

RESEARCH ARTICLE

Forecasting RON97 fuel prices in Malaysia using the Box-Jenkins methodology

Suhaila Bahrom^{1*} and Noratikah Abu²¹Department of Mathematics, Centre for Foundation Studies, International Islamic University Malaysia, 26300 Kuantan, Pahang, Malaysia²Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhr Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia

Abstract - Forecasting RON97 fuel prices is essential for economic planning and policy-making, as they influence transportation costs and the inflation rate. This study uses the Box-Jenkins method to develop a forecasting model for Malaysia's future four-week RON97 fuel price. The data was collected from Malaysia's official open data portal, covering the weekly period from March 30, 2017, to July 25, 2024. The R programming language was used for in-depth data analysis and model development. Using an 80:20 training-testing split, with residual adequacy assessed through the Ljung-Box test and forecasting accuracy evaluated using the Mean Absolute Percentage Error (MAPE), the ARIMA (2,1,0) model was identified as the most appropriate forecasting model. The model achieved an MAPE of 0.13%, indicating excellent predictive accuracy. The findings indicate that RON97 prices are forecast to remain relatively stable over the next four weeks. The findings provide valuable insights for policymakers and stakeholders in the energy sector, enabling more informed decision-making and strategic planning.

Article History

Received : 23 March 2026

Revised : 16 May 2026

Accepted : 4 June 2026

Published : 30 September 2026

Keywords

Box-Jenkins

RON97

ARIMA

Time series analysis

1. Introduction

In Malaysia, there are two types of petrol: Research Octane Number 95 (RON95) and Research Octane Number 97 (RON97). RON95 remains the most commonly used petrol in Malaysia due to its lower retail price, particularly among private vehicle users [1]. RON97 is a higher-grade fuel that comes at a higher price but can enhance the performance of certain vehicles [2]. The price of RON97 fuel in Malaysia has significant economic implications, directly affecting transportation costs, inflation, and overall financial stability [3]. Due to global crude oil prices surging, Malaysia's petrol prices have significantly increased, with RON95 and RON97 experiencing price hikes of 16% and 42%, respectively, since 2009 [4]. The volatility in fuel prices after the transition to a managed float system (MFS) in 2016 is primarily due to fluctuations in international crude oil prices, foreign exchange rates, and the reduction of subsidies [5]. Recent studies demonstrated the efficacy of time series models in predicting fuel prices. For instance, [6] utilised the ARIMA model to forecast RON97 prices in Malaysia. Univariate and multivariate time series models have been applied to forecast gasoline prices in the U.S. [7]. A comparison between the ARIMA and SARIMA models has been conducted to predict petrol and diesel prices in Ghana [8]. These studies underscore the relevance and importance of employing advanced statistical techniques for fuel price forecasting.

This study aims to analyse the price of RON97 fuel in Malaysia and develop a forecasting model using the Box-Jenkins methodology. The Box-Jenkins methodology, specifically the Autoregressive Integrated Moving Average (ARIMA) model, has been widely recognised as a robust approach for time series forecasting [9]. This methodology is particularly suitable for analysing and predicting economic and financial time series data due to its flexibility and effectiveness in capturing various temporal patterns. The ARIMA model's ability to handle non-stationary data through differencing, along with its reliance on past values and error terms, makes it a powerful tool for forecasting fuel prices [10].

2. Methodology

This study uses a comprehensive historical weekly dataset of RON97 fuel prices in Malaysian Ringgit, obtained from Malaysia's Official Open Data Portal website, covering the period from 30 March 2017 to 25 July 2024, and analysed using R software. The dataset provides sufficient observations to capture price movements, short-term fluctuations, and underlying trends over time. Such an extended time horizon is essential for developing a reliable forecasting model and improving the accuracy of future price predictions. The research framework for this study is shown in Figure 1.

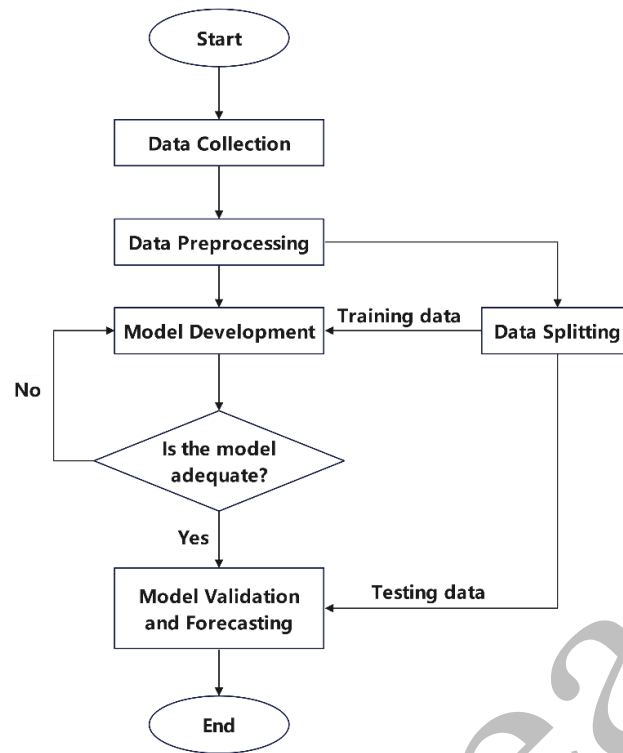


Figure 1. Research framework

2.1 Box-Jenkins model

Introduced by Box and Jenkins in 1968, the Box-Jenkins model is a highly effective approach to predictive analysis in time series research [11]. It is widely recognised for its systematic procedure in modelling historical data and producing accurate forecasts. The method is particularly suitable for analysing data that exhibit dependence over time and changing patterns. The methodology includes identifying the model, estimating parameters, diagnostic checking, and forecasting, as depicted in Figure 2.

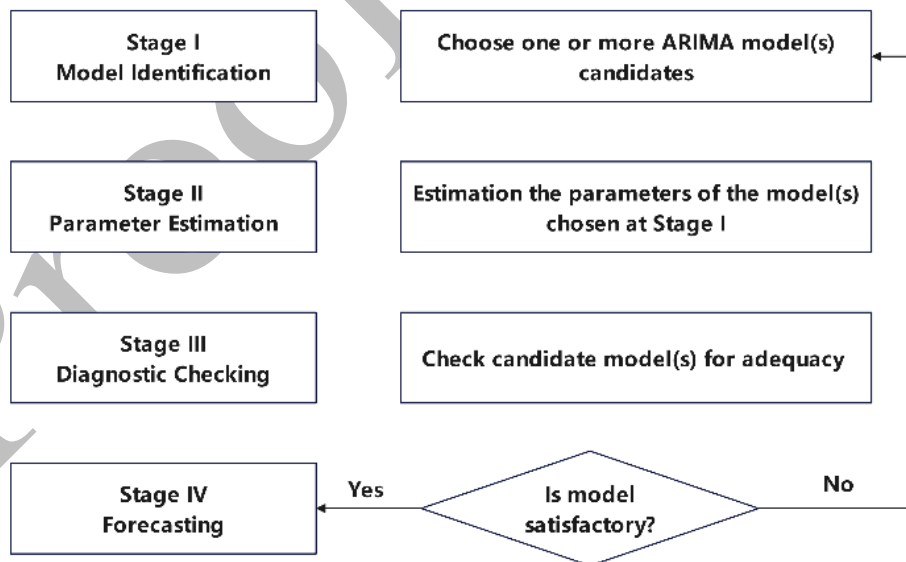


Figure 2. Box-Jenkins research framework

In model identification, the appropriate order of potential models is determined by visually inspecting the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. Additionally, tests such as the Augmented Dickey-Fuller (ADF) are used to assess the stationarity of the time series data. In the parameter estimation stage, the parameters of the selected ARIMA model are estimated using Maximum Likelihood Estimation (MLE) and AIC, and the model selection criteria. The null hypothesis indicates that parameters in the ARIMA model are equal to zero. A p -value below 0.05 suggests rejecting the null hypothesis, signifying the parameter's significance in the ARIMA model. The third stage, diagnostic checking, involves testing the residuals of the selected model against model

assumptions after fitting. The residuals of the fitted model are analysed to ensure they resemble white noise. The Ljung-Box test is used to check for autocorrelation. The final stage in the Box-Jenkins approach is forecasting. The best model was selected based on the lowest forecast errors and evaluated using the mean absolute percentage error (MAPE). For validation purposes, the data is split into training and testing sets using an 80:20 ratio. Then, the fitted ARIMA model forecasts RON97 fuel prices for the next four weeks.

2.2 ARIMA model

In general, the Box-Jenkins model consists of five models: $AR(p)$, $MA(q)$, and $ARMA(p,q)$ for the stationary models together with $ARIMA(p,d,q)$ and $SARIMA(p,d,q)(P,D,Q)$ for non-stationary models. Hence, in this study, $ARIMA(p,d,q)$ model was selected due to the presence of a trend in the fuel price data. The general form of the $ARIMA(p,d,q)$ model is presented in Eq. (1).

$$\varphi_p(B)\nabla^d y_t = c + \theta_q(B)a_t \quad (1)$$

In Eq. (1), p represents the order of the autoregressive (AR) model, indicating the number of lagged observations included in the model, q denotes the order of the moving average (MA) model, specifying the number of lagged forecast errors incorporated in the prediction equation. The order of differencing indicated by d is the number of times the data have been differenced to achieve stationarity. The backshift operator, denoted by B shifts the data back by k periods, such that $B^k y_t = y_{t-k}$. The white noise error term at the time t represent by a_t . Finally, the difference operator, represented by ∇ , is defined as $\nabla^d = (1-B)^d$, and it indicates the differencing needed to transform the time series into a stationary one.

3. Results and Discussion

This section presents and discusses all analyses and results obtained from the development of the ARIMA model. Before model development, the weekly RON97 fuel price series from 30 March 2017 to 25 July 2024 was plotted to examine its overall pattern, as shown in Figure 3. The series exhibited moderate fluctuations between 2017 and 2019, followed by a sharp decline in early 2020. Thereafter, prices recovered gradually during 2020–2021 and increased substantially to reach a peak in 2022. Following this surge, prices declined and remained relatively stable from late 2022 to July 2024. These changes in level and trend suggest that the original series is non-stationary.

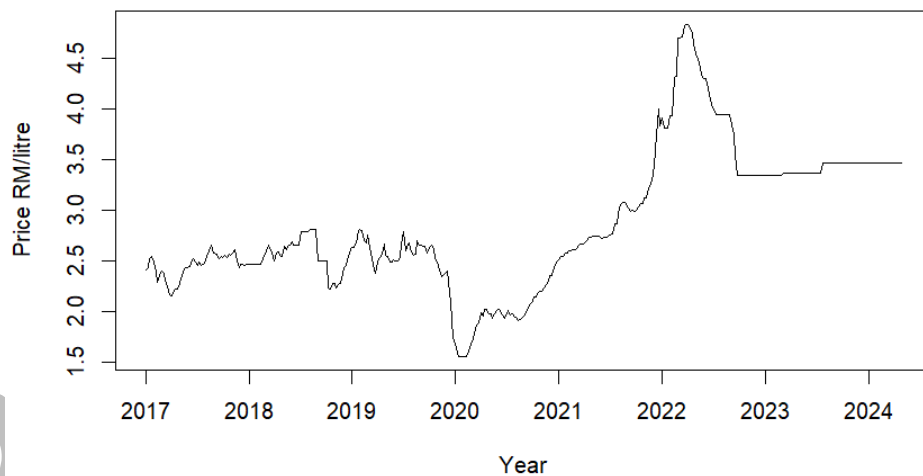


Figure 3. Time series plot of weekly RON97 fuel prices in Malaysia

Next, the dataset was split into two parts: training and testing using an 80:20 ratio, with the first 304 data points from 30 March 2017 to 16 February 2023 used for training purposes, whereas the remaining 76 observations from 23 February 2023 to 25 July 2024 were reserved for testing. The Box-Cox transformation and differencing were applied to the training dataset to satisfy the stationarity assumption. The estimated Box-Cox lambda value was 0.3681. As this value is close to 0.5, a square root transformation was applied to stabilise the variance of the series. The ADF test was applied to the transformed training series and yielded a test statistic of -1.7786 ($p = 0.6694$), indicating that the series was non-stationary. After first differencing, the ADF statistic was -5.5273 ($p < 0.01$), confirming that the differenced series was stationary and appropriate for ARIMA modelling. Figure 4 presents the ACF and PACF plots of the transformed non-stationary training series. The ACF plot on the left shows a slow decay in autocorrelation values, indicating that the dataset is non-stationary. Meanwhile, the PACF plot on the right displays a significant spike at lag 1, with most subsequent lags falling within the significance bands.

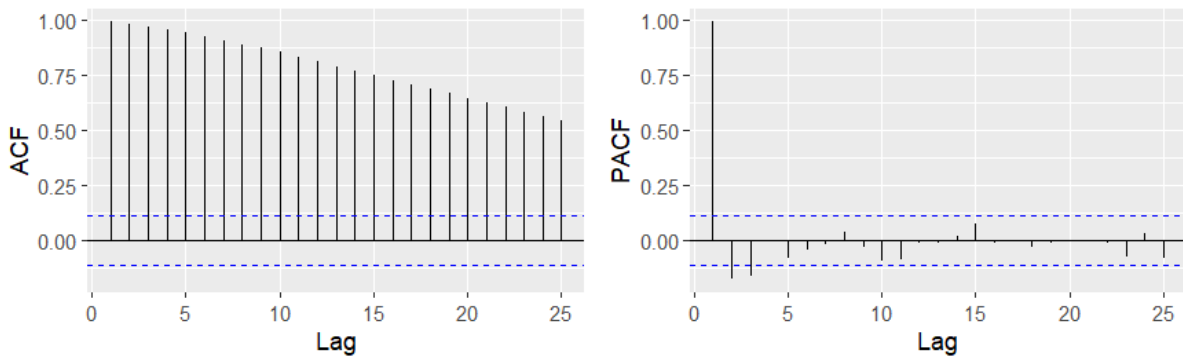


Figure 4. ACF (left) and PACF (right) for the non-stationary training dataset

First-order differencing was applied to stabilise the mean of the RON97 fuel price series by removing the trend present in the data. ARIMA model identification was then carried out using the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. Figure 5 shows the ACF and PACF plots after first differencing. The ACF displays significant spikes up to lag 2 followed by a gradual decay, suggesting a moving average component of order 2. The PACF shows significant spikes at lags 1 and 2, with the remaining lags within the confidence limits, indicating a possible autoregressive component of order 2. Based on these patterns, several ARIMA model specifications were identified for further evaluation.

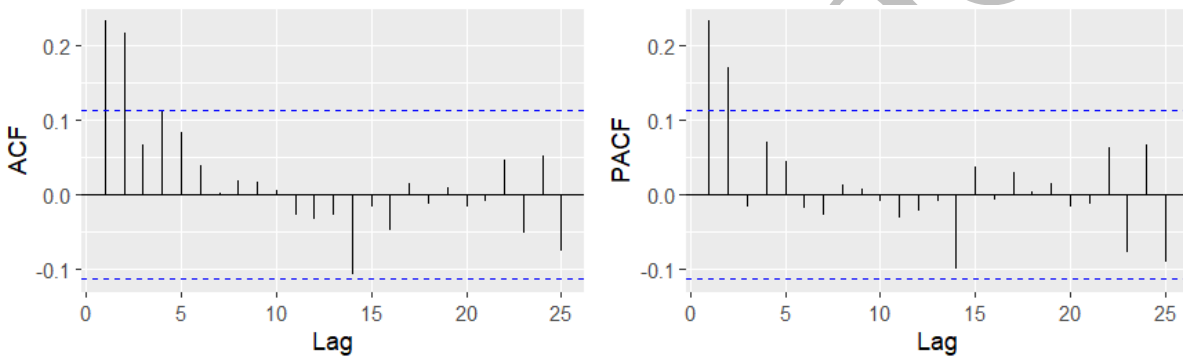


Figure 5. ACF and PACF plot training dataset after first differencing

The most suitable model was determined based on the forecasting accuracy and model selection criteria, namely the Mean Absolute Percentage Error (MAPE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). In general, lower AIC and BIC values indicate a better-fitting and more parsimonious model, as these criteria evaluate model adequacy while penalizing unnecessary model complexity [12]. As shown in Table 1, the ARIMA(2,1,0) model recorded the lowest AIC (-1124.80) and BIC (-1113.66) values among the candidate models, suggesting that it provided the best overall fit to the data. In terms of forecasting performance, both ARIMA(2,1,0) and ARIMA(2,1,2) achieved the lowest MAPE value of 0.13% while ARIMA(1,1,1) and ARIMA(0,1,2) produced slightly higher MAPE values of 0.14%. Although the difference in forecasting accuracy was relatively small, the ARIMA(2,1,2) model contained several parameters that were not statistically significant, indicating that the model may be more complex than necessary. Therefore, ARIMA(2,1,0) was selected as the preferred forecasting model as it combined high accuracy with lower information criterion values and a more parsimonious model structure. The summary of parameter estimation and diagnostic checking results for the ARIMA model is presented in Table 1.

Table 1. Summary of parameter estimation and diagnostic checking of ARIMA models

Box-Jenkins Model	Parameter Estimation		AIC	BIC	<i>p</i> -value of Ljung-Box Statistics	MAPE in The Testing Set
	Coefficient	<i>p</i> -value				
ARIMA (2,1,0)	$\phi_1 = 0.1928$	0.0001	-1124.80	-1113.66	0.9408	0.13%
	$\phi_2 = 0.1714$	0.0007				
ARIMA (1,1,1)	$\phi_1 = 0.7160$	0.0000	-1123.49	-1112.35	0.8634	0.14%
	$\theta_1 = -0.5133$	0.0000				
ARIMA (0,1,2)	$\theta_1 = 0.2009$	0.0000	-1123.36	-1112.22	0.7343	0.14%
	$\theta_2 = 0.1919$	0.0000				
ARIMA (2,1,2)	$\phi_1 = 0.8597$	0.8466	-1121.60	-1103.06	0.8748	0.13%
	$\phi_2 = 0.1047$	0.0683				
	$\theta_1 = 0.4198$	0.3725				
	$\theta_2 = 0.4248$	0.3631				

Table 1 shows that both ARIMA(2,1,0) and ARIMA(1,1,1) models have statistically significant estimated parameters, as all p -values are below the 0.05 significance level. Furthermore, the Ljung–Box test p -values exceed 0.05, indicating that residual autocorrelation is not statistically significant and that both models are adequate. However, ARIMA(2,1,0) achieved lower AIC, BIC, and MAPE values than ARIMA(1,1,1), suggesting superior model fit, parsimony, and forecasting accuracy. To further assess predictive performance, the forecasted values from the testing period were compared with the corresponding observed values, as illustrated in Figure 6.

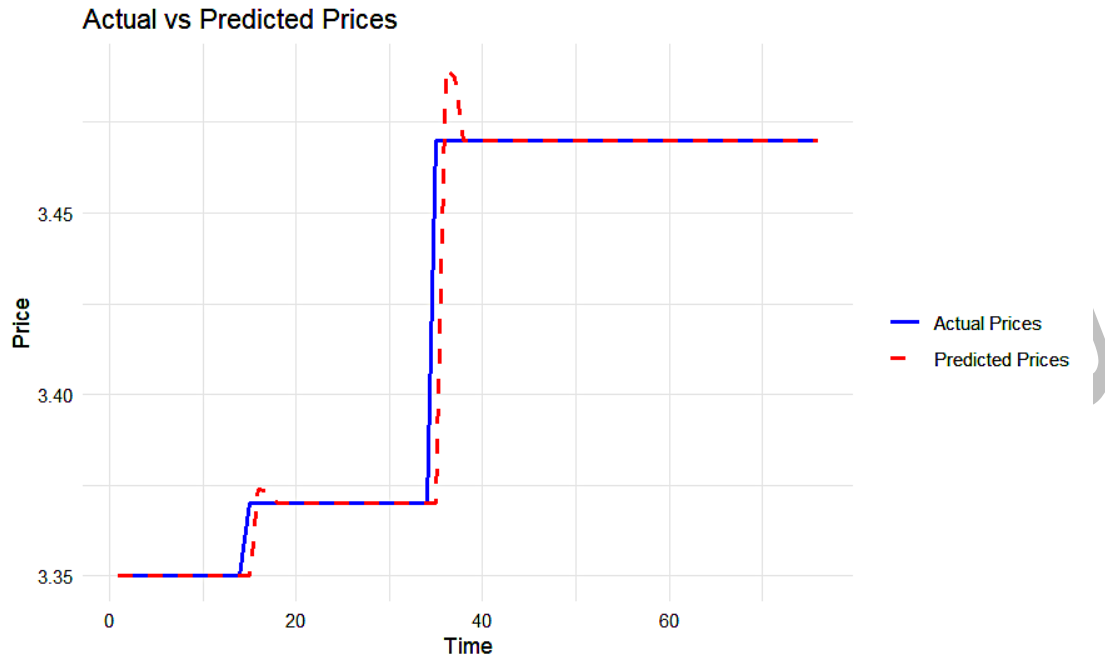


Figure 6. Comparison of actual vs predicted RON97 fuel prices in Malaysia

Figure 6 shows that the predicted RON97 fuel prices closely followed the actual prices, throughout the study period, indicating that the model performed well in capturing the overall movement of the data. Although slight delays can be observed during periods of price changes, the model was still able to reflect the general trend and short-term stability of the fuel prices. Based on the MAPE values in Table 1, the ARIMA(2,1,0) model achieved the lowest MAPE of 0.13%, making it the most suitable model for forecasting RON97 fuel prices. The ARIMA(2,1,0) model can be expressed as:

$$y_t = 1.1928y_{t-1} - 0.0214y_{t-2} - 0.1714y_{t-3} + a_t \quad (2)$$

where a_t represents the random error term. The coefficient values were obtained from the estimated ARIMA(2,1,0) model using the R statistical software. After selecting the most suitable forecasting model, diagnostic checking was carried out to evaluate the adequacy of the residuals. Figure 7 presents the diagnostic plots for the residuals of the ARIMA(2,1,0) model.

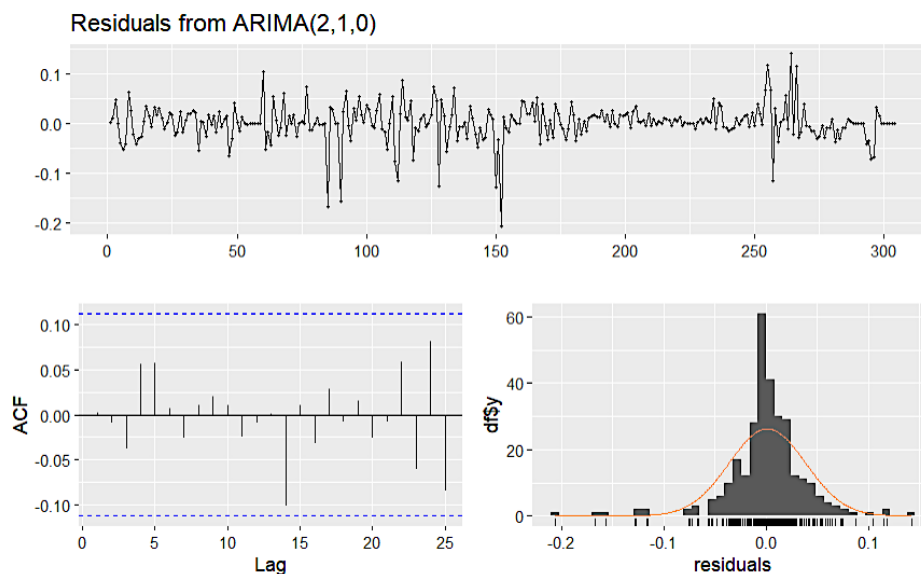


Figure 7. Diagnostic plots for the residuals from the ARIMA(2,1,0) model

The residual time series plot shows the residuals fluctuating randomly around zero, indicating no clear patterns or trends. The ACF plot and the histogram confirm the model's adequacy by showing no significant autocorrelations and a roughly normal distribution of the residuals, respectively. Since the model is adequate for time series forecasting, the next stage involved forecasting future RON97 fuel prices. In this study, a four-step ahead forecast after 25 July 2024 was generated using the ARIMA(2,1,0) model. Figure 8 illustrates the four-step ahead forecast results for RON97 fuel prices. The forecasted prices for the next four weeks were estimated to be approximately RM3.47, suggesting a relatively stable price trend throughout the forecast period.

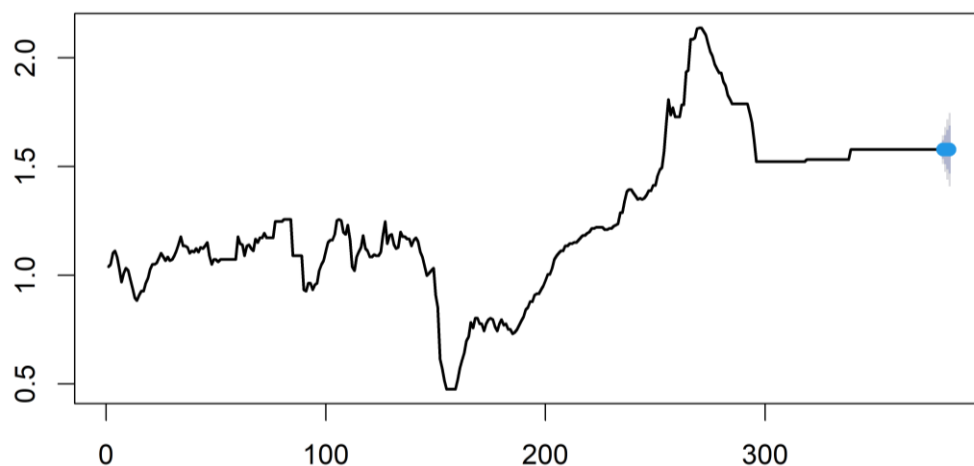


Figure 8. historical fuel prices and forecasted value plots for the ARIMA(2,1,0) model

Based on Figure 8, the solid line represents historical fuel prices, while the blue point and the shaded area represent the forecasted value and its corresponding confidence interval, respectively. The narrow confidence intervals suggest a relatively high level of confidence in the model's forecasts.

4. Conclusions

The findings showed that the ARIMA(2,1,0) model produced better forecasting performance than the ARIMA(1,1,0) model, as indicated by the lower MAPE value of 0.13%. Therefore, ARIMA(2,1,0) was selected as the most suitable model for forecasting weekly RON97 fuel prices in Malaysia. Since RON97 fuel prices in Malaysia are revised weekly, a four-week forecasting horizon was considered appropriate for short-term monitoring and planning. The forecasting results suggested that RON97 fuel prices would remain relatively stable throughout the forecast period. These findings may help policymakers, businesses, and consumers better understand short-term fuel price movements and support planning and decision-making activities. Overall, the study demonstrates that the Box-Jenkins approach can be effectively applied to forecasting fuel prices. Further studies may explore the incorporation of external factors or alternative forecasting approaches to improve accuracy.

Acknowledgement

The authors gratefully acknowledge Universiti Islam Antarabangsa Malaysia and Universiti Malaysia Pahang Al-Sultan Abdullah for their support of this work. The authors also express their sincere gratitude to the anonymous reviewers for their valuable comments and constructive suggestions.

Funding

This research was funded by Universiti Malaysia Pahang Al-Sultan Abdullah under the Geran Penyelidikan Kerjasama Strategik MTUN, grant number RDU233301.

Declaration of Competing Interest

The authors declare no conflict of interest.

CRedit Authorship Contribution Statement

Suhaila Bahrom (Conceptualisation; Methodology; Visualisation; Data curation; Writing - original draft)
Noratikah Abu (Formal Analysis, Funding acquisition; Project administration; Writing - review & editing; Supervision)

Availability of the Data and Materials

The data used in this study were obtained from Malaysia's Official Open Data Portal (data.gov.my), which is publicly accessible online.

Ethical Declarations

No artificial intelligence tools were used in the preparation of this manuscript. All content was developed manually by the authors. This study did not involve human participants or animals. Ethical approval was therefore not required.

Generative Artificial Intelligence Declarations

The authors claim that artificially intelligent-assisted technologies, such as generative AI, were not used to generate content, ideas, or theories. We have just utilised AI to enhance readability and refine the language. We used it under strict human control and oversight. The authors take full responsibility for reviewing and approving the content.

References

- [1] Yunus SM, Hodin NS, Zulkifli FH, Mustaffa N, Osman SA. The comparative analysis between gasolines RON95, RON97 and RON100 on engine performance. *Fuel, Mixture Formation and Combustion Process* 2019;1.
- [2] Rohani MM, Pahazri N. Survey on how fluctuating petrol prices are affecting Malaysian large city dwellers in changing their trip patterns. In *IOP Conference Series: Earth and Environmental Science* 2018 Apr 1 (Vol. 140, No. 1, p. 012085). IOP Publishing. <https://doi.org/10.1088/1755-1315/140/1/012085>.
- [3] Shaari MS, Yusuf AH, Abashah A, Pei TL. Can Retail Selling Prices of Petrol and Diesel Trigger Inflation in Malaysia? *MATEC Web of Conferences* 2018;150:05070. <https://doi.org/10.1051/mateconf/201815005070>.
- [4] Mohamad TI, Geok HH. Part-load performance and emissions of a spark ignition engine fueled with RON95 and RON97 gasoline: Technical viewpoint on Malaysia's fuel price debate. *Energy Convers Manag* 2014;88:928–35. <https://doi.org/10.1016/j.enconman.2014.09.008>.
- [5] Sarpong-Streeter RMNY, Sokkalingam R, Othman M, Daud H, Owusu DA. ARIMAX Modelling of Ron97 Price with Crude Oil Price as an Exogenous Variable in Malaysian. *Springer Proceedings in Complexity*, Springer Science and Business Media B.V.; 2021, p. 679–91. https://doi.org/10.1007/978-981-16-4513-6_59.
- [6] Sokkalingam R, Sarpong-Streeter RM, Othman M, Daud H, Owusu DA. Forecasting petroleum fuel price in malaysia by arima model. In *Proceedings of the 6th International Conference on Fundamental and Applied Sciences: ICFAS 2020* 2022 Jan 5 (pp. 671-678). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-4513-6_58.
- [7] He XJ. Forecasting gasoline price with time series models. *Communications of the IIMA*. 2023;21(1):1. <https://doi.org/10.58729/1941-6687.1440>
- [8] Agyare S, Odoi B, Wiah EN. Predicting petrol and diesel prices in Ghana, a comparison of ARIMA and SARIMA Models. *Asian Journal of Economics, Business and Accounting* 2024;24:594–608. <https://doi.org/10.9734/ajeba/2024/v24i51333>.
- [9] Abu N, Tukimat NN, Abu N. Forecasting of hydropower production using Box-Jenkins model at Tasik Kenyir, Terengganu. In *AIP Conference Proceedings* 2024 Mar 7 (Vol. 2895, No. 1, p. 050005). AIP Publishing LLC. <https://doi.org/10.1063/5.0193157>.
- [10] Montgomery DC, Jennings CL, Kulahci M. *Introduction to time series analysis and forecasting*. John Wiley & Sons; 2015. <https://doi.org/10.1007/978-3-319-29854-2>
- [11] Box GE, Jenkins GM. Some recent advances in forecasting and control. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 1968 Jan 1;17(2):91-109.
- [12] Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;19:716–23. <https://doi.org/10.1109/TAC.1974.1100705>.