

ARIMAX Modelling of Ron97 Price with Crude Oil Price as an Exogenous Variable in Malaysian



Richard M. N. Y. Sarpong-Streeter , Rajalingam Sokkalingam ,
Mahmod Othman , Hanita Daud , and Derrick Asamoah Owusu 

Abstract Malaysia's petroleum fuel pricing policy has shuffled between Automatic Adjusted Formula (APM) and Managed Float System (MFS). One major problem in forecasting fuel prices is the use of the structural model. The model uses an input or independent variable, the refined fuel price also known as Mean of Platts Singapore (MOPS). MOPS is published regularly but is sold, which makes it inaccessible to a section of forecasters in Malaysia. In this paper, we suggest an alternative method of forecasting Ron97 price by applying the time series method, Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) to model and forecast Ron97 price in Malaysia using publicly available data.

Keywords Automatic price mechanism (APM) · ARIMAX · ARIMA · Forecast · Fuel · Price · OPEC · Ron97

R. M. N. Y. Sarpong-Streeter (✉) · R. Sokkalingam · M. Othman · H. Daud
Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, 32610 Seri
Iskandar, Perak, Malaysia

e-mail: richard_17004029@utp.edu.my

R. Sokkalingam

e-mail: raja.sokkalingahm@utp.edu.my

M. Othman

e-mail: mahmod.othman@utp.edu.my

H. Daud

e-mail: hanitadaud@utp.edu.my

D. A. Owusu

Mathematics Department, Kwame Nkrumah University of Science and Technology, Kumasi,
Ghana

1 Introduction

Over the years the retail price of petroleum fuel in Malaysia, Ron95, Ron97 and Diesel have been controlled by the governments using the Automatic Price Mechanism (APM) which made the price of fuel in Malaysia relatively stable up until 2004. Beyond the year 2004, the price of petroleum fuel has been volatile even with APM, still being implemented. After changing the Policy to Managed float system in 2016 fuel price have still not been stable. Reasons that have been attributed to the volatilities are the international crude oil price and foreign exchange volatilities and reduction of subsidies to improve government fiscal space [1].

The unstable nature of the fuel price demands the need for forecasting of the fuel price. Modelling and forecasting fuel price for a section of forecasters have become difficult. This because the APM model, Eq. (1), the popular model used for fuel price forecasting uses an input variable MOPS, (A), which is published regularly but sold to forecasters, making it difficult to access. Equation (1), have the inputs (A) as the Refined fuel price, Mean of Platts Singapore (MOPS), published by Platts [2], (B) Alpha, the difference between the MOPS and actual purchasing price from the refinery's companies, (C) Tax/Subsidy [3, 4], (D) Operational cost at bulk storage for transportation and advertisement, (E) Bulk distribution company's margin and (F) Fuel station [5, 6].

$$P = A + B + C + D + E + F \quad (1)$$

In this paper, we apply the time series method Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) to model and forecast the petroleum fuel price for Ron97 using available public data.

2 Data

The primary data used in the paper are the published weekly price of Ron97 [7–9], daily crude oil price in barrels (WTI, BRENT and OPEC) [10, 11] and daily foreign exchange rate (Selling rate) of the Ringgit per US dollars [12]. The data is pre-processed to clean the data and standardize the data for the modelling process. The crude oil price and foreign exchange rates are converted to weekly averages. Daily missing data are replaced with weekly averages. The foreign exchange rates, (G) in Ringgits per US dollar (RM/USD) [12], is multiplied with the international crude oil price (X) to convert ringgits per barrel (RM/blur), (I), using Eq. (2). The crude oil price are converted to ringgit per litre using the Barrel to Litre Metric Conversion (M), [13], as shown in Eq. (3). The compiled data can be found on the doi link: <https://doi.org/10.17632/zxjnrpmwd8.1> [14].

$$I = X \times G \quad (2)$$

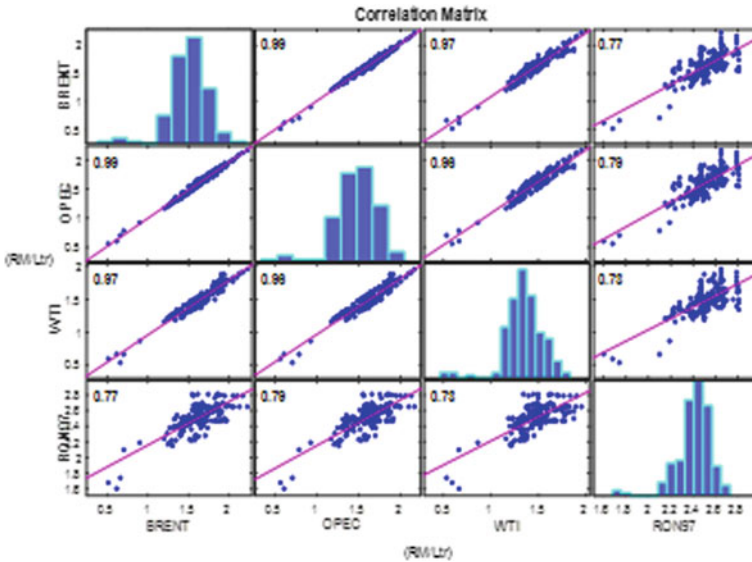


Fig. 1 Correlation matrix of Ron97 and crude oil price in Malaysia

$$Q = I \times M \tag{3}$$

Data used for the model, range from 7 April 2017 to 6 March 2020. Data used for the validation process, range from 13 March 2020 to 7 August 2020. We apply the correlation matrix to select the best crude oil price time series to model Ron97. Figure 1 is the Correlation Matrix of the three crude oil prices being assessed to model the Ron97 price. OPEC crude oil price have a better correlation of 0.79 with Ron97 compared with that of WTI and Brent crude oils. OPEC crude oil price is therefore selected as the exogenous variable to apply in modelling Ron97 price. Figure 2 are the selected Data for Ron97 price modelling.

3 Methodology

An Autoregressive Integrated Moving Average with Explanatory Variable model (ARIMAX), Eq. (4), can be viewed as a multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average (MA) terms.

$$\phi(1 + L)^p(1 + L)^d y_t = c + \sum_{i=1}^n x_i \beta_i + \theta(1 + L)^q \varepsilon_t \tag{4}$$

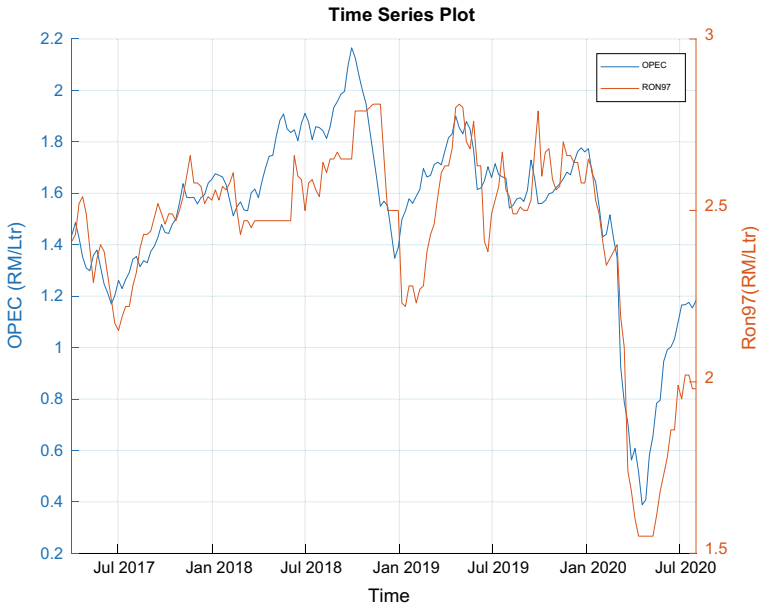


Fig. 2 Selected data for Ron97 modelling

This method is suitable for forecasting when data is stationary/nonstationary, and multivariate with any type of data pattern. ARIMAX is related to the ARIMA technique but, while ARIMA is suitable for datasets that are univariate, ARIMAX is suitable for analysis where there are additional explanatory variables (multivariate) in categorical and/or numeric format [15–18]. The ARIMAX modelling is in three stages, the Model Identification, Model estimation and Diagnostic checking [19]. The Modelling and 18-week forecast are implemented in MATLAB. The model is validated by plotting the forecast of the ARIMAX model with ARIMA, NARNET and actual data recorded over the time [20–22].

4 Results

The result has been sectioned into two parts the modelling and Evaluation of the forecast.

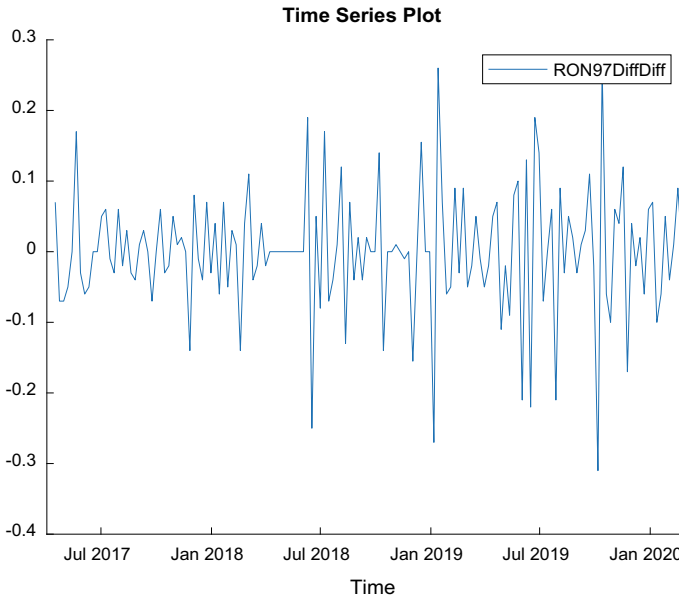


Fig. 3 Second difference of Ron97

4.1 ARIMAX Modelling

Ron97 fuel price has been modelled in a three-step procedure;

Model Identification

The precondition for the ARIMAX model to be identified is for the time series to be stationary or there exist no unit roots in the time series [23]. Thus for multiple time series they should be cointegrated [24]. We consider two Time Series, Ron97 and OPEC time series for the ARIMAX modelling. A suitable ARIMA model is identified for the dependent variable, Ron97, then the best regression is done on the ARIMA model and the independent variable, OPEC (Fig. 3). The Sample autocorrelation (SAC) of Ron97 dies down sharply at lag 1 on the second differencing as shown in Fig. 4; thus, stationarity is at difference order of 2 and moving average order is 1. The sample partial autocorrelation (SPAC) at the second differencing, Fig. 5 does not die down, which implies the autoregression order is 0 at the second differencing. The analysis deduced from the above information implies the ARIMA model have a differencing order of 2 and the ARIMA model is set at ARIMA (0, 2, 1) for Ron97. The ARIMA model is modified to ARIMAX by introducing the exogenous variable, OPEC (Fig. 6). OPEC must not be unit root time series. Based on Phillips-Perron tests [25], OPEC at lags 0, 1 and 2, (OPEC, OPECdiff, OPECdiffDiff) with significance level of (p -value) of 0.05 do not contain unit root as shown in Table 1. We will need the ARIMA model of OPEC to forecast the exogenous value in

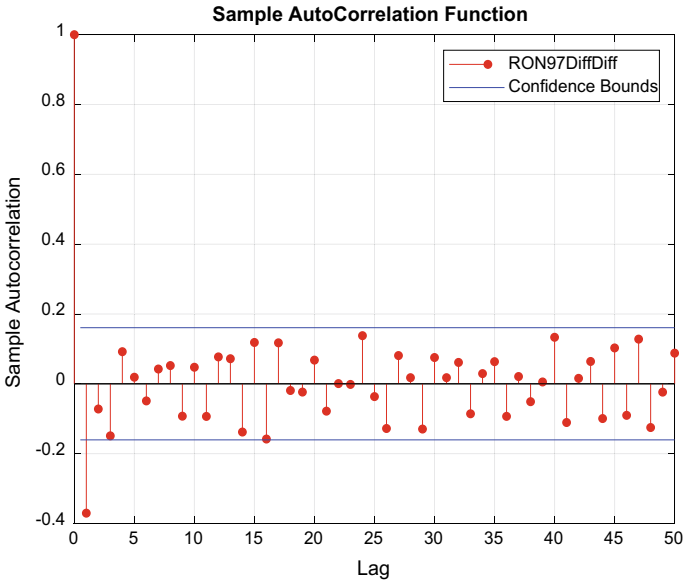


Fig. 4 Sample autocorrelation (SAC) of Ron97

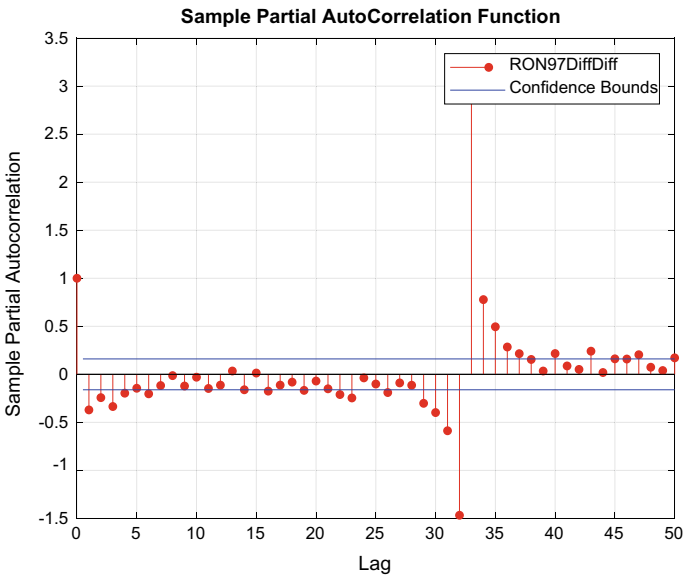


Fig. 5 Sample partial autocorrelation (SPAC) of Ron97

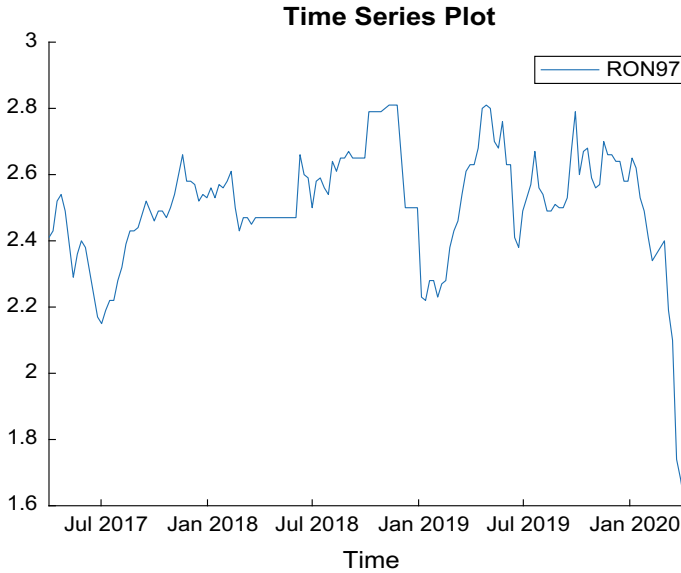


Fig. 6 Time series of OPEC

Table 1 Phillips-Perron test on OPEC time series

Null hypothesis: OPEC contains a unit root								
Test parameters					Test results			
S/N	Lags	Model	Test statistic	Sign. level	Null rejected	P-value	T-statistic	Critical value
1	0	ARD	t1	0.05	False	0.994	0.829	-2.881
2	1	ARD	t1	0.05	False	0.974	0.233	-2.881
3	2	ARD	t1	0.05	False	0.953	-0.037	-2.881

the ARIMAX model. Assessing the SAC and SPAC of OPEC as shown in Figs. 7 and 8 respectively, the SAC does not die down, but the SPAC is considered to die down at lag 2. Thus, the ARIMA (2, 0, 0) is chosen as the model for OPEC. The distributions of Ron97 and OPEC are also considered. Observing the histograms from the correlation matrix table in Fig. 1, Ron97 and OPEC have skewed tail on the left side which is characteristics of the *t*-distribution, hence the *t*-distribution is specified over the Gaussian for all ARIMA and ARIMAX models that will be estimated.

Model Estimation

MATLAB Econometric Medullar Application [26] is applied to estimate the model parameters. The Model Identification section identified the tentative model,

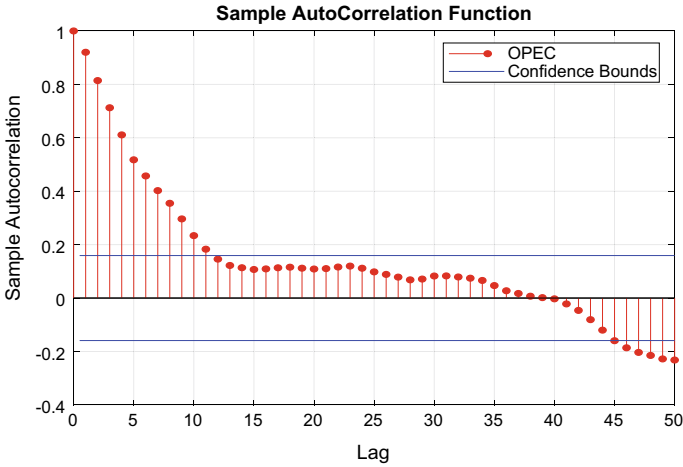


Fig. 7 Sample autocorrelation (SAC) of OPEC

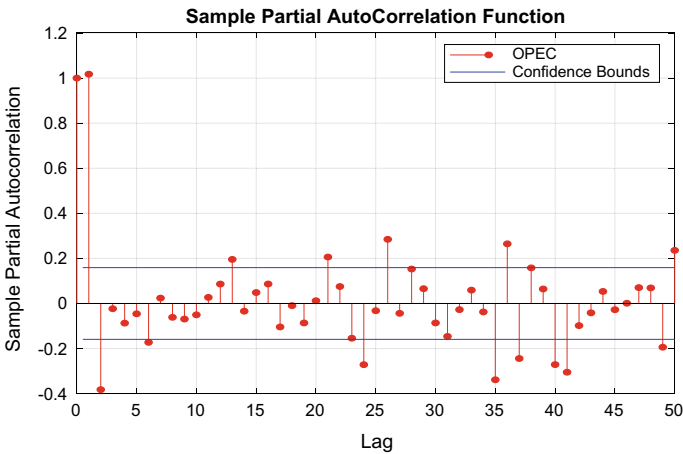


Fig. 8 Sample partial autocorrelation (SPAC) of OPEC

ARIMAX (0, 2, 1) with t -distribution and beta innovations of the exogenous predictors' time series (OPEC, OPECdiff, OPECdiffDiff). The ARIMA model associated with the ARIMAX is estimated first, then the exogenous predictors are introduced. Table 2 is the estimated parameters of ARIMA (0, 2, 1) model, Eq. (5). Equation (5) is expanded to give Eq. (6) and the estimated parameters is substituted into Eq. (6) to give Eq. (7)

$$(1 - L)^2 y_t = (1 + \theta_1 L) \varepsilon_t \tag{5}$$

Table 2 Estimation results ARIMA (0, 2, 1) (*t*-distribution) for Ron97

Parameter	Value	Standard error	T-statistic	<i>P</i> -value
Constant	0	0		
MA {1}	-0.9801	0.0126	-77.8371	0
DoF	2.031	0.5462	3.7183	0.0002
Variance	0.0910	1.5674	0.0581	0.9537

Table 3 Estimation results ARIMAX (0, 21) (*t*-distribution) for Ron97 with OPEC, OPECDiff and OPECDiffDiff as exogeneous variables

Parameter	Value	Standard error	T-statistic	<i>P</i> -value
Constant	0	0		
MA{1}	-1	0.017	-57.67	0
Beta (OPEC)	-0.00019	7.48e-05	-2.473	0.013
Beta (OPECDiff)	0.03361	0.020	1.661	0.097
Beta (OPECDiffDiff)	-0.04975	0.074	-0.668	0.504
DoF	2.42790	0.826	2.940	0.003
Variance	0.00945	0.013	0.702	0.483

$$y_t = 2y_{t-1} - y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{6}$$

$$y_t = 2y_{t-1} - y_{t-2} + \varepsilon_t - 0.9801\varepsilon_{t-1} \tag{7}$$

$$(1 - L)^2 y_t = X_1 \beta_1 + X_2 \beta_2 + X_3 \beta_3 + (1 + \theta_1 L) \varepsilon_t \tag{8}$$

Table 3 presents the tentative model estimates of the parameters of the ARIMAX which have the exogeneous time series' variables OPEC, OPECDiff and OPECDiffDiff with its associated model, is Eq. (8). We want the most parsimonious model for the ARIMAX Ron97 model. From Table 3, the *p*-value of the regression coefficients OPECDiff and OPECDiffDiff are statistically not significant at significance level of 5%. They are 9.7% and 50.4% respectively. On the other-hand, the *p*-value of OPEC is significant with a value of 1.3%. Estimating the ARIMAX model again with OPEC as the only exogenous variable Table 4, gives the *p*-value, 2.51%, Table 4.

Re-estimating the model, Eq. (8) as Eq. (9) makes the model more parsimonious. The Akaike information criterion (AIC) or the Bayesian information criterion (BIC) of the Eq. (9) is smaller than that of Eq. (8). The errors of Eq. (9) have been minimized more than that of Eq. (8), as shown. Figure 5. Hence the final model, ARIMAX (0, 2, 1) model with OPEC as the exogeneous variable is chosen as shown in Eq. (9).

$$(1 - L)^2 y_t = X_1 \beta_1 + (1 + \theta_1 L) \varepsilon_t \tag{9}$$

Table 4 Parameter estimation results, ARIMAX (0, 2, 1) (*t*-distribution) for Ron97 with OPEC as exogenous variable

Parameter	Value	Standard error	T-statistic	<i>P</i> -value
Constant	0	0		
MA{1}	-1	0.0138	-72.565	0
Beta (OPEC)	-0.00014	6.0668e-05	-2.24	0.0251
DoF	2.0983	0.6364	3.2972	0.0010
Variance	0.0313	0.1883	0.1663	0.8680

$$y_t = 2y_{t-1} - y_{t-2} + X_1\beta_1 + \varepsilon_t + \theta_1\varepsilon_{t-1} \tag{10}$$

$$y_t = 2y_{t-1} - y_{t-2} - 0.00014 X_1 + \varepsilon_t - \varepsilon_{t-1} \tag{11}$$

Note that the constant term was omitted when specifying the model. This helped achieved a more parsimonious model. Expanding Eq. (9) into Eq. (10), and substituting the parameters from Table 4 into Eq. (10), we obtain Eq. (11) (Table 5)

Lastly, we estimate the OPEC ARIMA model which is needed to update the exogenous variable X_t in the forecast horizon interval in the ARIMAX model. Table 6 is the model parameters and Eq. (12) is it associated equation. Expanding Eq. (12) and substituting the parameters from Table 6 into Eq. (13), we obtain Eq. (14), the exogenous variable simulator for the ARIMAX model.

$$(1 - \phi_1 L - \phi_2 L^2)y_t = \varepsilon_t \tag{12}$$

Table 5 Goodness of fit for ARIMAX

Model	Exogenous variable	AIC	BIC
ARIMAX (0, 2, 1)	OPEC, OPECdif, OPECdifdif	-399.98	-381.75
ARIMAX (0, 21)	OPEC	-406.45	-394.25

Table 6 Estimation results of ARIMA(2, 0, 0) (*t*-distribution) with OPEC

Parameter	Value	Standard error	T-statistic	<i>P</i> -value
Constant	0	0		
AR{1}	1.324	0.068	19.423	4.964e-84
AR{2}	-0.324	0.068	-4.785	1.706e-06
DoF	6.751	1.674	4.033	5.503e-05
Variance	0.003	0.0005	6.696	2.150e-11

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (13)$$

$$y_t = 1.324 y_{t-1} - 0.324 y_{t-2} + \varepsilon_t \quad (14)$$

In the current scenario a one step ahead forecast can be done using Eqs. (15) and updating the exogenous variable with Eq. (16). These models are deduced from Eqs. (13) and (14)

$$y_{t+1} = 2y_t - y_{t-1} - 0.00014 X_{1+1} + \varepsilon_{t+1} - \varepsilon_t \quad (15)$$

$$X_t = 1.324 X_{t-1} - 0.324 X_{t-2} + \varepsilon_t \quad (16)$$

Diagnostic Checking

After fitting the ARIMAX model we are left with residual Time series of Ron97, We assess the residual to confirm the adequacy of the ARIMAX model. We apply the residual autocorrelation plot tests and the Ljung Box Q test in the diagnostic check [27]. The sample autocorrelation function is very weak as the correlation plot does not spike. This is corroborated by the Ljung Box Q Test. The null Hypothesis is accepted, ‘The first m autocorrelations of the residuals of ARIMAX_RON97 are jointly zero ‘0’’. This implies the ARIMAX model is adequate for forecasting the fuel price in the neighbourhood of the period considered.

4.2 Model Forecast Performance and Validations

The ARIMAX was better at forecasting Ron97 for the first month than the benchmarks ARIMA and NARNET as can be seen in Fig. 9.

5 Conclusion

This paper has examined the ability of ARIMAX to model the fuel price of Ron97 using times series data of Ron97 and another exogenous time series, the crude oil price, in this case OPEC in Malaysia. It is possible to do the modelling and forecasting accurately using the ARIMAX.

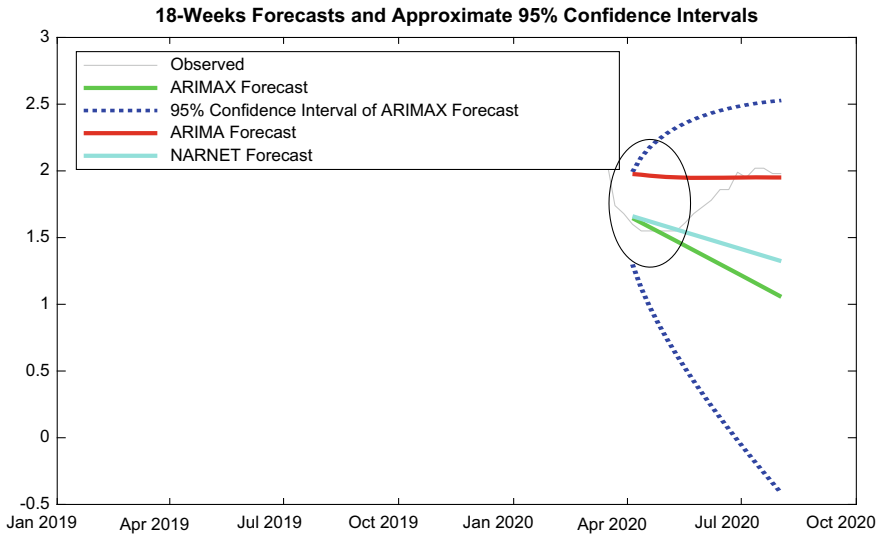


Fig. 9 Forecast performance of the ARIMAX models

Acknowledgements We would like to thank anonymous referees for many the helpful suggestions which improved the paper considerably. We appreciate the FUNDAMENTAL RESEARCH GRANT SCHEME (FRGS) Reference code FRGS/1/2018/STG06/UTP/02/4, Malaysia for sponsoring the research. Lastly, we appreciate Universiti Teknologi PETRONAS where the research was conducted. The views expressed are those of the authors and do not reflect any other authority.

References

1. Bridel, A., Lontoh, L.: Lessons Learned: Malaysia's 2013 Fuel Subsidy Reform (2014) [Online]. Available: <http://www.iisd.org/gsi>
2. S&P Global Platts: Methodology and Specifications Guide Asia Pacific and Middle East Refined Oil Products, p. 47. Platts (2017)
3. Royal Malaysian Customs Department: Guide on Goods and Service Tax on Downstream Petroleum, vol. 1, 2016 edn., p. 24 (2019)
4. Chambers, A.G.: Sales Tax Regulations, p. 86. Malaysia Sales and Service Tax' Royal, Malaysia (2018)
5. Wagner, A.: Fuel Price Policies Reading List Update 2001–2015. Dtsch. Ges. Int. Zusammenarbeit, 40 (2015) [Online]. Available: <http://www.giz.de/fuelprices>
6. Tan, P.: APM: How fuel prices are calculated in Malaysia. In: Malaysian Fuel Prices (2009). <https://paultan.org/2009/02/15/how-fuel-prices-are-calculated-in-malaysia/>
7. M. of F. Malaysia.: Retail Price of Petroleum Products (2021). <https://www.mof.gov.my/en/component/tags/tag/price-of-petroleum%0A>
8. CompareHero.my.: "Latest Petrol Price for Ron95, Ron97 & Diesel in Malaysia," Money Tips, Transportation (2020). <https://www.comparehero.my/transportation/articles/latest-petrol-price-ron95-ron97-diesel>
9. The Ministry of Domestic Trade and Consumer Affairs.: "Fuel Prices," The Ministry of Domestic Trade (2020). <https://www.kpdnhep.gov.my/ms/>

10. OPEC: OPEC basket reference price data. In: Organization of the Petroleum Exporting Countries (2020). http://www.opec.org/opec_web/en/data_graphs/40.htm
11. EIA: Spot prices for crude oil and petroleum products. In: Energy Information Administration (2020). <https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rbrte&f=w>
12. Bank Negara Malaysia: Exchange rate, ringgit to dollar, Selling rate @ 12 Noon (2020). https://www.bnm.gov.my/index.php?ch=statistic&pg=stats_spot_rates%0A
13. W. H. Ltd.: US Barrels (Oil) to Litres. In: Metric Conversions (2020). <https://www.metric-conversions.org/volume/us-oil-barrels-to-liters.html>
14. Sarpong-Streeter, R.M.N.Y.: ARIMAX Modelling of Ron97 Price with Crude Oil Price as an Exogenous Variable in Malaysian, Mendeley Data, V1 (2021). <https://doi.org/10.17632/zxjnrp mwd8.1>
15. Yang, M., Xie, J., Mao, P., Wang, C., Ye, Z.: Application of the ARIMAX model on forecasting freeway traffic flow. In: CICTP 2017 Transportation Reform and Change—Equity, Inclusiveness, Sharing, Innovation—Proceedings of the 17th COTA International Conference of Transportation Professionals, vol. 2018, pp. 593–602, July 2017. <https://doi.org/10.1061/9780784480915.061>
16. Anggraeni, W., Vinarti, R.A., Kurniawati, Y.D.: Performance comparisons between Arima and Arimax method in Moslem kids clothes demand forecasting: case study. *Proc. Comput. Sci.* **72**, 630–637 (2015). <https://doi.org/10.1016/j.procs.2015.12.172>
17. Andrews, B.H., Dean, M.D., Swain, R., Cole, C.: Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors sponsored by society of actuaries health section prepared by University of Southern Maine. *Soc. Actuar.* Aug 2013
18. Kongcharoen, C., Kruangpradit, T.: Autoregressive integrated moving average with explanatory variable (ARIMAX) model for Thailand export. In: 33rd International Symposium on Forecasting. South Korea, pp. 1–8, June 2013 [Online]. Available: [http://www.researchgate.net/publication/255731345_Autoregressive_Integrated_Moving_Average_with_Explanatory_Variable_\(ARIMAX\)_Model_for_Thailand_Export/file/9c9605209ac48013f6.pdf](http://www.researchgate.net/publication/255731345_Autoregressive_Integrated_Moving_Average_with_Explanatory_Variable_(ARIMAX)_Model_for_Thailand_Export/file/9c9605209ac48013f6.pdf)
19. Mathworks: ARIMA model including exogenous covariates. In: Mathworks Help Centre (2021). <https://ch.mathworks.com/help/econ/arima-model-including-exogenous-regressors.html>
20. Yang, Z., Mehmed, E.E.: Artificial neural networks in freight rate forecasting. *Marit. Econ. Logist.* **21**(3), 390–414 (2019). <https://doi.org/10.1057/s41278-019-00121-x>
21. Mathworks: narxnet. In: Deep Learning Toolbox Documentation (2010) <https://ch.mathworks.com/help/deeplearning/ref/narxnet.html;jsessionid=4351ba4802441e46dae0d9a3a7f1>. Accessed 30 June 2020
22. Durrani, A., Khurram, M., Khan, H.R.: Smart weather alert system for dwellers of different areas. In: Proceedings of 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST 2019), pp. 333–339 (2019). <https://doi.org/10.1109/IBCAST.2019.8667190>
23. Diebold, F.X., Kilian, L.: Unit-root tests are useful for selecting forecasting models. *J. Bus. Econ. Stat.* **18**(3), 265–273 (2014)
24. Escribano, Á., Wang, D.: Mixed random forest, cointegration, and forecasting gasoline prices. *Int. J. Forecast.* (2021). ISSN 0169-2070. <https://doi.org/10.1016/j.ijforecast.2020.12.008>. (<https://www.sciencedirect.com/science/article/pii/S0169207020301941>)
25. Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *J. Econom.* **54**(1–3), 159–178 (1992). [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
26. Manjon, J.: Econometric Modeler. MathWorks Inc., Massachusetts, USA
27. Ljung, G.M., Box, G.E.P.P.: On a measure of lack of fit in time series models. *Biometrika* **65**(2), 297–303 (1978). <https://doi.org/10.1093/biomet/65.2.297>