International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2015, 5(3), 716-724.

An Econometric Investigation of Forecasting Premium Fuel

Samuel Yeboah Asuamah[1]*, Joseph Ohene-Manu[2]

[1]Business School, Accra Institute of Technology, Accra, Ghana/Sunyani Polytechnic, Ghana, [2]Department of Economics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. *Email: nelkonsegal@yahoo.com

ABSTRACT

For a sustainable economic development, premium fuel forecasting is becoming increasingly relevant to policy makers and consumers. The current paper develops a structural econometric model of premium fuel using the autoregressive integrated moving average (ARIMA) to analyse and forecast premium demand. The results show that the ARIMA models (1, 1, 0); (0, 1, 1) and (1, 1, 1) are the appropriate identified order. The estimated models included a constant term. All the coefficients of the variables in the model except the constant term were significant. The diagnostic checking of the estimated model shows ARIMA (1, 1, 1) as the best fitted model since all the series were randomly distributed. The data for the forecast covers the period 2000:01 to 2011:12. The results indicated that the forecasted values fitted the actual consumption of the energy variables since the forecasted values insignificantly underestimate the actual consumption and thus indicate consistency of the results. The evaluation statistics indicate that the estimated models are suitable for forecasting. The model developed in the work is helpful to the energy sector and policy makers in making energy related decisions and investigating the changes in premium demand.

Keywords: Premium Fuel, Autoregressive Integrated Moving Average, Forecasting
JEL Classifications: C51, C52, C53, E17, Q47

1. INTRODUCTION

Forecasting energy demand and price has become the focus of many economists and decision makers in recent times (Ziel et al., 2015; Yeboah et al., 2012; Weiqi et al., 2011; Bajjalieh, 2010). Since premium is a necessary input for many industries, making timely decisions with regards to the supply of premium fuel can affect the bottom line for many business entities, especially if large amounts of premium fuel are involved in the usage.

An accurate forecast of energy demand provides information that plays an important role in policymaking in an economy (Ajith and Baikunth, 2001). In this respect, policymakers, the public, and academics have been interested in producing accurate premium fuel forecasts. Earlier realistic estimates of premium fuel are recommended and would be of considerable benefit, because of the significant impact of premium demand on the entire economy. Premium fuel is an important input for many industries in many economies. It is extensively used in the construction industry, transportation industry, and the agriculture sector. The biggest consumer of premium fuel in most economies is on-highway transportation in Ghana (Bajjalieh, 2010).

Various researchers (Weiqi et al., 2011; Fahimifard et al., 2009; Ediger and Akar, 2007) have used both quantitative (autoregressive moving average, autoregressive conditional heteroscedasticity) and qualitative (Delphi, belief networks, artificial neural network models and support vector machine regression models) forecasting models in assessing future changes in energy demand and energy prices. The autoregressive integrated moving average (ARIMA) is one of the popular models used in the literature in assessing future changes in energy. According to Yeboah et al. (2012), various works on forecasting are based on the use of ARIMA models since the models produce accurate forecast values. The findings are found in various empirical works (Wang and Meng, 2012; Ahmad and Latif, 2011; Albayrak, 2010; Mucuk and Uysal, 2009; Erdoğdu, 2007; Al-Fattah, 2006; Ediger et al., 2006). For comprehensive review of these works, refer to Yeboah et al. (2012). The results from these works indicate that energy demand is expected to increase in future.

To date, the majority of empirical studies of premium estimates focus on developed countries. Hence, the current study adds to the literature by developing premium fuel forecast model. Developing economies have suffered shortages of energy supply over the years and this called for works in the area of development of accurate forecast model since such shortages have serious consequences on the economic development of an economy (Yeboah et al., 2012). The study serves as a reference material for researchers and students. Policy makers, economist and energy experts are provided with useful guide in planning for the economy's energy demand to help ensure energy sufficiency in an economy. The objective of this work is to develop a forecast model that can provide improved monthly forecasts for premium fuel.

2. ECONOMETRIC METHODOLOGY

The premium fuel forecast is performed in two ways. First, the unit root properties are first examined using the Augmented Dickey-Fuller (ADF) (1981) and the Kwiatkowski et al. (1992, KPSS). Secondly, the forecast model is developed using the ARIMA model in four steps. The study is based on monthly time series data from the energy commission of Ghana for the period 2000-2011.

2.1. Unit Root Tests

The unit root test is conducted to determine whether the series in model are stationary. If the series are non-stationary theory should be made stationary through differencing before they are used in the estimation. If this is not done the regression results become spurious and not valid. In the current study the unit root test is performed using the ADF (1981) and the Kwiatkowski et al. (1992, KPSS). These tests (ADF and KPSS) have their strengths and their weaknesses and that is why the two tests are used in the study so that the weaknesses of one test are catered for by the strength of other tests. Among the weaknesses of the ADF test is the lack power of the test, since the test is not able to reject false null hypothesis always (Nanathakumar and Subramaniam, 2010). The ADF test is based on the null assumption (H_0) that there is a unit root in levels. The alternative assumption (H_1) is that the series are stationary in levels. The critical values are compared with the calculated values at 5%, 1% and 10% levels of significant. The KPSS test is used as a confirmatory test for the ADF test. The KPSS test is based on the null assumption that the series variables under investigation are stationary in levels against the alternative assumption that the series are non-stationary (Kwiatkowski et al., 1992).

2.2. ARIMA Model

The premium energy (P) forecast model is based on the ARIMA model. In using the ARIMA model, four steps are model identification, parameter estimation, model diagnostics, and forecast verification and reasonableness (Ajith and Baikunth, 2001). For the explanation of the steps, according to Ajith and Baikunth (2001), refer to Yeboah et al. (2012). The ARIMA model is specified as in Equation 1, following the works of Mucuk and Uysal (2009).

$$\Delta X_t = \mu + \theta_1 \Delta X_{t-1} + \theta_2 \Delta X_{t-2},\ldots,\theta_p \Delta X_{t-p}$$
$$u_t + \phi_1 \varepsilon_{t-1} + \phi \varepsilon_{t-2},\ldots,\phi_q \varepsilon_{t-q} \quad (1)$$

Where θ and ϕ are the coefficients of the ARIMA respectively.

3. EMPIRICAL WORK AND RESULTS

3.1. Unit Root Properties of the Variables

The nature of unit root of the variables series was first examined using the time series plot. The result is shown in Figure 1. The result from the plots indicates the series are not stationary in levels. The variables attained stationarity after first differencing. The result is shown in Figure 2.

The results from the plots indicate unit root in the series, which calls for scientific examination using the ADF model and the KPSS model. The results of the ADF and KPSS test results are shown in Tables 1-4.

3.2. Forecastings Results and Discussions of Premium (P)

3.2.1. Identification of the model

The series attained stationarity at first difference, which allows the introduction of the AR and MA portions into the model. An experiment was performed with ARIMA (1, 1, 0); (0, 1, 1) and (1, 1, 1) using their respective sample of ACF and PACF. The samples ACF and PACF of the monthly growth rate of premium

Figure 1: Time series plot for premium energy in levels

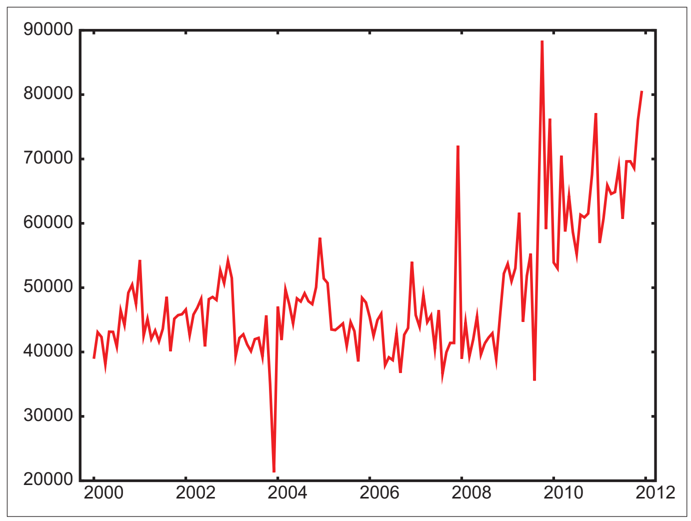
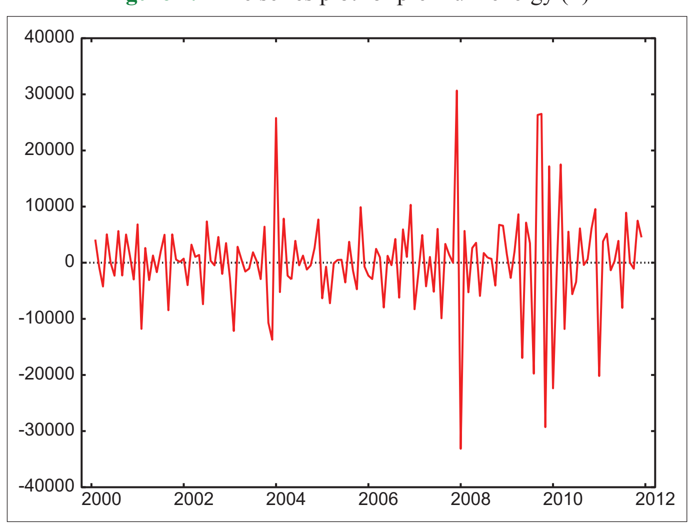


Figure 2: Time series plot for premium energy (P)



consumption using ARIMA (1, 1, 0); (0, 1, 1) and (1, 1, 1) are plotted on Figures 5-7.

An inspection of the samples ACF and PACF of ARIMA (1, 1, 0) shows that the ACF have significant spikes at lags 2 and 4 at 1% level of significance and the PACF have significant spikes at lags 2, 4 and 24 at 1% and at lag 21 at 5% level of significance with the rest of the spikes been insignificant. The inspection of the samples ACF and PACF of the ARIMA (0, 1, 1) model shows that the ACF have significant spikes at lag 4 at 1% significant level and lags 22 and 24 at 5% significant level.

The PACF have significant spikes at lags 4, 22 and 24 at 5% level of significance. The rest of the spikes are insignificant for the ACF and PACF samples. The inspection of the samples ACF and PACF of the ARIMA (1, 1, 1) model shows that the ACF and the PACF have significant spikes at lags 4 at 5% significant level for ACF and lag 4 at 1% significant level for the PACF. The rest of the spikes are insignificant at the 5% level of significance. The results suggest that ARIMA (1, 1, 1) is the best model for the forecasting.

Table 1: ADF unit root test with a constant

Variables	t statics	*P* value	Results	Lag length
$P_{\text{-level}}$	0.217	0.974	Accept H_0	12
$\Delta P_{\text{-first}}$ difference	−6.177	4.4e-08***	Reject H_0	12

Source: Author's computations, December, 2013/2014. ***Denote significance at 1% level. ADF: Augmented Dickey-Fuller

Table 2: ADF units root test with a constant and trend

Series variables	t statics	ADF *P* value	Results	Lag length
$P_{\text{-level}}$	−1.832	0.689	Accept H_0	12
$\Delta P_{\text{-first}}$ difference	−6.351	1.51e-07 ***	Reject H_0	12

Source: Author's computations, December, 2013/2014. ***Denote significance at 1% level. ADF: Augmented Dickey-Fuller

Table 3: KPSS unit root test with a constant

Series variables	t statics	Results	Max lag length
$P_{\text{-level}}$	0.730**	Reject H_0	12
$\Delta P_{\text{-first}}$ difference	0.173	Accept H_0	12

Source: Author's computations, December 2013/2014. **Denote significance at 5%. Critical values (0.464) 5% and (0.737) 1% for level test

Table 4: KPSS units root test with a constant and trend

Variables	t statics	Results	Lag length
$P_{\text{-level}}$	0.239***	Reject H_0	12
$\Delta P_{\text{-first}}$ difference	0.073	Accept H_0	12

Source: Author's computations, December, 2013/2014. ***Denote significance at 1% level. Critical values (0.148) 5% and (0.216) 1% for first difference test

3.2.2. Estimation of the ARIMA model

In the estimation stage of the ARIMA forecasting the AR model and the MA model are estimated in the research. The AR model predicts the change in premium consumption as an average, plus some fraction of the previous change, plus a random error. AR Part of the ARIMA model is estimated as in Equation 2 and the estimated results are as shown in Table 5.

$$\Delta P_t^* = \mu_0 + \theta_2 P_{t\text{-}2}^* + \theta_4 P_{t\text{-}4}^* + \varepsilon_t \tag{2}$$

The MA process of the ARIMA model is estimated as in Equation 3with the estimated results shown in Table 6.

$$\Delta P_t^* = \mu_0 + \varepsilon_t - \phi_4 \varepsilon_{t\text{-}4}^* \tag{3}$$

The ARIMA model is specified as in Equation 4. The estimated results are shown in Table 7.

$$\Delta P_t^* = \mu_0 + \theta_2 P_{t\text{-}2}^* + \theta_4 P_{t\text{-}4}^* + \varepsilon_t + \phi_4 \varepsilon_{t\text{-}4} \tag{4}$$

Figure 3: Plot of level correlograms of autocorrelation function and partial autocorrelation function

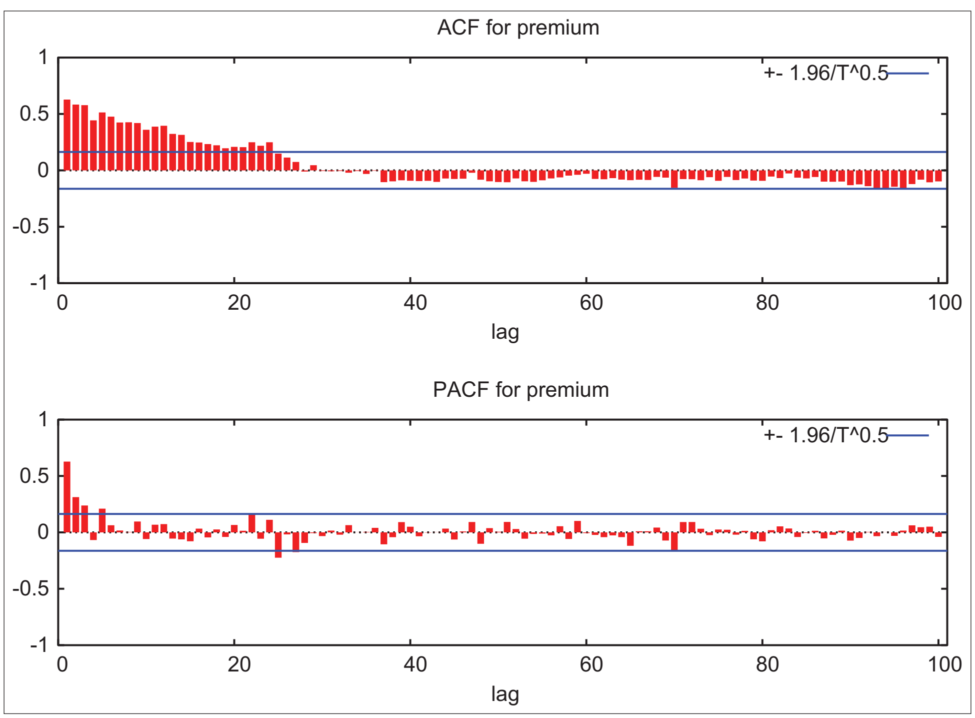


Figure 4: Plot of first difference correlograms of correlation function and partial autocorrelation function

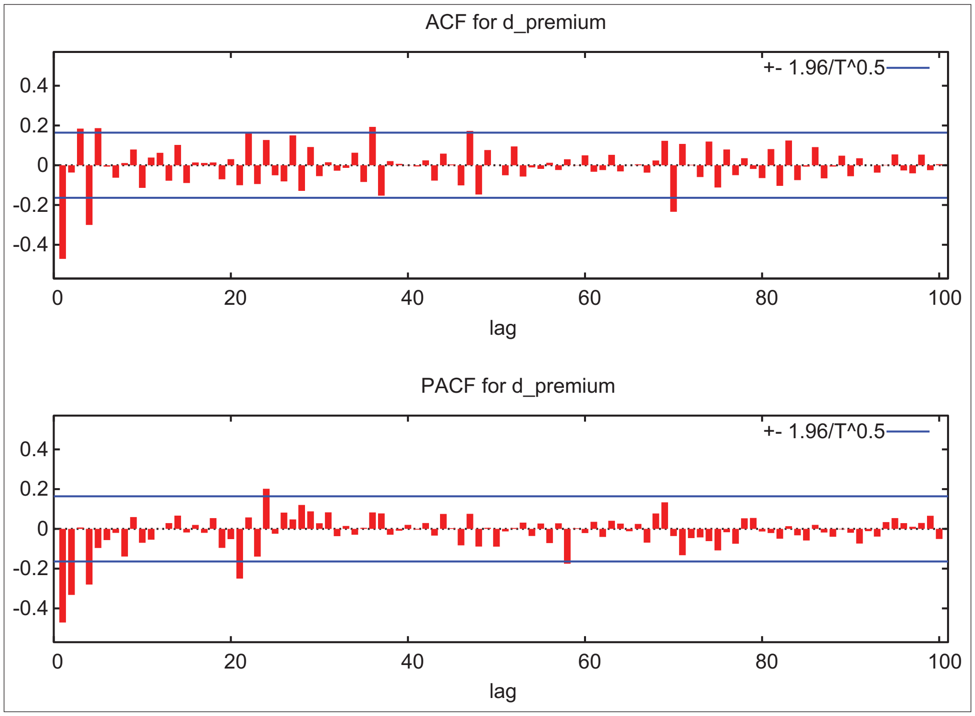
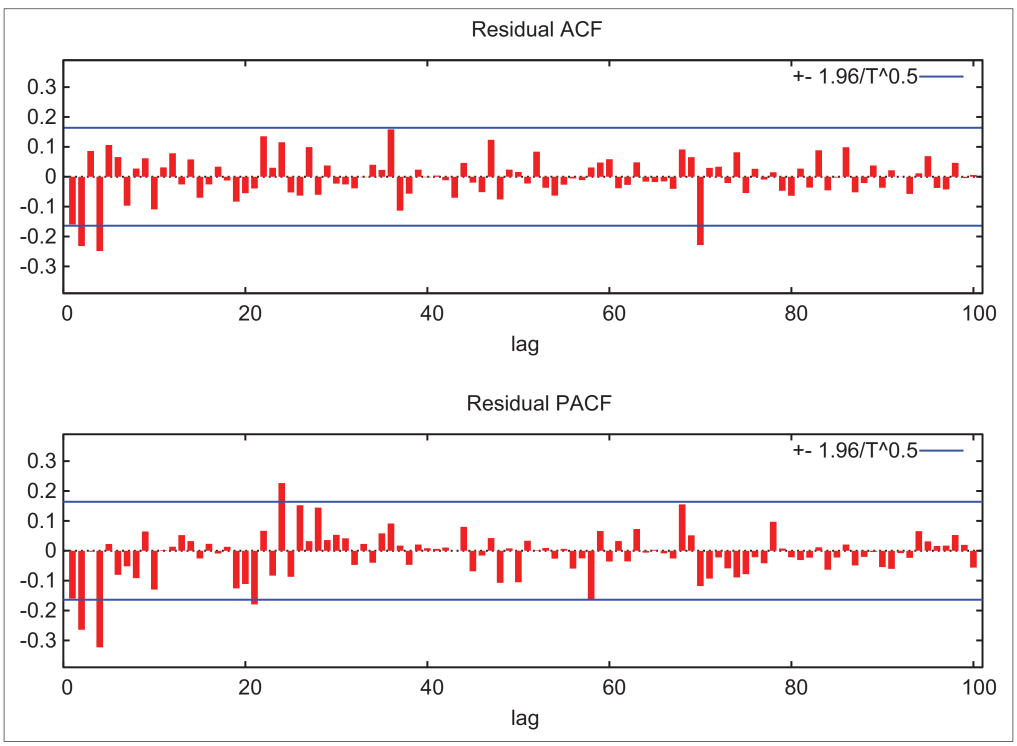


Figure 5: Autoregressive integrated moving average (1, 1, 0)



3.2.3. Diagnostic checking

The diagnostic checking involves generating residuals from Equations 2-4 as well as the ACF and PACF of these residuals up to 50 lags. The estimated ACF and PACF are presented in Figures 8-10. As indicated in Figure 8, all the autocorrelation and PACF are statistically insignificant for Equation 4. However, in Equation 2 there are significant spikes at lag 2 for the autocorrelation and PAC has significant spikes at lags 2 and 10. On the other hand, for Equation 3 there are significant lag at 34 for the autocorrelation and the PACF has significant spikes at lag 29. These are indications that the series are not randomly distributed. However, the correlogrames of both autocorrelation and PAC indicate that the residuals estimated using Equation 5.3 were purely random. Hence, the study adopted Equation 5.3 as the best model and for forecasting of premium consumption. There is thus no need to find another model and therefore ARIMA (1, 1, 1) is the best.

Figure 6: Autoregressive integrated moving average (0, 1, 1)

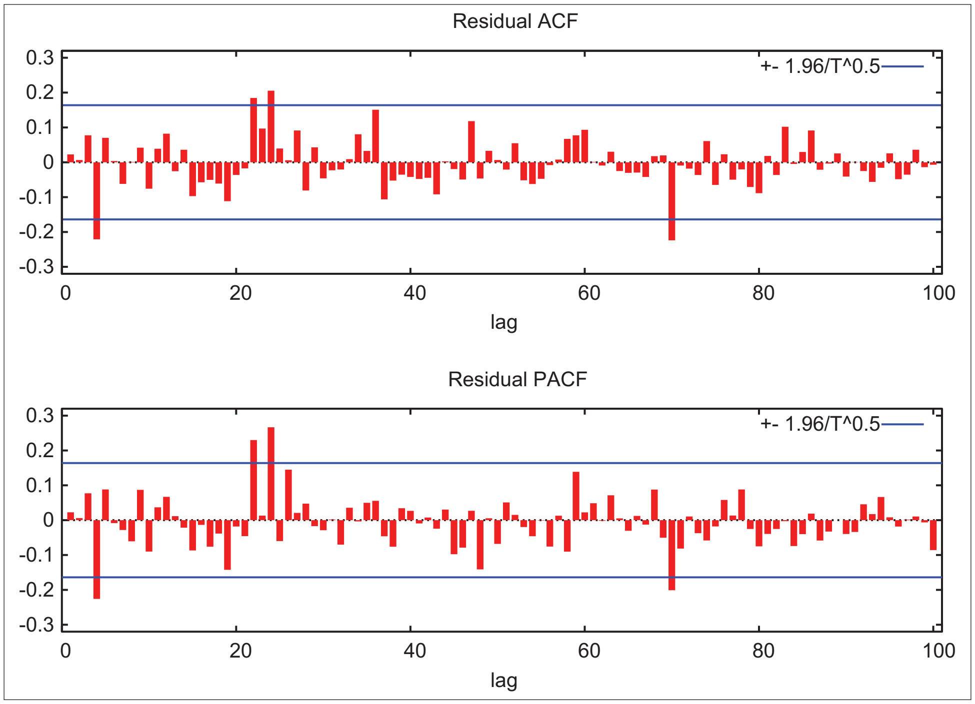
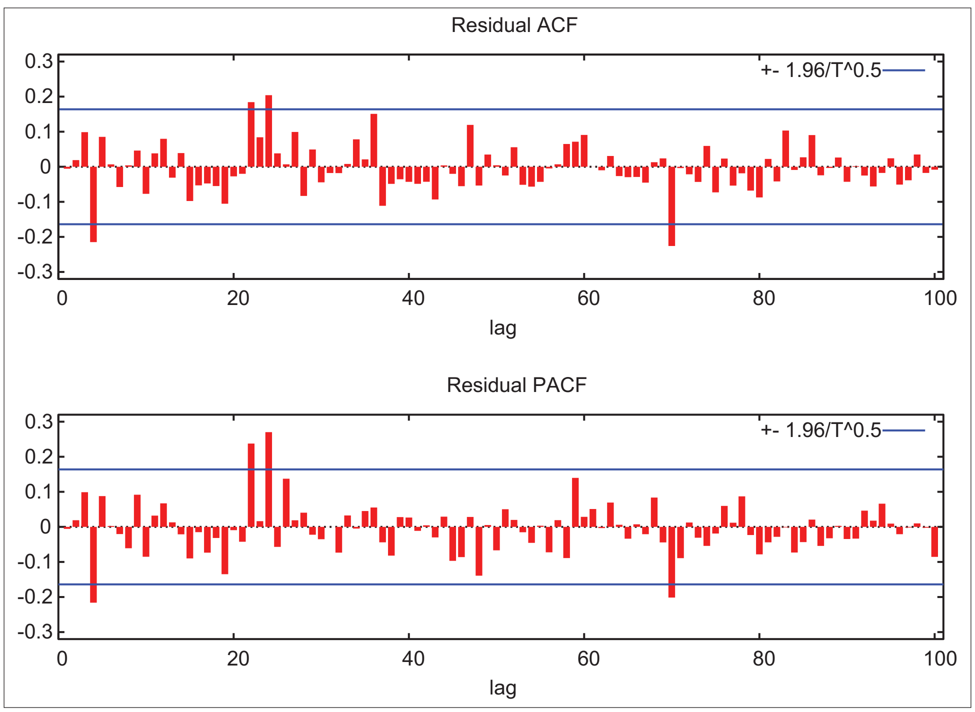


Figure 7: Autoregressive integrated moving average (1, 1, 1)



3.2.4. Forecasting

This is the final step of modeling premium consumption using ARIMA. The data for premium consumption covers the period 2000:01 to 2011:12 and based on Equation 4 the study forecasts premium consumption for the last 12 months of 2011. The results are shown in Table 8.

Figure 11 shows the correlogram of the ARIMA forecast model. The forecasted accurately fitted the actual consumption of premium since it insignificantly underestimates the actual consumption and thus indicates consistency of the results. The Forecast evaluation statistics are shown in Table 9.

Table 9 shows the results of forecast evaluation using the mean error (ME); mean squared error (MSE); root mean squared error (RMSE); mean absolute error (MAE); mean percentage error; mean absolute percentage error and the Theil's U. The 377.00 value of ME indicates the model is forecasting too low on average.

Figure 8: Residuals of autocorrelation function and partial autocorrelation function of Equation 2 for AR model

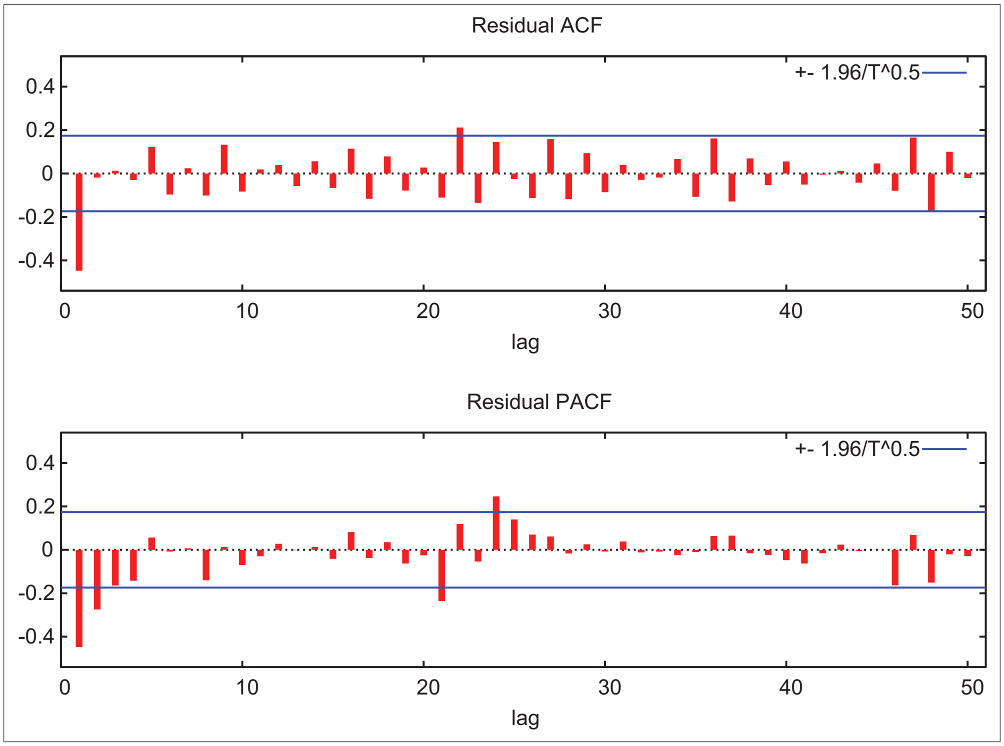
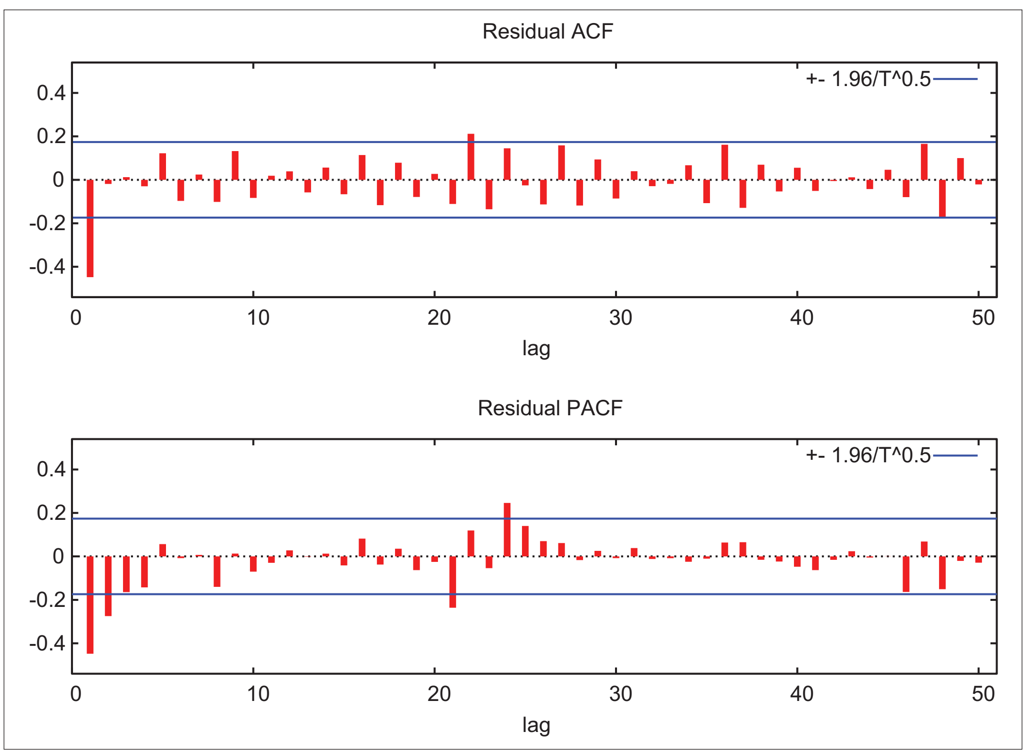


Figure 9: Residuals of autocorrelation function and partial autocorrelation function of Equation 3 for MA Model



The ME value of 377.00 is less than the 5510.20 value of the MAE. The MAE measure the closeness of the forecast values to the actual outcome. For a better and robust prediction of an estimated model, the value should be closer to zero. The value of MAE indicates the forecast values are not robust since the value is not close to zero.

The RMSE value of 7700.50 does not indicate accurate estimation methodologies in the forecasting. The 1.115 value of the Theil's statistics which is more than unity, is an indication that on average the selected model does not perform better than the simple "naïve" model. The Theil's U compares the RMSE of the chosen model to that of the 'naive' forecast model.

Figure 10: Residuals of autocorrelation function and partial autocorrelation function of Equation 4 of autoregressive integrated moving average (4, 1, 4)

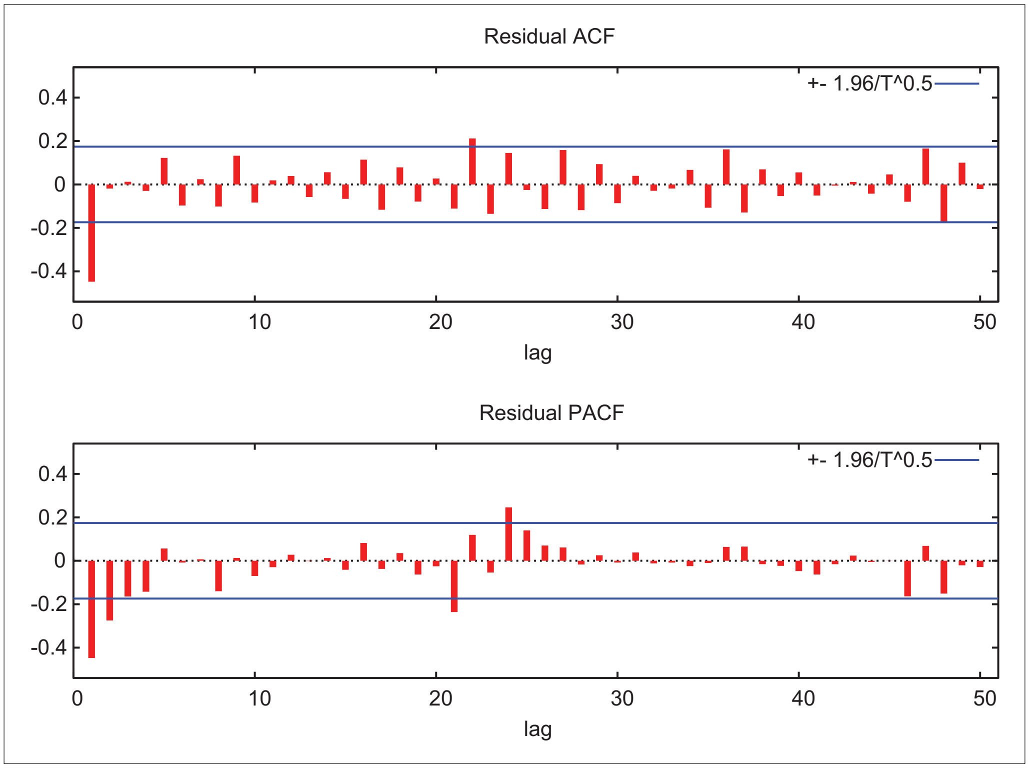
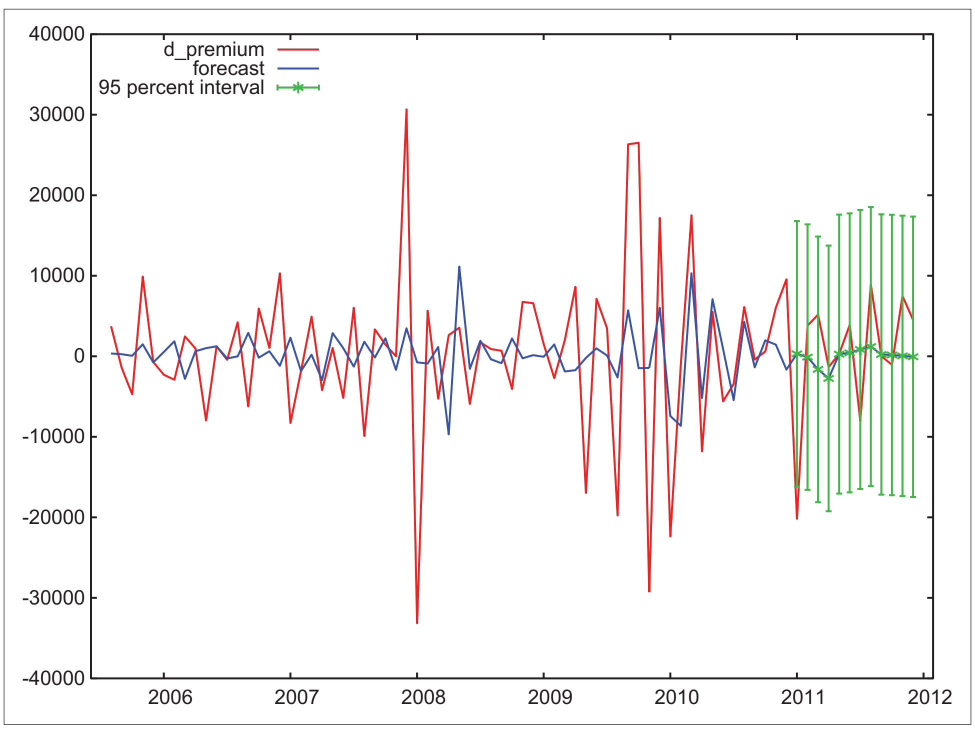


Figure 11: Correlogram of the autoregressive integrated moving average forecast



4. CONCLUSIONS

Accurate and early estimates of Premium fuel are important for decision-making processes. The ARIMA model is a popular forecast model in the literature since it provides accurate estimates. In the identification step of the forecasting of the ARIMA models, (1, 1, 0); (0, 1, 1) and (1, 1, 1) were identified as the appropriate order. The estimated models included a constant term. All the coefficients of the variables in the model except the constant term were significant. The diagnostic checking of the estimated model shows ARIMA (1, 1, 1) as the best fitted models since all the series were randomly distributed. The data for the forecast covers the period 2000:01 to 2011:12.

Table 5: Estimates of AR using observations 2000:04-2011:12 (T=141) (Dependent variable: premium [P])

Variables	Coefficient	Standard error	*t* ratio	*P* value
Constant	490.853	605.468	0.811	0.419
ΔP_{-2}	−0.634	0.080	−7.877	<0.000***
ΔP_{-4}	−0.339	0.081	−4.210	0.000***
Mean dependent variable=271.326				SD dependent variable=8588.760
Sum square residual=7.1e+09				SE of regression=7180.266
R^2=0.311				Adjusted R^2=0.301
F (2, 138)=31.156				*P* value (F)=6.8e-12
Log-likelihold=−1450.506				Akaike criterion=2907.012
Schwarz criterion=2915.858				Hannan-Quinn=2910.607
Rho=0.001				Durbin's h=0.049

Source: Author's computation, December, 2013/2014. ***Denotes 1% significant level. SD: Standard deviation, SE: Standard error

Table 6: Estimates of MA using observations 2000:06-2011:12 (T=139) (Dependent variable: Premium [P])

Variables	Coefficient	SE	*t* ratio	*P* value
Constant	336.096	700.698	0.479	0.632
ε_{-4}	−0.303	0.081	−3.714	0.000***
Mean dependent variable=269.338				SD dependent variable=8632.654
Sum square residual=9.3e+09				SE of regression=8258.387
R^2=0.091				Adjusted R^2=0.085
F (2, 138)=13.792				*P* value (F)=0.000
Log-likelihold=−1449.864				Akaike criterion=2903.728
Schwarz criterion=2909.597				Hannan-Quinn=2906.113
Rho=−0.445				Durbin's h=2.881

Source: Author's computation, December 2013/2014. ***denotes 1% significant level of significance. SD: Standard deviation, SE: Standard error

Table 7: Estimates of ARIMA using observations 2000:06-2011:12 (T=139) (Dependent variable: Premium [P])

Variables	Coefficient	SE	*t* ratio	*P* value
Constant	536.060	597.634	0.897	0.371
ΔP_{-2}	−0.587	0.080	−7.304	<0.000***
ΔP_{-4}	−0.323	0.079	−4.079	0.000***
ε_{-4}	−0.205	0.071	−2.899	0.004***
Mean dependent variable=269.338				SD dependent variable=8632.654
Sum square residual=6.7e+09				SE of regression=7035.097
R^2=0.350				Adjusted R^2=0.336
F (2, 138)=24.264				*P* value (F) =1.3e-12
Log-likelihold=−1426.558				Akaike criterion=2861.116
Schwarz criterion=2872.854				Hannan-Quinn=2865.886
Rho=−0.055				Durbin's h=−1.972

Source: Author's computation, 2013/2014. *** Denotes 1% significant level of significance. SD: Standard deviation, SE: Standard error, ARIMA: Autoregressive integrated moving average

Table 8: Forecast of premium from 2011:01 to 2011:12

Observations	Premium (First difference)	Forecast	SE	95% interval
2011:01	−20182.000	307.251	8330.910	(−16182.000, 16796.500)
2011:02	3798.000	−102.389	8330.910	(−16591.600, 16386.800)
2011:03	5178.000	−1625.800	8333.490	(−18120.100, 14868.500)
2011:04	−1342.000	−2748.240	8333.490	(−19242.600, 13746.100)
2011:05	289.000	269.517	8753.150	(−17055.400, 17594.400)
2011:06	3893.000	429.395	8753.150	(−16895.500, 17754.300)
2011:07	−8045.000	844.844	8754.160	(−16482.100, 18171.800)
2011:08	8905.000	1202.340	8754.160	(−16124.600, 18529.300)
2011:09	20.000	220.135	8796.210	(−17190.000, 17630.300)
2011:10	−1060.000	159.743	8796.210	(−17250.400, 17569.900)
2011:11	7491.000	50.412	8796.450	(−17360.200, 17461.000)
2011:12	4523.000	−63.213	8796.450	(−17473.800, 17347.400)

Source: Author's computation, December 2013/2014

The results indicated that the forecasted values fitted the actual consumption of the premium fuel since the forecasted values insignificantly underestimate the actual consumption and thus indicate consistency of the results. The evaluation statistics such as the ME; MSE; RMSE; MAE and Theil's statistic, indicate that the estimated model is suitable for forecasting. For example, the

Table 9: Forecast evaluation statistics

Diagnostic models	Value of statistics
Mean error	377.00
Mean squared error	5.9e+007
Root mean squared error	7700.500
Mean absolute error	5510.200
Mean percentage error	−13.445
Mean absolute percentage error	98.819
Theil's U	1.115
Bias proportion, UM	0.002
Regression proportion, UR	0.036
Disturbance proportion, UD	0.961

Source: Author's computation, December 2014

positive values of ME indicate the models is forecasting too low on average. The fitted model has higher predictive power since its REMSE value is smaller. The model has higher degree of accuracy since its MAE value is smaller. The value of the Theil's U statistics is more than unity, which is an indication that the chosen model does not outperform the simple "naïve" models.

Our results go in line with literature that suggests using ARIMA model produces accurate forecast (Wang and Meng, 2012; Ahmad and Latif, 2011; Albayrak, 2010). The findings of the study are additions to the growing body of literature that develop in-sample forecast models to forecast future changes in energy demand. Developing forecast models for other energy products such as liquefied petroleum gas is worth doing in future studies.

REFERENCES

Ahmad, S., Latif, H.A. (2001), Competencies Analysis of Box-Jenkins Method in Forecasting Electricity Demand, UMTAS, Empowering Science, Technology and Innovation towards a Better Tomorrow. p201-204.

Ajith, A., Baikunth, N. (2001), A neuro-fuzzy approach for modelling electricity demand in Victoria. Applied Soft Computing, 1(2), 127-138.

Albayrak, A.S. (2010), ARIMA forecasting of primary energy production and consumption in Turkey: 1923-2006. Enerji, Piyasa ve Düzenleme, 1(1), 24-50.

Al-Fattah, S.M. (2006), Time series modeling for U.S. natural gas forecasting. E Journal of Petroleum Management and Economics Petroleum Journals Online, 1(1), 1-17.

Bajjalieh, J.W. (2010), Forecasting Diesel Fuel Prices, MSc Thesis. Available from: htt://www.citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.6028.

Dickey, D.A., Fuller, W.A. (1981), Distribution of the estimators for autoregressive time series with a unit root. Econometrica, 49, 1057-72.

Ediger, V.S., Akar, S. (2007), ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy, 35, 8-1701.

Ediger, V.S., Akar, S., Ugurlu, B. (2006), Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. Energy Policy, 34, 46-3836.

Erdoğdu, E. (2007), Electricity demand analysis using cointegration and ARIMA modeling: a case study of Turkey. Energy Policy, 35, 1129-1146.

Fahimifard, S.M., Homayounifar, M., Sabouhi, M., Moghaddamnia, A.R. (2009), Comparison of ANFIS, ANN, GARCH and ARIMA techniques to exchange rate forecasting. Journal of Applied Sciences, 9(20), 3641-3641.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y. (1992), Testing the null hypothesis of stationarity against the alternative of a Unit Root. Journal of Econometrics, 54, 159-178.

Mucuk, M., Uysal, D. (2009), Turkey's energy demand. Current Research Journal of Social Science, 1(3), 123-128.

Nanathakumar, L., Subramaniam, T. (2010), Dynamic cointergation link between energy consumption and economic performance: empirical evidence from Malaysia. International Journal of Trade, Economics and Finance, 1(3), 261-267.

Wang, X., Meng, M. (2012), A hybrid neural network and ARIMA model for energy consumption forecasting. Journal of Computers, 7(5), 184-1190.

Weiqi, L., Linwei, M., Yaping, D., Pei, L. (2011), An Econometric Modeling Approach to Short-term Crude Oil Price Forecasting, Proceedings of the 30th Chinese Control Conference, Yantai, China, 1582-1585.

Yeboah, S.A., Ohene, M., Wereko, T.B. (2012), Forecasting aggregate and disaggregate energy consumption using arima models: a literature survey. Journal of Statistical and Econometric Methods, 1(2), 71-79.

Ziel, F., Steinert, R., Husmann, S. (2015), Efficient modeling and forecasting of electricity spot prices. Energy Economics, 47, 98-111.