAT, and
15. $ValueAddedTax(VAT)$: value added tax.

The retail prices of petrol price is calculated as

$$P_r = P_w + T_s + M_{ri} + M_{ro} + VAT$$

where

1. P_r is the highest price at which petrol should retail at,
2. T_s is transporting cost of prices of petrol from secondary storage area to the retailing location,
3. M_{ri} is the retail margin that corresponds to the costs incurred by a benchmark retail dispensing location,
4. M_{ro} is the retail margin that matches the benchmark retail dispensing site's operational expenses, and
5. P_w: is the highest possible wholesale cost.

Reliable forecasts of petrol prices are of interest for a wide variety of applications. For example, central bank and private sector forecast consider the price of petrol to be one essential variable when developing macroeconomic projections and analysis of macroeconomic risks. An estimate of petrol prices which is more accurate can increase macroeconomic forecasts and policy response accuracy [6].

Predicting is a key tool utilized to forecast future values using past values [7]. In modelling petrol prices, several models have been used. First is the ARIMA model which is utilized when data has stationary properties and autocorrelation. Secondly is the SARIMA model that is an extended ARIMA model to capture the seasonality component in the data. Thirdly is the Vector Error Correction Model (VECM). This model is used when there is a cointegration between several variables. Moreover, we have the ARIMA-X model which makes it possible to incorporate the external factors for example the prices of global oil, but does not account for seasonality component in the data [8].

This study made use of SARIMAX model to fit the best model for petrol prices and also made forecasts.

2. Materials and Methods

2.1. Introduction to SARIMAX Model

The AutoRegressive (AR) model describes a series according to the association among current values and their own past observations [9]. On the other hand, Moving Average (MA) describes a series as a combination of present and past errors. These two models deal with data which is stationary and linear. Nevertheless, in most situations the data is non-stationary hence requiring differencing to become stationary. This is where the ARIMA model comes in which is an amalgamation of three components: the AR model, MA and the I component for differencing operation. The ARIMA model is denoted as ARIMA (p, d, q) whereby p is the order of AR part, d is the number of differences applied to create a stationary series and q is the MA's order of the model.

Seasonality might be present in data which needs to be accounted for though the ARIMA model do not handle seasonality. In this case the Seasonal ARIMA model is used since it handles the seasonality component by utilizing the seasonal MA, AR and seasonal differencing. The model is identified as SARIMA $(p, d, q)(P, D, Q)_s$ where p is the AR order, q the MA order, d the non-seasonal differences, P the seasonal AR order, Q the seasonal MA order, D the seasonal differences and s the number of periods in a season.

The SARIMAX model is an extended SARIMA model to involve the exogenous variables to improve the performance of the model. Exogenous variables are those that have an impact on a model but do not get any influence from it. For instance the exchange rates is an exogenous variable in analyzing petrol prices. The SARIMAX model incorporates the external impacts of influences affecting the variable of interest being investigated in a time series.

2.2. Box-Jenkins Methodology

Box-Jenkins Methodology was utilized for this study which involves model identification, estimation, diagnostic checking and forecasting[10]. Below are the steps:

2.2.1. Model Identification

This process determined which subclass of models best fits the data by utilizing the data and the knowledge about how it was developed. This step is subdivided into two;

a. Data Preparation.

To guarantee the stability of the variance, a plot on the time series was created and data processing utilizing the offset logarithm was carried out. Secondly, the data was tested for stationarity. The researcher first assessed if or not the series was stationary by examining the ACF graph. A stationary series is defined as one without a trend, with mean and variance constant across time, making it simple for predicting values [11].

Assuming that the graph of the time series values either cut off fairly quickly, therefore the series will be regarded stationary. Assuming a graph of ACF decrease particularly slowly, formerly the series will be regarded non-stationary. The researcher additionally used the Dickey-Fuller unit root test for stationarity with an augmentation. The null hypothesis was that the series is non-stationary while the alternative hypothesis was that the series is stationary. If the time series is non stationary, differencing transforms it into a stationary series. Differencing had to be conducted since SARIMAX works with stationary data.

b. Model Selection.

In model selection, the plots of ACF and PACF were utilized to identify potential models. The Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC) were utilized in choosing the best models based on their minimum values. Though BIC is regarded less effective with big data sets [12]. AIC penalizes complexity while gauging a model's goodness of fit, or how well it matches the data. Consequently, AIC lowers the chance of both underfitting and overfitting. Using AIC the best model was determined as Equation.

$$\mathrm{AIC} = -2 \log L + 2m \tag{1}$$

where $m = p + q$ and L is the likelihood function,

p is the number of AR parameters,

q is the number of MA parameters.

Comparing the AIC values of competing models allows for the inference of the superior model. The model that has the lowest value of AIC is the best.

2.2.2. Model Estimation

After the best model was chosen, the model parameters were estimated by utilizing the Maximum Likelihood Estimation (MLE) approach. The MLE approach was chosen due to the fact that it is the most practical and appropriate approach given that measurement errors can give rise to error terms that can be categorized as random components. Suppose that $X_1, X_2, \cdots, X_n$ are random variables from a density function given by $f(x, \theta)$, then the likelihood function is obtained as

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta) \tag{2}$$

In relation to θ, the likelihood function is differentiated and equated to zero in order to find the MLE of θ as

$$\frac{dL(\theta;x)}{d(\theta)} = 0$$

2.2.3. Diagnostic Checking

This aided in determining whether or not the projected model was adequate [13]. To find out the goodness of fit of the chosen model, the autocorrelation test was performed. It was carried out using the Ljung-Box test. The null hypothesis(H_0) was that no correlation between residual errors and the alternative hypothesi(H_1) was that there is correlation between residual errors. The p-value from this test was compared with a specified level of significance. The residuals' ACF was also plotted and examined if the residuals were white noise.

Histogram and normal Q-Q plots were generated and examined if the Kernel Distribution Estimation(KDE) line demonstrate the residuals approximately follow a normal distribution. The residuals from the best selected model should be white noise and conform to a normal distribution.

3. Results and Discussion

3.1. Exploratory Data Analysis (EDA) Techniques

The data was decomposed into its individual components and the results are shown in Figure 1.

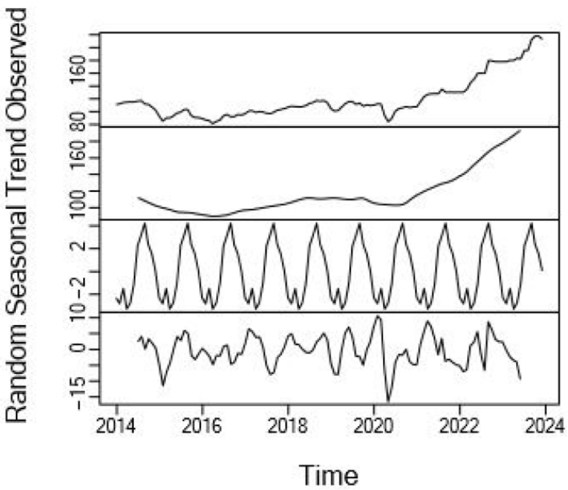

Figure 1. *Decomposed Petrol Price Data.*

Figure 1 indicates the isolated petrol price components. The figure shows that there was a downward trend followed by an upward trend. There was also seasonal component and the random component which is due to measurement errors.

The ACF and PACF of the petrol prices data were plotted and the results are given in Figure 2.

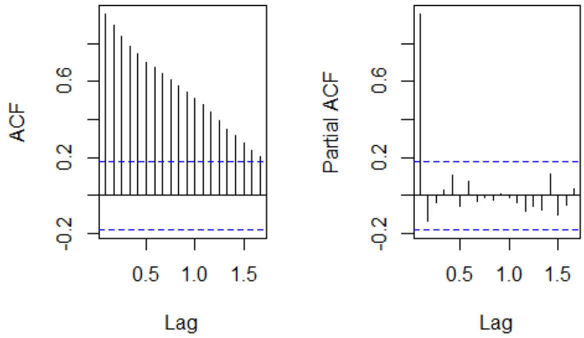

Figure 2. *ACF and PACF of Petrol Price Data.*

The slow decay in the ACF plot from Figure 2 shows that the data is non stationary. As there are numerous non-zero spikes in the ACF plot, the data showed strong positive correlations, indicating that the time series data was not random. The PACF plot in Figure 2 indicates that there was a high level of autocorrelation in the data.

3.2. Confirmatory Statistics

The confirmatory statistics for the petrol prices data obtained from Jan 2014 to Dec 2023 were calculated. The average price of petrol per month was Ksh 120.09 and the variance was 1030.59. The median was Ksh 110.26 and the skewness was 1.478936 which was positive meaning that the tail was stretching to the right. The p-value from the Shapiro Wilk test was 3.674×10^{-11} which was less than 0.05 showing that the data did not follow normal distribution.

Table 1. *Confirmatory Statistics of Petrol Price Data.*

Statistic	Value
Mean	120.09
Variance	1030.59
Median	110.26
Skewness	1.478936
Kurtosis	1.342780
Shapiro-Wilk test	p-value = 3.674×10^{-11}

3.3. Stationarity Test

The stationarity of the data was tested using the ADF stationarity test which also helped to identify whether or not differencing was necessary. The test results are given in Table 2.

The test results in Table 2 produced a p-value of 0.7397 which is greater than 0.05, the chosen level of significance. This led to failing to reject the null hypothesis that the series is non-stationary. Since the series was non-stationary, differencing was needed to make it stationary. Differencing was conducted and the ADF stationarity test was performed again and the results are given in Table 3.

Table 2. *Augmented Dickey-Fuller Test Results.*

Description	Value
Data	log(ts_petrol)
Dickey-Fuller Statistic	-1.6073
Lag Order	4
p-value	0.7397
Alternative Hypothesis	Stationary

Table 3. *Augmented Dickey-Fuller Test Results after Differencing.*

Description	Value
Data	differenced
Dickey-Fuller Statistic	-5.7556
Lag Order	4
p-value	0.01
Alternative Hypothesis	Stationary

A p-value of 0.01 in Table above was gotten which was less than 0.05 level of significance. This led to rejecting the null hypothesis and concluding that the series was stationary after carrying out regular differencing.

Seasonal stationarity test was carried out using the seasonal unit root test and the results are given in the Table below.

Table 4. *Seasonal Unit Root Test.*

Test	p-value
Seasonal unit root test	0.1075

The p-value obtained from this test was 0.1075 which was greater than 0.05 hence we fail to reject null hypothesis and conclude that the series is non-stationary. Since the series was non-stationary, seasonal differencing was conducted and the results from the seasonal ADF test are shown below.

Table 5. *Seasonal Unit Root Test after Seasonal Differencing.*

Test	p-value
Seasonal unit root test	0.0001891

A p-value of 0.0001891 was obtained. This p-value is less than 0.05, a chosen significance level and hence we reject the null hypothesis and conclude that the series is stationary after seasonal differencing.

3.4. Training SARIMAX Model

A grid search was done on 79 distinct combinations of seasonal (P, D, Q) and nonseasonal (p, d, q) characteristics in order to determine the optimal SARIMAX $(p, d, q)(P, D, Q)_s$ model. The p ,q ,P and Q parameters took the values 0 to 2 and d and D parameters took the value 1 since regular and seasonal differencing were both 1. Twelve (12) was the value of the seasonal parameter used in order to extract yearly patterns.

The AIC values of the different SARIMAX models were compared and SARIMAX $(0, 1, 1)(2, 1, 2)_{12}$ model was chosen since it had the least AIC value of 656.3733. SARIMAX $(0, 1, 1)(2, 1, 2)_{12}$ model can be interpreted as follows;

1. $p = 0$: Model does not utilize the petrol price of the month before.
2. $d = 1$: Model needs one regular differencing to make the data stationary.
3. $q = 1$: Model uses one lagged (one year) forecast errors.
4. $P = 2$: Model requires the petrol prices for the two previous seasons that is two years.
5. $D = 1$: Model needs one seasonal differencing to achieve stationarity.
6. $Q = 2$: Model applies two seasonally lagged (two years) forecast error.

Table 6 gives the SARIMAX $(0, 1, 1)(2, 1, 2)_{12}$ model coefficients and their standard errors.

Table 6. *Model Coefficients and Standard Errors.*

Term	Coefficient	Standard Error
MA(1)	0.2670	0.0924
SAR(1)	0.2138	0.2638
SAR(2)	-0.4641	0.1313
SMA(1)	-1.2077	0.3421
SMA(2)	0.6299	0.3187
Exchange rates	0.3417	0.3971

The significance (weight) of each term is displayed in the coefficient column. A coefficient's negative sign denotes that the term and the current value have an opposite association. The larger absolute value of the coefficient indicates the greater influence that variable has on the price of petrol.

The coefficient's standard deviations is displayed as the standard error. An estimate is more accurate if its standard error is lower [14]. For this case MA(1) is more accurate since it has the least standard error. A more accurate model of petrol price is produced by a more accurate coefficient.

4. Diagnostic Checks

The diagnostic of the residuals of the SARIMAX $(0, 1, 1)(2, 1, 2)_{12}$ model was checked using the plot of standardized residuals, ACF, histogram and Q-Q plot for normality and autocorrelation test using the Ljung Box test.

Table 7. *Box-Ljung test Results.*

Test	p-value
Box-Ljung test	0.6372

A p-value of 0.6372 (>0.05) level of significance was obtained. These led to failing to reject the null hypothesis. We then conclude that the residuals were not correlated.

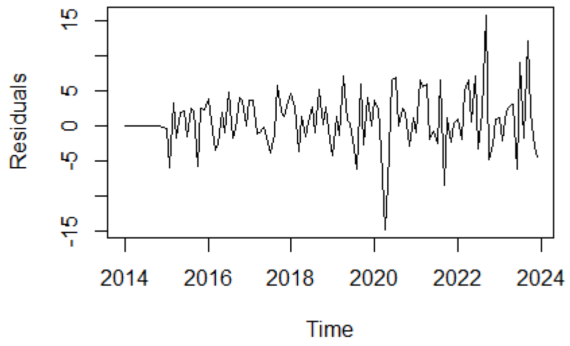

***Figure 3.** Standardized Residuals.*

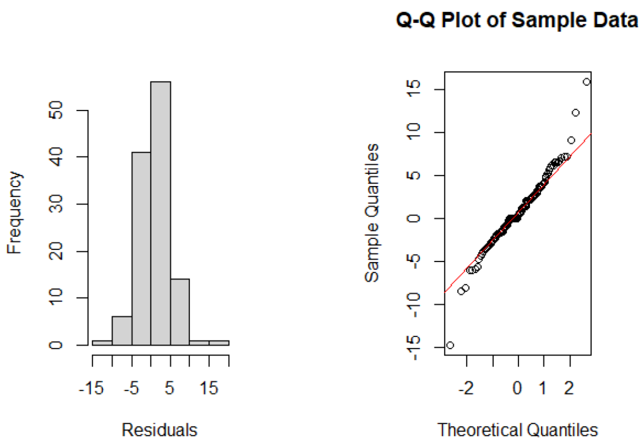

***Figure 4.** Histogram and Q-Q Plot of the Residuals.*

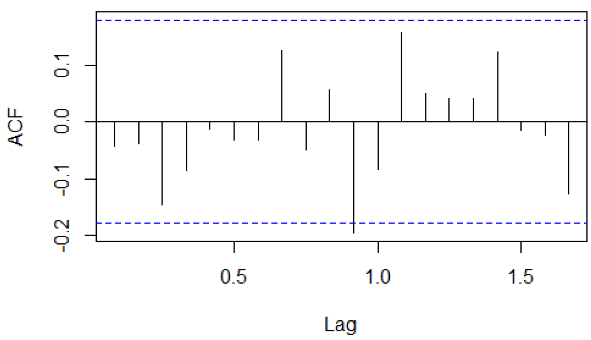

***Figure 5.** ACF of Residuals.*

We then plotted the ACF of residuals and the plot is given above. The residuals had zero autocorrelation even if the ACF did not show complete zero autocorrelation according to the Box-Ljung test. As a result, the residuals were independent of one another, meeting the assumption of the error terms' independent and identical distribution [15].

5. Conclusions

Petrol prices have been changing from time to time and currently the prices have been increasing. Therefore there is a strong desire to ascertain whether the prices will continue increasing going forward or not. SARIMAX model was fitted to the data. The confirmatory statistics of petrol prices and the results of each specific objective were discussed. The first specific objective was achieved by fitting the SARIMAX model, the second by validating the chosen model and the third by forecasting into future.

The variance of the data was found to be non-stable and stabilization was performed using log transformation. The plots of ACF and PACF suggested that the data was non-stationary and had yearly seasonality. The ADF test of stationarity was also conducted and it confirmed that the data was non-stationary. Regular differencing was conducted to remove the trend component and the ADF test was performed again and it revealed that the data was stationary after first differencing. Seasonal differencing was also performed to eliminate the seasonal component.

The ACF and PACF of regular differenced data suggested possible non-seasonal orders while the possible seasonal orders were suggested by the ACF and PACF plots of seasonally differenced data. The possible set of SARIMAX models were fitted and SARIMAX $(0,1,1)(2,1,2)_{12}$ emerged to be the best since it had the least AIC. The MLE approach was utilized to estimate the parameters of the chosen model. Diagnostics checking were also performed on the selected model. The standardized residuals indicated that the residuals were white noise. The histogram and the KDE line on the Q-Q plot showed that the residuals followed a normal distribution. The residuals' ACF showed that they had zero autocorrelation and the Box-Ljung test, with a p-value of 0.6372, verified that the residuals were white noise.

ORCID

0009-0002-0558-4116 (Fidelis Nyamai),
0009-0000-8033-9175 (Joseph Esekon),
0000-0002-8320-5874 (Edwine Atitwa)

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Azam, Muhammad. "Energy and Economic Growth in Developing Asian Economies." Journal of the Asia Pacific Economy 25, no. 3 (September 18, 2019): 447-71. https://doi.org/10.1080/13547860.2019.1665328

[2] Greene, David L., Charles B. Sims, and Matteo Muratori. “Two trillion gallons: Fuel savings from fuel economy improvements to US light-duty vehicles,” Energy Policy 142 (2020)

[3] Vochozka, Marek, Zuzana Rowland, Petr Suler, and Josef Marousek. “THE INFLUENCE OF THE INTERNATIONAL PRICE OF OIL ON THE VALUE OF THE EUR/USD EXCHANGE RATE.” Journal of Competitiveness 2 (2020).

[4] Eliasson, Jonas. “Will We Travel Less after the Pandemic” Transportation Research Interdisciplinary Perspectives 13 (March 2022): 100509. https://doi.org/10.1016/j.trip.2021.100509

[5] Ngare, Lucy W., and Okova W. Derek. “THE EFFECT of FUEL PRICES on FOOD PRICES in KENYA.” International Journal of Energy Economics and Policy 11, no. 4 (June 8, 2021): 127-31. https://doi.org/10.32479/ijeep.10600

[6] Kilian, Lutz, and Xiaoqing Zhou. “Oil Prices, Gasoline Prices, and Inflation Expectations.” Journal of Applied Econometrics, May 30, 2022. https://doi.org/10.1002/jae.2911

[7] Hewamalage, Hansika, Christoph Bergmeir, and Kasun Bandara. “Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions.” International Journal of Forecasting 37, no. 1 (January 2021): 388-427. https://doi.org/10.1016/j.ijforecast.2020.06.008

[8] Fang, Xinyu, Wendong Liu, Jing Ai, Mike He, Ying Wu, Yingying Shi, Wenqi Shen, and Changjun Bao. “Forecasting Incidence of Infectious Diarrhea Using Random Forest in Jiangsu Province, China.” BMC Infectious Diseases 20, no. 1 (March 14, 2020). https://doi.org/10.1186/s12879-020-4930-2

[9] Staffini, Alessio, Thomas Svensson, Ung-il Chung, and Akiko Kishi Svensson. “Heart Rate Modeling and Prediction Using Autoregressive Models and Deep Learning.” Sensors 22, no. 1 (December 22, 2021): 34. https://doi.org/10.3390/s22010034

[10] Hadwan, Mohammad, Basheer M. Al-Maqaleh, Fuad N. Al-Badani, Rehan Ullah Khan, and Mohammed A. Al-Hagery. “A Hybrid Neural Network and Box-Jenkins Models for Time Series Forecasting.” Computers, Materials & Continua 70, no. 3 (2022): 4829-45. https://doi.org/10.32604/cmc.2022.017824

[11] Dama, Fatoumata, and Christine Sinoquet. “Time series analysis and modeling to forecast: A survey.” arXiv preprint arXiv: 2104.00164 (2021).

[12] Tandon, Chahat, Sanjana Revankar, Hemant Palivela, and Sidharth Singh Parihar. “How Can We Predict the Impact of the Social Media Messages on the Value of Cryptocurrency” Insights from Big Data Analytics? International Journal of Information Management Data Insights 1, no. 2 (November 2021): 100035. https://doi.org/10.1016/j.jjimei.2021.100035

[13] Jewell, Nicholas P., Joseph A. Lewnard, and Britta L. Jewell. “Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections.” JAMA, April 16, 2020. https://doi.org/10.1001/jama.2020.6585

[14] Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. “When Should You Adjust Standard Errors for Clustering” The Quarterly Journal of Economics 138, no. 1 (October 6, 2022). https://doi.org/10.1093/qje/qjac038

[15] Lee, Myoung-jae. “Instrument Residual Estimator for Any Response Variable with Endogenous Binary Treatment.” Journal of the Royal Statistical Society Series B (Statistical Methodology) 83, no. 3 (July 1, 2021): 612-35. https://doi.org/10.1111/rssb.12442