Journal of Econometrics and Statistics
Vol. 1, Issue 1, 2021, pp. 75-91
© ARF India. All Right Reserved
URL : www.arfjournals.com/jes

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) AND SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (SARIMA) MODEL OF CRUDE OIL PRODUCTION IN NIGERIA

Oseni, Bolanle. Azeez[1]
[1]*Department of Mathematics and Statistics, The Polytechnic, Ibadan, Oyo-State, Nigeria.*
E-mail: bolaoseni2007@yahoo.com

Article History
Received : 23 February 2021
Revised : 16 March 2021
Accepted : 6 April 2021
Published : 2 September 2021

To cite this paper
Oseni, Bolanle. Azeez (2021). Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated moving Average (SARIMA) Model of Crude Oil Production in Nigeria. *Journal of Econometrics and Statistics.* 1(1), 75-91.

Abstract: In this study, crude oil production in Nigeria was constructed using SARIMA model. The study was carried out to estimate the parameters of the model, to identify a model that best fits the production of crude oil, and to check for the model adequacy on the quarterly production of crude oil. Test of normality of the data was done via Anderson Darling test statistics. Augmented Dickey-Fuller test was employed to test for stationarity of the data. Adequacy of the model was carried out using Ljung-Box chi-square amongst others. Finally, the results show that Seasonal Autoregressive Integrated Moving Average (SARIMA) model best fits the production of crude oil in Nigeria.

Keywords: Autoregressive, stationarity, normality, seasonal and invertibility.

1. Introduction

Models for time series data can have many forms and represent different stochastic process which itself is defined as a statistical phenomenon that involves with time according to some probabilistic laws. When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the Integrated (I) models and the Moving Average (MA) models. These three classes depend linearly on previous data points. The combinations of these ideas produce autoregressive integrated moving average (ARIMA) models. ARIMA models are a class of models that have capabilities to represent stationary as well as non-stationary time series and to produce accurate forecasts based on a description of historical data of single variable.

However, since it does not assume any particular pattern in the historical data of the time series that is to be forecast, this model is very different from other models used for forecasting.

Crude oil exports positively contribute to economic growth through various ways; it promotes specialization in the production of export commodities when there is an increase in export which may lead to an increase in the productivity of the export sector.

Hence, there is need to derive a model for the production of crude oil, and forecast future production of crude oil. Specifically, in this study, a model for the production of crude oil in Nigeria was built, carried out diagnostic checking for the model adequacy and forecasts for future crude oil production in Nigeria amongst others.

Crude oil production is a mixture of hydrocarbon that exists in liquid phase in natural underground reservoirs and remains liquid at atmospheric pressure after passing through surface separating facilities. Crude oil is refined to produce a wide array of petroleum products including heating oil's, gasoline, diesel and jet fuels, lubricants asphalt, ethane, propane, butane and many other products used for their energy or chemical content.

Oil is for apparent reasons one of the global commodities most studied by economists. Key areas of interest include price formation, that is, the role of speculation versus fundamental drivers; the interaction of prices with other economic variables such as exchange rates and Gross Domestic Product (GDP); and the drivers of oil supply. While there has been much empirical work on the determinants of OPEC production, less effort has been devoted to a systematic investigation of global oil production.

In this study, we take a naïve approach and estimate models to identify the relationship between country-specific oil production decision and world oil market prices as well as price volatility, while controlling for other important determinants of oil production decisions. In general, oil production has been analyzed from two perspectives. Since a major feature of fossil fuels is their nature, namely their exhaustibility and their geologic attributes, one stream of literature investigates whether oil production develops according to economic models of exhaustible resources based on, Hotelling (1931), or whether oil production more closely relates to the question of worldwide oil depletion as suggested by Hubbart (1956). This stream of literature has produced mixed results given that the assumptions required in each model, such as the geographic scope of production, determine the predicted production pattern to some degree.

A second steam of literature examines the strategic behavior of the major oil producers. For example, the focus is on competition, MacAvoy (1982) and revenue targets, Teece (1982) in examining OPEC. Followers of the cartel hypothesis test if OPEC is a monopoly, an oligopoly, or if it acts as a dominant firm. Griffin's seminar paper (1985) is the starting point for numerous contributions to the cartel hypothesis and also analyses the potential mechanisms used to steer production, mostly based on current price and production data. One conclusion in non-OPEC and OPEC countries reacts differently to current price changes, yet there are also differing interpretations about the exact nature of the potentially strategic interactions.

Some authors (Griffin, 1985; Jones, 1990) claim that OPEC acts as a cartel or a bureaucratic syndicate (Smith,2005; Alhaji and Huettner, 2000) find that market results can be explained by

Saudi Arabia's dominant role, and a few researchers promote the 'target revenue hypothesis" (Griffin, 1985; Ramcharran, 2002; Alhaji and Huettner, 2000), and the existence of a quota system (Kaufmann *et al.*., 2008). The empirical steam of literature disregards the importance of a range of prices prior to the current period for future oil production. However, oil production follows physical investment with a significant time lag of seven to ten years (Wurzel *et al.*., 2009).

Most remarkable is the increase of oil production from non-OECD/non-OPEC countries which increased from share in global production from 29% in 1994 to 34% in 2009, at the same time the share of OECD producers decreased from 32% to 25%. We are interested in the major determinants of production in all countries, that is the high prices triggering exploration activities, financial crisis implying economic downturns and hence negative growth in oil consumption; terrorist attacks delaying or even alienating investments and so on.

Meyler *et al.* (1998) drew a framework for ARIMA time series model for forecasting Irish inflation. In their work, they emphasized heavily on optimizing forecast performance while forecasting on minimizing out 'of-sample forecast errors rather than maximizing in-sample"goodness of fit". Stergiou (1999) in his research used ARIMA model technique on a 17 years' time series data (1964 to 1980) with 204 observations of monthly catches of pilchard (Saedinapilchardus) from Greek waters for forecasting up to 12 months ahead and forecasts were compared with actual data for 1981 which was not used in the estimation of the parameters. The research found that mean error as 14% suggesting that ARIMA procedure was capable of forecasting the complex dynamics of the Greek pilchard fishery, which otherwise was difficult to predict because of the year to year changes in Oceanographic and biological conditions.

Conteras *et al.* (2003) in their study, using ARIMA model provided a method to predict next-day electricity prices both for short markets and long-term contracts for mainland Spain and Californian markets. Many researchers restricted their studies using only one or two models to World Crude Oil prices. For example, Siti *et al.*, (2011) uses trend and volatility models, long memory and volatility models (Kyongwook *et al.*, 2013). Approaches like modeling seasonality, volatility, long memory and trend etc may capture all the possible attributes in crude oil prices. The application of one method in analyzing world crude oil production data is sometimes misleading because behaviours of data such as seasonality, trend (upward and downward), volatility and long memory etc are not taken into account by single approach. This study examines the production of crude oil in Nigeria rather than crude oil prices using ARIMA and SARIMA models.

In fact, a plethora of research studies are available to justify that a careful and precise selection of ARIMA model can be fitted to the time series data of single variable (with any kind of pattern in the series and with auto-correlations between the successive values in the time series) to forecast with better accuracy, the future values in the series.

The structure of the paper is as follows: Section 1 introduces the work. Section 2 presents methodology on SARIMA. Section 3 presents the data, empirical results as well as model scenarios. Section 4 gives the discussion of results, conclusion and recommendations/ policy implications.

2. Methodology

2.1 The Autoregressive Moving Average (ARMA) Models

An *ARMA*(p, q) model is a combination of *AR*(p) and *MA*(q) models and is suitable for univariate time series modeling. In a *AR*(p) model, the future value of a variable is assumed to be a linear combination of p past observations and a random error together with a constant term.

Mathematically, the *AR*(p) model can be expressed as:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \tag{1}$$

Given

$$\sum_{i=1}^{p} \phi_i y_{t-i} = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

So that equation (1) above then becomes;

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \tag{2}$$

where, y_1 and ε_t are respectively the actual value and random error (random shock) a white noise process at time t: $\phi_i(i = 1, 2, \ldots, p)$ are model parameters and c is a constant. The integer constant p is known as the order of the model. At times, the constant term is omitted in the model for simplicity. For estimating parameters of a *AR* process using the given time series, it is done via the Yule-Walker equation. Just as a *AR*(p), a *MA*(q) model uses past errors as the explanatory variables.

The *MA*(q) model is given by

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{3}$$

Given,

$$\sum_{j=1}^{q} \theta_j \varepsilon_{t-j} = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

So that equation (3) above then becomes

$$y_t = \mu + \phi_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{4}$$

where, μ is the mean of the series: $\theta_j(i = 1, 2, \ldots, q)$ are the model parameters and q is the order of the model, random shocks (ε_i) are assumed to be a white noise process that is; a sequence of independent and identically distributed (i.i.d) random variables with mean zero and a constant variance σ^2.

Generally, the random shocks are assumed to follow the typical normal distribution. Thus, conceptually a moving average model is a linear regression of the current observation of the time series against the random shocks of one or more prior observations. Fitting an *MA* to a time series is more complicated than fitting an *AR* model because in the former one the random error terms are not fore-seeable. Autoregressive (AR) and moving average (MA) models can be effectively combined together to form a general and useful class of tie series model known as the ARMA model.

Mathematically, a $ARMA(p, q)$ model is represented by

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-1} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{5}$$

Usually, ARMA models are manipulated using the lag operator notation. The lag or backshift operator is defined as $Ly_t = y_{t-1}$. Polynomials of lag operator are used to represent ARMA model as follows:

$$ARMA(p, q) : \phi(L)y_t = \theta(L)\varepsilon_t \tag{6}$$

Given,

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i \text{ and } \theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j.$$

Where, $AR(p)$ process can be always be written in terms of a $MA(q)$ whereas, for a $MA(q)$ process to be invertible, all the roots of the equation $\theta(L) = 0$ must be outside the unit circle. This condition is known as the invertibility condition for an *MA* process. The advancement of ARMA model is Autoregressive Integrated Moving Average (ARIMA) model in which the integration aspect is introduced to accommodate non-stationary data. This model has been used by many researchers in analysis of time series data.

In the work of Raymond (1997) suggested that the following two questions must be answered to identify the data series in a time series analysis viz:

(i) whether the data are random and

(ii) have any trends?

This is followed by another three steps of model identification parameter estimation and testing for model validity. If a series is random, the correlation between successive values in a time series is close to zero. If the observations of time series are statistically dependent on each another, then the ARIMA is appropriate for the time series analysis.

2.2 Autoregressive Integrated Moving Average (ARIMA) Model

The ARIMA model described above can be used for stationary time series data. However, in practice many time series such as those related to socio-economic and business show non-stationary behavior. Time series which contain trend and seasonal patterns are also non-stationary in nature. Thus, from application view point ARMA model are inadequate to properly describe non-stationary time series, which are frequently encountered in practice. For this reason we proposed ARIMA model which is a generalization of an ARMA model to include the case of non-stationarity as well.

In ARIMA model, a non-stationary time series is made stationary by applying finite differencing of the data points. The mathematical formulation of the ARIMA (p, d, q) model using lag polynomials is given as;

$$\phi(L)(1 - L)^d y_t = \phi(L)\varepsilon_t \tag{9}$$

The equation (9) can be re-written as:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d y_t = \left(1+\sum_{j=1}^{q}\theta_j L^j\right)\varepsilon_t \tag{10}$$

Here, p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated and moving average parts of the model respectively. The integer d controls the level of differencing. Generally, $d = 1$ is enough in most cases. When $d = 0$, then it reduces to an $ARMA(p, q)$ model.

An $ARIMA(p, 0, 0)$ is nothing but the $AR(p)$ model and $ARIMA(0, 0, q)$ is the $MA(q)$ model. $ARIMA(0, 1, 0)$, that is, $y_t = y_{t+1} + \varepsilon_t$ is a special one and known as random walk model. It is widely used for non-stationary data in economic data and stock price series. At the model identification stage, one needs to decide how many autoregressive (p) and moving average (q) parameters are necessary to yield an effective but still parsimonious model of the process *i.e.* it has the fewest parameters and greatest number of degrees of freedom among all models that fit the data. Least square and maximum likelihood estimator is usually used for the parameter estimation which is done after identified the specific number and type off ARIMA parameters to be estimated. The major tool used in the identification phase and plots of the series, correlograms of autocorrelation, ACF and PACF.

The data used in this study was collected from the statistical bulletin of organization of petroleum exporting countries (OPEC). It is a secondary data because it was extracted from their records and the period considered is from 1985-2015 and the statistical package used was R-Program version 2.3.

2.3 Box-Jenkin (ARIMA) Model

The approach of Box-Jenkins methodology in order to build ARIMA model, normality tests are used to determine if a data set is well modeled by a normal distribution and to compare how likely it is for a random variable underlying the data set to be normally distributed.

2.4 Anderson-Darling Test

This is a statistical test of whether a given sample of data is drawn from a given probability distribution: the test assumes that there are no parameters to be estimated in the distribution being tested when applied to testing if normal distribution adequately describes a set of data. It is one of the most powerful statistical tools for detecting most departures from. The test statistic is given as:

$$A^2 = n - s. \tag{7}$$

where

$$s=\sum_{k=1}^{n}2k-1\left[\ln(F(Y_k))+\ln(1-F(1-F(Y_n+1-k)))\right]$$

The test statistic is now compared with the critical values of the theoretical distribution; the null hypothesis is rejected if p-value is less than the significance level.

2.5 Test for Stationarity

A stationarity time series is one whose statistical properties such as mean, variance, autocorrelation and so on are all constant over time. Stationarizing a time series through differencing (where needed) is an important part of the process of fitting an ARIMA model. Thus, finding the sequence of transformations needed to stationarize a time series often provides important clues in the search for an appropriate forecasting model.

2.6 Augmented Dickey-Fuller Test (ADF)

This is used to test whether a unit root is present in an autoregressive model. The test is defined for the hypothesis $H_0 : \sigma = 0$; the data has a unit root (not stationary) against $H_1 : \sigma \neq 0$; the data has no unit root (is stationary). The Dickey-Fuller unit root test is based on the following three regressive forms:

(i) $\Delta Y_t = \sigma Y_{t-1} + U_{tt}$ (without constant and trend)
(ii) $\Delta Y_t = \alpha + \sigma Y_{t-1} + U_t$ (with constant) (8)
(iii) $\Delta Y_t = \alpha + \beta T + \sigma Y_{t-1} + U_t$ (with constant and trend)

If the p-value is less than the level of significance α, we reject the null hypothesis and conclude that there is no unit root in the data–it is stationary. If unit root exist, the data can be differenced to remove unit root from the data, if such data is differenced once and there is no unit root, then it becomes an ARMA model of order one ARI(1)MA, if twice it becomes ARI(2)MA and so on.

2.7 Autocorrelation and Partial Autocorrelation Function (ACF and PACF)

To determine a proper model for a given time series data, it is necessary to carry out the ACF and PACF analysis. These statistical measures reflect how the observations in a time series are related to each other. The ACF is a way to measure the linear relationship between an observation at time t and the observations at previous times. If we assume an $AR(k)$ model, then we may wish to only measure the association between y_t and y_{t-1} and filter out the linear influence of the random variables that lies in between (*i.e.*, $y_{t-1}, y_{t-2}, \ldots, y_{t-(k-1)}$) which requires a transformation on the time series. By calculating the correlation of the transformed time series, we obtain the partial autocorrelations (PACF). The PACF is most useful for identifying the order of autoregressive model. Specifically, single partial autocorrelations that are significantly different from zero indicate lagged terms of y that are useful predictors of y_t.

2.8 Diagnostic Check

The Ljung-Box statistic test is a diagnostic tool used to test whether any group of autocorrelations of a time series are different from zero. Instead of testing randomness at each distinct lag, it tests the overall randomness based on a number of lags. The Ljung-Box test is commonly used in ARIMA modeling and it is applied to the residuals of a fitted ARIMA model, not only the original series. The Akaike Information Criterion (AIC) estimates the quality of each model relative to each of the other models and it provides a means for model selection. Another straight form and common measure of the reliability of the model is the accuracy of its forecasts generated based on a partial data so that the forecast can be compared with the original observations.

However, a good model should not only provide sufficiently accurate forecasts, it should be fit and produce statistically independent residuals that contain only noise and no systematic components.

3. Data Analysis and Empirical Results

The data used in this study was collected from the statistical bulletin of organization of petroleum exporting countries (OPEC). It is a secondary data because it was extracted from their records and the period considered is from 1985-2015 and the statistical package used was *R*-Program version 2.3.

The descriptive statistics analysis results are presented in the table 1.

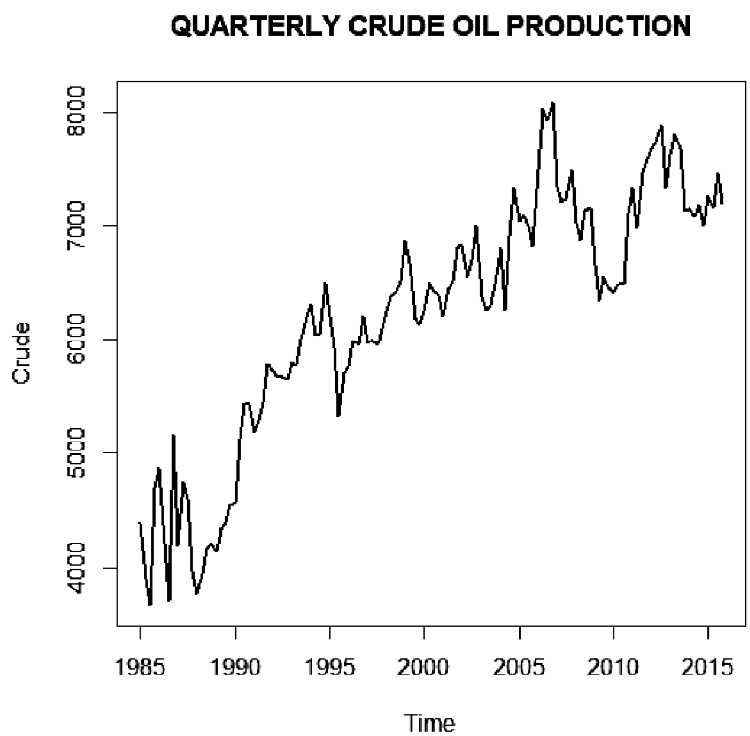

Figure 1: The plot of Crude Oil production

Table 1: Descriptive Statistics of Crude Oil Production in Nigeria.

Year	*Minimum*	*Maximum*	*Mean*	*Std. Dev.*	*Skewness*	*Kurtosis*
1985	1125	1565	1.3888E3	159.1901	–0.267	–1.341
1986	1016	1789	1.4967E3	259.0073	–0.759	–0.671
1987	1197	1761	1.4589E3	170.7771	0.158	–1.049
1988	1183	1451	1.3396E3	87.1055	–0.791	–0.347
1989	1360	1560	1.4500E3	61.7547	0.122	–0.565
1990	1474	1879	1.7147E3	141.7452	–0.562	–0.515
1991	1731	1929	1.8094E3	88.2275	0.414	–1.782
1992	1809	1930	1.8919E3	33.3206	–1.682	2.656
1993	1900	2100	1.9812E3	64.0712	0.484	–0.939
1994	1975	2175	2.0500E3	58.8527	0.971	0.325
1995	1545	2085	1.9322E3	143.6187	–1.840	4.761
1996	1855	2110	1.9925E3	69.9838	–0.244	0.137
1997	1965	2038	2.0006E3	23.8916	0.103	–1.087
1998	2026	2208	2.1296E3	45.9831	–0.683	1.643
1999	1960	2380	2.1541E3	125.3856	0.004	–0.707
2000	2010	2190	2.1292E3	54.0132	–1.320	0.898
2001	2030	2287	2.1648E3	77.5779	0.092	–0.417
2002	2140	2360	2.2564E3	77.0802	–0.161	–1.105
2003	2050	2200	2.1178E3	43.6601	0.038	–0.296
2004	1994	2479	2.2741E3	163.2619	–0.512	–1.151
2005	2210	2395	2.3290E3	50.0854	–1.003	2.008
2006	2430	2695	2.6267E3	91.1127	–1.307	0.746
2007	2370	2560	2.4396E3	64.4719	0.563	–0.690
2008	2230	2430	2.3500E3	66.3668	–0.962	–0.426
2009	2060	2330	2.1651E3	75.7093	0.701	0.604
2010	2051	2450	2.2080E3	122.1043	0.508	0.042
2011	2310	2580	2.4467E3	88.7625	–0.270	–0.927
2012	2400	2640	2.5510E3	88.9883	–0.839	–0.804
2013	2280	2640	2.5200E3	111.5185	–1.223	0.905
2014	2260	2420	2.3667E3	56.6221	–0.984	–0.187
2015	2320	2520	2.4236E3	60.5692	–0.563	–0.084

Anderson-Darling Test

H_0 : The variable from which the sample was extracted follows a Normal distribution

H_1 : The variable from which the sample was extracted does not follow a Normal distribution

The Anderson-Darling test-statistic is 5.019 with *p*-value $1.973e^{-12}$

Since *p*-value is less than 0.05, we therefore conclude that the data is not normally distributed.

Test for Stationarity (Augmented Dickey-Fuller Unit Root Test)

The hypotheses are:

H_0 : Crude oil production has a unit root (is not stationary)

H_1 : Crude oil production is stationary

Augmented Dickey-Fuller test statistic is –6.5834 with probability 0.01.

Since the *p*-value (0.01) is less than 0.05, we reject null hypothesis, hence conclude that the data does not have a unit root, the data is stationary.

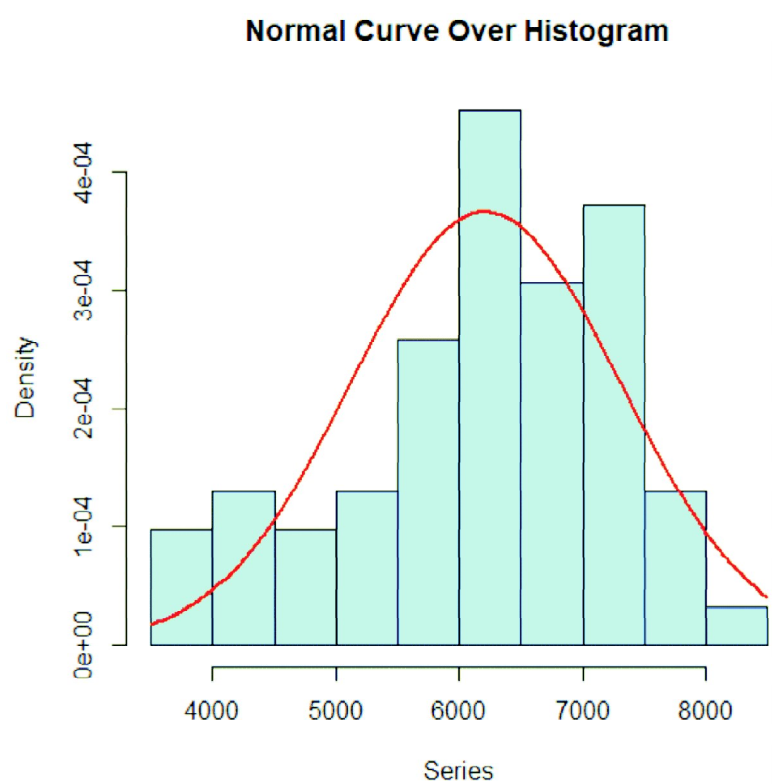

Figure 2: Histogram of quarterly Crude Oil Production in Nigeria

Table 2: Sample ACF and PACF for Crude Oil Production in Nigeria

Lag	*ACF*	*PACF*
1	1	0.749812
2	0.749812	0.040936
3	0.580139	–0.13043
4	0.390878	–0.16371
5	0.193207	–0.02161
6	0.058479	0.150027
7	0.028187	–0.04259
8	–0.01668	–0.16853
9	–0.08383	–0.11676
10	–0.14468	–0.10801
11	–0.23190	–0.06363
12	–0.31161	–0.01967
13	–0.34176	–0.07021
14	–0.35317	–0.17035
15	–0.38142	–0.08571
16	–0.37286	–0.09220
17	–0.37466	–0.11513
18	–0.38917	–0.04564
19	–0.35785	–0.15420
20	–0.33998	–0.12723
21	–0.30858	–0.06846
22	–0.24412	–0.02439
23	–0.15696	–0.08952
24	–0.09140	–0.15096
25	–0.03072	–0.21338

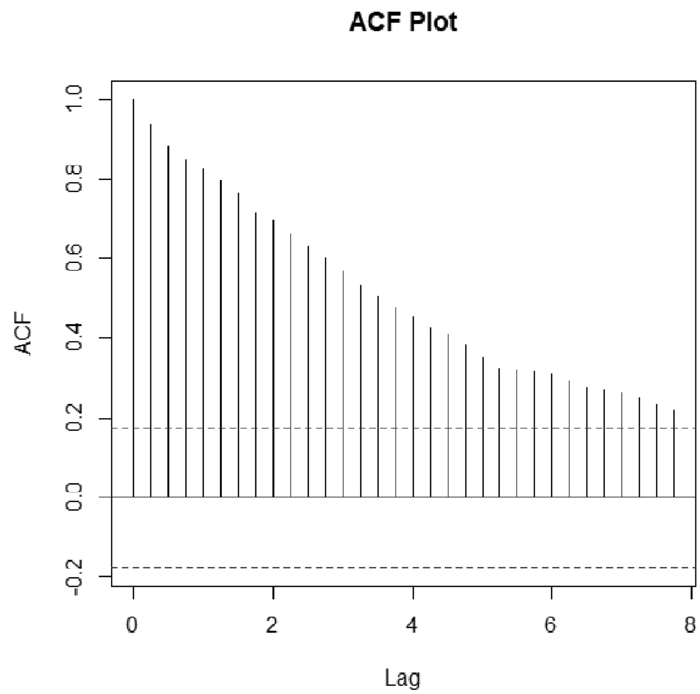

Figure 3: ACF Plot for Crude Oil Production in N igeria

Fig. 4 missing (NA in word file)
Figure 4: PACF Plot for Crude Oil Production in Nigeria

Fitting ARIMA model to the Crude Oil Data

Table 3

Table 3: Fitting ARIMA (1, 0, 1) model to the Crude Oil Data

Model	*Coefficient*	*Standard Error*
AR(1)	0.7763	0.0416
MA(1)	–0.0407	0.0594
Intercept	2057.1809	53.5966

AIC = 4989.64

Table 4: Fitting ARIMA (1, 0, 2) model to the Crude Oil Data

Model	*Coefficient*	*Standard Error*
AR(1)	0.7211	0.0504
MA(1)	–0.0378	0.0655
MA(2)	0.1798	0.0684
Intercept	2057.114	50.774

AIC = 4984.64

Table 5: Fitting ARIMA (1, 0, 3) model to the Crude Oil Data

Model	*Coefficient*	*Standard Error*
AR(1)	0.5323	0.0718
MA(1)	0.1488	0.0717
MA(2)	0.3345	0.0517
MA(3)	0.2906	0.0608
Intercept	2057.8118	45.9822

AIC = 4964.28

Using autoarima function in *R*-Programming it generated a SARIMA model; this had been indicated in the time plot that there is seasonality in the data set.

Table 6: Fitting SARIMA (3, 0, 1)(1, 0, 2)^4 model to the Crude Oil Data

Model	*Coefficient*	*Standard Error*
AR(1)	1.5828	0.0529
AR(2)	–0.4325	0.0983
AR(3)	–0.2152	0.0544
MA(1)	–0.9667	0.0133
SAR(1)	–0.4891	0.1385
SMA(1)	0.5320	0.1401
SMA(2)	0.2438	0.0667
Intercept	2065.4283	7.7326

AIC = 4949.96

Diagnostic Checking

Over fitting is a method of diagnostic checking. In this study the method of over fitting additional parameters were included in the autoregressive function in order to fit the ARIMA model.

Table 7: Table for diagnostic checking using AIC

Model	ARIMA (1, 0, 1)	ARIMA (1, 0,2)	ARIMA (1, 0, 3)	SARIMA (3,0, 1)(1, 0, 2)^4
AIC	4989.64	4984.64	4964.28	4949.96

From the table above, it can be seen that the seasonal ARIMA model has the lowest AIC.

Table 8: Ljung-Box Statistics

Model	*Chi-Square*	*p – value*
ARIMA (1, 0, 1)	0.035824	0.8499
ARIMA (1, 0, 2)	0.163270	0.6862
ARIMA (1, 0, 3)	0.051439	0.8206
SARIMA (3, 0, 1)(1, 0, 2)^4	0.020134	0.8872

Table 9: Forecasts crude oil production for January to December 2017 are:

Months	*Forecasts*
January	2380.215
February	2283.373
March	2210.253
April	2237.723
May	1966.295
June	1957.151
July	1792.439
August	1787.543
September	1834.456
October	1893.858
November	1876.441
December	1848.832

Table 10: Forecasts crude oil productions for January to December 2018 are:

Months	*Forecasts*
January	1891.523
February	1950.396
March	1904.028
April	1953.245
May	2053.476
June	2089.145
July	2089.145
August	2138.162
September	2185.854
October	2168.741
November	2154.112
December	2163.208

4. Discussion of Results

From the time plot, it can be seen that the production of crude oil in Nigeria is seasonal in nature. Using Box-Jenkins approach, the model for the quarterly crude oil production in Nigeria is achieved by estimating the autocorrelation (ACF) and the partial autocorrelation function (PACF). The selection of a tentative time series is frequently accomplished by matching estimated autocorrelations with the theoretical autocorrelation. The matching of the first 25 estimated sample autocorrelations and partial autocorrelations of the underlying stochastic processes suggested that the series were stationary with the ACF, PACF.

The estimated ACF and PACF are shown in table 2. From the autocorrelation function plot (figure 3), it was shown that there is a gradual decay in the correlation between various lags. Also, from the PACF plot (figure 4), there was a cutoff at lag 1 indicating that the model is an MA(1) for the quarterly crude oil production in Nigeria. Several related ARIMA model was fitted for the data *ARIMA*(1, 0, 1), *ARIMA*(1, 0, 2), and *ARIMA*(1, 0, 3). Using auto arima package in *R* software to select the best fitted model, seasonal *ARIMA SARIMA* $(3, 0, 1)(1, 0, 2)^4$ was fitted. Diagnostic checking shows that amongst the fitted model, the SARIMA model is best as it has the least AIC values. The $SARIMA(3, 0, 1)(1, 0, 2)^4$ means $ARIMA(p, d, q) \times (P, D, Q)S$ which "*p*" is non seasonal *AR* order, "*d*" is non seasonal differencing, and "*q*" is non seasonal *MA* order, "*P*" is seasonal *AR* order, "*D*" is seasonal differencing, "*Q*" is seasonal *MA* order and "*S*" is time span of repeating seasonal pattern. $SARIMA(3, 0, 1)(1, 0, 2)^4$ has a non seasonal *AR*(3), non seasonal *MA*(1), a seasonal *AR*(1), a seasonal *MA*(2), no differencing and the seasonal period is *S*(4).

The Ljung-Box statistics for all the models gives a non-significant *p*-value which indicates that the residual are not correlated, which implies that the models are adequate. Thus, the four (4) models fitted are:

5. Conclusion and Recommendation

In conclusion, the $SARIMA(3, 0, 1)(1, 0, 2)^4$ has the lowest AIC of 4949.96 and chi-square *p*-value of 0.8872 which make it better than the other three (3) models. It was seen that the production of crude oil can be seasonal. In such a case, *SARIMA* model is recommended for modeling crude oil production in Nigeria. Therefore, policy can be made to follow when there is high volume of production and when there is low volume of production and it will help the government of Nigeria planning purposes and for making future policies in order to generate more income from crude oil production sector. The major drawback of this work is non-availability of recent data in the study area.

References

Alhaji, A f., Huettner, D. (2000). OPEC and World Crude Oil Markets from 1973 t0 1994: Cartel, Oligopoly or Competitive? *The Energy Journal* 21 (3), 31-60.

Contreras, J., Espinola, R., Nogales, F. J., and Conejo, A. J. (2003). ARIMA Models to Predict Nextday Electricity Prices, *IEEE Transactions on Power Systems*, vol. 18 No. 3, 1014 1020..

Griffin, J. M (1985). OPEC Behavior: A Test of Alternative Hypotheses. *American Economic Review* 75 (5), 954-963.

Hotelling, H. (1931). The Economics of Exhaustible Resources, *Journal of Political Economy* 39,

Hubbert, M. K. (1956). Nuclear Energy and the Fossil Fuels. American Petroleum Institute Drilling and Production Practice, Proceedings of Spring Meeting, San Antonio, 7-25.

Kaufmann, R. K., Bradford, A., Belanger, L. H., Mclaughlin, J. P and Miki, Y (2008). Determinants of OPEC Production: Implications for OPEC Behaviour, *Energy Economics* 30, 333-351.

Kyongwook, C., Wei-Choun, Y. and Eric Z. (2-13). Long Memory versus Structural Breaks in Modeling and Forecasting Realized Volatility. Journal of Business and Economic Statistics 29: 234-257.

MacAvoy, P (1982). Crude Oil Prices as Determined by OPEC and Market Fundamentals.

Meyler, Aidan., Kenny, Geoff nd Quinn, Terry (1998). Forecasting Irish inflation using ARIMA models, Central Bank and Financial Services Authority of Ireland Technical Paper Series, Vol. 1998, No. 3/TR/98, 1-48.

Raymond, Y. C. Tse (1997). An application of the ARIMA model to real-estate prices in Hong-Kong, *Journal of Property Finance*, Vol.8, No. 2, 152-163.

Ramcharran, H. (2002). Production Responses to Price Changes: An Emperical Application of the Competitive Model to OPEC and Non-OPEC Countries. *Energy Economics* 24, 97-106.

Stergiou, K. I (199). Modeling and forecasting the fishery for pilchard (Sardinapilchardus) in Greek waters using ARIMA time-series models, *ICES Journal of Marine Science*, Volume 46, No.1, 16-23.

Smith, J. L (2005). Inscrutable OPEC? Behavioral Tests of the Cartel Hypothesis. *The Energy Journal* 26 (1), 51-82.

Teece, D (1982). OPEC Behavior: An Alternative View in Griffin, J. M, OPEC Behavior and World Oil Prices, London: Allen & Unwin, World Bank, Worldwide Governance Indicators. Data and extensive documentation available at http://info.worldbank.org/governance/wgi//index.asp.

Wurzel, E., Willard, L and Ollivaud, P (2009). Recent Oil Price Movements-Forces and Policy Issues OECD Economics Department Working Paper No.737.

Siti, R. Y., Maizah, H. A., Lee, C. N. and Noryanti, M. (2011). “A comparative Study on Box-Jenkins and Garch Models in Forecasting Crude Oil Prices”. Journal of Applied Sciences 11(7): 1129-1135, Asian Network for Information.