IASET: Journal of Applied Mathematics
& Statistical Sciences (IASET: JAMSS)
ISSN(P): Applied; ISSN(E): Applied
Vol. 2, Issue 1, Jan - Jun 2017; 1-8
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

PREDICTIVE MODEL WITH SQUARE-ROOT VARIANCE STABILIZING TRANSFORMATION FOR NIGERIA CRUDE OIL EXPORT TO AMERICA

O. D. ADUBISI[1], T. T. MOM[2], C. E. ADUBISI[3] & P. LUKA[4]

[1,2,4] Department of Mathematics and Statistics, Federal University Wukari, Taraba State, Nigeria

[3]Department of Physics, University of Ilorin, Ilorin, Kwara State, Nigeria

ABSTRACT

In the last few decades, crude oil claims to be in the topmost position in the Nigerian export list, constituting a very fundamental change in the structure of Nigerian international trade. In this study, secondary data on monthly crude oil export to the United States was obtained, from the Energy Information Administration (EIA) database. Using the Box-Jenkins (ARIMA) methodology, the results showed that, Seasonal ARIMA (0, 1, 1) $(1, 0, 1)_{12}$ model had the least information criteria, after the data was Square-Root transformed and non-seasonally first differences, in order to achieve series stationarity. The diagnostic tests on the selected model residuals, using the Ljung-Box, Shapiro-Wilk Normality and ARCH-LM tests revealed that, the residuals are Gaussian white noise.

KEYWORDS: Transformation, SARIMA, Unit Root, Crude Oil Export, ARCH-LM

1. INTRODUCTION

Crude oil is considered as the major source of energy in Nigeria and the world, in general. Crude oil, being the mainstay of the Nigerian economy, plays a vital role in shaping the economic and political destiny of the country. Nigeria exports most of its crude oil to countries like India, United States of America, Brazil, The Netherland, United Kingdom and Spain (Central Intelligence Agency, 2013). As recently as 2010, Nigeria provided about 10 percent of over-all United States oil imports, and ranked as the fifth-largest source of oil imports in the United State. However, Nigeria's crude oil export to the United States has recently been declining, as a result of the boom in shale oil and the U.S. Senate lifting its 40-year ban on crude oil exports (Vanguard-Nigeria, 2016). Therefore, the ultimate aim of this study is to construct a statistical model, that could be used to monitor the export pattern of crude oil export from Nigeria to the United States. Using this model, forecast of future values of crude oil export to the United States can be obtained. A lot of studies have been carried out using time series ARIMA model approach, to identify patterns and appropriate models. Adubisi (2016) used ARIMA procedure, in modelling the growth pattern of reserve currency in Nigeria, Smart (2013) explored the feasibility for application of Box-Jenkins Approach (ARIMA), in modelling and forecasting maternal mortality Ratios (MMR), ARIMA modelling approach was used to model yearly exchange rates between USD/KZT, EUR/KZT and SGD/KZT, and the actual data compared with developed forecasts by Daniya (2014), Kumar and Anand (2012) used ARIMA modelling approach, to forecast sugarcane production in India. Adubisi and Jolayemi (2015), used ARIMA-intervention analysis modelling approach, to evaluate and estimate the impact of the financial crisis on Nigeria's Crude oil export. Bakari et al. (2013) used ARIMA modelling procedure, to build a model for annual production and utilization of gas from Nigeria National petroleum company (N.N.P.C) and Adubisi et al; (2017) used the seasonal ARIMA to model the Nigeria money in circulation series, and also produced a three years forecast values using the fitted model.

2. MAIN RESEARCH

2.1. Material

The data used for this study are secondary data, on Nigeria's monthly crude oil export to the United States, obtained from the Energy Information Administration (EIA) database for twenty-three consecutive years, from January 1993 to December 2015. (www.eia.gov)

2.2. Series Transformation

The parametric family of transformations from the original series, to a transformed series was originally proposed by Box and Cox, 1964. The power transformer is a continuously varying function, with respect to the power parameter λ, in a piecewise function form, that makes it continuous at the point of singularity $(\lambda = 0)$. Suppose, we observe a $(n \times 1)$ vector of observations $(x_1, \dots, x_n)$ in which each $x_i > 0$, the power transform is given as

$$x_t^{(\lambda)} = \begin{cases} \dfrac{(x_t^{\lambda} - 1)}{\lambda(GM(x))^{\lambda - 1}} & , \lambda \neq 0 \\ GM(x) \log x_t & , \lambda = 0 \end{cases} \quad (1)$$

$GM(x) = (x_1, \dots, x_n)^{\frac{1}{n}}$ is the Geometric mean of the data. It generalizes both the square root and the log transformation, and admits a likelihood ratio test to select the best fitting parameter. The one-parameter Box-Cox transformation is expressed as

$$x_t^{(\lambda)} = \begin{cases} \dfrac{(x_t^{\lambda} - 1)}{\lambda} & , \lambda \neq 0 \\ \log x_t & , \lambda = 0 \end{cases} \quad (2)$$

For more details on Box-Cox variance stabilization transformation procedures, see Box and Cox (1964), Yan (2015), Carroll and Ruppert (1987), Nishili (2001), Sakia (1992) and Bickel et al; (1981). The various transformation parameter lambda (λ) values and the appropriate transformation attached to each are summarized in the Table 1.

Table 1: Transformation for some Values of Parameter Lambda (λ)

S/No.	1	2	3	4	5	6	7
Lambda (λ)	0	0.5	-0.5	-1	1	-2	2
Transformation	$\log x_t$	$\sqrt{x_t}$	$1/\sqrt{x_t}$	$1/x_t$	No Transformation	$1/x_t^2$	x_t^2

Source: Minitab (2010)

2.3. Box-Jenkins Methodology

The Autoregressive integrated moving average (ARIMA) model procedure, popularized by Box and Jenkins (1976) and Box *et al.* (1994). The ARIMA (p, d, q) model, which contains both the non-seaso parameters and seasonal parameters is written as

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \quad (3)$$

The observed values is Y_t, (B) represent the Backshift operator, t is the time, (ϕ_p, Φ_P) represent the non-seasonal and seasonal autoregressive coefficient parameters with the roots within the unit circle

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

(4)

$$\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \ldots - \Phi_P B^{PS} \quad \textbf{(5)}$$

(θ_q, Θ_Q), represent the non-seasonal and seasonal moving-average coefficients parameters with the roots within the unit circle

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q \quad \textbf{(6)}$$

$$\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots - \Theta_Q B^{QS} \quad \textbf{(7)}$$

While $(1-B)^d$ being the regular differencing, which is applied to remove the stochastic trend in the series, $(1-B^S)^D$ is the seasonal differencing, applied to remove the series seasonal effects and ε_t is the white noise error i.e. $\varepsilon_t \sim WN(0, \sigma^2)$. More details on seasonal ARIMA can be found in Box and Jenkins (1976), Box *et al*. (1994), Pankratz (1983).

2.4. Stationarity Tests

The Augmented Dickey-Fuller (ADF) test and the KPSS tests are performed to determine, if the series contains a unit root. The tests are based on the assumption that, the time series data follows a random walk. The Augmented Dickey-Fuller (ADF) test, corresponding to modelling a random walk pattern, with drift around a stochastic trend

$$Y_t = \alpha + \rho Y_{t-1} + \sum_{i=1}^{p-1} \partial_i \nabla y_{i-i} + \beta t + \varepsilon_t \quad \textbf{(8)}$$

The expression $\rho y_{t-1} + \sum_{i=1}^{p-1} \partial_i \nabla y_{i-i}$ is the augmented part, y_{t-1} is the lagged term, ∇y_{t-1} shows the lagged change, t and α represent the deterministic trend and drift components, respectively, the ε_t is the error term and ρ, ∂ are coefficients to be estimated. If $\rho = 1$ the model is said to be non-stationary. The null hypothesis is ($\rho = 1$ or $\partial = 0$) against the alternative ($\rho < 1$ or $\partial < 0$). When the p-value is greater than the alpha, this would lead to none rejection of the null hypothesis. The KPSS test with a random walk $\alpha_t = \alpha_{t-1} + \varepsilon_t$ allowed is expressed as

$$Y_t = \alpha_t + \beta_t + \varepsilon_t \quad \textbf{(9)}$$

The procedure has a null hypothesis of stationary series and an alternative of non-stationary series. A p-value less than alpha (α) at 5% of significance, from the result of the KPSS test would be enough to reject the null hypothesis. Details on ADF and KPSS see, Dickey and Fuller (1979) and Kwiatkowski, et al; (1992).

3. RESULTS AND DISCUSSIONS

The series plot in Figure 1, depicts sharp peaks and troughs in the crude oil export series over the periods in-view, suggesting some influence and also the variance was observed and found not stable over the periods.

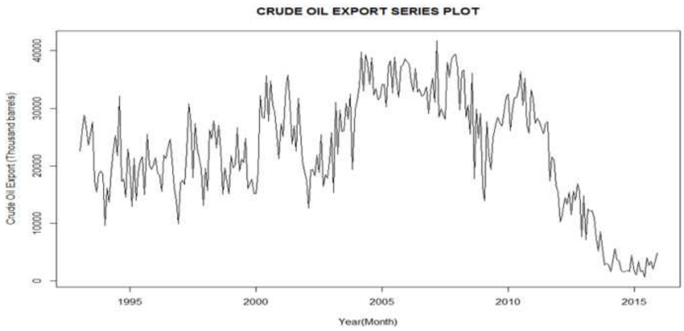

Figure 1: Crude Oil Export Series Plot

The non-stationarity claim from the series plot was affirmed from the slow decay, in the autocorrelation function (ACF) of the data, and a significant spike at lag 1 and 2 of the partial-autocorrelation function (PACF) in Figure 2. Therefore, the data requires a variance, stabilizing transformation and differencing, to achieve stationarity.

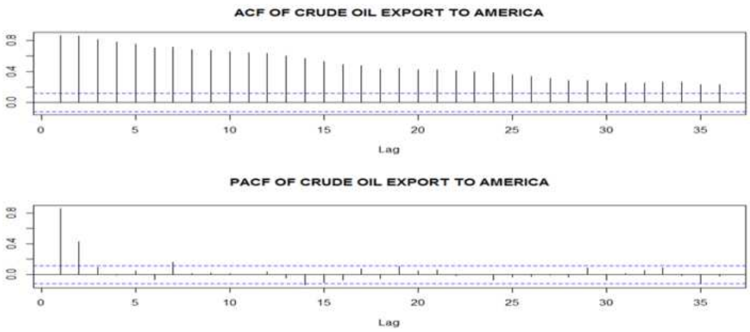

Figure 2: ACF and PACF of the Actual Data

The Square-Root transformation was used to transform the series, based on the result of the computed transformation lambda value $(\lambda = 0.5)$, using equation (2). The series was also non-seasonal differences to achieve stationarity. Figure 3, depicts the non-seasonal first-order differences, of crude oil export series.

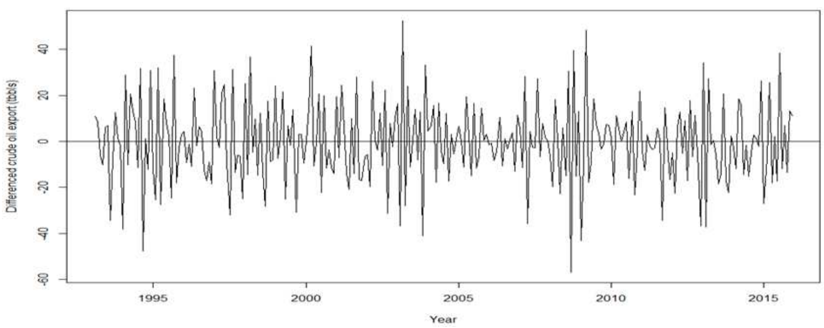

Figure 3: Non-Seasonal Differenced Crude Oil Export Series

The results of the ADF and KPSS unit root tests in Table 2, also indicates that, the data is stationary after the square root transformation and the non-seasonal first-order difference was applied to the series.

Table 2: The Stationarity Tests for Differenced Data

Summary of Test Statistics			
Test Type	**Test Statistics**	**Lag Order**	**P-Value**
ADF	-8.4569	**6**	**0.01**
KPSS	0.0843	**3**	**0.1**

The decay in the correlogram of the differences transformed data in Figure 4, when compared with the 95% confidence limits $\left(\pm\frac{2}{\sqrt{n}}=\pm 0.35\right)$, the PACF decays and the ACF cuts off, after the lag 1 with a significant spike at lag 12.

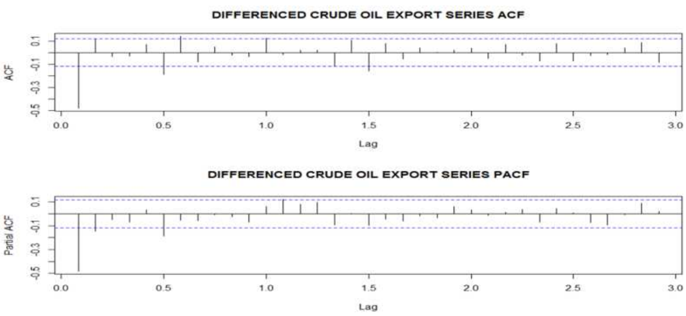

Figure 4: ACF and PACF of Transformed Differenced Series

Various tentative Seasonal ARIMA model structures, extracted from the correlogram plots in Figure 4 are presented in Table 3.

Table 3: Tentative Seasonal ARIMA Models

	Tentative Models		
MODELS	AIC	AICc	BIC
ARIMA (2, 1, 1) $(0, 0, 1)_{12}$	2279.15	2279.37	2297.24
ARIMA (2, 1, 1) $(1, 0, 1)_{12}$	2281.20	2281.51	2302.90
ARIMA (0, 1, 1) $(0, 0, 1)_{12}$	2279.58	2279.67	2290.43
ARIMA (1, 1, 1) $(1, 0, 1)_{12}$	2273.64	2273.86	2291.72
ARIMA (1, 1, 2) $(1, 0, 1)_{12}$	2275.24	2275.55	2296.94
*ARIMA (0, 1, 1) $(1, 0, 1)_{12}$	2273.19	2273.34	2287.66

"*" Means the best fit based on the selection criteria

The Seasonal ARIMA (0,1,1) $(1,0,1)_{12}$ was found to fit the series, based on the AIC, AICc and BIC selection criteria values. The model parameter estimates in Table 4 are statistically significant with t-values, and satisfies the stationarity and invertibility conditions.

Table 4: Estimates of ARIMA (0, 1, 1) $(1, 0, 1)_{12}$ Model

		Model Fit Statistics		
Parameter	Coefficients	Standard Error	t-statistics	p-value
MA1	-0.5507	0.0504	-10.926	0.00001
SAR1	0.9518	0.0569	16.727	0.00001
SMA1	-0.8690	0.0953	- 9.118	0.00001

Hence, the ARIMA (0, 1, 1) $(1, 0, 1)_{12}$ model in back shift is expressed as

$$(1-\Phi_1 B^{12})(1-B)y_t=(1+\theta_1 B)(1+\Theta_1 B^{12})\varepsilon_t \quad \textbf{(10)}$$

The fitted model in terms of the transformed series is

$$(1 - 0.9518\,B^{12})\,y_t = (1 - 0.5507\,B)(1 - 0.8690\,B^{12})\,\varepsilon_t$$

$$\hat{y}_t = 0.9518\,y_{t-12} - 0.5507\,\varepsilon_{t-1} - 0.8690\,\varepsilon_{t-12} + 0.4785583\,\varepsilon_{t-13} + \varepsilon_t \qquad \textbf{(11)}$$

The diagnostic adequacy check is performed using the correlogram plots of the model residuals, coupled with other objective diagnostic tests like the Box-Ljung test, Shapiro-Wilk Normality test and the ARCH-LM test. The test results in Table 5, failed to reject the null hypothesis at the 5% level of significance confirming that, the residuals are normally distributed with no autocorrelation and no conditional homoscedasticity (ARCH) effects. It implies that, the residuals are Gaussian white noise.

Table 5: ARIMA (0, 1, 1) (1, 0, 1)$_{12}$ Residuals Diagnostic Test

Ljung-Box, ARCH-LM and Shapiro-Wilk Test Statistics		
Test Type	**Test Statistics**	**P-Value**
Ljung-Box	14.407	0.2113
ARCH-LM	6.9373	0.7313
Shapiro-Wilk	0.99443	0.4104

The Gaussian white noise residuals are clearly portrayed in Figure 5, by the randomness of the residuals, non-significant spikes in the ACF residuals plot and the probability values, falling above the 0.05 limit in the probability plot.

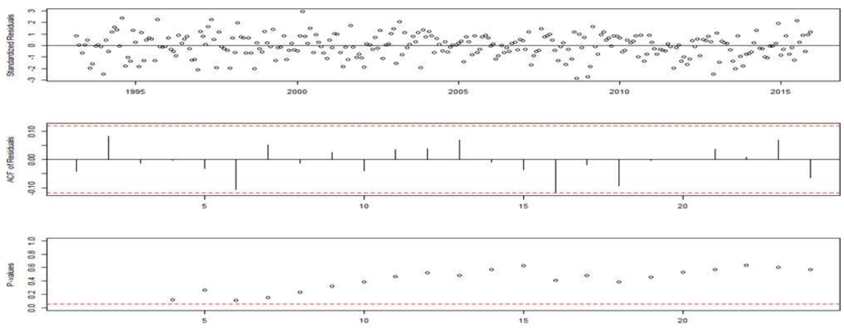

Figure 5: Fitted Model Residuals Diagnostic Plots

The plot of the fitted values against the actual crude oil export to the United States, is displayed in Figure 6, which shows that, the fitted model fits the series.

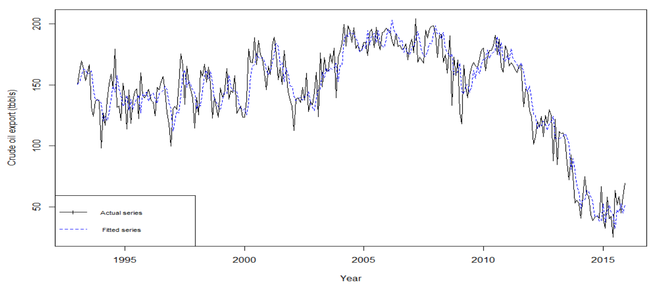

Figure 6: Actual Observed Values versus Fitted Model Values

4. CONCLUSIONS

In this study, the crude oil export to America was modelled using the ARIMA modelling procedures. The data evaluation results for the assumptions of ARIMA models showed that, the data required square-root transformation for series distributional normality and variance stability. The difference series was then subjected to the Box and Jenkins iterative procedure, for ARIMA model building. The results from analysis showed that, the appropriate model for the difference series is the Seasonal ARIMA (0, 1, 1) $(1, 0, 1)_{12}$ model. The model adequacy tests confirmed that, the model residuals are normally distributed uncorrelated random shocks (Gaussian white noise). This model is therefore, recommended for use in the forecast of Nigeria crude oil export, to the United States, until proven otherwise.

CONFLICT OF INTEREST

The authors declare that, there is no conflict of interest regarding the publication of this article; "Predictive Model with Square-Root Variance Stabilizing Transformation, for Nigeria Crude Oil Export to America".

REFERENCES

1. Akaike, H, (1974). A new look at the statistical model identification. IEEE transaction on automatic control, 19 (6): 716 723.
2. Bickel PJ and Doksum KA (1981): An analysis of transformation revisited. Journal of the American statistical Association. 76(374): 296 – 311.
3. Box GEP, Jenkins GM & Reinsel GC (1994). Time series Analysis, Forecasting and Control, 3rd edn. Prentice Hall, Englewood Cliffs, New Jersey-USA.
4. Box GEP & Jenkins GM (1976). Time series Analysis, Forecasting and Control, Revised edn. Holden-Day, San Francisco-USA. ISBN: 0816211043
5. Box GEP and Cox DR (1964): An analysis of transformations. Journal of the Royal statistical society, series B 26 (2): 211 – 252.
6. Bumham KP & Anderson DR (2002). Model selection and multimodel inference. A practical information-theoretical approach, 2nd edn. Springer-Verlag, ISBN: 0-387-95364-7, New-York-USA.
7. Carroll RJ and Ruppert D (1960): On prediction and the power transformation family. Biometrika 68: 609 – 615.
8. Central Intelligence Agency (2013). United States central Intelligence Agency (CIA) World Fact-book, www.cia.gov.
9. DeGroot MH (1987): A conversation with George Box statistical science. 2(3): 239 – 258.
10. Dickey DA & Fuller WA 1979. Distribution of the Estimators for Autoregressive time series with a unit root. Journal of the American statistics Association. 7(366): 427 431. JSTOR: 2286348.
11. EIA (2009). Energy Information Administration (EIA); World oil statistics. OPEC, Zurich, Switzerland. Obtained from www.Eia.org on 24/8/2016.
12. Engle, R: (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom

inflation, Econometrica, (50): 987-1008.

13. George E.P and David R. Cox
14. Kumar, Manoj and AnandManhu (2012). An application of time Series ARIMA Forecasting
15. A model for predicting sugarcane production in India. Studies in Business and Economics.81-94.
16. Kwiatkowski, .D, Phillips, P.C.B, Schmidt, .P, and Shin,. Y (1992). Testing the Null
17. Hypothesis of Stationarity against the Alternative of a unit Root. Journal of Econometrics 54: 159-178.
18. Nishii R (2001): Box-Cox transformation Encyclopaedia of Mathematics, Springer, ISBN: 978-1-55-608-010-4.
19. Obinna Adubisi and E.T. Jolayemi (2015). Estimating the impact on the Nigeria Crude oil Export from 2002 to 2013. (An ARIMA-Intervention Analysis). International Journal of Scientific and Engineering Research. 6(10): 878-886.
20. O. Adubisi (2016). Modelling the growth pattern of reserve currency in Nigeria. FUW Trends in Sciences & Technology Journal. 1(1): 130 – 133.
21. O.D. Adubisi, C. C. Eleke, T. T. Mom and C. E. Adubisi (2017). Application of SARIMA to modelling and forecasting money circulation in Nigeria. 6(1): 1–10. DOI: 10.9734/ARJOM/2017/35555
22. Pankratz, A. (1983). Forecasting with Univariate Box-Jenkins model: concepts and case, John Wiley and Sons, ISBN: 0471090239, New-York-USA
23. Sakia RM (1992): The B0x-Cox transformation technique: A review. The statistician. 41: 169 – 178.
24. Shapiro, S. S., and Wilk, M.B. (1965): “An analysis of variance test for normality (complete samples). Biometrika 52 (3-4): 591-611 doi10. 1093/biomet/52.3-4.591. JSTOR 2333709.MR205384.p.593.
25. Vanguard, Nigeria (2016): Importing Nigeria crude oil. http://www.vanguardngr.com/2016/us-importing-nigeria-crude-oil.
26. Yan L (2015): Variance stabilizing properties of Box-Cox transformation for dependent observations. Advances in science, Technology and Environmentology. B12 (3): 63 – 70.