ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 6 No. 1 (2026) 566-571
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Forecasting Natural Gas Prices in Malaysia Using Double Exponential Smoothing, Arima and XGBoost Models

Kang Yuan Chin[1], Kamil Khalid[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600, Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: kamil@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2026.06.01.062

Article Info

Received: 4 January 2026
Accepted: 6 June 2026
Available online: 9 July 2026

Keywords

Natural Gas Price Forecasting, ARIMA, Double Exponential Smoothing, XGBoost, Time series Analysis, Malaysia

Abstract

Natural gas is an integral component of Malaysia's energy structure, accounting for a significant share of overall primary energy consumption, which makes the market prone to price volatility due to the influence of changes in policies, demand, and global factors. As reliable natural gas price forecasts play a pivotal role in making informed energy policies and other related decisions, this research work examines and compares the ability and performance of three models: Double Exponential Smoothing (DES), Autoregressive Integrated Moving Average (ARIMA), and eXtreme Gradient Boosting (XGBoost) in making accurate forecasts of the prices of natural gas in Malaysia using historical price time series data on the basis of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) on the training and testing datasets. Findings indicate the poor performance of the DES model in terms of accuracy in conditions of price volatility, while XGBoost yields an impressive performance in terms of accuracy on the training dataset but performs poorly on the testing dataset; hence, the ARIMA(2,1,0) model is shown to provide the lowest forecast errors on the testing dataset, making it the model of superior performance on the test datasets in terms of precision and accuracy. The forecasts of the selected model indicate a stabilisation of prices towards the end of the forecast time period, ranging between the values of RM 37-38/MMBtu.

1. Introduction

Natural gas plays a crucial role in Malaysia's energy mix, supporting electricity generation, industrial production, and economic growth. However, natural gas prices are subject to significant volatility due to factors such as global supply–demand imbalances, geopolitical tensions, and fluctuations in energy markets. Accurate forecasting of natural gas prices is therefore essential for effective energy planning, cost management, and policy decision-making.

Accurate forecasting of natural gas prices has been widely recognised as a critical component of effective energy planning and economic decision-making. Time series–based forecasting approaches are commonly applied in energy economics research to model price dynamics and volatility, particularly for commodities such as natural gas [1], [2]. More recently, machine learning techniques have been introduced to enhance forecasting performance by capturing non-linear patterns in energy price data [3].

This study aims to address this gap by comparing the forecasting performance of Double Exponential Smoothing (DES), ARIMA, and XGBoost models using Malaysian natural gas price data. The specific objectives are to compare the forecast performance between Double Exponential Smoothing, ARIMA, and XGBoost by using MAPE, RMSE, MAE and, to forecast the natural gas pricing in Malaysia by using the most accurate model.

2. Problem Statement

The energy structure of Malaysia is quite vulnerable to gas as it currently contributes more than 40% of the total primary energy intake of the nation. This has made the energy market extremely volatile due to geopolitical issues, post-pandemic demand revival, and the transition to market-oriented pricing systems. Traditional models in energy price forecasting, such as the Double Exponential Smoothing (DES) approach and the Autoregressive Integrated Moving Average (ARIMA) model, are currently being used; however, it has been seen in some latest researches that they are not effective in energy market predictions, taking into consideration the non-linear behaviour and external shocks in the volatile energy market. The ARIMA(1,1,0) model produced a mean absolute percentage error of 14.23%, in the prediction of oil prices for Malaysia, outperformed by the artificial neural network models [4]. Similarly, models based on DES have shown relatively higher values of energy forecasting errors in petroleum-based models, with Holt's DES model showing a mean absolute percentage error of 26.1%, diminishing in efficacy under volatile market conditions [5].

Recent studies have illustrated the capabilities of machine learning approaches and especially the use of eXtreme Gradient Boosting (XGBoost) in forecasting the cost of energy [6]. Although its effectiveness has been observed in similar commodity markets, there has been little research on using XGBoost for natural gas price forecasting in Malaysia, taking into consideration regional factors like subsidy policies and domestic consumption [7]. Furthermore, none of these models (DES, ARIMA, or XGBoost) has been compared in a similar Malaysian natural gas market in any existing literature, creating a need for this study to determine an effective forecasting model for natural gas in Malaysia.

3. Materials and Methods

In this study, secondary time series data on the price of gas in Malaysia has been used to build and evaluate different models for making gas price forecasts. The data has price measurements over time, which vary accordingly due to market forces. Before embarking on the analysis, an examination of the data was carried out to determine its completeness for conducting a time series approach. It was necessary to carry out the first level of analysis, which would identify the type of trend that is present in the natural gas price series. Additionally, differencing was performed for the purpose of meeting the stationary criterion needed for forecasting, following time series forecasting methodology [1]. In order to check the accuracy of the forecasting algorithms, the original dataset was split into training and testing examples. Three forecasting methods were applied and they were Double Exponential Smoothing, ARIMA, and XGBoost. Model performance was assessed using MAPE, MAE, and RMSE to ensure accuracy and reliability across techniques.

3.1 Double Exponential Smoothing

The double exponential smoothing process proposed by Holt was applied in the representation of the level and trend components of the natural gas price series. The method is popular in the analysis of time series when the trend exists but the seasonality is not present [8]. The parameters of level and trend were determined using the training data in an effort to reduce the forecasting errors.

Equation 1 presents the forecast equation while Equation 2 and Equation 3 show the level and trend equation respectively.

$$\boldsymbol{\hat{y}_{T+P}(T) = \ell_T + pb_T} \tag{1}$$

$$\boldsymbol{\ell_T = \alpha y_T + (1-\alpha)(\ell_{T-1} + b_{T-1})} \tag{2}$$

$$\boldsymbol{b_T = \beta(\ell_T - \ell_{T-1}) + (1-\beta)b_{T-1}} \tag{3}$$

where, ℓ_T is the estimate of the level of the series at time T, b_T is the estimate of the trend of the series at time T, α is the smoothing constant for the level ($0 \leq \alpha \leq 1$), and $\boldsymbol{\beta}$ is the smoothing constant for the trend ($0 \leq \beta \leq 1$).

3.2 ARIMA

The ARIMA model was developed following the Box & Jenkins approach, which involved model specification, estimation, and checking [1]. Differencing was done to ensure stationarity, while the AR and MA orders were identified using the Autocorrelation Function and Partial Autocorrelation Function. Finally, the ARIMA mode was fitted to the dataset, and the residuals checked to ensure that the ARIMA mode assumption had been met.

Table 1 shows the identification of ARIMA model while Equation 4 presents the equation of non-seasonal ARIMA model [2].

Table 1 Identification of ARIMA model

ACF	PACF	Model
Decay to zero with exponential pattern	Cuts off after lag *p*	AR (*p*)
Cuts off after lag *q*	Decay to zero with exponential pattern	MA (*q*)
Decay to zero with exponential pattern	Decay to zero with exponential pattern	ARIMA (*p, q*)
Cuts off after lag *q*	Cuts off after lag *p*	AR (*p*) or MA (*q*)

$$\boldsymbol{\phi_p(B)\nabla^d y_t = \theta_q(B)\alpha_t} \tag{4}$$

where, $\phi_p(B)$ is the polynomials in B of order p, $\theta_q(B)$ is the polynomials in B of order q, B is the backshift operator, ∇^d is the differencing order, and α_t is the random errors.

3.3 XGBoost

A tree-based ensemble learning method, XGBoost, was used to explore the nonlinear relationship available in the natural gas price time series. The training process was conducted using the previous values of the natural gas price as input variables. The use of XGBoost has proven successful in making predictions related to energy prices due to its ability to exploit complex interactions, as well as improving the performance of a model via gradient boosting techniques [3]. Hyperparameters of the model were also optimized using the training set.

Equation 5 presents the mathematical model of XGBoost model [3].

$$\boldsymbol{\hat{y}_i = \phi(x_i) = \textstyle\sum_{k=1}^{K} f_k\,(x_i)} \tag{5}$$

where, $\hat{y}_i$ *is the* final predicted value for the i^{th} data point, K is the number of trees in the ensemble, and $f_k(x_i)$ is the prediction of the K^{th} tree for the i^{th} data point.

The objective function in XGBoost consists of two parts: a loss function and a regularization term. The loss function measures how well the model fits the data, and the regularization term simplifies complex trees. Equation 6 presents the general form of the loss function [3].

$$\mathcal{L}(\theta) = \textstyle\sum_i \ell(\hat{y}_i, y_i) + \sum_k \Omega f_k \tag{6}$$

where, ℓ is the differentiable convex loss function that measures the difference between the prediction, $\hat{y}_i$ and the target y_i, Ωf_k is the regularization term which discourages overly complex trees, T is the number of trees in the tree, γ is the regularization parameter that controls the complexity of the tree, λ is the parameter that penalizes the squared weight of the leaves, and w is the weight of the leaves.

4. Results and Discussion

Table 2 Overall Comparison of Different Forecasting Methods on Training and Testing Set

Forecasting Method	Training Set			Testing Set		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Double Exponential Smoothing	1.85638	6.01017	2.64285	2.386	5.709	2.7203
ARIMA (2,1,0)	0.1624	2.902	0.2126	0.1453	2.203	0.17776
XGBoost	0.028418	0.079194	0.037940	1.352485	3.236701	1.773731

According to Table 2, it compares the results of the Double Exponential Smoothing method, ARIMA(2,1,0), and XGBoost algorithms regarding the MAE, MAPE, and RMSE metrics for both the training and testing data. There are differences among the results obtained by the different algorithms.

DES has the maximum error values for all metrics on both the train and test datasets, and this indicates low capability of modelling the dynamics of the Malaysian natural gas prices. This result is supported by previous evidence that exponential smoothing techniques can be inefficient when modelling energy markets characterized by high variability.

The ARIMA(2,1,0) model performs well and consistently with lower values of error on both training and test datasets. To be precise, on the test dataset, ARIMA gives the lowest MAE of 0.1453, MAPE of 2.203%, and RMSE of 0.1778 compared to other models and algorithms. This indicates that this method is capable of identifying linear patterns present in natural gas prices effectively while retaining a good level of generality.

On the other hand, the XGBoost algorithm displays a remarkably small error margin regarding forecasting on the training data, denoting its excellent aptness for the historical series. Notwithstanding its capability to handle non-linear modelling tasks, its error on the testing series escalates remarkably compared to ARIMA models, portending possible cases of overfitting.

In general, the results show that although machine learning algorithms like XGBoost can produce very good in-sample results, more accurate and stable out-of-sample results for the price of natural gas in Malaysia might be obtained by the traditional statistical approach of ARIMA.

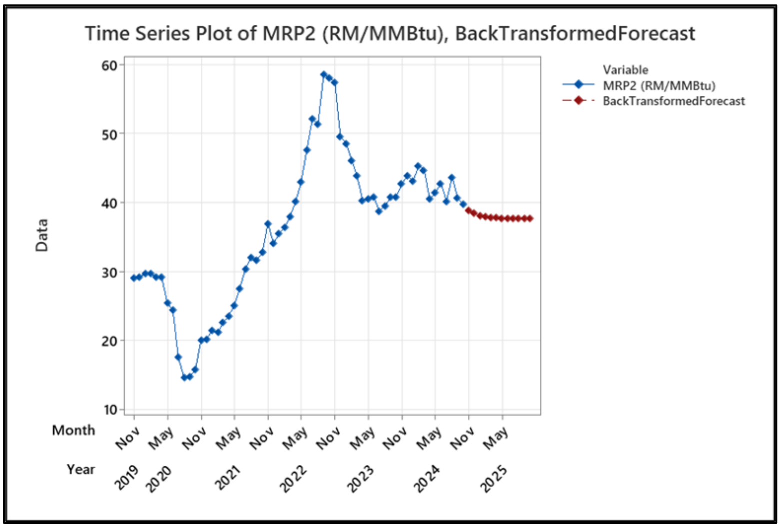

Fig. 1 Time Series Plot of Future Forecasted Historical MRP of Natural Gas in Malaysia

Based on Figure 1, it shows the historical Malaysian natural gas price series (MRP2, RM/MMBtu) and the forecasted values obtained using the selected ARIMA(2,1,0) model. Looking at the historical values, there is a large degree of volatility, including a sharp fall in 2020, a sharp increase in 2021-2022, and then a steady stabilization of prices in the later periods.

The result of the forecast indicates a stable price trend over the forecast period. There is convergence of the projected values within a small range compared to the previous periods of high variability. The trend indicates that the ARIMA model is able to capture the structure of variability of the series and dampen short-run variability when making projections. The lack of large changes in both directions of the forecasted values indicates low variability of the prices within the short run, assuming that no major shocks have occurred.

In broad terms, the forecast appears realistic regarding future price movements, being in keeping with recent historical data rather than making unrealistic projections. Taken together with its strong performance on out-of-sample data, as is clear from the model evaluation study, the ARIMA(2,1,0) model is considered appropriate for short-term forecasting on Malaysian natural gas prices.

5. Conclusion

The performance of these models in terms of their ability to forecast Malaysian natural gas prices is evaluated in this study. Based on the comparison results using MAE, RMSE, and MAPE errors, the model ARIMA(2,1,0) was found to possess the most consistent and reliable performance for out-of-sample forecasts. Although the Double Exponential Smoothing model was able to identify the overall pattern in the prices in the form of a time series, its performance was compromised for conditions associated with higher volatility. The error in XGBoost for the training model was remarkably low, but its generalization performance for the testing model was compromised.

Applying the chosen ARIMA(2,1,0) model, the forecasted natural gas prices in the future are expected to converge at the end of the forecast cycle, stabilizing at around RM 37-38/MMBtu. By doing so, this forecasted trend reflects past observed behaviours, indicating the expected natural gas price observations to be less volatile in the absence of any extraordinary events being witnessed in the present and past periods. Furthermore, it can be concluded that ARIMA(2,1,0) is an appropriate and reliable model for short-term forecasting of natural gas prices in Malaysia.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Kang Yuan Chin;* ***data collection:*** *Kang Yuan Chin;* ***analysis and interpretation of results:*** *Kang Yuan Chin, Kamil Khalid;* ***draft manuscript preparation:*** *Kang Yuan Chin, Kamil Khalid. All authors reviewed the results and approved the final version of the manuscript.*

Appendix A: Overall Comparison of Different Hyperparameter Tuning of XGBoost Model

Model	Train MAE	Test MAE	Train RMSE	Test RMSE	Train MAPE	Test MAPE
GridSearch	0.007882	1.394674	0.011095	1.922559	0.021009	3.360092
Optuna	0.028418	1.352485	0.037940	1.773731	0.079194	3.236701
Bayesian	0.000391	1.632182	0.000573	2.144843	0.001053	3.931954

Appendix B: Future Forecasted Values from Period 60 (12 months ahead)

Period	Forecast	95% Lower Limit	95% Upper Limit
61	38.8300	33.8205	44.1854
62	38.4334	30.8396	46.8622
63	38.1088	27.8513	49.9713
64	37.9425	25.4408	52.9344
65	37.8248	23.3190	55.8218
66	37.7583	21.4977	58.5683
67	37.7143	19.8930	61.1874
68	37.6883	18.4735	63.6817
69	37.6716	17.2017	66.0637
70	37.6615	16.0537	68.3446
71	37.6552	15.0093	70.5361

72	37.6513	14.0531	72.6480

References

[1] Wilson, G. T. (2016). Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1. *Journal of Time Series Analysis*, *37*(5), 709–711. https://doi.org/10.1111/jtsa.12194

[2] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice. Otexts.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System, 785–794. https://doi.org/10.1145/2939672.2939785

[4] Zamri, N. F., & Nor, M. E. (2024). Forecasting oil price in Malaysia using ARIMA and artificial neural network. Enhanced Knowledge in Sciences and Technology, 4(2), 273–282. https://doi.org/10.30880/ekst.2024.04.02.031

[5] Sam, N. S., & Kamarudin A.N. (2023). Forecasting petroleum fuel price in Malaysia by ARIMA model. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 7(2), 256–263.

[6] Tissaoui, K., Zaghdoudi, T., Hakimi, A., & Nsaibi, M. (2022). Do gas price and uncertainty indices forecast crude oil prices? Fresh evidence through XGBOOST modeling. Computational Economics, 62(2), 663–687. https://doi.org/10.1007/s10614-022-10305-y

[7] Tachev, V. (2025, February 17). Natural gas Price Forecast 2025: Asia to drive global demand. Energy Tracker Asia. https://energytracker.asia/natural-gas-price-forecast-2025/

[8] Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015