Stochastic Forecasting and Modeling of Volatility Oil Prices in Ghana using ARIMA Time Series Model

Godfred Kwame Abledu[1*] and Agbodah Kobina[2]

1. School of Applied Science and Technology, Koforidua Polytechnic, PO Box 981, Koforidua, Ghana
2. Department of Applied Mathematics, Koforidua Polytechnic, PO Box 981, Koforidua, Ghana

* E-mail of the corresponding author: godfredabledu@gmail.com (Tel: 233-244705093)

Abstract

Ghana's demand for crude oil and refined petroleum products has been growing over the past decade. This growth has been driven by socio-economic and technical factors that have influenced each category of final energy use. The growing urban population is demanding new vehicles and new roads, raising the demand for energy in the transportation and all other sectors of the economy. Consequently, Oil prices rose from 2004 to historic highs in mid-2008, only to fall precipitously in the last four months of 2008 and lose all the gains of the preceding four and a half years. The steep price increase was challenging for all economies including Ghana. The high price of oil will invariably affect revenue mobilisation, expenditure, and therefore the fiscal position of government and inflation.

The study is an attempt to forecast and analyse the macroeconomic impact of oil price fluctuations in Ghana using annual data from 2000-2011. It focuses on studying the feasibility forecast using nested conditional mean (ARIMA) and conditional variance (GARCH, GJR, EGARCH) family of models under such volatile market conditions. A regression based forecast filtering simulation is proposed and studied for any improvements in the forecasted results.

Keywords: ARIMA model, stochastic, volatility, forecast, crude oil, price shocks

1. Introduction

Oil is very important in Ghana's economy. It runs in every industry and every utility device in our modern technology all over the world. It is obvious that any increase in oil prices will trickle down to each and every part of the economy causing inflation in every sector. Oil policies form an important part of national policy in all oil consuming countries. Some countries provide subsidy on oil and petroleum products to promote domestic industries and check inflation, whereas some countries impose tax on oil consumption to check demand. Economies all over the world constantly monitor oil price movements (WAMA, 2008)..

In Ghana, demand for oil and refined petroleum products has been growing over the past decade. This growth has been driven by socio-economic and technical factors that have influenced each category of final energy use. The growing urban population is demanding new vehicles and new roads, raising the demand for energy in the transportation and all other sectors of the economy. Consequently, Oil prices rose from 2004 to historic highs in mid-2008, only to fall precipitously in the last four months of 2008 and lose all the gains of the preceding four and a half years. The steep price increase was challenging for all economies including Ghana. The high price of oil will invariably affect revenue mobilisation, expenditure, and the fiscal position of government and inflation(Banapurmath, et. al., 2011; WAMA, 2008).

The objective of the study is to determine model(s) that explain(s) the observed data and allow(s) extrapolation into the future to provide a forecast. In this regard, we will be looking at the family of ARIMA models. Auto Regressive Integrated Moving Average (ARIMA) models have been already applied to forecast commodity prices (Weiss, 2000), such as oil (Morana, 2001) or natural gas (Buchananan, et. al.,2001). In power systems, ARIMA techniques have been used for load forecasting (Gross and Galiana, 1987), (Hagan and Behr, 1987) with good results. Currently, with the restructuring process that is taking place in many countries, simpler Auto Regressive (AR) models are also being used to predict weekly prices, like in the Norwegian system(Fosso, et. al., 1999)

2. Materials and method

2.1. The ARCH Model

The first model that provides a systematic framework for volatility modeling is the ARCH model of Engle (Gujarati, 2006). The basic idea is the ARCH model is that the shock a_t of an asset return is serially uncorrected but dependent; also the dependence of a_t can be described by a simple quadratic function of its lagged values. Specifically, an ARCH (m) model assume that

$$a_t = \sigma_t \varepsilon_t \ , \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_{t-m}^2$$ ((Gujarati, 2006).

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (iid) random variables with mean zero and variance 1, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$. The coefficient α_i must satisfy some regularity condition to ensure that the unconditional variance of a_t is finite. In practices, ε_t is often assumed to follow the standard normal or a standardized student $-t$ distribution or a generalized error distribution.

From the structure of the model, it is seen that large past squared shocks $\{a_{t-i}^2\}_{i=1}^m$ imply a large conditional variance σ_t^2 for the innovation a_t. Consequently, a_t tends to assume a large value (in modulus). This means that, under the ARCH framework, large shock tend to be followed by another shock; because a large variance does not necessarily produce a large realization. It only says that the probability of obtaining a large variate is greater than that of a smaller variance.

To understand the ARCH models, it pays to carefully study ARCH (I) model

$$a_t = \sigma_t \varepsilon_t \ , \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2,$$

Where $\alpha_0 > 0$ and $\alpha_l \geq 0$.

The unconditional mean of a_t remains zero because

$$E(a_t) = E\left[E(a_t / F_{t-1})\right] = E\left[\sigma_t E(\varepsilon_t)\right] = 0$$

The conditional variance if a_t can be obtained as

$$\operatorname{var}(a_t)=E(a_t^2)=E\left[E\left(a_t^2/F_{t-1}\right)\right]=E\left(\alpha_0+\alpha_1 a_{t-1}^2\right)=\alpha_0+\alpha_1 E\left(a_{t-1}^2\right).$$

Because a_t is a stationary process with $E\left(a_{t-1}^2\right)=0$

$$\operatorname{var}(a_t)=\operatorname{var}(a_{t-1})=E\left(a_{t-1}^2\right)$$

Therefore, we have $\operatorname{var}(a_t)=\alpha_0+\alpha_1\operatorname{var}(a_t)$ and $\operatorname{var}(a_t)=\dfrac{\alpha_0}{1-\alpha_1}$. Since the variance of a_t must be positive, we require $0\le\alpha_1\le 1$. In some applications, we need higher order moments of a_t to exist and, hence, α_1 must also satisfy some additional constraints. For instance, to satisfy its tail behavior, we require that the fourth moment of a_t is finite. Under the normality of in we have

$E\left(a_t^4/F_{t-1}\right)=3\left[E\left(a_t^2/F_{t-1}\right)\right]^2=3\left(\alpha_0+\alpha_1 a_{t-1}^2\right)^2$ (Brockwell and Davis, 1996).

Therefore , $E\left(a_t^4\right)=E\left[E\left(a_t^4/F_{t-1}\right)\right]=3E\left(\alpha_0+\alpha_1 a_{t-1}^2\right)^2=3E\left(\alpha_0^2+2\alpha_0\alpha_1 a_{t-1}^2+\alpha_1^2 a_{t-1}^4\right)$

If a_t is fourth – order stationary with

$m_4=E\left(a_t^4\right)$, then we have

$$m_4=3\left[\alpha_0^2+2\alpha_0\alpha_1\operatorname{var}(a_t)+\alpha_1 m_4\right]$$

$$=3\alpha_0^2\left(1+2\frac{\alpha_1}{1-\alpha_1}\right)+3\alpha_1^2 m_4$$

Consequently

$$m_4=\frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)\left(1-3\alpha_1^2\right)}.$$

This result has two important implications: since the fourth moment of a_t is positive, we see that α_1 must also satisfy the condition $1-3\alpha_1^3>0$; that is, $0\le\alpha_1^2\le 1/3$; and the unconditional Kurtosis of a_t is

$$\frac{E\left(a_t^4\right)}{\left[\operatorname{var}(a_t)\right]^2}=3\frac{\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)\left(1-3\alpha_1^2\right)}\times\frac{(1-\alpha_1)^2}{\alpha_0^2}=3\frac{1-\alpha_1^2}{1-3\alpha_1^{2>3}}$$

Thus, the excess of a_t is positive and the tail distribution of a_t is heavier than that of a normal distribution. In other words, the shock a_t of a conditional Gaussian ARCH (I) model is more likely than Gaussian white noise series to produce "outcome". This is in agreement with the empirical finding that "outlines" appear more often in asset returns than that implied by an iid sequence of normal random variates. These properties continue to hold for general ARCH models, but the formula becomes more complicated for higher order ARCH models.

The condition $\alpha_i\ge 0$ in $\sigma_t^2=\alpha_0+\alpha_1 a_{t-1}^2+....+\alpha_m a_{t-m}^2$ can be related. It is a condition to ensure that the conditional variance σ_t^2 is positive for all t. The model has some weakness: it assume that positive and negative shocks have the same effects on volatility because it depends on the square of the previous shocks. In practices it is well known that price of a financial asset responds differently to positive and negative shocks.

The ARCH model is rather restrictive. For instance, α_1^2 of an ARCH (I) model must be in the interval $\left[0\ \frac{1}{3}\right]$ if the series has a finite fourth moment. The constraint becomes complicated for higher order ARCH models. In limits, the ability of ARCH models with Gaussian innovations is to capture excess kurtosis. The ARCH model does not provide any new insight for understanding the sources of variation of a financial time series. It merely provides a mechanical way to describe the behavior of the conditional variation. It gives no indications of what causes such behavior to occur. ARCH models are likely to over predict the volatility because they respond slowly to large isolated shocks to the return series(Brockwell and Davis, 2002)..

2.2. The GARCH Model

Although the ARCH is simple, it often requires many parameters to adequately describe the volatility process of an asset return some alternative models must be sought. Shrivastava, et al. (2010) and Hull(2006) proposed a useful extension known as the generalized ARCH (GARCH) model. For a long return series γ_t , let $a_t = \gamma_t - u_t$ be the innovation at time t. Then a_t follows a GARCH (M,S) model if $a_t = \sigma_t \epsilon_t$,$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i\, a_{i-1}^2 + \sum_{j-1}^{s} \beta_j \sigma_{t-j}^2$

Where again $\{\epsilon_t\}$ is a sequence of iid random variables with mean 0 and variance 1.0,

$\alpha_0 > 0;\ \alpha_i\ \geq 0\ ;\ \beta_j\ \geq 0$ and $\sum_{i=1}^{\max\,(m,s)} \alpha_i + \beta_i\ < 1$

2.3. The EGARCH Model

This model is used to allow for symmetric effects between positive and negative asset returns. An EGARCH (m, s) model can be written as (Dhar, et. Al. , 2009).

$$a_t = \sigma_t \varepsilon_t\ , \quad In\left(\sigma_t^2\right) = \alpha_0 \frac{1+\beta_1 B + ... + \beta_{s-1} B^{s-1}}{1-\alpha_1 B....\alpha_m B^m} g\left(\varepsilon_{t-1}\right)$$

wher α_0 is a constant, B is the back-shift (or lag) operator such that $Bg(\epsilon_t) = g(\epsilon_{t-1})$ and $1 + \beta_{1B+\cdots+\beta_{s-1}B^{s-1}}$ *and* $1 - \alpha_{1B-\cdots-\alpha_m B^m}$ are polynomials with zeros outside the unit circle and have no common factors. By outside the unit circle, we mean that absolute values of the zeros are greater than 1. Here, it is understood that $\alpha_i = 0\ for\ i > m$ and $\beta_j = 0\ for\ j > s$. The latter constraint on α_i and β_i implies that the unconditional variance α_t is finite, whereas its conditional variance σ_t^2 evolves over time, and ϵ_t is often assumed to be a standard normal standardized student-t distribution or generalized error distribution.

$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-1}^2 + \sum_{i=1}^{m} \beta_j\, \sigma_{t-j}^2$

reduces to a pure ARCH (m) model if S=0.

The α_i *and* β_j are referred to as ARCH and GARCH parameters respectively. The unconditional mean of $\ln(\sigma_t^2)$ is α_0. It uses logged conditional variance to relax the positiveness constraint of model coefficients. The use of $g(\epsilon_t)$ enables the model to respond asymmetrically to the positive and negative lagged values of α_t .

The model is nonlinear if $\theta \neq 0$. Since negative shocks tend to have larger impacts, we expect θ to be negative. For higher order EGARCH model, the nonlinearity becomes much more complicated. This model can be used to obtain multistep ahead volatility forecasts.

2.4. The ARIMA Models

ARIMA processes are a class of stochastic processes used to analysis time series. The application of the ARIMA methodology for the study of time series analysis is due to Box and Jenkins. The proposed general formulation is

$\emptyset(B)P_t = \theta(B)\epsilon_t$ (Shrivastava, et, al. , 2010; Abu and Behrooz, 2011)

where

P_t is the price of at time t,

$\emptyset(B)$ *and* $\theta(B)$ are functions of the back-shift operator B: $B_{P_t}^{\ell} = P_t - \ell$, ϵ_t is the error term.

Functions $\emptyset(B)$ *and* $\theta(B)$ have special forms. They contain factors of polynomial functions of the form $\emptyset(B) = 1 - \sum^{\Phi} \emptyset_{\ell} B^{\ell}$ and /or $\theta(B) = 1 - \sum_{\ell=1}^{\alpha} \theta_{\ell} B^{\ell}$ and/or $(1 - B^s)$. Where several values of ϕ_{ℓ} and θ_{ℓ} can be set to 0. The error term ϵ_t is assumed to be a randomly drawn series from a normal distribution with zero mean and constant variance σ^2, that is a white noise process. Frequently, a diagnosis check is used to validate this assumption.

2.5. Fitting the Parameters of the Model

Once a model is selected and data are collected, it is the job of the researcher to estimate the parameters of the Model. These are values that best fit the historical data. It is hypothesized that the resulting model will provide a prediction of the future observation. It is also hypothesized that all values in a given sample are equal.

The time series model includes one or more parameters. We identify the estimated values with a hat. For instance, the estimated value of β is denoted $\hat{\beta}$ The procedure also provide estimates of the standard deviation of the noise, σ_{ε}.

2.6. Forecasting from The Model

The main purpose of modeling a time series is to make a forecast which are then used directly for making decisions. In this analysis, we let the current time be T, and assume that the prices for periods 1 through T are known. We now want to forecast the price for the period (T+ς). The unknown price is the random variable $X_{(T+\varsigma)}$., and its realization is $x_{(T+\varsigma)}$.. Our forecast for the realization is $\hat{x}_{T+\varsigma}$.

2.7. Measuring the Accuracy of the Model

The forecast error is the difference between the realization and the forecast. Thus

$$e_{\varsigma} = x_{(T+\varsigma)} - \hat{x}_{T+\varsigma}.$$

Assuming the model is correct, then we have

$$e_{\varsigma} = E[X_{T+\varsigma}] + \varepsilon_{\varsigma} - \hat{x}_{\varsigma}$$

We investigate the probability distribution of the error by computing its mean and variance. One desirable characteristics of the forecast $\hat{x}_{T+\varsigma}$ is that it is unbiased. For an unbiased estimate, the expected value of the forecast is the same as the expected value of the time series. Because ε_t is assumed to have a mean of zero, an unbiased forecast implies $E[\varepsilon_{\varsigma}]$. The fact that the noise in independent from one period to the next period means that the variance of the error is:

$Var[\varepsilon_t] = Var\{E[X_{T+\varsigma}] - \hat{x}_{T+\varsigma}\} + Var[\varepsilon_{T+\varsigma}]$ and $\sigma_{\varepsilon^2}(\varsigma) = \sigma_{E^2}(\varsigma) + \sigma^2$.

3. Data analysis and results

The data employed in this study are the yearly Ghana National Petroleum Authority(GNPA) oil prices from, 1999 to 20011.

Figure 1 displays the yearly Ghana National Petroleum Authority(GNPA) oil prices from, 1999 to 2011. Figure 1 depicts a sturdy upward rise in oil prices. There is also an indication of a positive trend, implying that future oil prices will continue to increase. Table 2 presents descriptive summary of the yearly Ghana National Petroleum Authority oil prices from, 1999 to 2011..

A kurtosis value of -0.162 is an indication of the presence of platy kurtosis in the probability distribution of the series. This means that the presence of tiny tails in the probability distribution of oil price series. In a tiny-tailed distribution, there is a lower-than-normal likelihood of a big positive or negative realization which represents non-symmetry. The autocorrelation and partial autocorrelation structures provide a summary of the dynamics of oil price series.
The standard deviation value of GHc 7,096.29 is an indication of the large dispersion of oil price about the mean value of GHc13,866.00. This value shows how the various oil prices differ from the mean value. It could be realized that about half of oil prices are below GHc10,284.00 per litre. It can be noticed that the coefficient of skewness is positive (0.879). This is an indication that oil price distribution is positively skewed. That is, it has a long tail tapering to the right which confirms the fact that the mean is larger than the median.

3.1. Forecasting

Figure 2 shows the plot of the differenced series and it is clear from the plot that the assumption of stationarity is reasonably.

To confirm this and to suggest possible models, we need to examine the ACF and PACF functions. These are shown in Figures 3 and 4. The model here is less certain indicating that the ACF is showing a rapid decay and the PACF is showing a sharp cutoff after one lag. This suggests that the original series can be modeled as ARIMA (1,1,0) series. The fitted model in this case is $\nabla Y_t = 0.457\nabla Y_t + \varepsilon_t$, where

ε_t is an estimated variance of 84.23948. The model is used to predict the oil price from 2012 to 2016. Tables 3 and 4 show the results of the in-sample and out-sample forecasts for the ARIMA model.

4. Conclusion

To study the long-term memory, DFA method was employed to GNPA oil price series, and evidence was found that there existed comparatively strong long-term memory in it. An ARIMA (p,d,q) model was selected and estimated automatically using the Hyndman-Khandakar algorithm to select p and q and the Haslett and Raftery algorithm to estimate the parameters including d. the best model was ARIMA (1,1,0) which was used to predict the of oil prices in GNPA till the end of 2016.

The study has an important implication for managers of the OMC. Decision-makers choosing an appropriate model for their monthly and quarterly arrivals forecast should include an ARIMA-based model in their consideration. The results of the study would serve as a possible aid for policymakers in responding to oil price shocks.

References

Abu H. S. N. and Behrooz, G., (2011): Application of Dynamic Models for Exchange Rate Prediction. International Journal of Innovation, Management and Technology, Vol. 2, No. 6, December.

Banapurmath, N. R. et. Al.(2011): Combustion characteristics of a four-stroke CI engine operated on Honge and Jatropha oil methyl ester-ethanol blends when directly injected and dual fuelled with CNG induction. *International Journal of Sustainable Engineering 4, no. 2 : 145-152*

Brockwell, P.J. and R.A. Davis, (2002): *Introduction to time series and forecasting*, second edition Springer-Verlag, New York.

Brockwell, P.J. and R.A. Davis, (1996): *Time series: theory and methods*, second edition Springer-Verlag, New York.

Buchananan W. K., Hodges P., and Theis, J. (2001): "Which way the natural gas price: An attempt to predict the direction of natural gas spot price movements using trader positions," *Energy Economics*, vol. 23, no. 3, pp. 279–293.

Dhar, J. et. al. (2009): "Simulative approach to Constant Mean and Conditional Variance Heteroscedastic model selection analysis using Likelihood Ratio Test for Indian Market Returns", Proc. *IEEE Advance Computing Conference, Patiala*.

Fosso, O. B., Gjelsvik, A., Haugstad, A., Birger, M., and Wangensteen, I.(1999):"Generation scheduling in a deregulated system. The norwegian case," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 75–81.

Gujarati, D.(2006): *Essentials of Economics(3rd edition).* McGraw Hill International.

Gross G. and Galiana, F. D. (1987): "Short-Term load forecasting," *Proc. IEEE*, vol. 75, no. 12, pp. 1558–1573.

Hagan M. T. and. Behr S. M(1987): "The time series approach to short term load forecasting," *IEEE Trans. Power Syst.*, vol. 2, pp. 785–791,

Hull, J.(2006): *Options, Futures and Other Derivatives*. New Delhi: Pearson Education

Javier C., Rosario ,E., Francisco J.. Nogales, and Antonio J. C. (2003): ARIMA Models to Predict Next-Day Electricity Prices. *IEEE Transactions On Power Systems,vol.18, NO. 3,*

Morana, C.(2001): "A semiparametric approach to short-term oil price forecasting," *Energy Economics*, vol. 23, no. 3, pp. 325–338.

Shrivastava, et, al. , (2010):Regression Based Approach to Filter Conditional Mean and Variance Model Forecast of Stock Market Returns. International Research Journal of Finance and Economics ISSN 1450-2887 Issue 5.

WAMA(2008):Impact of Petroleum Price Fluctuations on Key Convergence Criteria in ECOWAS Member States. *West African Monetary Agency*.

Weiss, E. (2000): "Forecasting commodity prices using ARIMA," *Technical Analysis of Stocks & Commodities*, vol. 18, no. 1, pp. 18–19.

Table 1: Retail Prices Of Petroleum Produts:1999 - 2006 (Prices in Cedis per litre)

Year	LPG	Premium Gasoline	Kerosene	Diesel	Residual Fuel Oil	Total
1999	2,070	1,420	1,325	1,325	1,031	7,171
2000	2,070	1,420	1,325	1,325	1,031	7,171
2001	2,200	2,321	2,464	1,956	1,343	10,284
2002	2,200	2,320	2,464	1,956	1,343	10,283
2003	3,800	4,444	3,889	3,889	1,927	17,949
2004	3,800	4,444	3,889	3,889	1,927	17,949
2005	5,384	6,852	5,036	6,133	2,850	26,255
2006	6,200	8,056	5,610	7,253	3,311	30,430
2007	6244	7522	6156	7390	3312	32192
2008	6862	8380	6802	8263	3638	35955
2009	7480	9239	7448	9135	3964	39719
2010	8097	10097	8094	10007	4291	43483
2011	8715	10955	8739	10879	4617	47247

Table 2: Descriptive Statistics of yearly GNPA oil prices

Value	Mean	Median	Minimum	Maximum	Standard Deviation	Skewness	Kurtosis
Statistic	13,866.00	10,284.00	7,171.00	26,255.00	7,096.29	0.879	-0.162

Table 3: Forecast for the ARIMA Model from the Sample Data

Date	**Actual**	**Forecast**	**Error**
1999	7171	2081.25	5089.75
2000	1717	5845.107	-4128.11
2001	10284	9608.964	675.036
2002	10283	13372.82	-3089.82
2003	17949	17136.68	812.32
2004	17949	20900.54	-2951.54
2005	26255	24664.39	1590.61
2006	30430	28428.25	2001.75
2007	32192	32192.11	-0.11
2008	35955	35955.96	-0.96
2009	39719	39719.82	-0.82
2010	43483	43483.68	-0.68
2011	47247	47247.53	-0.53

Table 4: Forecast for the ARIMA Model from 2012 to 2015

Date	Forecast
2012	51011.39
2013	54775.25
2014	58539.11
2015	62302.96
2016	66066.82

Figure 1: GNPA oil price data from 1999 to 2011.

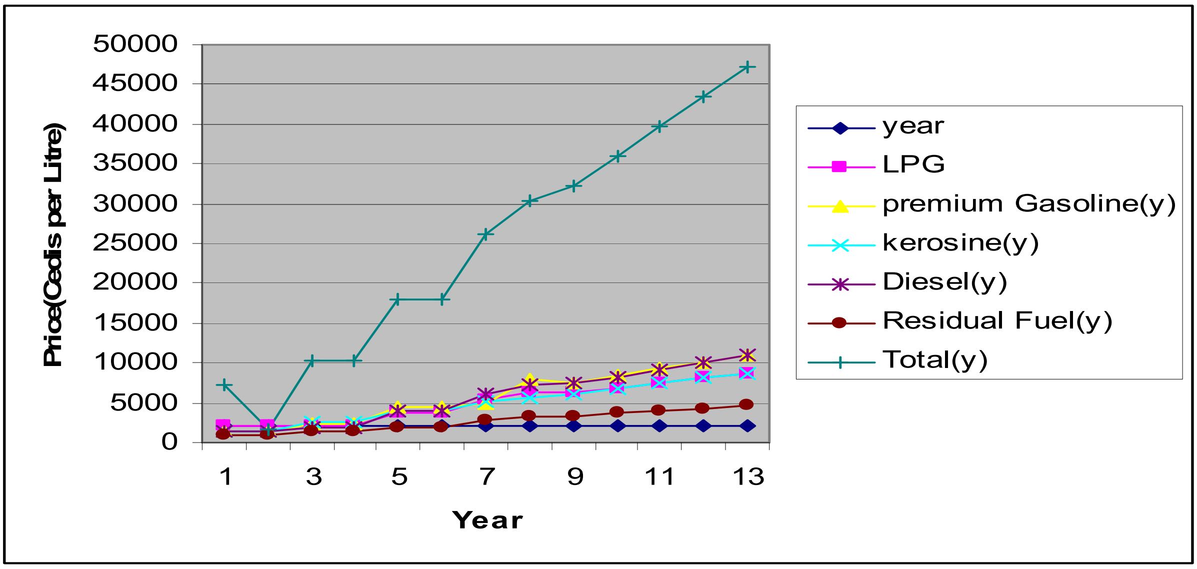
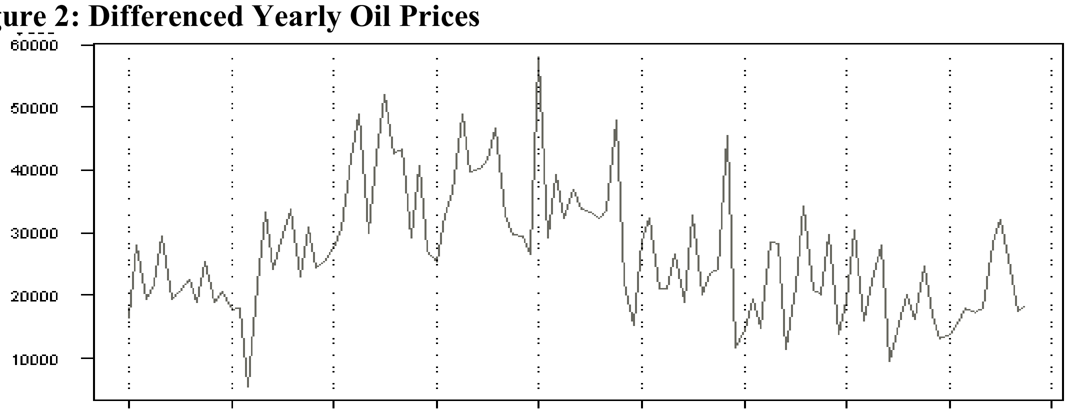


Figure 2: Differenced Yearly Oil Prices



About the Authors:

1. Godfred Kwame Abledu holds BEd and MPhil degrees in Mathematics Education in 1994 and 2000 respectively from University of Cape Coast, Ghana and a PhD in Statistics in 2012 from the Atlantic International University, USA. Currently, he is a lecturer in the School of Applied Science and Technology, Koforidua Polytechnic, Koforidua, Ghana.

2. Agbodah Kobina holds a Master of Science degree in Statistics from Regent University College Accra - Ghana, and a Bachelor of Science in Statistics from the University of Cape Coast, Ghana He is currently a lecturer at the Department of Applied Mathematics.