An econometric investigation of forecasting liquefied petroleum Gas in Ghana

Yeboah Asuamah, Samuel

Business School, Accra Institute of Technology, Accra, Ghana, Business School, Sunyani Polytechnic, Ghana

11 July 2015

An econometric investigation of forecasting liquefied petroleum gas in Ghana

Samuel Yeboah Asuamah

Business School, Accra Institute of Technology, Accra, Ghana/Business School, Sunyani Polytechnic, Ghana
Phone number: +233244723071
Email: nelkonsegal@yahoo.com

Abstract

The aim of the paper is to contribute to the body of knowledge in the area of forecasting using Autoregressive Integrated Moving Average (ARIMA) modelling for liquefied petroleum gas (LPG) for Ghana using monthly data for the period 2000-2011. The ARIMA (1, 1, 1) model was identified as suitable model. The findings show that the forecasted values insignificantly underestimate the actual consumption and thus indicate consistency of the results. The values of the evaluation statistics such as the ME; MSE; RMSE; MAE, and Theil's statistic, on the accuracy of the model indicate that the estimated model is suitable for forecasting LPG. The findings support the continuous use of the ARIMA model in forecasting, in econometric time series forecast. Future study should consider modelling other energy sources that are used in Ghana and other developing economies such as kerosene.

Keywords: Liquefied petroleum gas, autoregressive integrated moving average, Forecasting, Diagnostic statistics
JEL Classifications: C51, C52, C53, E17, Q 47

1 Introduction

Sufficient supply of liquefied petroleum gas (LPG) has always been a matter of concern for policy makers in economies that import LPG products. This concern results from the fact that LPG as energy sources plays a key role in the economic growth of a country such as Ghana, which is small but open. Energy is an input in all sectors of the economy, and as such forecasting the consumption has attracted attention in the literature (Yeboah & Ohene-Manu, 2015; As'ad, 2012). It is used in transportation, at the industry level, and at the domestic level. Researchers such as Yeboah and Ohene-Manu (2015), and Ajith and Baikunth, (2001) indicate that timely and accurate availability of forecast values help in policy making decisions in managing energy consumption.

Various models have been used to produce accurate forecast data of many products. Some of these models are qualitative and quantitative (see Yeboah & Ohene-Manu, 2015). The autoregressive integrated moving average (ARIMA) is one of the popular forecasting models that have been used in producing accurate forecast (As'ad, 2012; Yeboah, Ohene-Manu, & Wereku, 2012; Wang & Meng, 2012; Ahmad & Latif, 2011; Albayrak, 2010; Kumar, Kumara, Mallik, & Shuklaa, 2009, Mucuk & Uysal, 2009; Erdoğdu, 2007; Al-Fattah, 2006; Lloret, Lleonart, & Sole, 2000). For example, Yeboah and Ohene-Manu (2015) used the ARIMA model to produce accurate premium forecast for Ghana. They identified ARIMA (1, 1, 1) as the suitable model for forecasting premium energy. As'ad (2012) used ARIMA model to forecast daily peak electricity demand from New South Wales, Australia. The results indicated that the ARIMA model is the best model in term of forecasting two to seven days ahead. Abdullah (2012) used the ARIMA model to forecast gold bullion coin selling prices for Malaysia. He identified ARIMA (2, 1, 2) as the suitable model for forecasting the prices. Abledu and Agbodah (2012) used ARIMA to Forecast and Modelled Volatility Oil Prices and identified ARIMA (1, 1, 0) as the suitable model.

The objective of the current paper is to develop an appropriate model to forecast monthly LPG to aid policy makers in managing LPG demand, using ARIMA model. The paper is motivated by the fact that less empirical works exist on ARIMA forecasting of LPG in a small but open economy such as Ghana. Previous works have been focused on developing economies. The findings provide further understanding on the theories of forecasting in econometrics and serves as a reference material to researchers interested in forecasting. The rest of the paper considers the methodology, results, and conclusions.

2. Econometric Models and Estimation Methods

Two steps are used in the forecast process. The first is to examine the unit root properties using the Augmented Dickey- Fuller (ADF) (1981) and the Kwiatkowski et al. (1992, KPSS) as a confirmatory test to the ADF. The second step deals with the forecasting of LPG using the ARIMA model in four main steps.

2.1 Stationarity Test

The test of stationarity is based on the ADF and KPSS. Since there are detailed discussions on the theory and the use of the two tests, the theories are not provided here. The KPSS test is based on the null assumption that the variable under investigation is not unit root in levels against the alternative assumption that the variable is unit root (Kwiatkowski et al., 1992). The KPSS test is specified as in equation (1).

$$Z_t = \xi t + r_t + \varepsilon_t \text{..}(1)$$

Where Z_t is the series variable under investigation, (t) is deterministic trend, (r_t) is a random walk, and ε_t is stationary error term. The ADF test is based on the null assumption that the variable under investigation not stationary in levels. The ADF test is as specified in equation (2).

$$\Delta Z_t = \alpha_t + \beta_t T_t + \rho Z_{t-1} + \sum_{i=1}^{q} \partial \Delta Z_{t-1} + \varepsilon_t \text{....................}(2)$$

Where T = time trend, Z= time series variable in the model, ε_t = error term or stochastic error term.

2.2 Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) model is used to forecast LPG in four main steps. The steps are model identification, parameter estimation, model diagnostics, and forecast verification and reasonableness. Graphs, statistics, autocorrelation function, partial autocorrelation functions, transformations, to achieve stationarity and tentatively identify patterns and model components are used in the model identification step. The examination of the model coefficients using the method of least squares and maximum likelihood methods are done at the Parameter estimation step. The validity of the model estimated model is examined at the diagnostic step. The last step, which is the forecast step, produces the forecast. The ARIMA model for forecasting LPG is specified as in equation (3), where θ and Φ are the coefficients of the model, and 'Z' the series been forecasted.

$$z_t = \mu + \theta_1 \Delta z_{t-1} + \theta_2 \Delta z_{t-2}, \ldots, \theta_p \Delta z_{t-p}$$
$$u_t + \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2}, \ldots, \Phi_q \varepsilon_{t-q} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

2.2 Data

The paper is based on monthly time series data of LPG consumption, from Energy Commission (Ghana) database. The data span 2000-2011 period.

3 RESULTS

3.1 Unit Root Properties of the Variables

Time series plots are used to examine the stationarity properties of the series. The results are shown in figures 1 and 2. The results from the plots indicate the series are not stationary in levels (figure 1). The variables attained stationarity after first differencing (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.

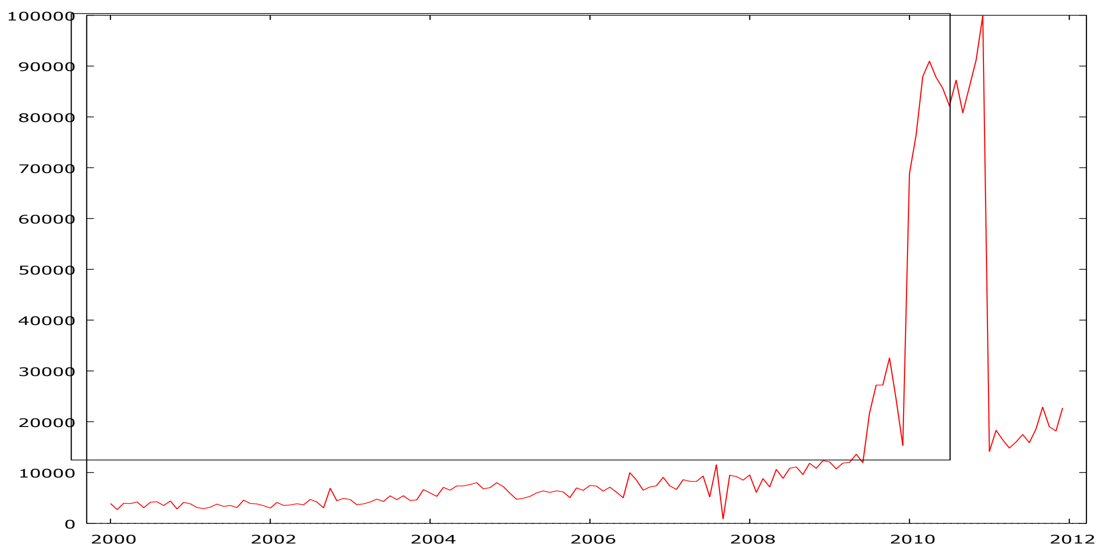

Figure 1. Time series Plots of LPG (levels)

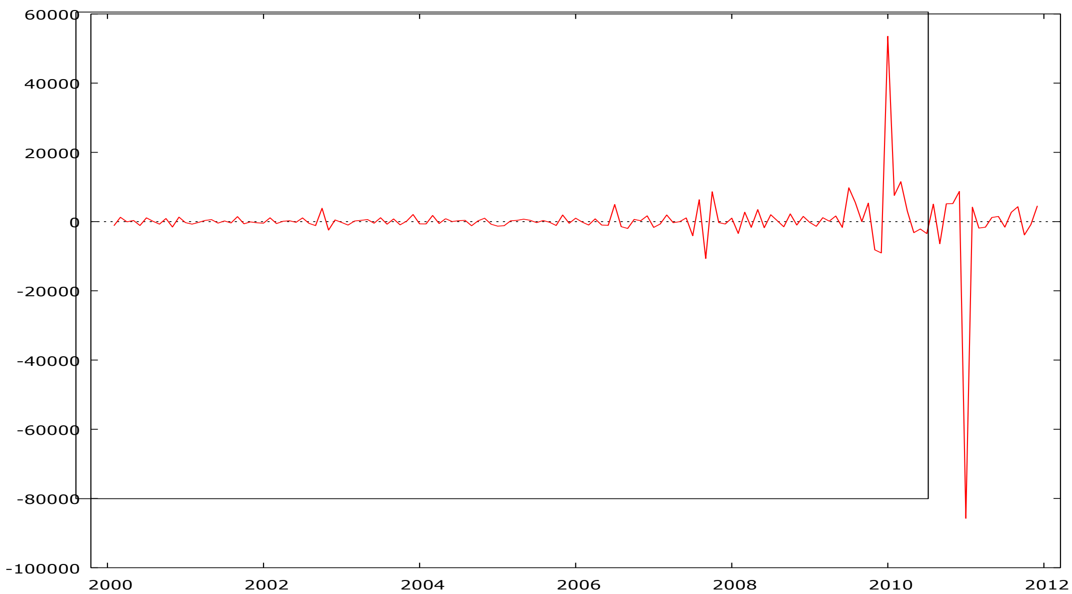

Figure 2. Time Series Plot for LPG (First difference)

3.2 The ADF and KPSS

The results of the ADF test, and KPSS test results are shown in Tables 1 to Table 4. The series attained stationarity at first differenced.

Table 1 ADF unit root test with a constant

Variables	t-stats	P-value	Results	Lag length
L_{-level}	1.330	0.999	Accept Ho	12
ΔL_{-first} difference	-10.030	2.73e-019 ***	Reject Ho	12

Source: Author's computations, 2014: Note: *** denote significance at 1% level

Table 2 ADF units root test with a constant and trend

Series Variables	t-stats	ADF P-value	Results	Lag length
L_{-level}	-0.434	0.986	Accept Ho	12
ΔL_{-first} difference	-8.475	4.7e-014 ***	Reject Ho	12

Source: Author's computations, 2014:
Note: *** denote significance at 1% level

Table 3 KPSS unit root test with a constant

Series Variables	t-stats	Results	Max Lag length
L_{-level}	0.557**	Reject Ho	12
ΔL_{-first} difference	0.045	Accept Ho	12

Source: Author's computations, 2014: 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-stats**	**Results**	**Lag length**
$L_{\text{-level}}$	0.110	Accept Ho	12
$\Delta L_{\text{-first}}$ difference	0.046	Accept Ho	12

Source: Author's computations, 2014:
Critical values (0.148) 5% and (0.216) 1% for first difference test

3.3 Forecasting Results and Discussions of LPG
3.3.1 Identification of the LPG Model

This section deals with the identification of the suitable ARIMA (p, d, q) model for LPG consumption. It is based on the use of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) with the associated correlogram against the lag length 100 and presented the results in Figure 3 and Figure 4. In Figure 3, ACF is statistically significant at lag 1 to 11 at 1% significant level and lags 12 to 14 at 5% level of significance. The rest of the spikes are not significant. Secondly, the PACF experienced significant spike at lags 1 and 13 at 1% significant level. The rest of the spikes are insignificant. The results indicate the series are unit root and may be stationary after differencing by either one or more times.

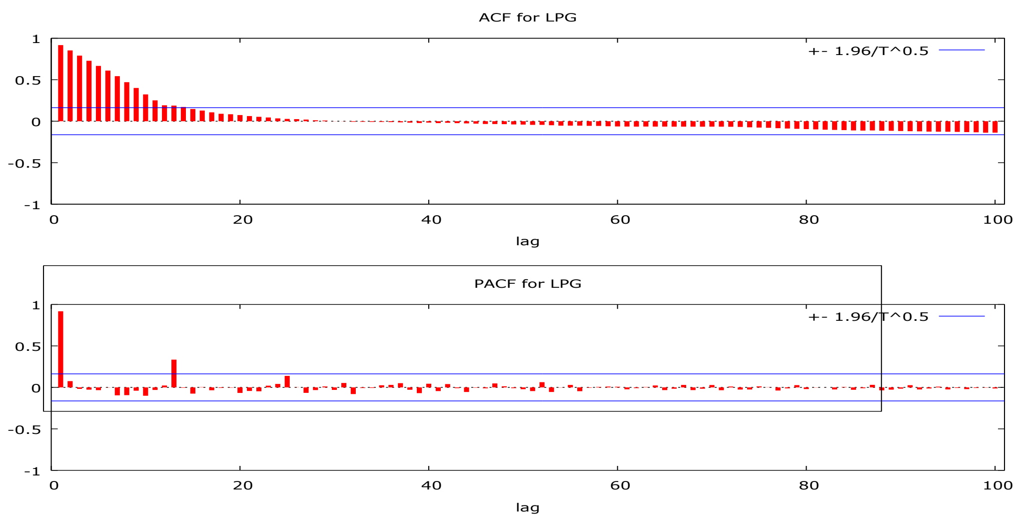

Figure 3. A Plot of series levels correlograms of ACF and PACF for LPG

The results of the first difference of the ACF and PACF sample and the associated plots of the correlograms of the first differencing up to 100 lags are shown in Figure 4. The results indicate the series attained stationarity after first difference since the series assumed irregular pattern. Only few spikes are significant at 1% (lag 1) levels and 5% (lag 12) for ACF and lag 24 at 5% level of significance for the PACF sample. The rest of the spikes are insignificant.

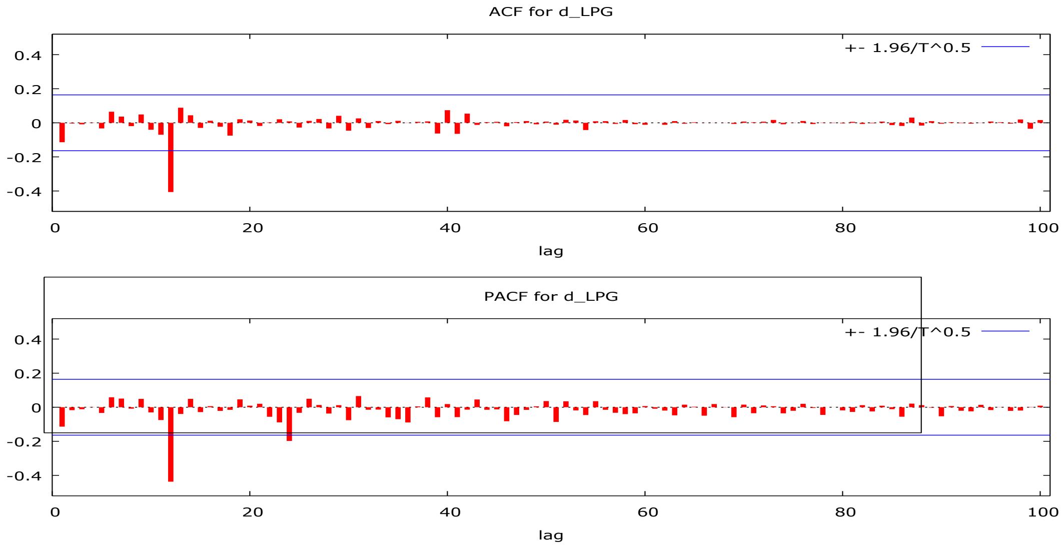

Figure 4. Plot of first difference correlograms of ACF and PACF for LPG

The series variables 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 samples of ACF and PACF. The samples ACF and PACF of the monthly growth rate of LPG consumption using ARIMA (1, 1, 0); (0, 1, 1) and (1, 1, 1) are plotted on Figure 5; 6 and 7.

An inspection of the samples ACF and PACF of ARIMA (1, 1, 0) shows that the ACF have significant spikes at lags 2 at 5% level of significance and the PACF have significant spikes at lags 2 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 and the PACF have insignificant spikes at 1% and 5% significant levels. The inspection of the samples ACF and PACF of the ARIMA (1, 1, 1) model shows that the ACF and the PACF have no significant spikes at 1% and 5% significant levels. The results suggest that ARIMA (1, 1, 1) is the better model for the forecasting.

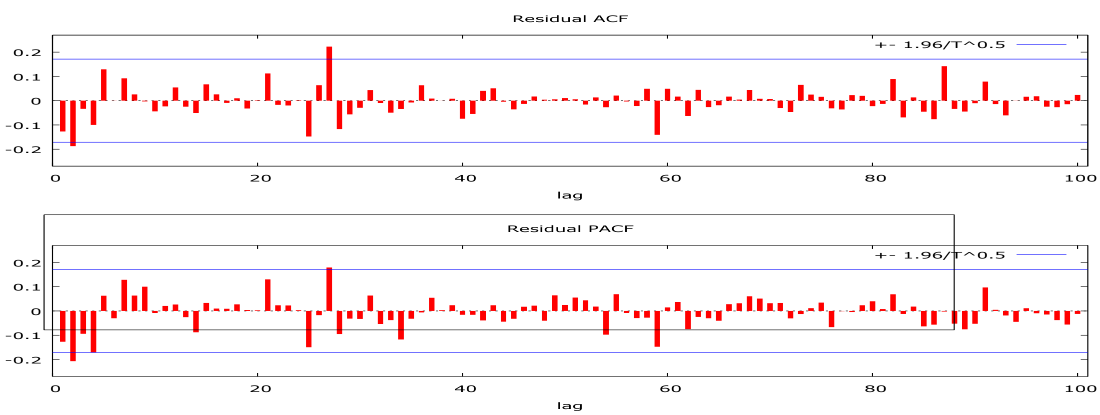

Figure 5 ARIMA (1, 1, 0)

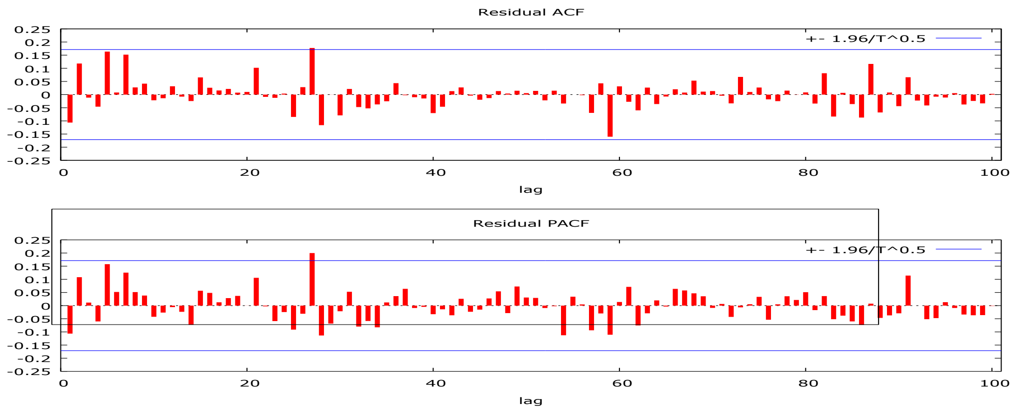

Figure 6 ARIMA (0, 1, 1)

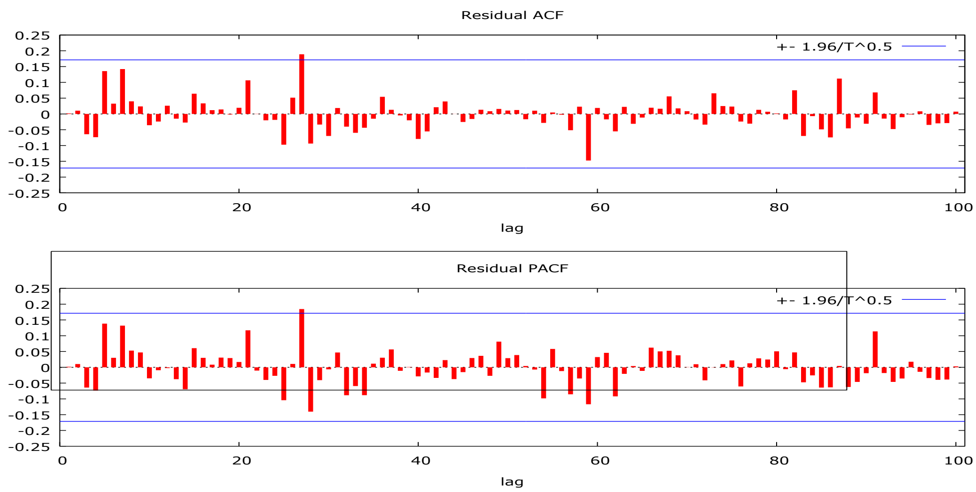

Figure 7 ARIMA (1, 1, 1)

3.3.2 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 LPG consumption as an average, plus some fraction of the previous change, plus a random error. The AR Part of the ARIMA model is specified as equation (4) with the estimated results reported in Table 5. This is a pure AR process since there is no MA process.

$$LPG_t^* = \alpha_0 + \alpha_2 LPG_{t-2}^* + \mu_t \text{..(4)}$$

Table 5 Estimates of AR, using observations 2000:04-2011:12 (T = 129)
Dependent variable: log difference of LPG

Variables	Coefficient	Std. Error	t-ratio	P-value
Constant	0.021	0.035	0.602	0.548
ΔldLPG_2	0.157	0.087	1.798	0.075
Mean dependent variable = 0.025		S.D. Dependent var. =0.404		
Sum square residual =20.401		S.E. of regression =0.401		
R-square =0.025		Adjusted R-square =0.017		
F(2, 138) = 3.233		P-value (F) =0.075		
Log-likelihold = -64.089		Akaike criterion =132.179		
Schwarz criterion =137.899		Hannan-Quinn =134.503		
Rho =-0.491		Durbin's h =2.982		

Source: Author's computation, 2014:

3.3.3 Diagnostic Checking

The diagnostic checking involves generating residuals from equations (4) as well as the ACF and PACF of these residuals up to 50 lags. The estimated ACF and PACF are presented in Figure 8. This is a pure AR process since there is no MA process. As indicated in Figure 8 for equation (4), for the ACF sample there is significant spike at lag 1 at 1% level of significance with the rest of the spikes been insignificance. For the PACF there are statistically significant spikes at lags 1, 2 and 4 at 1% level of significance. The rest of the spikes are insignificant for equation (4).

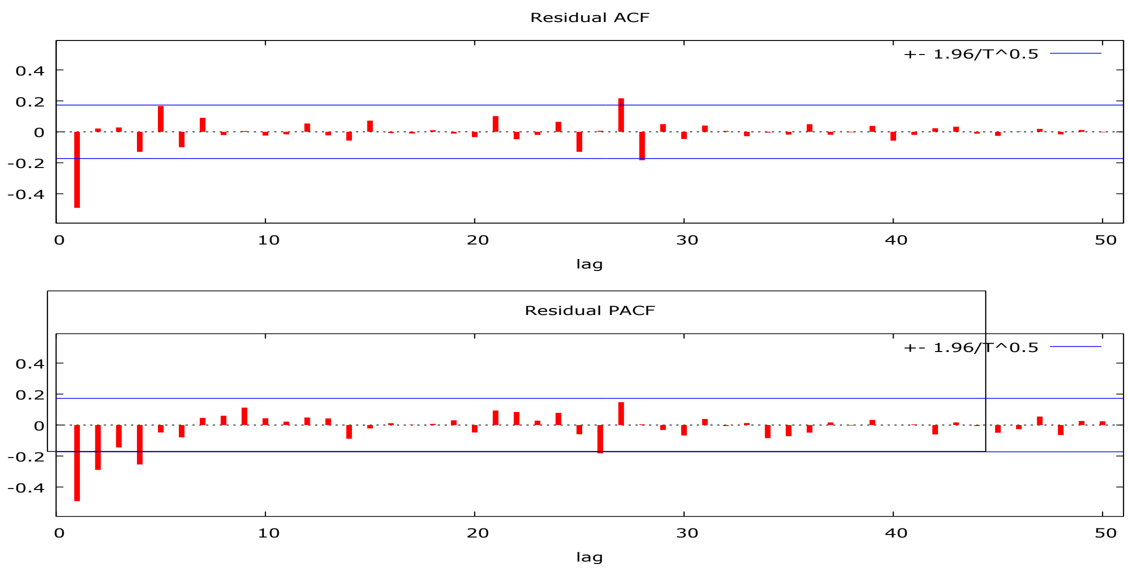

Figure 8 Residuals of ACF and PACF of equation 4 for AR Model

3.3.4 Forecasting

This is the final step of modeling LPG consumption using ARIMA. The data for LPG consumption covers the period 2000:01 to 2011:12 and based on equation (4), the study forecasts LPG consumption for the last 12 months of 2011. The results are shown in Table 6.

Table 6 Forecast of LPG from 2011:01 to 2011:12

Observations	LPG (First difference)	Forecast	Std. error	95% interval
2011:01	-1.955	0.030	0.401	(-0.763, 0.823)
2011:02	0.259	0.036	0.401	(-0.758, 0.829)
2011:03	-0.108	0.026	0.406	(-0.778, 0.829)
2011:04	-0.104	0.026	0.406	(-0.776, 0.829)
2011:05	0.077	0.025	0.406	(-0.778, 0.828)
2011:06	0.089	0.025	0.406	(-0.778, 0.828)
2011:07	-0.096	0.025	0.406	(-0.778, 0.828)
2011:08	0.157	0.025	0.406	(-0.778, 0.828)
2011:09	0.208	0.025	0.406	(-0.778, 0.828)
2011:10	-0.184	0.0252	0.406	(-0.778, 0.828)
2011:11	-0.045	0.025	0.406	(-0.778, 0.828)
2011:12	0.223	0.025	0.406	(-0.778, 0.828)

Source: Author's computation, 2014

Figure 9 shows the correlogram of the ARIMA forecast model. The forecasted accurately fitted the actual consumption of LPG since it insignificantly underestimates the actual consumption and thus indicates consistency of the results.

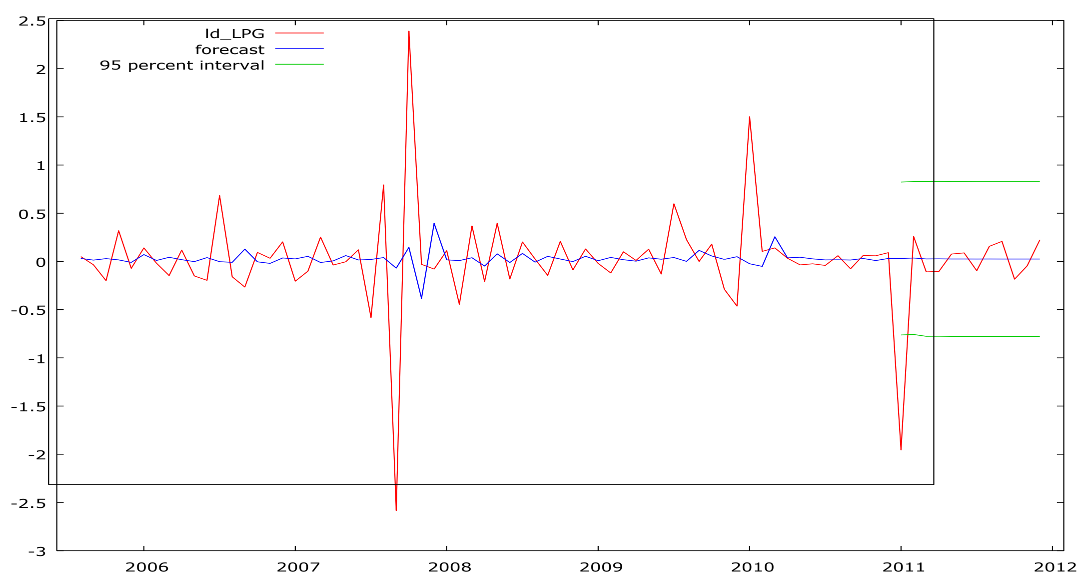

Figure 9: Correlogram of the ARIMA forecast.

Table 7 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 (MPE); Mean Absolute Percentage Error (MAPE) and the Theil's U. The -0.15014 value of ME indicates the model is forecasting too high on average. The ME value of -0.15014 is less than the 0.29185 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 in Table 7 indicates the forecast values are robust since the value is close to zero. The RMSE value of 0.59073 does not indicate accurate estimation methodologies in the forecasting. The 0.70134 value of the Theil's statistics which is less than unity, is an indication that on average the selected model performs better than the simple 'naive' model. The Theil's U compares the RMSE of the chosen model to that of the 'naive' forecast model.

Table 7 Forecast evaluation statistics

Diagnostic Models	**Value of statistics**
1. Mean Error (ME)	-0.150
2.Mean Squared Error (MSE)	0.349
3. Root Mean Squared Error (RMSE)	0.591
4. Mean Absolute Error (MAE)	0.292
5. Mean Percentage Error (MPE)	102.660
6. Mean Absolute Percentage Error (MAPE)	-21.831
7. Theil's U	0.701
8. Bias proportion, UM	0.065
9. Regression proportion, UR	0.069
10. Disturbance proportion, UD	0.866

Source: Author's computation, 2014

4 Conclusions

The aim of the paper has been achieved. The best model to forecast LPG has been identified as ARIMA (1, 1, 1). The findings show that, the forecasted values insignificantly underestimate the actual consumption and thus indicate consistency of the results. The findings of the research show that the forecasted values fitted the actual consumption of the LPG. The values of the evaluation statistics such as the ME; MSE; RMSE; MAE and Theil's statistic, on the accuracy of the model indicate that the estimated model is suitable for forecasting LPG. The results are in support of previous studies that reported that the ARIMA model is suitable to produce accurate estimates (Yeboah & Ohene-Manu, 2015; As'ad, 2012; Yeboah et al., 2012, Wang & Meng, 2012, Abdullah, 2012). The findings support the continuous use of ARIMA model in time series forecasting. Policy makers in the area of energy management should provide timely forecast values using the ARIMA model. Future study should consider modelling other energy sources that are used in Ghana, and other developing economies such as kerosene.

References

Abdullah, L., 2012. ARIMA Model for Gold Bullion Coin Selling Prices Forecasting. International Journal of Advances in Applied Sciences (IJAAS), 1(4), 153-158.

Ahmad, S., and Latif, H. A. 2001. Competencies Analysis of Box-Jenkins Method in Forecasting Electricity Demand, UMTAS, Empowering Science, Technology, and Innovation towards a Better Tomorrow. 201-204.

Ajith, A., and 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.

As'ad, M., 2012. "Finding the Best ARIMA Model to Forecast Daily Peak Electricity Demand", Applied Statistics education and Research Collaboration (ASEARC)-Conf. Papers, Uni. of Wollongong, Australia, 2012, Paper 11.

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

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

Kumar, M., Kumara, A., Mallik, N. C., C., and Shuklaa, R. K., 2009. "Surface flux modelling using ARIMA technique in humid subtropical monsoon area. Journal of Atmospheric and Solar-Terrestrial Physics, 71(12), 1293-1298.

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

Lloret, J., Lleonart, J., and Sole, I., 2000. "Time series modeling of landings in North Mediterranean Sea. ICES Journal of Marine Science, 57 (1), 171-184.

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

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

Yeboah A. S., and Ohene-Manu, J. 2012. An Econometric Investigation of Forecasting Premium Fuel. International Journal of Energy Economics and Policy, 5(3), 716-724. Available at available at http: www.econjournals.com: Retrieved on 30/11/2015.

Yeboah, A. S., Ohene-Manu, J., and Wereko, T. B. 2012. Determinants of Energy Consumption: A Review. International Journal of Management Sciences, 1(12), 482-487.