African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

TIME SERIES MODELS OF CRUDE OIL PRODUCTION AND EXPORT IN NIGERIA (1999-2015)

Etaga Harrison O[1], Igwebuike Emeka K[1] and Etaga Cecilia N[2]

[1]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

[2]Department of Educational Foundation, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

ABSTRACT: *This study discussed the time series models of crude oil production and export in Nigeria from January 1999- December 2015. The Augmented Dickey-Fuller unit root test employed in the analysis to test for stationarity of the two series indicated that the crude oil production series was stationary at no differencing while crude oil exportation was stationary after first order differencing. The ACF and PACF of both crude production and export identified possible model for both series. Based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) the best model for Crude Production and Export were SARIMA (1,0,1) $(2,0,0)_{12}$ and SARIMA (2,1,0) $(1,0,1)_{12}$. Residual analysis and Box-Ljung test proved the adequacy of the models for both series.*

KEYWORDS: Crude Oil, AIC, BIC, SARIMA, Time Series, Crude Oil Export Data, Nigeria

INTRODUCTION

Crude oil is one of the natural resources to mankind and a major source of energy in Nigeria. It is a vital commodity in world market despite the campaign for green energy and other sources of power. Crude oil is one of the expensive commodities in the international market. Nigerian economy is heavily dependent on crude oil. The United States remains Nigeria's largest customer for crude oil, accounting for about 40% of the country's total crude oil exports. (Nwogu,2010).

The history of oil exploration in Nigeria dates back to 1907 when Nigerian Bitumen Corporation conducted exploratory work in the country; however, the firm left the country at the onset of World War I. Thereafter, a new license covering 357,000 sq. miles was given to a new firm called Shell D'arcy Petroleum Development Company of Nigeria. The new firm was a consortium of Shell and British Petroleum (then known as Anglo-Iranian). Shell-BP in the pursuit of commercially available petroleum found oil in Oloibiri, Nigeria in 1956. Production of crude oil began in 1957 and in 1960, a total of 847,000 tonnes of crude oil was exported. After that, the economy of Nigeria should have seemingly have experienced a strong increase. In the 70s, Nigeria realised enormous revenue from crude oil, which was regarded as oil boom. As a result of the tremendous revenue realized from crude oil, other exploration companies were attracted to the industry (Udosen et al, 2009).

Nigeria was able to reap instant riches from its oil production (Odularu,2008). Also, NNPC report had it that Nigeria petroleum industry is the main generator of Gross domestic product (GPD) with statistics that oil revenue has totalled $ 340 billion in exports since the 1970s. With this great achievement in the oil sector, government concentrated solely on oil for foreign exchange earnings lead to the neglect on agriculture which had been the source of revenue to the country.

In 2000, oil and gas exports accounted for more than 98% of export earnings and about 83% of federal government revenue, as well as generating more than 14% of its GDP. It also provides 95% of foreign exchange earnings, and about 65% of government budgetary revenues. Nigeria's proven oil reserves are estimated by the United States Energy Information Administration (EIA) at between 16 and 22 billion barrels (3.5×10^9 m^3), but other sources claim there could be as much as 35.3 billion barrels (5.61×10^9 m^3). Its reserves make Nigeria the tenth most petroleum-rich nation, and by the far the most affluent in Africa. In mid-2001 its crude oil production was averaging around 2.2 million barrels (350,000 m^3) per day. It is expected that the industry will continue to be profitable based on an average bench mark oil price of $85-$90 per barrel.

As recently as 2010, Nigeria provided about 10% of overall U.S oil imports and ranked as the fifth-largest source for oil imports in the U.S. However, Nigeria ceased exports to the US in July 2014 because of the impact of shale production in America; India is now the largest consumer of Nigerian oil. There are six petroleum exportation terminals in the country. Oil companies in Africa investigate offshore production as an alternative area of production. Angola and Nigeria are the largest oil producers in Africa. In Nigeria, the deep-water sector still has a large avenue to expand and develop. The Agbami oilfields hit full production in 2005, at 250,000 barrels a day. The amount of oil extracted from Nigeria was expected to expand from 15,000 bbl/d (2,400 m^3/d) in 2003 to 1.27 Mbbl/d (202,000 m^3/d) in 2010.

Nigeria being a mono- product economy that relies on the income generated from crude oil for day to day running of governmental functions, provision of infrastructure and other social amenities for the citizenry but the income generated depends on the quantity of crude oil produced as well as the price per barrel. Therefore, it is necessary to provide a time series model that can be used to forecast quantity of crude oil produced for the purpose of making reliable budget for the sustenance of the economy.

Time series data often exhibit seasonal variation, especially economic and business data. Box and Jenkins (1976) developed a time series model called Seasonal Autoregressive Integrated Moving Average (SARIMA) specifically for series that are seasonal in nature. Nigerian monthly crude oil production and export, being an economic time series data tend to exhibit some seasonality, their volatility notwithstanding.

In the past, several models were fitted by researchers to make forecast for crude oil production, export and price using Box-Jenkins ARIMA modelling approach.

Box and Jenkins (1976), Madsen (2008) and Boubaker (2011) are a few of authors that have written extensively on seasonal ARIMA models which are specially articulated for seasonal time series. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models. A seasonal ARIMA process is referred to as an ARIMA $(p,d,q)(P,D,Q)_s$, process. The lower-case letters (p,d,q) indicate the non-seasonal orders and the upper-case letters (P, D, Q)denote the seasonal orders of the process while s denotes the period.

Box and Jenkins (1976) gave Seasonal Autoregressive Integrated Moving Average (SARIMA) of a time series model as

$$\phi_P(\mathrm{B})\Phi_{\mathrm{P}}(\mathrm{B}^{\mathrm{S}})\nabla^{\mathrm{d}}\nabla_S^{\mathrm{D}}\,\mathrm{Z_t} = \Theta_{\mathrm{Q}}(\mathrm{B}^{\mathrm{S}})\theta_{\mathrm{q}}(\mathrm{B})\mathrm{a_t} \quad (1.1)$$

where

$\phi_P(\mathrm{B})$ is the non seasonal AR operator, $\theta_{\mathrm{q}}(\mathrm{B})$ is the non seasonal MA operator ,$\Phi_{\mathrm{P}}(\mathrm{B}^{\mathrm{S}})$ is the seasonal AR operator, $\Theta_{\mathrm{Q}}(\mathrm{B}^{\mathrm{S}})$ is the seasonal MA operator, ∇^{d} is the non-seasonal d^{th} order differencing, $\nabla_{\mathrm{S}}^{\mathrm{D}}$ is the seasonal D^{th} order differencing

When there is no seasonal effect, a SARIMA model reduces to pure ARIMA (p,d,q). SARIMA method applies only to stationary data. A stationary time series has a mean, variance, and autocorrelation function that are essentially constant through time. SARIMA modelling procedure by Box-Jenkins outlined three stages in finding a good model which are as follow; model identification, model fitting or estimation and model verification or diagnostics. It is important that any identified model be subjected to a number of diagnostic checks (usually based on checking the residuals). If the diagnostic checks indicate problems with the identified model one should return to the model identification stage. Once a model or selection of models has been chosen; the stability of the estimated parameters should be tested with respect to time frame chosen. SARIMA model has been used in various fields for making forecast.

In the past, several models were fitted by researchers to make forecast for crude oil production and export using Box-Jenkins SARIMA modelling approach. Etuk and Amadi, (2013) modelled Nigeria monthly crude oil domestic production from January 2006 to August 2012 using (SARIMA) approach. Omekara, et al (2015) modelled Nigeria monthly crude oil domestic production from January 2006 to March2015 using (SARIMA) approach. Kayode and Habib (2013) modelled Nigeria monthly crude oil exportation from January 2006 to August 2012 using (SARIMA) approach. Balogun and Ogunleye (2015) modelled Nigeria monthly crude oil exportation using (SARIMA) approach.

The study intends to fit an appropriate time series models of crude oil production and export in Nigeria (1999-2015) and also compare the statistical properties of these models. The data employed in this study based on Box-Jenkins SARIMA model, comprise of 408 monthly observations of the Crude Oil Production and Exportation in Nigeria spanning from January, 1999 to December, 2015.

LITERATURE REVIEW

Crude oil being the mainstay of Nigerian economy has attracted research interest in modelling of monthly crude oil production, export and price. Several research works have modelled monthly crude oil production and export using Box-Jenkins SARIMA method. This chapter reviewed the following literature which indicated gap and as such motivated the choice of the study.

Etuk, et al (2013) in their work modelled Nigeria monthly crude oil domestic production from January 2006 to August 2012 using Seasonal Autoregressive Moving Average (SARIMA) approach and found the Multiplicative SARIMA (0,1,1) $(0,1,1)_{12}$ model as the best model for crude oil production data.

Omekara, et al (2015) modelled Nigeria monthly crude oil domestic production from January 2006 to March 2015 using Seasonal Autoregressive Integrated Moving (SARIMA) approach. The visual inspection of time plot showed that the series was not stationary. The autocorrelation function (ACF) showed that there was trend in the series. Since the ACF had large values at lags 12, 24 and 36 it therefore implied that the data was seasonal. Therefore, they carried out a non-seasonal differencing of order 1and a seasonal differencing at lag 12 to remove both trend and seasonality. From the ACF and PACF plot of the non-seasonal and seasonally differenced data, it was observed that there is a cut off at lag 1 and lag 12 with negative value. They proposed the multiplicative SARIMA (1,1,1) $(0,1,1)_{12}$ model as the best model to be fitted to the crude oil production data.

Kayode and Habib (2013) in their research work modelled Nigeria monthly crude oil export from January 2006 to August 2012 using Seasonal Autoregressive Moving Average (SARIMA) approach. Result reveals an upward trend of the series which became stationary at 1st difference, a sharp drop between 2007 and 2009 and autocorrelation function with significant spikes at lag 1, 7 and 12 suggesting the presence of seasonality in the series. Based on Akaike Information Criterion (AIC), Schwartz Bayesian Information Criterion (SBIC) and Hannan-Quinn Information Criterion (HQC), the best model was SARIMA $(1,1,1)(0,1,1)_{12}$. The diagnosis on such model was confirmed, the error was white noise, presence of no serial correlation and a forecast for current and future values within 24 months period was made which indicates that the crude oil exportation is fairly unstable.

Nwogu and Iwu (2010) in their work, modelled Nigeria monthly crude oil exportation using Autoregressive Moving Average (ARIMA) approach and proposed ARIMA (1,1,1) as the best model for crude oil exportation data.

Balogun and Ogunleye (2015) in their work modelled Nigeria monthly crude oil exportation using Autoregressive Moving Average (ARIMA) approach. The research result showed there was a general increase in production over time (production was mostly below 60 million barrel per month in 2002 while in 2011 no production per month was less than 64 million barrel) showing that trend exists in the data, the simple average method was used to check for seasonality, the minimum value was 93.34 while the maximum was 104.92 none of which was too far from 100% showing there was no seasonal variation. The AIC as well as Mean Square Error (MSE) settled for ARIMA (1,1,1) out of the four initial picks and its was therefore chosen as the best model fit for the data collected for forecasting purposes and for policy formulation. ARIMA (1,1,1) was used to forecast for the year 2011.

Chiamaka and Omowunmi (2014) modelled crude oil production in Nigeria using Hubbert's model approach and developed an empirical model to describe and explain competing factors underlying crude oil production patterns. The production model equations were formulated with a non-linear curve fitting method to estimate the Hubbert's model parameters. The model was used to forecast future production outlook for Nigeria.

The model results suggest that production rate should have peaked at 2.70 MMSTB/D in year 2010, and forecasted that the estimated ultimate recovery, at year end 2050, will be 65 billion barrels (~ cumulative production of 31.25 billion barrels up till 2012 plus current proven reserves of 37.2 billion barrels)

Suleiman et al (2015) examined empirically the best ARIMA and GARCH models to forecast crude oil prices in Nigeria. The best ARIMA and GARCH models were selected using the criteria such as AIC, HQC, and SIC. The model for which the values of criteria are smallest is considered as the best model. They proposed ARIMA (3, 1, 1) and GARCH (2, 1) as the best models for forecasting the crude oil price data series.

Akpanta and Okorie (2015) in their work modelled and forecast Nigeria Crude Oil Prices obtained from 1982 to 2013 using Box-Jenkins ARIMA method. From their analysis, they proposed ARIMA (2, 1,2) as the best model to forecast crude oil prices. The descriptive statistics obtained showed, among other statistics, the mean to be 40.63USD/Barrel with a standard deviation of 32.28. The Augmented Dickey -Fuller test revealed that the time series data was unit root non-stationary. First order differencing was done to coerce the non-stationary time series into a stationary one - a condition that allowed the use of the univariate Box- Jenkins modelling approach.

Bakari et al (2013) in their work developed models for Annual Gas Production and Utilization in Nigeria using Box-Jenkins ARIMA approach. They proposed ARIMA (1,1,1) as the best model for production and for utilization was ARIMA (0,1,1). These models were used to forecast the production and utilization of gas for the upcoming 4 years to help decision makers establish priorities in terms of gas demand management.

Omorogbe (2016) in his work analyzed the Nigeria's crude oil export series using state space level model with stochastic and deterministic seasonal and SARIMA to model the dynamic features in the Nigeria crude oil export. His results clearly indicated that the local level model with deterministic seasonal is the most parsimonious model between the two state space models considered in his study.

Also, a parsimonious SARIMA model was also fitted to the data. He compared the forecasting performance of the two parsimonious models and evaluates their forecasts using ex-post indicators such as mean absolute percentage error (MAPE), root mean square percentage error (RMSPE) and the Theil's U statistic. The forecast analysis and evaluation results indicated that the state space local level model with deterministic seasonal outperforms the Box-Jenkins model in shorter and medium – range forecasting horizons. However, the forecast of the SARIMA model improves in the longer horizon. The Theil's U statistic also indicated that the state space local level model with deterministic seasonal and SARIMA model outperform the naive model at most of the forecasting horizons. He recommended that the state space model with deterministic seasonal component should be used in shorter and medium range forecasting horizons of the Nigeria's monthly crude oil export, for longer forecasting horizon, ten months and above, the seasonal ARIMA model should be considered.

However, we noted in the course of carrying out this research that literature on model comparison between crude oil production and exportation using Seasonal Autoregressive Integrated Moving (SARIMA) method was scarce. This indicated a gap in literature and as such motivated the choice of the study.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24) www.abjournals.org

METHODOLOGY

Data Collection

The data used for this study are secondary data on crude oil production and export in Nigeria from January 1999 to December 2015. The data were collected from the NNPC Annual Statistical Bulletin. The conceptual framework adopted for this work is Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The multiplicative seasonal autoregressive integrated moving average model was used in analyzing the data.

Preliminary Analysis

The analysis involves the calculation of some statistics for the two data series; mean, median, standard deviation, kurtosis, skewness, decomposition of the two series trend analysis and seasonal index error.

Autoregressive process (AR)

The autoregressive structure is a stochastic process that assumes that current data can be modelled as weighted summation of previous values plus a random term. The process X_t is regressed on past values of itself and this explains the prefix 'auto' in the regression process. Assume the random term e_t is purely random that is the term with mean zero and variance $\sigma^2 > 0$. Then the auto regressive process of order p or AR (p) is given by $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots + \phi_p X_{t-p} + e_t$ (3.1)

where $\phi_1, \phi_2, \ldots, \phi_p$ are constants.

In particular, the first AR process, AR (1) can be stated as $X_t = \phi_1 X_{t-1} + e_t$, which is sometimes called markov first order property of the AR, provided $|\phi| < 1$, then the AR (1) may be written as an infinite moving average process. In effect $E(X_t) = 0$.

The autocorrelation function of AR process is given by

$$\rho_k = \begin{cases} 1 & k=0 \\ \phi_1 & k=1 \\ \phi_1 \rho_{k-1} & k=2,3,\ldots \end{cases} \tag{3.2}$$

In general AR (p) an autoregressive process with order p can be written as

$$Z_t = (1 - \phi_1\beta - \phi_2\beta^2 - \phi_3\beta^3 - \phi_p\beta^p)X_t \tag{3.3}$$

Moving Average process (MA)

Moving average structure is a stochastic process which is defined as a type of finite impulse response filter process used to analyze a set of data points by creating a series of averages of different subsets of the full data set. The moving average process of order q or MA (q) is given by

$$X_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} \ldots\ldots + \theta_q e_{t-q} + e_t \tag{3.4}$$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

where $\theta_1, \theta_2 \ldots, \theta_q$ are constants, and e_t is purely a random process with mean zero and variance $\sigma^2 > 0$. In MA process, the auto covariance function is given by

$$\gamma(\tau) = cov(X_t, X_{t+\tau}) = cov(\theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \theta_\tau Z_{t+\tau} + \ldots + \theta_q Z_{t+k-q}) \tag{3.5}$$

In general MA (q) process with order q can be written as

$$X_t = (\theta_1 B + \theta_2 B^2 + \theta_1 B + \ldots + \theta_q B^q) Z_t = \theta(B) Z_t \tag{3.6}$$

The finite order moving average process is always stationary. For invertibility of the MA(q) process, it is required that all the roots of θ(B)=0 must lie outside the unit circle.

The autocorrelation function of MA process is given by

$$\rho_k = \begin{cases} 1 & k=0 \\ \frac{\theta}{1+\theta^2} & k=1 \\ 0 & k>1 \end{cases} \tag{3.7}$$

Differencing

Differencing is a relatively simple operation that involves calculating successive changes in the values of a data series. Differencing is used when the mean of a series is changing over time. It is frequently applied to time-series data to induce a stationary mean.

The letter "I" in the acronym ARIMA is integration step, and it corresponds to the number of times (d) the original series has been differenced, to make it stationary.

Autoregressive Integrated Moving Average processes (ARIMA)

The general ARIMA (p,d,q) process can be written in the general form:

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(1-B)^d X_t = (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q) e_t \tag{3.8}$$

The general ARIMA (p,d,q) process could also be written as

$$\phi_p(B) \nabla^d X_t = \theta_q(B) e_t \tag{3.9}$$

where:

$\nabla^d = (1-B)^d$

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

$\theta_q(B) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$

From the above expression

$\phi_p(B)$ is the autoregressive operator; which is assumed to be stationary if root of $\phi_p(B) = 0$ lie outside the unit circle.

∇^d is called the differencing operator.

$\theta_q(B)$ is called the moving average operator; it is assumed to be invertible if root of $\theta_q(B) = 0$ lie outside the unit circle.

Seasonal Autoregressive Integrated Moving Average (SARIMA)

More often than not, practical time series contain a seasonal component which repeats itself periodically, for instance, every s period, where s denotes the seasonal period. For example, with s = 12 in this case we may expect the series, X_t to depend on values at annual lags such as X_{t-12} and X_{t-24} and on more recent non-seasonal values as x_1 and x_2.

However, SARIMA models are capable of modelling a wide range of seasonal data which is formed by including additional seasonal terms in the ARIMA models. A seasonal ARIMA process is referred to as an ARIMA (p,d,q)(P,D,Q)$_s$, process. The lower-case letters (p,d,q) indicate the non-seasonal orders and the upper-case letters (P,D,Q) denote the seasonal orders of the process while s denotes the period. The parentheses mean that the seasonal and non-seasonal elements are multiplied.

Stationarity Test

Augmented Dickey Fuller (ADF) Test

Augmented Dickey Fuller (ADF) is mostly used to test for stationarity of time series data.

The regression equation by Dickey and Fuller (1979), is given by

$$\Delta yt = \mu 0t + \mu 1t + \phi yt\text{-}1 + \sum_{j=1}^{t} \alpha j \Delta yt\text{-}j + \varepsilon t \tag{3.10}$$

t = p+1, p+2,. . .T

where μ0 is the intercept, μ1t represents the trend in case it is present, ϕ is the coefficient of the lagged dependent variable. yt-1 and p lags of Δyt-j with coefficient αj are added to account for series correlation in the residuals. The null hypothesis HO: ϕ = 0 is that the series is not stationary while the alternative hypothesis H1: ϕ ≠ 0 is that the series is stationary. The ADF test statistic is given by

$$\text{ADF} = \frac{\phi}{SE\ (\phi)} \tag{3.11}$$

where SE (ϕ) is the standard error for denotes estimate.

The null hypothesis of the unit root is accepted if the ADF test statistic is greater than the critical value for $\alpha = 0.05$.

Model Selection Criterion

Akaike Information Criterion

The Akaike information criterion (AIC) is a measure of the relative goodness of fit of a statistical model. Akaike suggests measuring the goodness of fit for some particular model by balancing the error of the fit against the number of parameters in the model. It provides the measure of information lost when a given model is used to describe reality. AIC values provide a means for

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

model selection and cannot say anything how well a model fits the data in an absolute sense. If the entire candidate models fit poorly, AIC will not give any warning of that.

The AIC is defined as

$$AIC = 2k - 2In\,(L) \tag{3.12}$$

where k is the number of parameters in the statistical model and L is the maximized value of the likelihood function for the estimated model. The AIC is applied in model selection in which the model with the least AIC value is selected as the best candidate model.

Bayesian Information Criteria

In statistics, the Bayesian information Criterion (BIC) or Schwarz Bayesian Information Criterion (SBIC) is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based on the likelihood function and it is closely related to Akaike Information Criterion (AIC). When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in over fitting. Both BIC and AIC attempt to resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC.

The BIC is defined as

$$BIC = \ln(n)k - 2In(L) \tag{3.13}$$

Where k is the number of parameters in the statistical model, L is the maximized value of the likelihood function for the estimated model and n is the number of observations.

Model Checking

Ljung-Box Test

The Ljung-Box test is a diagnostic tool used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an ARIMA (*p*,d,*q*) model to the data. The test examines m autocorrelations of the residuals. If the autocorrelations are very small, we conclude that the model does not exhibit significant lack of fit. In general, the Box-Ljung test is defined as:

H_0: The model does not exhibit lack of fit.

H_a: The model exhibits lack of fit.

Test Statistic: Given a time series Y of length n, the test statistic is defined as:

$$Q = (n + 2)\sum_{k=1}^{m} \frac{\hat{r}_k^2}{n-k} \tag{3.14}$$

where $\hat{r}_k$ is the estimated autocorrelation of the series at lag k, and m is the number of lags being tested. The Box-Ljung test rejects the null hypothesis (indicating that the model has significant lack of fit) if

$$Q > X^2_{1-\alpha,h} \tag{3.15}$$

where $X^2_{1-\alpha,h}$ is the chi-square critical value with h degrees of freedom and significance level α.

Because the test is applied to residuals, the degrees of freedom must account for the estimated model parameters so that h=m-p-q, where p and q indicate the number of

parameters from the ARMA (p,q) model fit to the data.

Residual Analysis

Once a model has been fitted, the final step is the diagnostic checking of the model. One of the steps to be carried out is by studying the autocorrelation plots of the residuals to see if further structure (large correlation values) can be found. If the autocorrelations and partial autocorrelations residual show at least one significant cut-off in twenty lags, the model is considered adequate and forecasts are generated. If otherwise, model is re-estimated. This process of checking the residuals and adjusting the values of p and q continues until the resulting residuals contain no additional structure. Once a suitable model is selected, the fitted model may be used to generate forecasts.

Co-integration Test

Co-integration is a statistical property of a collection (X_1, X_2, ..., X_k) of time series variables. If a group of variables are individually integrated of the same order and there is at least one linear combination of the variables that is stationary, then the variables are said to be co-integrated. A common example is where the individual series are first-order integrated **I(1)** but some co-integrating vector of coefficients exists to form a stationary linear combination of them. First, all of the series must be integrated of order 1. Co-integration has become an important property in contemporary time series analysis.

Engle–Granger two-step Method Co-integration Test

If x_t and y_t are non-stationary and co-integrated, then a linear combination of them must be stationary. In other words:

$$y_t - \beta x_t = \mu_t \tag{3.16}$$

where μ_t is stationary.

If we knew μ_t, we could just test it for stationarity with something like a Dickey–fuller test, Phillips-Perron test and be done. But because we do not know μ_t, we must estimate this first, generally by using ordinary least squares, and run our stationary test on the estimated μ_t series, often denoted $\hat{\mu}_t$. A second regression is then run on the first differenced variables from the first regression, and the lagged residuals $\hat{\mu}_{t-1}$ is included as a regressor.

Johansen Co-integration Test

The Johansen test is a test for co-integration that allows for more than one co-integrating relationship, unlike the Engle–Granger method, but this test is subject to asymptotic properties, i.e. large samples. If the sample size is too small, then the results will not be reliable and one should use Auto Regressive Distributed Lags (ARDL).

Phillips–Ouliaris Co-integration test

Phillips-Ouliaris test shows that residual-based unit root tests applied to the estimated co-integrating residuals do not have the usual Dickey–Fuller distributions under the null hypothesis of no-co-integration. Because of the spurious regression phenomenon under the null hypothesis, the distribution of these tests has asymptotic distributions that depend on (1) the number of deterministic trend terms and (2) the number of variables with which co-integration is being tested. These distributions are known as Phillips–Ouliaris distributions and critical values have been tabulated. In finite samples, a superior alternative to the use of these asymptotic critical values is to generate critical values from simulations.

RESULTS AND DISCUSSION

This chapter deals with the analysis of monthly crude oil production and export from January, 1999- December, 2015. The monthly data were collected from the NNPC Annual Statistical Bulletin.

Time Series Plot of Data

Time series plots which display observations on the y-axis against equally spaced time intervals on the x-axis were used to evaluate patterns and behaviours in data over time for both series are displayed in Fig 1 and 2

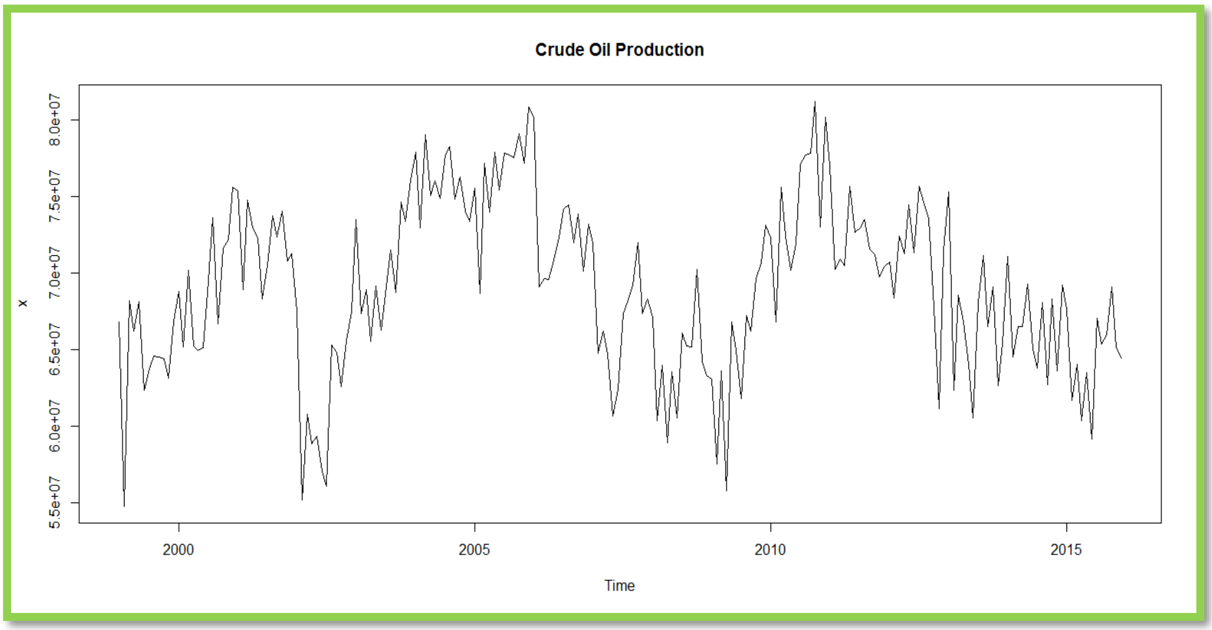

Fig 1 Plot of Crude Oil Production (Original Data)

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

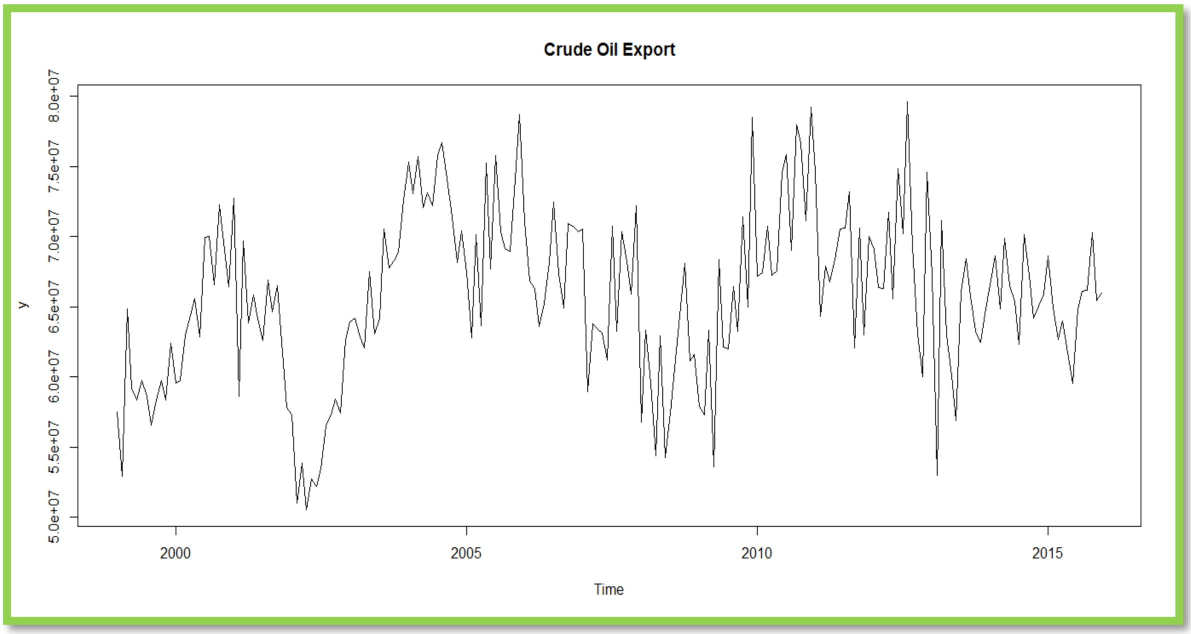

Fig 2 Plot of Crude Oil Export (Original Data)

Fig 1 and 2 showed the presence of seasonal variations in both time series data

Preliminary Analysis of the two Data Series

This provides descriptive statistics of the two series; the mean, median, standard deviation, kurtosis and skewness as contained in Tables 1 and 2 below.

Table 1 Descriptive Statistics for Crude Oil Production

Mean	Median	Standard Deviation	Skweness	Kurtosis
69068130	69083107	5494652	-0.15	-0.37

Table 2 Descriptive Statistics for Crude Oil Export

Mean	Median	Standard Deviation	Skweness	Kurtosis
65751252	65943257	5974012	-0.16	-0.19

Decomposition

Decomposing of both series into seasonal, trend and irregular components are shown in figures 3 and 4

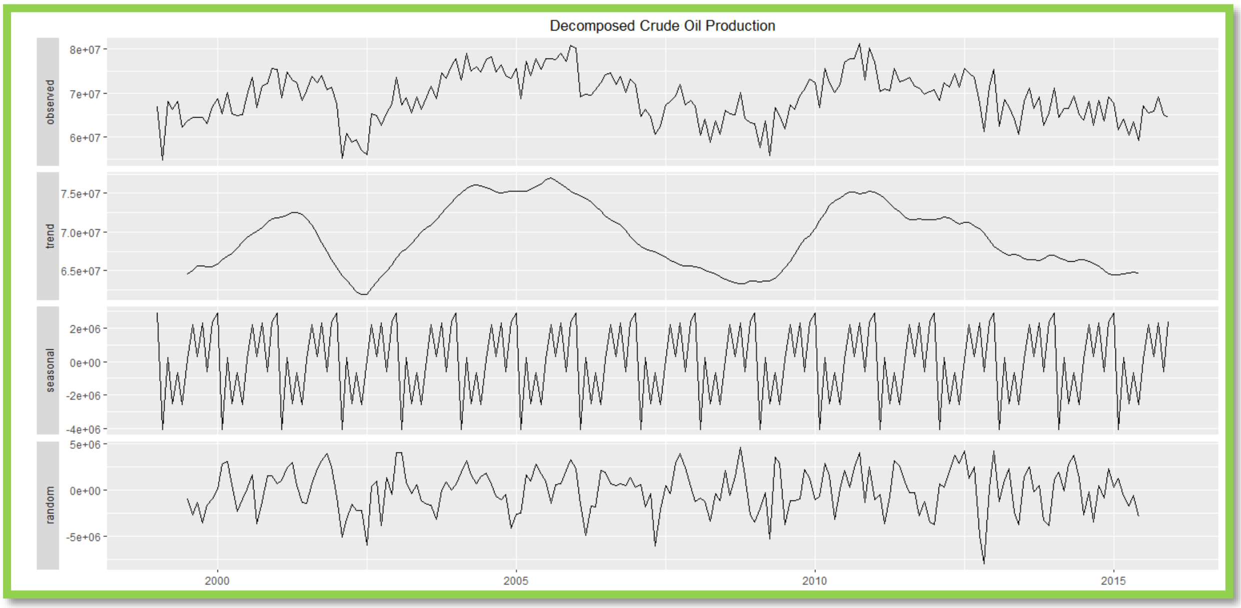

Fig 3 Decomposition of Crude Oil Production

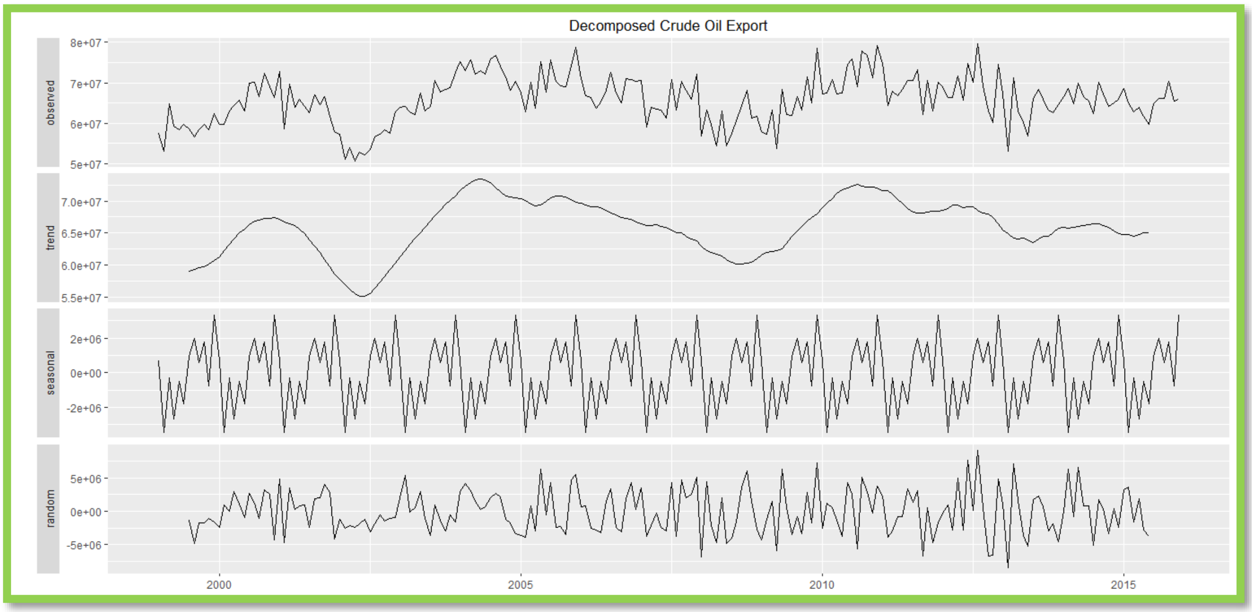

Fig 4 Decomposition of Crude Oil Export

Test for Stationarity

Augmented Dickey-Fuller test was used to test for stationarity of time series data for crude production and export. It computes test for the null hypothesis that a time series data is not stationary against the alternative.

ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)

Table 3 Stationarity Test for Crude Oil Production

Dickey-Fuller	Lag order	p-value	Alternative hypothesis
-3.5988	5	0.03474	Stationary

Since the p-value is less than 0.05, the alternative hypothesis is accepted, which proved that time series data for crude production is stationary.

Table 4 Stationary Test for Crude Oil Export

Dickey-Fuller	Lag order	p-value	Alternative hypothesis
-3.3882	5	0.05801	Stationary

Since the p-value is greater than 0.05, the null hypothesis is accepted, which proved that time series data for crude export is non-stationary.

Differencing of Crude Oil Export Data

Since Box-Jenkins SARIMA modelling approach applies to only stationary data. The crude export series which is non stationary was differenced to make it stationary as shown in Table 5 below.

Table 5 Differenced series for Crude Oil Exportation

Year	Months											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1999		-4587685	11908661	-5597210	-897187	1412891	-1109286	50897	-3461313	5719931	-3238021	-2620743
2000	-2810074	189632	3219203	1419248	1152519	-2664608	7082788	4314339	-2237501	1878842	-4803317	-3876191
2001	6289589	-14103701	11082729	-5863053	2032511	-1717838	-1571981	4314339	-2237501	1878842	-4803317	-3876191
2002	-573377	-6268839	2818125	-3281215	2154086	-529650	1276781	3076382	777926	1106293	-999480	5285420
2003	1227238	286292	-1323748	-821904	5465245	-4496790	1184997	6360087	-2805820	497054	610865	3693107
2004	2749110	-2244136	2615363	-3623822	1032331	-900737	3587656	932666	-2625291	-2692903	-3272080	2290635
2005	-2969573	-4677862	7377709	-6460239	11527291	-7504157	8051372	-5387890	-1224996	-178469	5139167	4596523
2006	-7652181	-4233214	-539751	-2712717	1681574	2929513	4272045	-5119159	-2423715	6049021	-239039	-347174
2007	172509	-11597308	4848215	-346699	-326911	-1883128	9493534	-7490683	7103951	-2139420	-2283489	6264381
2008	-15400947	6507334	-3718988	-5206257	8592326	-8703053	3152323	3178466	3765477	3711406	-6936254	470744
2009	-3686820	-640554	6022031	-9680376	14745776	-6250641	-122488	4456514	-3208190	8199772	-6469701	13509960
2010	-11336780	322710	3299710	-3508355	290231	7041013	1252463	-6834750	8953877	-1230732	-5570977	8047289
2011	-4540349	-10327340	3590881	-1136935	1568928	2164312	122197	2523413	-11119906	8532995	-7610063	7038147
2012	-946861	-2719216	-33071	5419433	-6132316	9277789	-4666146	9423428	-10312327	-6279658	-3011816	14530344
2013	-7338566	-14225427	18130718	-8107230	-2952022	-3179021	9246471	2307053	-2647394	-2514247	-790759	2042771
2014	2012375	2071570	-3744674	5000860	-3394075	-1105813	-3052676	7813695	-2920446	-2994742	813030	898121
2015	2728327	-3760580	-2224666	1311286	-2267999	-2150587	5294461	1273426	20272	4156124	-4829461	468121

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24) www.abjournals.org

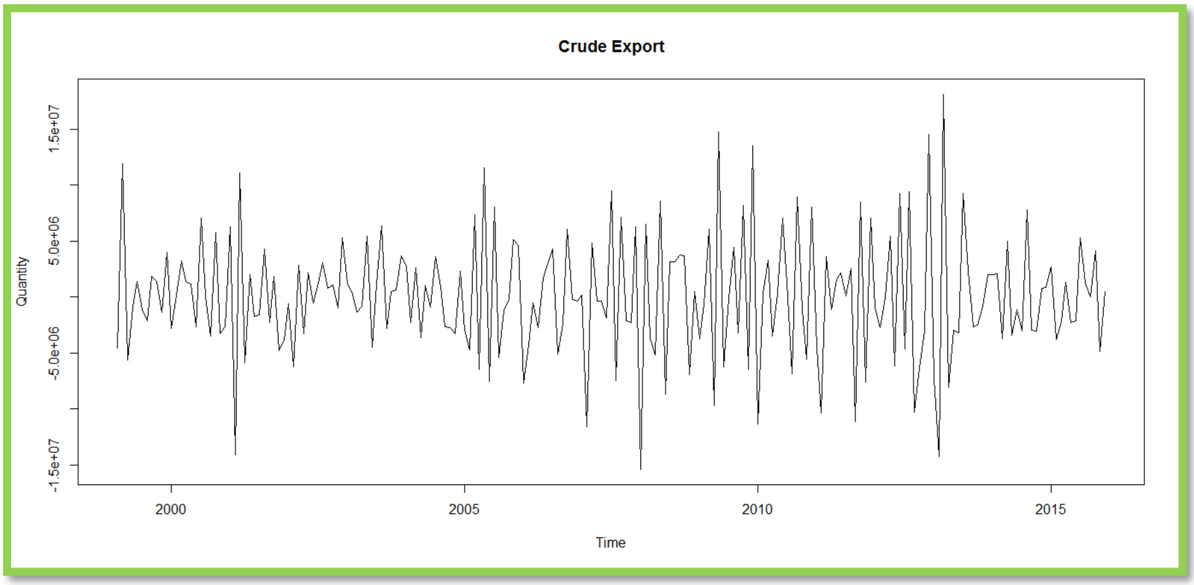

Fig 5 Plot of the Differenced series for Crude Oil Exportation

Model Identification for both Series

To select the appropriate ARIMA model, which means finding most appropriate values of p and q for an ARIMA (p,d,q) model. To do this, we usually need to examine the autocorrelation and partial autocorrelation of the stationary time series data.

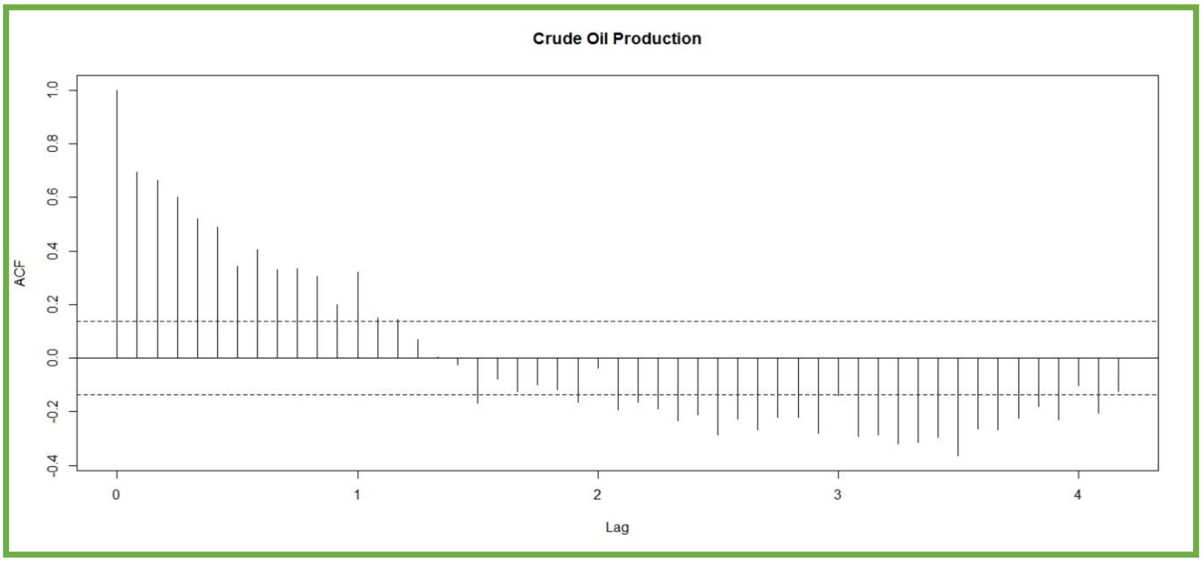

Fig 6 ACF plot of Crude Oil Production

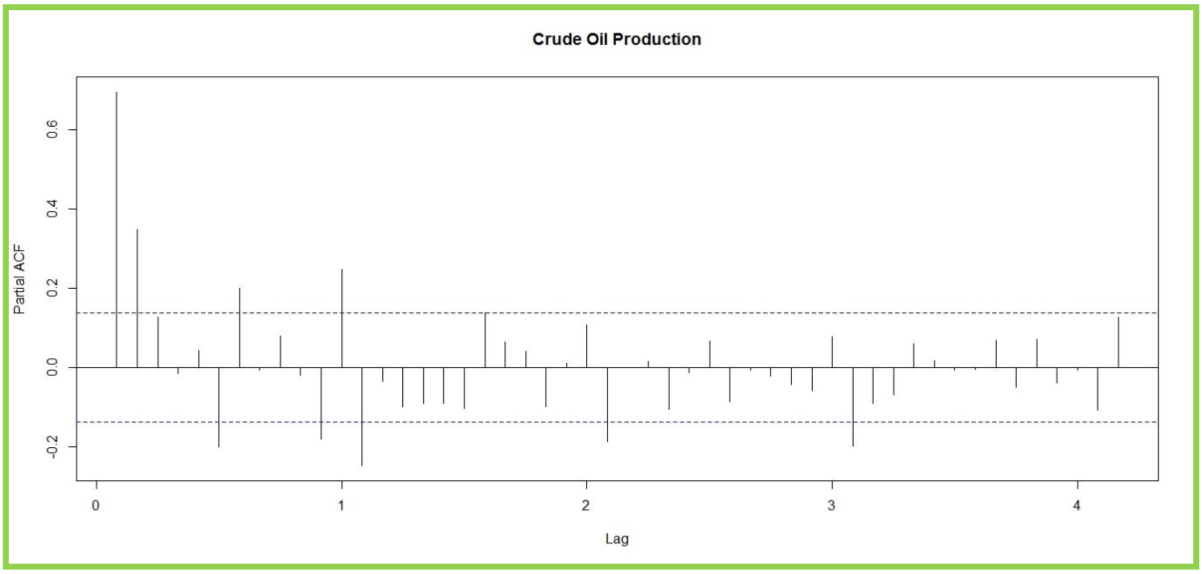

Fig 7 PACF plot of Crude Oil Production

From the ACF and PACF of Crude Oil Production, the spikes tailed off with mixed exponential

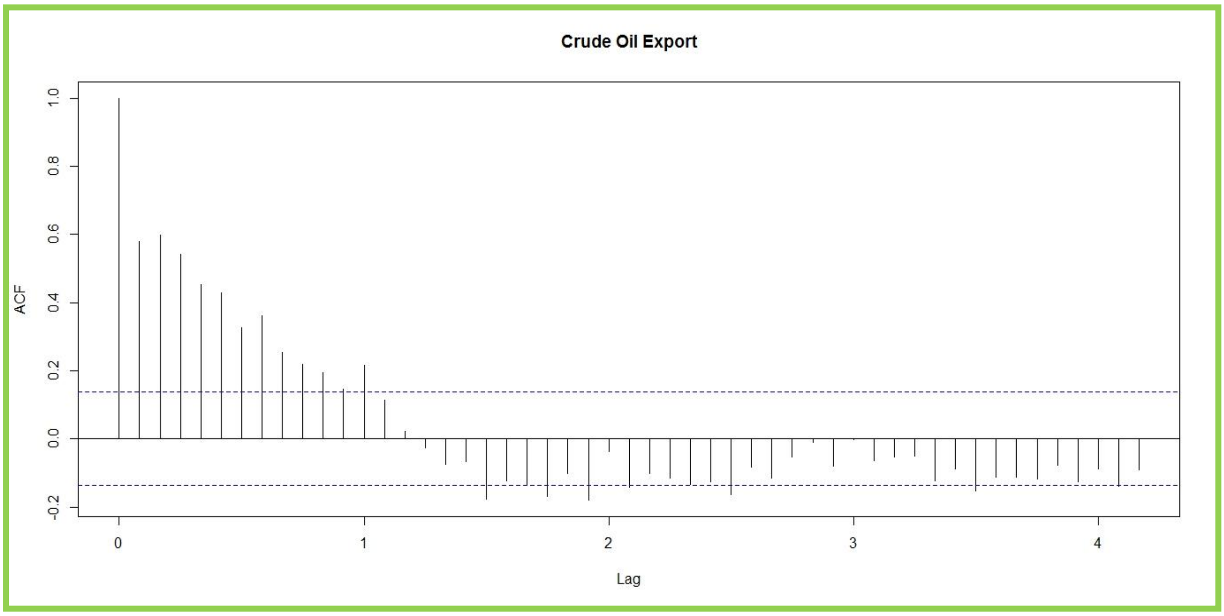

Fig 8 ACF of Crude Oil Export

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

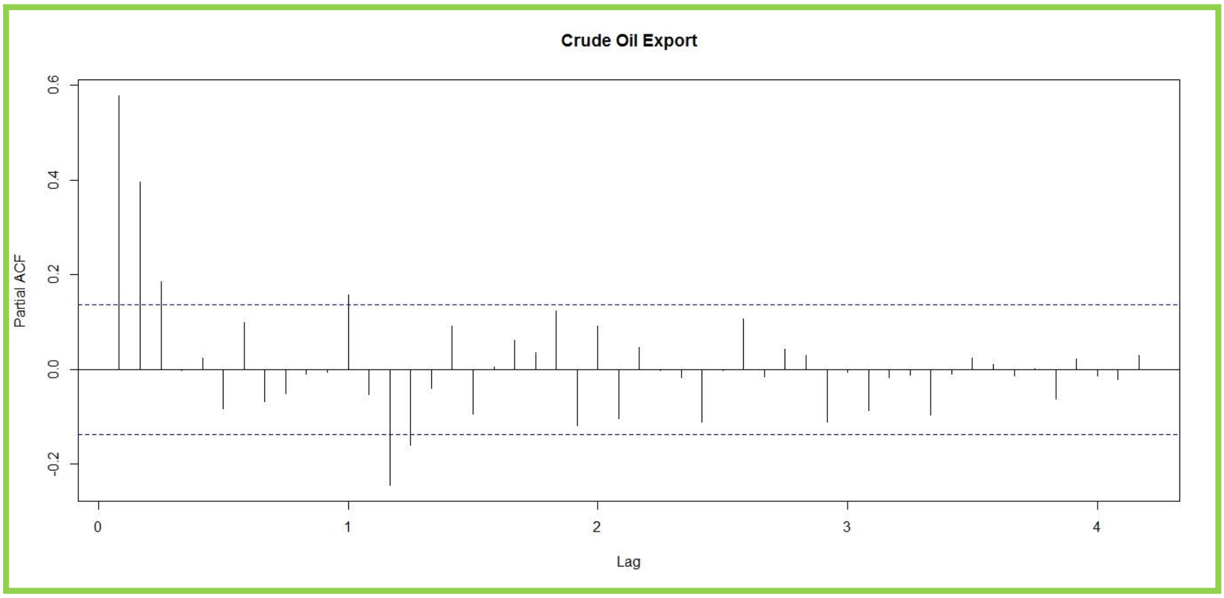

Fig 9 PACF of Crude Oil Export

From the ACF and PACF of Crude Oil Export, the spikes tailed off with mixed exponential

Model Estimation

After choosing the appropriate model for both series, the coefficients of the parameters selected were estimated in the tables below.

Table 6 Model Estimation for Crude Oil Production

SARIMA (1,0,1) $(2,0,0)_{12}$ with non-zero mean

	AR1	MA1	SAR1	SAR2	Intercept
Coefficients	0.9085	-0.3218	0.3172	0.2904	67539868
S.E	0.0337	0.0814	0.0693	0.0734	3277067
Sigma^2 estimated as 1.033e+13: log likelihood=-3346.56					
AIC=6705.11 AICc=6705.54 BIC=6725.02					

This shows that the Model is a Seasonal ARIMA: SARIMA (1,0,1) $(2,0,0)_{12}$ where p=1, d=0, q=1, P=2, D=0, Q=0

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)

Table 7 Model Estimation for Crude Oil Export

SARIMA (2,1,0) $(1,0,1)_{12}$

	AR1	AR2	SAR1	SMA1
Coefficients	-0.6609	-0.2607	0.7798	-0.5837
S.E	0.0703	0.0683	0.2463	0.3359
Sigma^2 estimated as 1.827e+13: log likelihood=-3386.54				
AIC=6783.07 AICc=6783.38 BIC=6799.64				

This shows that the Model is a Seasonal ARIMA: SARIMA (2,1,0) $(1,0,1)_{12}$ where p=2, d=1, q=0, P=1, D=0, Q=1

Table 8 Possible SARIMA Models for Crude Oil Production with AIC_C values

Models	AIC_C values
SARIMA(1,0,0)$(1,0,0)_{12}$	6729.57
SARIMA(1,0,1)$(1,0,0)_{12}$	6718.05
SARIMA(1,0,0)$(2,0,0)_{12}$	6715.70
SARIMA(1,0,1)$(2,0,0)_{12}$	6705.54

Table 9 Possible SARIMA models for Crude Oil Export with AIC_C values

Models	AIC_C values
SARIMA(2,1,0)$(1,0,1)_{12}$	6783.38
SARIMA(2,1,1)$(1,0,1)_{12}$	6784.91
SARIMA(2,1,1)$(2,0,1)_{12}$	6787.66
SARIMA(2,1,0)$(2,0,1)_{12}$	6785.73

ARIMA model with the lowest AIC_C gives a better forecast, from the two tables above SARIMA (1,0,1)$(2,0,0)_{12}$ and SARIMA (2,1,0)$(1,0,1)_{12}$ are chosen as the best models for crude oil production and export.

Model Checking

Residual Test

After a model has been fitted, the final step is the diagnostic checking of the model. This is carried out is by studying the autocorrelation plots of the residuals to see if further structure (large correlation values) can be found. If the autocorrelations and partial autocorrelations residual show at least one significant cut-off in twenty lags, the model is considered adequate and forecasts are generated. If otherwise, model is re-estimated.

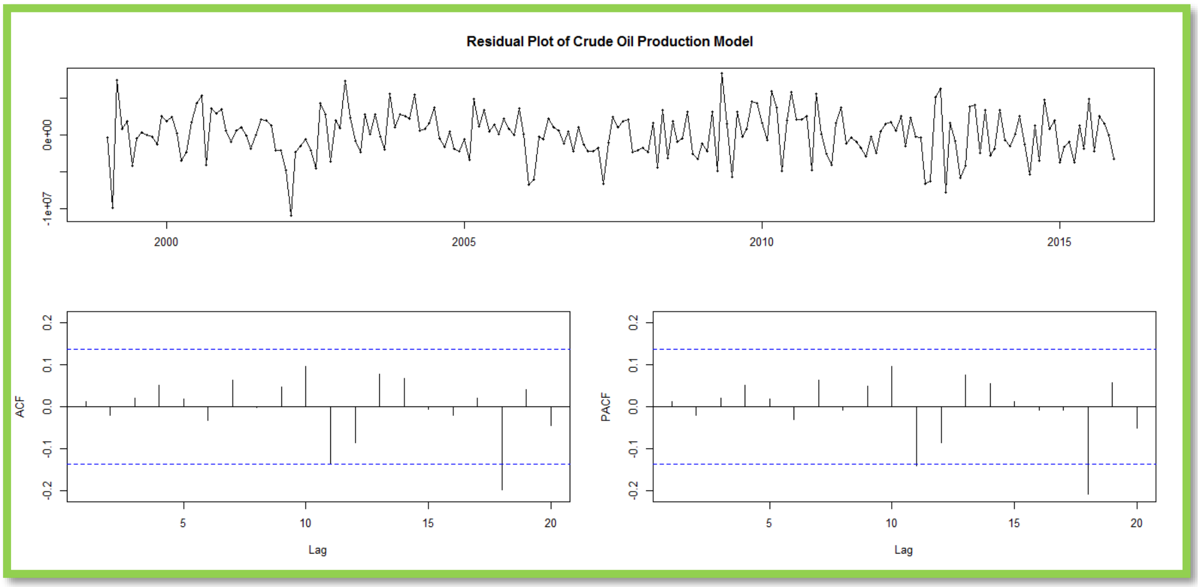

Fig 10 Residual plot of Crude Oil Production Model

From Figure 10 above there is one significant cut-off in both ACF and PACF in twenty lags, and then the model SARIMA $(1,0,1)(2,0,0)_{12}$is considered adequate for forecasting.

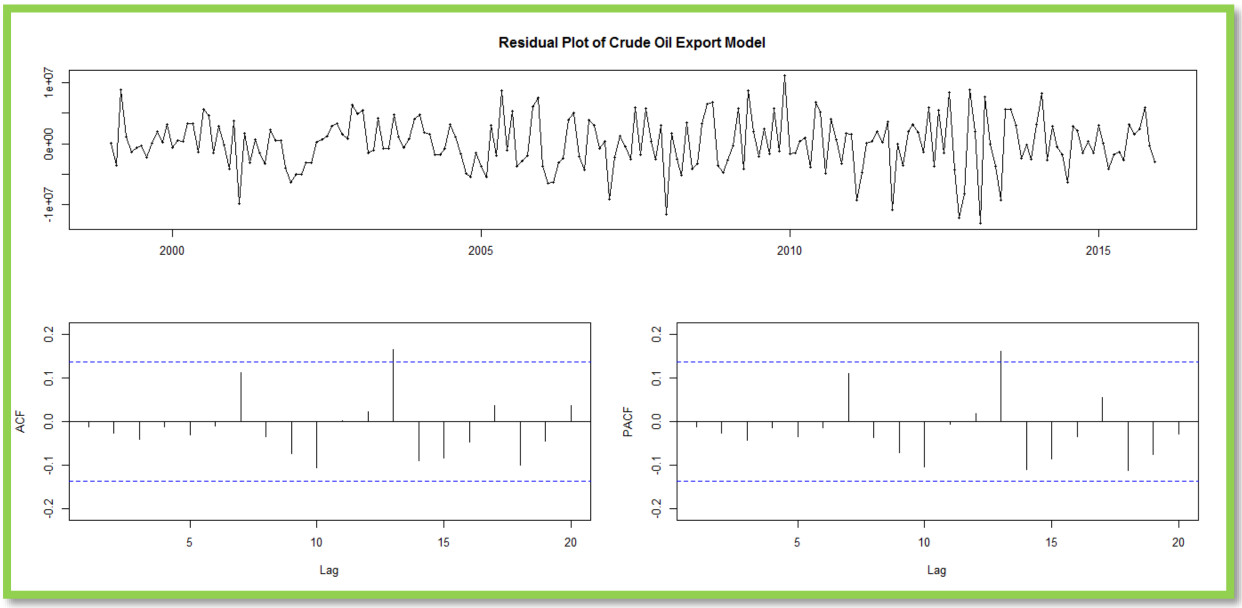

Fig 11 Residual plot of Crude Oil Export Model

From Figure 11 above there is one significant cut-off in both ACF and PACF in twenty lags, and then the model SARIMA $(2,1,0)(1,0,1)_{12}$ is considered adequate for forecasting.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

Box-Ljung Test

The Ljung-Box test is a diagnostic tool used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an ARIMA(p,d,q) model to the data.

Table 10 Box-Ljung Test for Crude Oil Production Model SARIMA $(1,0,1)(2,0,0)_{12}$

Data: Residuals of Crude Oil Production Model		
Chi-square	df	p-value
0.027287	1	0.8688

Since the p-value>0.05, we accept the null hypothesis that the model does not exhibit lack of fit.

Table 11 Box-Ljung Test for Crude Oil Export

Model SARIMA $(2,1,0)(1,0,1)_{12}$

Data: Residuals of Crude Export Model		
Chi-square	df	p-value
0.035721	1	0.8501

Since the p-value>0.05, we accept the null hypothesis that the model does not exhibit lack of fit.

Co-integration Test

Time series variables are said to be co-integrated if the variables are individually integrated of the same order and there is at least one linear combination of the variables that is stationary. From the Augmented Dickey-Fuller unit root test for both series, it shows that the Crude Oil Production series is stationary at integration order d=0, while the Crude Oil Exportation is stationary at integration order d=1. Therefore, there is no co-integration between the two series.

Forecasting

With the fitted models for both Crude Oil Production and Export, twelve months forecast for both series were shown in Tables 12 and 13 below.

Table 12 Twelve months forecast for Crude Oil Production in Barrels

Months	Point Forecast	Lower 95 Percent Limit	Upper 95 Percent Limit
Jan 2016	66898776	60600323	73197229
Feb 2016	63265826	55963184	70568467
Mar 2016	64727630	56690043	72765216
Apr 2016	63696770	55099721	72293818
May 2016	65601646	56568852	74634440
June 2016	63125469	53748227	72502712
July 2016	65353135	55700819	75005450
Aug 2016	66155329	56281724	76028935
Sept 2016	64836957	54784354	74889560
Oct 2016	67542555	57344564	77740547
Nov 2016	64987112	54670651	75303573
Dec 2016	66452290	56039053	76865528

Table 13 Twelve months forecast for Crude Oil Export in Barrels

Months	Point Forecast	Lower 95 Percent Limit	Upper 95 Percent Limit
Jan 2016	67955018	59576834	76333202
Feb 2016	65155060	56308331	74001788
Mar 2016	65778265	55934671	75621859
Apr 2016	66283931	55339057	77228806
May 2016	65069073	53384764	76753383
June 2016	64766893	52271935	77261851
July 2016	65971964	52729664	79214264
Aug 2016	67606685	53675866	81537504
Sept 2016	66576635	51974647	81178623
Oct 2016	66895293	51656755	82133831
Nov 2016	65536665	49687523	81385806
Dec 2016	66833020	50394456	83271584

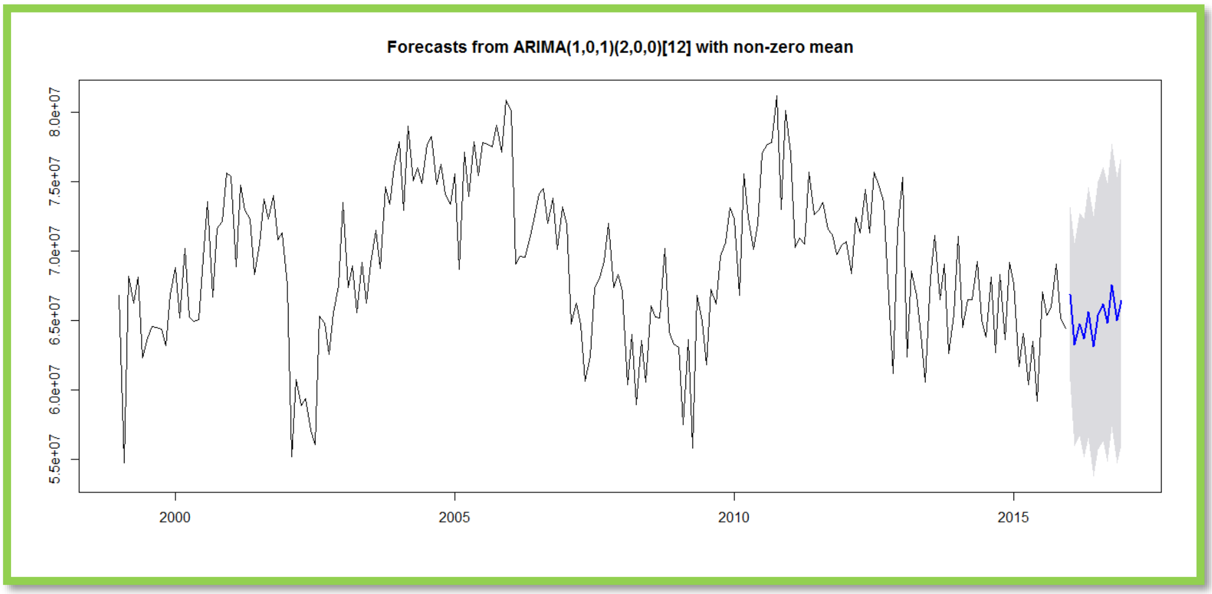

Fig 12 Plot of Crude Oil Production Model

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

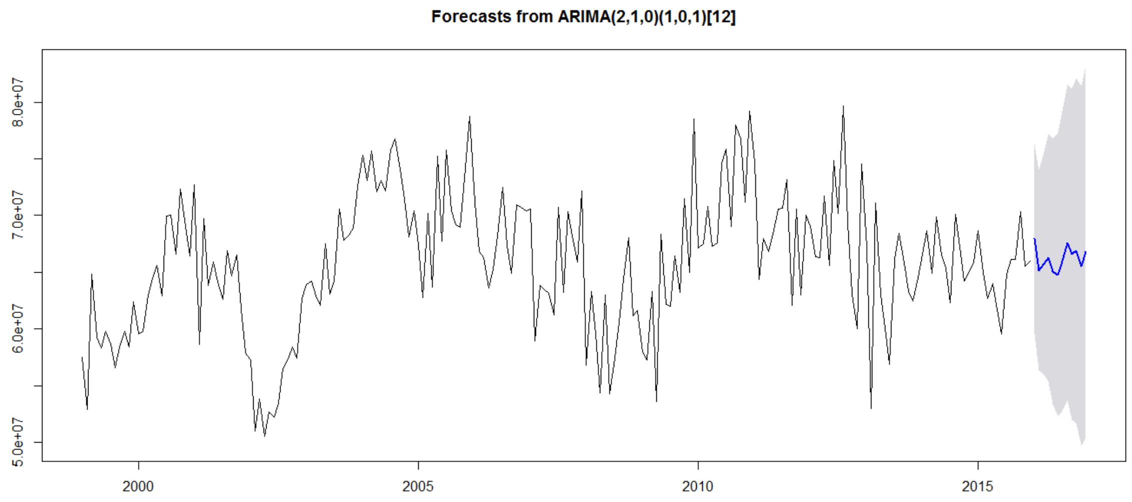

Fig 13 Plot of Crude Oil Export Model

From Figures 12 and 13, the shaded portion shows the forecast for the year 2016 for both series.

The equation for crude oil production model, SARIMA $(1,0,1)(2,0,0)_{12}$

$(1-\phi B)(1-\Phi_{12}B^{12}-\Phi_{13}B^{13})Y_t=(1-\theta B)e_t$

$Y_t = C + \phi y_{t-1} + \Phi y_{t-12} + \Phi y_{t-13} - \theta e_{t-1} + e_t$

$Y_t = C + 0.9085y_{t-1} + 0.3172y_{t-12} + 0.2904y_{t-13} + 0.3218e_{t-1} + e_t$

The equation for crude oil export model, SARIMA $(2,1,0)(1,0,1)_{12}$

$(1-\phi_1 B-\phi_2 B^2)(1-B)(1-\Phi_{12}B^{12})Y_t=(1-\theta B)e_t$

$Y_t = y_{t-1} + \phi y_{t-1} + \phi y_{t-2} + \Phi y_{t-12} - \theta e_{t-12} + e_t$

$Y_t = Y_{t-1} - 0.6609Y_{t-1} - 0.2607Y_{t-2} + 0.7798Y_{t-12} + 0.5837e_{t-12} + e_t$

The Augmented Dickey-Fuller unit root test employed in the analysis to test for stationarity of the two series indicated that the crude oil production series was stationary at no differencing while crude export was stationary after first differencing.

The ACF and PACF of both crude oil production and oil export identified possible model for both series. SARIMA $(1,0,1)(2,0,0)_{12}$ was fitted to crude production while SARIMA $(2,1,0)(1,0,1)_{12}$ to crude export. Residual analysis and Box-Ljung test proved the adequacy of the models for both series. Twelve months forecast for both series were made.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 1-24)
www.abjournals.org

CONCLUSION

The best model for Crude Production and Export were SARIMA $(1,0,1)(2,0,0)_{12}$ and SARIMA$(2,1,0)(1,0,1)_{12}$based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). There is no co-integration between the two series, since the Crude Oil Production series is stationary at integration order d=0, while the Crude Oil Export is stationary at integration order d=1, from the Augmented Dickey-Fuller unit root test for both series.

Further Studies

It is therefore, recommended for further studies the use of other time series model approach for model comparison for Crude Oil Production and Export in Nigeria.

Implication of Research

The study fitted an appropriate time series models of crude oil production and export in Nigeria (1999-2015) which provided a useful forecast for quantity of crude oil production and export for the purpose of making reliable budget for the sustenance of the economy.

REFERENCES

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B.N. Petrov and F. Csaki (eds.)

Akpanta, A.C and Okorie (2014) Application of Box-Jenkins Techniques in Modelling and Forecasting Nigeria Crude Oil Prices. International Journal of Statistics and Applications, 4(6): 283-291.

Alan Pankratz (1983): Forecasting with Univariate Box-Jenkins Models, Concepts and Cases, John Wily and Sons Inc, New York, USA.

Bakari, H.R, Chamalwa, H.A (2013) Time Series Analysis Model for Production and Utilization of Gas (A Case Study of Nigeria National Petroleum Corporation "NNPC") IOSR Journal of Mathematics (IOSR-JM) e-ISSN: 2278-5728, p-ISSN:2319-765X. Volume 9, Issue 1, PP 17-23

Balogun, O.S and Ogunleye, O.M, (2015). On The Time Series Modelling of Crude Oil Exportation in Nigeria. European Journal of Academic Essays 2(3): 1-11.

Boubaker, H.B.H, (2011) "The forecasting performance of seasonal and nonlinear models"Asian Economic and Financial Review 1(1): 26-39.

Box, G. E. P. and Jenkins, G. M. (1976): Time Series Analysis, Forecasting and Control, Revised Edition, Holden Day, San Francisco, California, USA.

Chatfield (1978), The Analysis of Time Series: An Introduction. London Chapman and Hall (Sixth Edition)

Chiamaka Kingsley-Akpara and Omowunmi O. Iledare, (2014) Modelling Crude Oil Production Outlook: A Case Study of the Oil and Gas Industry in Nigeria Society of Petroleum Engineers (SPE172381) pp 1-7

Dickey, D.A and Fuller, W.A. (1979) "Distributions of the Estimators for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association" 74(366), 427-431.

Etuk, E.H and Amadi, E.H. (2013) Modelling Nigeria Monthly Crude Oil Domestic Production. Journal Applied Mathematical Sciences Vol 7.

Hannan E. J., QuinnB.G. (1979) “The Determination of the Order of an Auto regression. Journal of the Royal Statistical Society B, 41(2), 190-195.

Harris, Richard and Robert, Sollis (2003), Applied Time Series Modelling and Forecasting, John John Wiley and Sons, Chichester.

Iheanyichukwu, S.I, Eleazer, C.N. and Valentine U.N. (2013): Time Series Modelling of Nigeria External Reserves. CBN Journal of Applied Statistics. Vol.4 No.2 (December, 2013)

Jonathan D. Cryer, Kung-Sik Chan (2008): Time Series Analysis with Application in R, Second Edition, Springer.

Kayode Ayinde and Habib Abdulwahab (2013) Modelling Nigeria Crude oil export. Journal of Mathematical Sciences Vol 9. Pages 23 32.

Kendall, M., and Ord, K. (1990) “Time Series Analysis”, 3rd Edition. New York: Oxford University Press, pp. 110.

Madsen, H., (2008). Time series analysis, London: Chapman & Hall/CRC

Medugu, N. I. (2012). Crude Oil Discovery in Nigeria and Matters Arising.

NNPC (1999-2015) Annual Statistical Bulletin

Nwogu, E.C, and Iwu, H.C, (2010) Modelling Nigeria Crude Oil export. Journal of Applied Mathematical Sciences. Vol. 42.

Omekara, C.O., Okereke, O. E, et al (2015). ARIMA Modelling of Nigeria Crude Oil Production. Journal of Energy Technologies and Policy Vol.5, No.10, pp 1-5

Omorogbe J.A (2016) State Space Versus SARIMA Modelling of the Nigeria’s Crude Oil Export. Sri Lankan Journal of Applied Statistics, Vol. (172) pp 87-108

Priestley, M. B (1981) “Spectral Analysis and Time Series”, Academic Press, London.

Suleiman. S, Alabi, M.A (2015). Modelling and Forecasting the Crude Oil Price in Nigeria. International Journal of Novel Research in Marketing Management and Economics Vol. 2, Issue 1, pp: (1-13)