JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3362-3370
https://publishoa.com
ISSN: 1309-3452

Stochastic Modelling and Forecasting for LPG Prices: SARIMA Approach

T. Jai Sankar, I. Angel Agnes Mary *and* **P. Sameerabanu**

Department of Statistics, Bharathidasan University, Tiruchirappalli, Tamilnadu, INDIA

tjaisankar@gmail.com, angelagnes709@gmail.com *and* imashameera@gmail.com

Received 2022 March 25; **Revised** 2022 April 28; **Accepted** 2022 May 15.

Abstract

Liquefied Petroleum Gas (LPG) is a by product of crude oil and natural gas that is rich in hydrocarbons. LPG can be used for cooking, water heating, and space heating and drying. LPG allows the user to control the level of heat and ensures a clean cooking environment. This study examines to predict and forecast analysis for Chennai's monthly LPG prices from December 2013 to November 2021 through the Seasonal Autoregressive Integrated Moving Average (SARIMA) approach. The use of Mean Absolute Percentage Error (MAPE) and Bayesian Information Criterion (BIC) criteria for the fitted model considered appropriate LPG pricing. The best model to forecast the Gas Prices is SARIMA$(0,1,1)(1,1,0)_{12}$. Based on the chosen model, it could be predicted that LPG price would increase to Rs. 1,308.75 in November 2023 from Rs. 900.20 in December 2021 in Chennai.

Key Words: LPG Price forecasting, PACF, SARIMA.

Introduction

LPG is a multipurpose fuel and used worldwide as an alternative fuel for various applications; residential, commercial, agribusiness, industrial, and auto gas. In the absence of natural gas, LPG is the best alternative fuel. Cooking with other fuels such as wood or kerosene results in pollution, accumulation of soot, and the heat generated is difficult to control. LPG has been in India since 1950 and was used only by the middle and upper class, despite government subsidies for all. In recent years, the Government of India has sought to change these access and utilization practices through a series of target policies. From 2009 to 2012, Rajiv Gandhi Gramin LPG Vitaran Yojana provided 1.5 million new LPG connections to rural areas (Jain, 2016). Since 2015, the Government of India, in collaboration with three major oil companies, has launched three major projects to improve LPG for poor and rural households. LPG prices in India are based on the Import Parity Price (IPP) of the international market and the fuel is imported into the country.

Materials and Methods

As the aim of the study was to predict and forecast analysis for Chennai's monthly LPG prices from December 2013 to November 2021 through the Seasonal Autoregressive Integrated Moving Average (SARIMA) approach, and data collected from the Annual Report (2020-21) of Indian Oil Corporation Limited. Akaike (1970) discussed the stationary time series by an AR (p), where p is finite and bounded by the same integer. Moving Average (MA) models were used by Slutzky (1973). Box and Pierce (1976) considered the distribution of residual autocorrelations in ARIMA. SARIMA model introduced by Box and Jenkins (1976), was frequently used for discovering the pattern and predicting the future values of the time series data. Hannan and Quinn (1979) suggested obtaining the order of a time series model by minimizing the errors for pure AR models, and Hannan (1980) for ARMA models. A second order determination method could be considered as a variance of Schwarz's Bayesian Criterion (SBC) which gives a consistent estimate of the order of an ARMA model. Hosking (1981) introduced a family of models, called fractionally differenced autoregressive integrated moving average models, by generalizing the ‘d’ fraction in ARIMA (p,d,q) model. Pindyck (1994) explored the relationship between short-term stocks and prices for copper, wood, and heating oil. Solomon (2001) sought to predict future oil and gas prices by triggering a linear regression analysis of the impact of oil and natural gas prices on oil industry activity. Stephane et al., (2003) attempted to estimate the actual oil price. Zamani (2004) proposed an economic forecasting model for short-term oil spot prices. Japlonowski et al., (2007) proposed a conclusion analysis model for evaluating the crude oil price forecast. Ogwo (2007) proposed an equivalent gas price model. Mogopadhyay et al., (2012) suggested that fuel stacking practiced for cooking methods in rural homes with

LPG. Holoda et al., (2017); Troncoso and da Silva, (2017) explained an important task of the fuel stacking procedures for the popular and clean option.

The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is

$$\text{ARIMA }(p,d,q)\times(P,D,Q)_S$$

with p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order and S = time span of repeating seasonal pattern, Φ = Autoregressive polynomial Order and Θ = Moving average polynomial Order.

Without differencing operations, the model could be written more formally as

$$\Phi(B^S)\phi(B)(x_t-\mu)=\Theta(B^S)\theta(B)\omega_t \text{ ------- (1)}$$

The non-seasonal components are:

$$\text{AR: } \phi(B)=1-\phi_1 B-.....-\phi_p B^p$$

$$\text{MA: } \theta(B)=1+\theta_1 B+.....-\theta_q B^q$$

The seasonal components are:

$$\text{Seasonal AR: } \Phi(B^S)=1-\Phi_1 B^S-.....-\Phi_P B^{PS}$$

$$\text{Seasonal MA: } \Theta(B^S)=1+\Theta_1 B^S-.....-\Theta_Q B^{QS}$$

On the left side of equation (1) the seasonal and non-seasonal AR components multiply each other, and on the right side of equation (1) the seasonal and non-seasonal MA components multiply each other.

SARIMA Analysis:

Table 1: Actual LPG Price (Rs.) in Chennai

Month	**Amount**	***Month***	**Amount**	***Month***	**Amount**	***Month***	**Amount**	***Month***	**Amount**
Dec 2013	1014	*Aug 2015*	604	*Apr 2017*	732	*Dec 2018*	827	*Aug 2020*	611
Jan 2014	1234	*Sep 2015*	577	*May 2017*	639	*Jan 2019*	705	*Sep 2020*	610
Feb 2014	1132	*Oct 2015*	532	*Jun 2017*	560	*Feb 2019*	673	*Oct 2020*	610
Mar 2014	1082	*Nov 2015*	560	*Jul 2017*	574	*Mar 2019*	717	*Nov 2020*	610
Apr 2014	981	*Dec 2015*	623	*Aug 2017*	533	*Apr 2019*	722	*Dec 2020*	685
May 2014	929	*Jan 2016*	672	*Sep 2017*	609	*May 2019*	728	*Jan 2021*	710
Jun 2014	905	*Feb 2016*	587	*Oct 2017*	657	*Jun 2019*	753	*Feb 2021*	760
Jul 2014	924	*Mar 2016*	526	*Nov 2017*	750	*Jul 2019*	653	*Mar 2021*	835

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3362-3370
https://publishoa.com
ISSN: 1309-3452

Aug 2014	922	*Apr 2016*	521	*Dec 2017*	756	*Aug 2019*	591	*Apr 2021*	825
Sep 2014	903	*May 2016*	538	*Jan 2018*	751	*Sep 2019*	607	*May 2021*	825
Oct 2014	883	*Jun 2016*	560	*Feb 2018*	746	*Oct 2019*	620	*Jun 2021*	825
Nov 2014	864	*Jul 2016*	551	*Mar 2018*	700	*Nov 2019*	696	*Jul 2021*	851
Dec 2014	750	*Aug 2016*	500	*Apr 2018*	664	*Dec 2019*	714	*Aug 2021*	876
Jan 2015	705	*Sep 2016*	478	*May 2018*	663	*Jan 2020*	734	*Sep 2021*	901
Feb 2015	600	*Oct 2016*	501	*Jun 2018*	713	*Feb 2020*	881	*Oct 2021*	916
Mar 2015	606	*Nov 2016*	539	*Jul 2018*	771	*Mar 2020*	826	*Nov 2021*	916
Apr 2015	614	*Dec 2016*	594	*Aug 2018*	806	*Apr 2020*	762		
May 2015	609	*Jan 2017*	595	*Sep 2018*	839	*May 2020*	570		
Jun 2015	620	*Feb 2017*	661	*Oct 2018*	896	*Jun 2020*	607		
Jul 2015	628	*Mar 2017*	747	*Nov 2018*	960	*Jul 2020*	611		

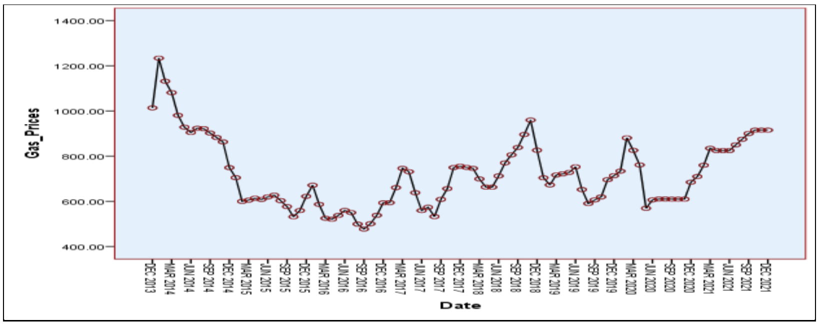

Figure1: Time Plot for Month-wise LPG Prices

Figure 1 depicts that the data used were non-stationary.The time series plot of monthly Gas Prices from December 2013 to December 2021 with *d*=1 was created as shown in Figure 2.

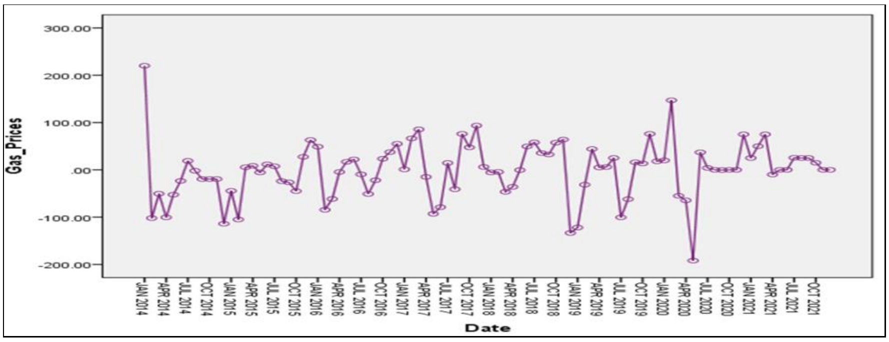

Figure 2: Time Plot for Month-wise LPG Prices with d=1

Table 2: ACF and PACF of LPG Prices with d=1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic			Partial Autocorrelation	Std. Error
			Value	df	Sig.[b]		
1	0.221	0.100	4.843	1	0.028	0.221	0.102
2	0.007	0.100	4.848	2	0.089	-0.044	0.102
3	-0.241	0.099	10.706	3	0.013	-0.245	0.102
4	-0.142	0.099	12.772	4	0.012	-0.039	0.102
5	-0.109	0.098	13.995	5	0.016	-0.073	0.102
6	-0.073	0.098	14.559	6	0.024	-0.104	0.102
7	0.017	0.097	14.591	7	0.042	0.014	0.102
8	0.137	0.097	16.611	8	0.034	0.100	0.102
9	0.139	0.096	18.710	9	0.028	0.046	0.102
10	0.126	0.096	20.450	10	0.025	0.084	0.102
11	-0.061	0.095	20.857	11	0.035	-0.066	0.102
12	-0.031	0.094	20.966	12	0.051	0.050	0.102
13	-0.162	0.094	23.953	13	0.032	-0.106	0.102
14	-0.087	0.093	24.813	14	0.036	-0.033	0.102
15	0.105	0.093	26.090	15	0.037	0.176	0.102
16	0.123	0.092	27.871	16	0.033	0.015	0.102

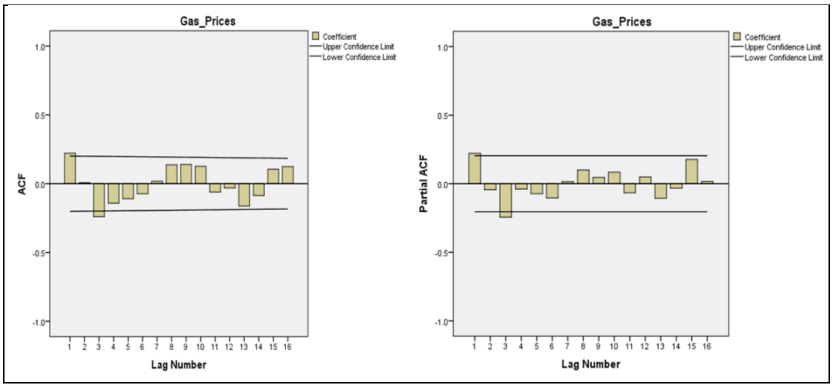

Figure 3: ACF and PACF plot with *d*=1

There is a significant autocorrelation at 3 lags which indicates that seasonality of data needed to be considered (Figure 3). First seasonal differencing needed to be done. The time series plot of monthly Gas Prices with *d*=1 and *D*=1 was created as shown in Figure 4.Stationary and normal time series plot with seasonality proof after the transformation of difference, *d*=1 and seasonal difference, *D*=1. The ACF and PACF plots with transformation *d*=1 and *D*=1 were depicted as shown in Figure 5.

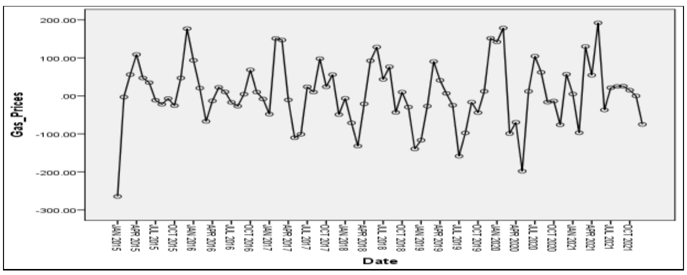

Figure 4: Time Plot for Month-wise LPG Prices with d=1 and D=1

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3362-3370
https://publishoa.com
ISSN: 1309-3452

Table 3: ACF and PACF of LPG Prices with d=1 and D=1

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic			Partial Autocorrelation	Std. Error
			Value	df	Sig.[b]		
1	0.312	0.107	8.493	1	0.004	0.312	0.109
2	-0.051	0.107	8.725	2	0.013	-0.165	0.109
3	-0.401	0.106	23.102	3	0.000	-0.377	0.109
4	-0.287	0.105	30.548	4	0.000	-0.066	0.109
5	-0.135	0.105	32.221	5	0.000	-0.087	0.109
6	-0.047	0.104	32.423	6	0.000	-0.197	0.109
7	0.029	0.103	32.500	7	0.000	-0.076	0.109
8	0.171	0.103	35.293	8	0.000	0.108	0.109
9	0.300	0.102	43.977	9	0.000	0.176	0.109
10	0.202	0.101	47.954	10	0.000	0.050	0.109
11	-0.146	0.101	50.068	11	0.000	-0.179	0.109
12	-0.443	0.100	69.764	12	0.000	-0.260	0.109
13	-0.214	0.099	74.425	13	0.000	0.146	0.109
14	-0.031	0.098	74.521	14	0.000	-0.067	0.109
15	0.216	0.098	79.412	15	0.000	-0.015	0.109
16	0.151	0.097	81.825	16	0.000	-0.005	0.109

[a] The underlying process assumed is independence (white noise).

[b] Based on the asymptotic chi-square approximation.

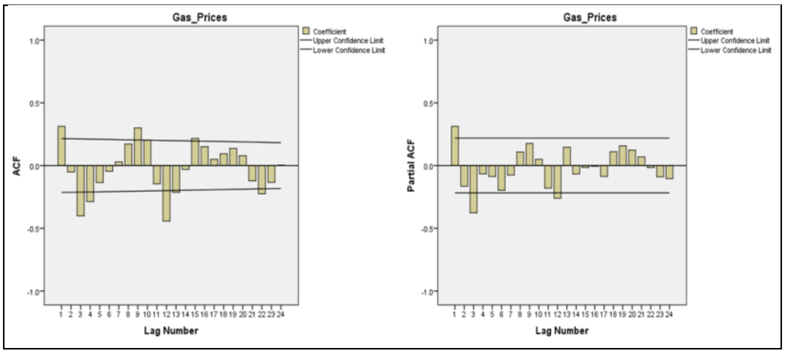

Figure 5: ACF and PACF plot with *d*=1 and D=1

From the ACF plot (Figure 5), there is significant autocorrelation at Lags 1 and 3 which will be the value of *q*. While the significant autocorrelation at Lag 3 indicates that the value of *Q*=1. From the PACF plot, there is significant autocorrelation at Lags 1 and 3 which will be the value of *p*. While the autocorrelation at Lag 3 presents but less significant, hence, it indicates that the value of *P* = 0 or *P* = 1. In addition, the PACF trails off after a lag and has a hard cut-off in the ACF after the lag *q*, at *p*=0 needed to be considered in the models.

In this study, the lags 12 on the order AR and MA were ignored. Table 4 displays parameter estimator (column 3) and the significance test of α value (column 4). The following hypothesis will be used to identify significant models:

H_0: $\alpha > 0.05$: Model insignificant

H_1: $\alpha < 0.05$: Model significant

There were 6 initial models being identified from Table 4, it is clearly shown that only SARIMA $(0,1,1)(1,1,0)_{12}$ has reached the assumptions of the model diagnostics.

Table 4: Significant parameters of identified models

Model: SARIMA	Parameters	Estimation	Significant	MAPE	Normalized BIC
$(1,1,0)(1,1,0)_{12}$	AR SAR	0.373 -0.564	0.001 0.000	7.703	8.692
$(1,1,1)(1,1,1)_{12}$	AR MA SAR SMA	0.216 -0.100 -0.224 0.525	0.552 0.784 0.274 0.009	7.318	8.749
$(1,1,1)(1,1,0)_{12}$	AR MA SAR	0.303 -0.095 -0.564	0.290 0.745 0.000	7.277	8.674
$(0,1,1)(1,1,0)_{12}$	MA SAR	-0.318 -0.554	0.003 0.000	7.352	**8.625**
$(0,1,1)(1,1,1)_{12}$	MA SAR SMA	-0.266 -0.181 0.613	0.014 0.349 0.002	6.885	8.584
$(1,1,0)(1,1,1)_{12}$	AR SAR SMA	0.301 -0.228 0.526	0.006 0.250 0.008	7.349	8.687

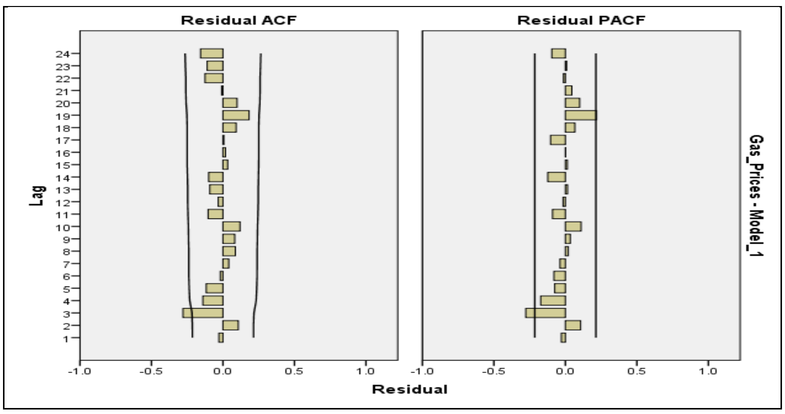

Figure 6: Residuals of ACF and PACF

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 3362-3370
https://publishoa.com
ISSN: 1309-3452

The residual plots of ACF and PACF in Figure 6 shows that, the model is adequate a random variation from the origin zero (0), the points below and above are all uneven, hence the model fitted is adequate.

Model parameters and fit statistics were estimated and the results of estimation are presented in Tables 5. BIC value was 8.625. Hence, the most suitable model for LPG priceswas SARIMA (0, 1, 1) $(1, 1, 0)_{12}$, as this model had the lowest normalized BIC value.

Table 5: ARIMA Model Parameters

		Estimate	SE	t	Sig.
Constant		4.010	6.636	0.604	0.547
Difference		1			
MA	Lag 1	-0.312	0.104	-3.002	0.004
AR, Seasonal	Lag 1	-0.551	0.100	-5.510	0.000
Seasonal Difference		1			

The forecasted plot of SARIMA$(0,1,1)(1,1,0)_{12}$ through Box-Jenkins method is shown in Figure 7 and the forecasted results were tabulated in Table 6.

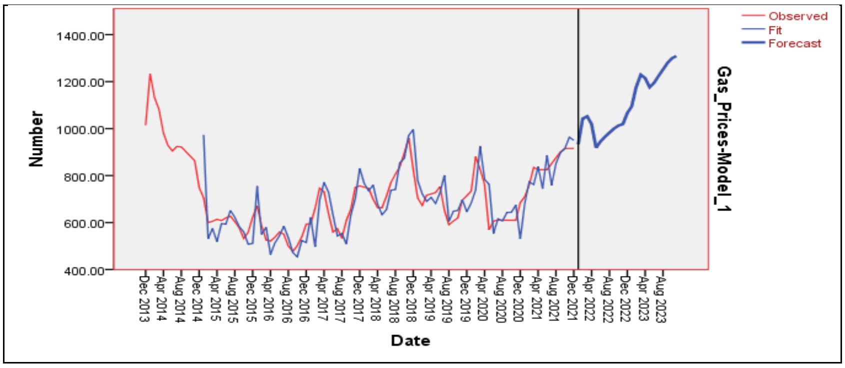

Figure 7: Actual and Estimated of LPG Prices

Table 6: Forecasting for LPG Prices

Year	Forecast	UCL	LCL	*Year*	Forecast	UCL	LCL
Dec 2021	900.20	995.53	725.50	*Dec 2022*	1066.72	1674.10	459.34
Jan 2022	933.13	1069.19	797.08	*Jan 2023*	1094.64	1747.54	441.74
Feb 2022	1042.85	1267.27	818.42	*Feb 2023*	1177.64	1879.86	475.42
Mar 2022	1052.38	1339.12	765.64	*Mar 2023*	1229.50	1977.80	481.20
Apr 2022	1018.55	1356.30	680.80	*Apr 2023*	1215.03	2006.73	423.33
May 2022	918.89	1300.90	536.88	*May 2023*	1176.55	2009.40	343.70

Jun 2022	945.51	1367.16	523.87	***Jun 2023***	1194.71	2066.76	322.66
Jul 2022	965.38	1423.24	507.52	***Jul 2023***	1223.91	2133.47	314.34
Aug 2022	982.81	1474.23	491.39	***Aug 2023***	1251.73	2197.33	306.14
Sep 2022	999.97	1522.80	477.15	***Sep 2023***	1279.44	2259.74	299.13
Oct 2022	1012.92	1565.37	460.47	***Oct 2023***	1299.74	2313.56	285.92
Nov 2022	1019.14	1599.71	438.58	***Nov 2023***	1308.75	2355.02	262.48

Conclusion

The results showed that prices would not remain stable throughout the year. The most appropriate SARIMA model for LPG price forecasting of data was found to be SARIMA(0,1,1)(1,1,0)$_{12}$. From the temporal data, it can be found that forecasted prices would increase to Rs. 1,308.75 in November 2023 from Rs.900.20 in December 2021 in Chennai. The results will be useful to the farming community and other relevant stakeholders. Using time series data from December 2013 to November 2021 on LPG prices, this study provides an evidence on future gas prices in the country, which can be considered for future policy making and formulating strategies for augmenting and sustaining prices in India.

References

1. C. Jablonowski and R. MacAskie, The value of oil and gas price forecasts, presented at SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, April 1-3, 2007, pp. 1-8, SPE Publication # 107570.
2. E. Slutzky, The summation of random causes as the source of cyclic processes, *Econometrica*, (1973), no. **5**, 105-146.
3. E.J. Hannan, The estimation of the order of an ARMA process, *Annuals of Statistics*, (1980), no. **8**, 1071-1081.
4. E.J., Hannan and B.G. Quinn, The determination of the order of an autoregression, *Journal of Royal Statistical Society*, (1979), no. **B(41)**, 190-195.
5. G.E.P. Box and D.A. Pierce, Distribution of Residual Autocorrelations in ARIMA Models, *J. American Stat. Assoc.*, (1970), no. **65**, 1509-1526.
6. G.E.P. Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control, San Francisco, Holden-Day, California, USA, 1976.
7. H. Akaike, Statistical Predictor Identification, *Annals of Institute of Statistical Mathematics*, (1970), no. **22**, 203-270.
8. Hollada Jacqueline, N. Williams Kendra, H. Miele Catherine, Danz David, A. Harvey Steven and Checkley William, Perceptions of Improved Biomass and Liquefied Petroleum Gas Stoves in Puno, Peru: Implications for Promoting Sustained and Exclusive Adoption of Clean Cooking Technologies, *International Journal of Environmental Research and Public Health*, (2017), no. **14(2)**, 182.
9. I.S. Agbon and J.C. Araque, Predicting oil and gas spot prices using chaos Time Series Analysis and Fuzzy Neural Network model, *SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas*, (2003).
10. J.R.M. Hosking, Fractional differencing, *Biometrika*, (1981), no. **68(1)**, 165-176.
11. Jain Abhishek, LPG for every Indian household, *The Hindu*, (2016), URL: http://www.the hindu.com/opinion/op-ed/LPG-for-every-Indian- household/article14169579.ece
12. M Zamani, An Econometrics Forecasting Model of Short Term Oil Spot Price, Paper presented at the 6th IAEE European Conference, Zurich, 2004.
13. Mukhopadhyay Rupak, Sambandam Sankar, Pillarisetti Ajay, Jack Darby, Mukhopadhyay Krishnendu, Balakrishnan Kalpana, Vaswani Mayur, N. Bates Michael, L. Kinney Patrick and Arora Narendra, Cooking Practices, Air Quality, and the Acceptability of Advanced Cook-stoves in Haryana, India: An Exploratory Study to Inform Large-Scale Interventions. *Global Health Action*, (2012), no. **5(1)**, 19016.

14. O.I. Solomon, M.K. Kunju, and O.I. Omowunmi, The responsiveness of global E&P industry to changes in Petroleum prices: Evidence from 1960-2000, *SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas*, (2001). pp 1-12. SPE68587.
15. S. Pindyck Robert, Inventories and the Short-Run Dynamics of Commodity Prices, *RAN Journal of Economics,* (1994), no.**25**, 14.
16. Stephane Dees, Sanchez, Marcelo, Karadeloglou, Pavlos, and Kaufmann, K. Robert, Modeling the world oil market: Assessment of a quarterly econometric model, *Energy Policy*, (2003) no. **35(1)**, 178-191.
17. Troncoso Karin, and da Silva Agnes Soares, LPG Fuel Subsidies in Latin American and the Use of Solid Fuels to Cook. *Energy Policy*, (2017), no. **107**, 188–196.
18. U.J.O. Ogwo, Equivalent Gas Pricing Model. SPE 111897, presented at 31st Nigerian Annual International Conference and Exhibition held in Abuja, Nigeria, (2017).