European Journal of Academic Essays 2(3): 1-12, 2015
ISSN (online): 2183-1904
ISSN (print): 2183-3818
www.euroessays.org

On the Time Series Modelling of Crude Oil Exportation in Nigeria

<**Balogun, Oluwafemi Samson**>[*1], <**Ogunleye, Opeoluwa Mercy**>[2]

[1]Department of Statistics and Operations Research, Modibbo Adama University of Technology, P.M.B. 2076, Yola, Adamawa State, Nigeria.

stapsalms@yahoo.com

[2] Department of Statistics, University of Ilorin, P.M.B.1515, Ilorin, Kwara State. Nigeria.

Abstract: The research result showed there was a general increase in production over time (production was mostly below 60million 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 and its showed that by January the production was 75.02 million barrel and by december production was about 67.25 million barrel, though its a wide difference from the actual value but was still chosen because of its simplicity to use and that its also has the lowest standard error after the forecast. ARIMA (1,1,1) was used to predict for short-term purpose, therefore the year 2012 was predicted and it revealed that the minimum production for 2012 should be around 66.63 million barrel and the maximum production should be around 68.83 million barrel.
Keywords: Autocorrelation, Autoregression, ARIMA, AIC and Jarque–Bera test

---

1. Introduction

The petroleum industry in Nigeria, Africa, is the largest industry and main generator of GDP in the continent's most populous nation. Historically, the Nigerian petroleum scene opened as far back as 1908, when a German company, the Nigerian Bitumen Corporation, was attracted to what is now known as the south-western Nigerian Tar Sand deposit. After World War 1, Shell-D'Arcy, a consortium of Shell and Royal Dutch, resumed oil exploration in 1937, this time in Owerri, on the northern frame of the Niger Delta. On June 5, 1956, after drilling 28 wells and 25 core holes, all dry, the new operator, Shell-BP, struck oil at Oloibiri in what is now Bayelsa State.

The Niger Delta became Nigeria's bumper scene of feverish exploration and production. From an initial output of 5,100 barrels per day in 1956, the nation steadily rose to the sixth position on the production chart of the Organization of Petroleum Exporting Countries (OPEC). By the mid-1970s, Shell, the leading producer had exceeded the one million barrels a day production mark. After over three decades during which the oil industry was dominated by foreign companies, a private indigenous oil company, Consolidated Oil, recorded its first discovery, Bella - 1, in 1991. Since 1992, following the release of new concessions in the Niger Delta to indigenous exploration and production companies, the number of indigenous companies has increased to 12. So far, out of some 870 oil fields discovered in Nigeria, only 120 fields are currently producing. Most of the fields are not producing because the country has to abide by OPEC's production quota of 1.8 million barrels per day to Nigeria. Violence in the oil-producing communities has also disrupted production, causing the shut-up of most land and swamp wells. Production is sustained by offshore fields.

Certain notable achievements and developments in Nigeria's petroleum exploration and production ventures deserve to be highlighted. In order to raise the country's proven petroleum reserves from 23 billion barrels to the target 25 billion barrels set for. 1995, the Federal Government not only opened new acreages for exploration, but also offered a package of fiscal incentives to petroleum companies. Among the incentives is the reduction of petroleum tax to boost exploration in the deeper offshore. Potential reserves in billion barrels are estimated for the new blocks which hold good prospects for smaller fields with less than 50 million barrels. Generally, in the Niger Delta, about 73 per cent of crude oil discoveries are in fields having less than 50 million barrels of proven reserves. The overall wildcat success ratio is 42 per cent. However, in some years the success ratios of exploratory and appraisal/development wells are substantially higher (83.5 per cent in 1989).

Petroleum prospects in the offshore Niger Delta are most attractive, with a potential 1.10 billion barrels

***Corresponding Author:** Balogun Oluwafemi Samson, Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola, Adamawa State, *stapsalms@yahoo.com*, +2348035217339

of crude awaiting discovery in recently awarded Oil Prospecting Licenses (OPLs). Oil Prospecting Licenses in the deeper Offshore (200-150m water depth) have received highly competitive bids, which extensive regional seismic and geochemical surveys have shown to be quite attractive. A new development in Nigeria's petroleum prospecting is the unitization scheme. Under this arrangement, petroleum prospecting companies, in order to reduce cost, conduct joint exploration and development of undeveloped oil fields which straddle their common concession boundaries. The exportation of crude oil in Nigeria has been on the increase since its discovery six decades ago. As of 2000, oil and gas export 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. The oil industry has been marred by political and economic strife due to a long history of corrupt military regimes and the complicity of multinational corporations, notably Royal Dutch Shell.

This research work is on time series modelling of crude oil exportation in Nigeria. Nigeria is a developing country faced with the challenged of corruption especially in the oil and gas industry, therefore posing a threat to the economic sustainment of the country. The model arrived at can be used to forecast, telling the possible income in the nearest future.

The aim of this research work is to to identify the trend, cyclic and seasonal variations (if they exists), to model exportation of crude oil and to forecast especially short-term ahead, using the best model derived.

The data used for this research is a secondary data obtained from the Nigerian National Petroleum Corporation (NNPC) bulletin, which described the aggregate monthly production.

Time series was originated in 1807 by French Mathematician name FOURIER, who claimed that any Series could be approximated as the sum of the Sine and Cosine terms.

According to [20] he defines time series as a set of observations taken at a specified time usually at equal interval.

According to[10] he define time series as a statistical series which tell us how data has been behaving in the past.

According to [7], he defines time series as a collection of observation segmental in time at regular intervals.

The usage of time series models is in twofold:

(1) To obtain an understanding of the underlying forces and structure that produced the observed data, and
(2) To fit a model and proceed to forecasting, monitoring or even feedback and feed forward control.

Time Series Analysis's includes: Economic Forecasting, Sales Forecasting, Budgetary Analysis, Stock Market Analysis, Yield Projections, Process and Quality Control, Inventory Studies, Workload Projections, Utility Studies, Census Analysis, and many, many more...

1.1 Types of Time Series

There are 3 types of time series which are:

(1) **Continuous Time Series:** This involves Hydrological parameters which are often continuously recorded. This occurs either on the record sheet of a chart recorder, or a data logger is used. A data logger typically records the data either at fixed time intervals or after a certain change in the Y-value has taken place. Despite this sampling, the data are interpreted as if they were continuous data. The data are recorded so that the information content due to the continuity is retained. (E.g. a precipitation event or precipitation free).

(2) Interval Time Series: An interval time series does not contain values for points in time but rather for particular intervals of time. These time intervals can be equidistantly or randomly distributed in time. Equidistant in terms of years or months still means that the actual intervals have different lengths. A typical equidistant time series is a daily total series, where each value is for an interval of 24 hours.

(3) Momentary Time Series: The momentary time series is the rarest form of time series. In contrast to the other time series, a momentary time series is only defined for a discrete set of points in time. The time series does not contain any information for the time between these points. Interpolation is not meaningful, and the value function thus has the value undefined for these points. An example of a momentary time series is the series of local maxima of a precipitation time series. The set of points in time is made up of randomly distributed points in time. There is no information for all other points in time.

1.2 Mathematical Models for Time Series

The following are the two models used for the decomposition of time series into its components:

1. **Additive Model or Decomposition by Additive Hypothesis** which can be expressed as:

$$Y_t = T_t + S_t + C_t + I_t$$

where Y_t denotes response
T_t denotes trend component
S_t denotes Seasonal component
C_t denotes cyclical component

This model assumes that all components of time series are independent of one another. For example, it assumes that trend has no effect on the seasonal and cyclical components, nor does seasonal swing have any influence on cyclical variation and vice versa. This model also assumes that the different components are absolute quantities expressed in original units and can take positive and negative values.

2. **Multiplicative Model or Decomposition by Multiplicative hypothesis** which can be expressed as:

$$Y_t = T_t \times S_t \times C_t \times I_t$$

In this model only trend is expressed in terms of original values, which the seasonal and cyclical components are expressed as relatives or percentages. This model assumes that the four components of a time series are due to different causes, but they are necessary and they can affect each another.

As most of the time series conform to the multiplicative model, the additive model is rarely used. In this research work, the multiplicative model would be employed.

1.3 Components of Time Series

The various forces affecting the values of a phenomenon in a time series may be broadly classified into the following four categories:

- Secular Trend or Long-Term Movement (T)
- Periodic movement of Short-term Fluctuations which can be *seasonal variations (S) or cyclical variations (C)*
- Random or Irregular Variations (R or I)

These components are used for descriptive purposes.

Trend

Trend also called secular or long-term trend, is the basic tendency of a series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movements over a long time.

Measurement of Trend

The following are the four methods which are generally used for the study and measurement of a trend component in time series:

1. Graphic (or Free-hand Curve Fitting) Method
2. Method of Semi-Averages
3. Method of Curve Fitting by the Principle of Least Squares
4. Method of Moving Averages

Seasonal Variation

Seasonal variation is a component of a time series which is defined as the repetitive and predictable movement around the trend line in one year or less. It is detected by measuring the quantity of interest for small time intervals such as days, weeks, months or quarter.

Detecting Seasonality. The following graphical techniques can be used to detect seasonality.

1. A run sequence plot will often shows seasonality.

2. A seasonal subseries plot is a specialized technique for showing seasonality.

3. Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.

4. The autocorrelation plot can help identify seasonality.

Measurement of Seasonal Variation

The following are different methods of measuring seasonal variation:

1. Methods of Simple Averages
2. Ratio to Trend method
3. Ratio to Moving Average method
4. Link Relative Method

In this research, the methods of Simple Averages was used and can be obtained using

$$Seasonal\ Index = \frac{Monthly\ Average}{Total\ Average} \times 100$$

Cyclical Variation

It is the oscillatory movements in a time series with period of oscillation greater than one year. These variations in a time series are due to ups and downs recurring after a period greater than one year.

Measurement of Cyclical Variation

An approximate or crude method of measuring cyclical variation is the 'Residual Method' which consists in first estimating trend (T) and seasonal (S) components, and then eliminating their effects on the given time series.

For a multiplicative time series, cyclical variation is obtained thus

$$\frac{Y}{T \times S} = \frac{TSCI}{TS} = CI$$

1.4 Types of Time Series

There are two major types of time series: Stationary and Non Stationary time series.

Stationary Time Series: A time series is said to be stationary if there is no observed change in periodic variation or mean, it implies that there is no trend since the time series is stagnant i.e. change in time series does not affect the trend line, no increase or decrease. A stationary time series has a constant mean, a constant variance and the covariance is independent of time. Stationarity is essential for standard econometric theory. Without it one cannot obtain consistent estimators.

A quick way of telling if a process is stationary is to plot the series against time. If the graph crosses the mean of the sample many times, chances are that the variable is stationary; otherwise that is an indication of persistent trends away from the mean of the series.

1.5 Estimation of Seasonal Variation/Seasonal Index

To determine the seasonal component in the time series data, we must estimate how the data in the time series vary from month to month throughout a typical year. Various methods are available for computing a seasonal index.

The Average-percentage Method: In this method, we express the data for each as percentages of the averages of the year. The percentages for corresponding month of different years are then averaged, the resulting 12 percentages gives the seasonal index.

The percentage Trend, or Ratio-To-Trend Method: In this method, the trend for each month is calculated by the method of least squares. We express the data for each month as percentage of monthly trend value. An appropriate average of the percentage for corresponding months given the required index.

The Percentage Moving-Average or Ratio-to-Moving Average Method: In this method, we compute a 12-month moving average. Since the result thus obtained fall between successive months instead of the middle of the month (which is where the original data fall), we compute a 2 – month moving average of this 12-month moving average. After doing this, we express the original data for each month as a percentage of the 12-month centered moving average that corresponds to the original data. The percentages for the corresponding months are then average, giving the required index.

1.6 Autocorrelation

Autocorrelation is a test that the residuals from a linear regression or multiple regressions are independent. The autocorrelation of a random process describes the correlation between values of the process at different times, as a function of the two times or of the time lag. Let X be some repeatable process, and i be some point in time after the start of that process. (i may be an integer for a discrete-time process or a real number for a continuous-time process.) Then X_i is the value (or realization) produced by a given run of the process at time i. Suppose that the process is further known to have defined values for mean μ_i and variance σ_i^2 for all times i.

Autocorrelation refers to correlation of values of time series variables with values of the same variable lagged one or more time period back.

Method: Because most regression problems involving time series data exhibit positive autocorrelation, the hypotheses usually considered in the Durbin-Watson test are

H_0: ρ = 0
H_1: ρ > 0

The test statistic is

$$d = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i{}^2}$$

Where $e_i = y_i - \hat{y}_i$ and y_i and $\hat{y}_i$ are, respectively, the observed and predicted values of the response variable for individual i.e. d becomes smaller as the serial correlations increase. Upper and lower critical values, d_u and d_L have been tabulated for different values of k (the number of explanatoryy variables) and n.

If d <dL, reject H_0: ρ = 0
If d >dU, do not reject H_0: ρ = 0
If dL< d <dU, test is inconclusive.

1.7 Correlogram Analysis

It is a group of r_k (autocorrelation coefficient at lag k) plotted against lag k. the objective is to give a clear interpretation of a set of autocorrelation coefficient. Correlogram analysis reveals trends and seasonal variation if present in the series. If the presence of variation is detected, the plotted autocorrelation coefficient require higher lags to tend to zero, it implies that the series is not stationary the correlogram tends to zero faster.

1.8 Unit Root Test

Unit root is a feature of processes that evolve through time that can cause problems in statistical inference involving time series models. In statistics, a **unit root test** tests whether a time series variable is non-stationary using an autoregressive model. A well-known test that is valid in large samples is the augmented Dickey–Fuller test.

A unit root test is a statistical test for the proposition that in an autoregressive statistical model of a time series, the autoregressive parameter is one. In a data series y(t), where t a whole number modeled by: y(t+1) = ay(t) + other terms, where a is an unknown constant, a unit root test would be a test of the hypothesis that a=1, usually against the alternative that |a| is less than 1.

Properties and characteristics of Unit-root processes

- Shocks to a unit root process have permanent effects which do not decay as they would if the process were stationary
- As noted above, a unit root process has a variance that depends on t, and diverges to infinity
- If it is known that a series has a unit root, the series can be differenced to render it stationary. For example, if a series is Y_t is I(1), the series $\Delta_t = Y_t - Y_{t-1}$ is I(0) (stationary). It is hence called a difference stationary series

1.9 Time Series Models

A time series consists of observations at discrete equi-spaced intervals of time. For example, Accidents in month "t" could be denoted as X_t and in the previous month by X_{t-1}. Typically, the objective of time series analysis is to forecast future values of X (such as X_{t+1}) based on present and past values of X and perhaps also on explanatory variables such as accidents.

A model in which future values are forecast purely on the basis of past values of the time series is called an Autoregressive (AR) process.
A model in which future values are forecast purely on the basis of past shocks (or noise or random disturbances) is called a Moving average (MA) process.
A model that uses both past values of the time series and past shocks is called an autoregressive-moving average (ARMA) process.
These models assume that the time series is stationary - that is the series fluctuates around a time invariant mean, and the variance and autocovariance i.e. covariance between X_t and X_{t-s} (for all values of s) do not vary with time. In practice, most time series need to be transformed to achieve stationarity. To stabilize variance a logarithm transform is often used - appropriate where the variance of the series increases in proportion to the mean. To stabilize the mean, differencing is usually employed. For example, first order differencing is $Z_t = X_t - X_{t-1}$. First order differencing eliminates "drift" but it often needs to be applied twice to eliminate trend. Seasonal differencing is often necessary too. An ARMA model of a differenced series is called an ARIMA model, where the 'I' stand's for Integrated because the output needs to be anti-differenced or integrated, to forecast the original series.
Collectively, these models along with the process of identification, fitting, and diagnostic checking are called Box Jenkins models [12].

One fundamental goal of statistical modeling is to use the simplest model possible that still explains the data. This is known as principle of parsimony [14].

ARIMA model

Early attempts to study time series, particularly in the 19th century, were generally characterized by the Idea of a deterministic world. It was the major Contribution of [23] which launched the notion of stochasticity in time series by postulating that every time series can be regarded as the realization of a stochastic process. Based on this simple idea, a number of time series methods have been developed since then.

Workers such as Slutsky, Walker, Yaglom, and Yule first formulated the concept of autoregressive (AR) and moving

average (MA) models. Wold's decomposition theorem led to the formulation and solution of the linear forecasting problem of [13].
Since then, a considerable body of literature has appeared in the area of time series, dealing with parameter estimation, identification, model checking, and forecasting; e.g., Newbold (1983) for an early survey. The publication Time Series Analysis: Forecasting and Control by Box and Jenkins (1970) integrated the existing knowledge.

AUTOREGRESSIVE

Autoregressive (AR) models were first introduced by [22]. They were consequently supplemented by [19] who presented Moving Average (MA) schemes. It was [21]), however, who combined both AR and MA schemes and showed that ARMA processes can be used to model all stationary time series as long as the appropriate order of p, the number of AR terms, and q, the number of MA terms, was appropriately specified. This means that any series x_t can be modeled as a combination of past x(t) values and/or past e(t) errors.
The utilization of the theoretical results suggested by Wold, to model real life series did not become possible until the mid 1960s when computers, capable of performing the required calculations became available and economical. [3] in the original edition [2] popularized the use of ARMA models through the following:
(a) Providing guidelines for making the series stationary in both its mean and variance,
(b) Suggesting the use of autocorrelations and partial autocorrelation coefficients for determining appropriate values of p and q (and their seasonal equivalent P and Q when the series exhibited seasonality),
(c) Providing a set of computer programs to help users identify appropriate values for p and q, as well as P and Q, and estimate the parameters involved and
(d) Once the parameters of the model were estimated, a diagnostic check was proposed to determine whether or not the residuals e(t) were white noise, in which case the order of the model was considered final (otherwise another model was entertained in (b) and steps (c) and (d) were repeated). If the diagnostic check showed random residuals then the model developed was used for forecasting or control purposes assuming of course constancy that is that the order of the model and its non-stationary behavior, if any, would remain the same during the forecasting, or control, phase.
The approach proposed by Box and Jenkins came to be known as the Box-Jenkins methodology to ARIMA models, where the letter "I", between AR and MA, stood for the word "Integrated". ARIMA models and the Box-Jenkins methodology became highly popular in the 1970s among academics, in particular when it was shown through empirical studies ([8]; [17]; [9]; [16]; [15] for a survey see [2]) that they could outperform the large and complex econometric models, popular at that time, in a variety of situations.

2. Methodology

If future values can be predicted exactly from past values, then a series is said to be deterministic. One fundamental goal of statistical modeling is to use the simplest model possible that still explains the data. This is known as principle of parsimony [14].

A model for a Stochastic time series is always called a Stochastic process and can be said to be a random variables family indexed by time (i.e., X_1, X_2, …) or generally (X_t) in discrete time space.

More precisely, $\{X_t, t \in T\}$ where T is the index of times on which the process is defined. The notation is necessary when observations are not equally spaced through time, but we restrict attention to the equally spaced case when the index set consisting of positive integers is commonly used [6].

Time series plot: - The time plot is the graphical representation of data. The first step in any time series analysis process is to plot the observed variables against time. The time plot reveals the presence of the likely component in the data.

2.2 Time Series Analysis Process

The first thing to do is to **Test for Stationarity of the series** using three different approaches. The approaches are (i) observing the graph of the data to determine whether it moves systematically with time (ii) the ACF (Autocorrelation Function) and the PACF (Partial Autocorrelation Function) of the stochastic processs (iii) Fit the unit root test using Argumented Dickey Fuller test on the series by considering different assumptions such as under constancy, along with no drift or along a trend and a drift term (iv) do the variable lag selection. If found that the series is not stationary at initial level, we first identify the order of diffencing needed to stationarizes the series and remove gross features of seasonality, perphaps in conjuction with a variance-stabilizing transformation and there are ways of carrying out transformation namely log transformation (carried out when variances are not stable) and difference transformation. The ACF and PACF are approximately normally distributed about their population and have standard deviation of about $1/\sqrt{T}$, where T is the length of the series.

2.3 Model Identification: When the series is identified stationary,the next thing is to derive the different values of p and q, and then estimate the parameters of ARIMA model. Since we know that sample autocorrelation and sample partial autocorrelations are compared with the theoretical plots, but it's very hardly to get the patterns similar to the theoretical plots one. So therefore an iterative method on the number of components for the Moving Average (MA) and Autoregressive (AR) was employed considering the information criteria i.e. AIC (Alkaike information criteria) and BIC (Bayesian information criteria), and relatively small SEE (standard error of estimate).

Equation for ARIMA model: $\Delta_d y_t = \mu + \sum_{i=1}^{\varphi} a_i y_{t-i} + \sum_{j=1}^{q} b_j y_{t-j} + U_t$

ARIMA models are designed to squeeze all autocorrelation out of the original time series; a systematic procedure exists for identifying the best ARIMA model for any given time series; ARIMA (p,d,q) models often provide a good fit to highly plentiful data.

p is the number of autoregressive terms, d is the number of nonseasonal differences and q is the number of lagged forecast errors in the prediction terms.

2.4 Akaike Information Criteria (AIC)

The Akaike information Criteria is generally regarded as the first model selection criterion, the most widely known and used model selection tool among practitioners. It was introduced by [1] in his Seminar paper "Information Theory and an Extension of the Maximum Likehood Principle". Thus Akaike proposed a framework wherein both model estimation and selection could be simultaneously accomplished. It measures the relative goodness of fit of a statistical model, describes the tradeoff between bias and variance and estimates which the Kullback-Leibler information lost by approximating full reality with the fitted model.

AIC= -2log (L) + 2p

where p is the number of model parameters, L is the maximum value of the likehood function for the estimated model. According to Wikipedia, the free encyclopedia, The **Akaike information criterion** (**AIC**) is a measure of the relative quality of a statistical model, for a given set of data. As such, AIC provides a means for model selection.

AIC deals with the trade-off between the goodness of fit of the model and the complexity of the model. It is founded on information entropy: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data.

AIC does not provide a test of a model in the sense of testing a null hypothesis; i.e. AIC can tell nothing about the quality of the model in an absolute sense. If the entire candidate models fit poorly, AIC will not give any warning of that.

2.5 Model Estimation: The parameters for the Box-Jenkins models is a quite complicated non-linear estimation problem. For this reason, the parameter estimation should be left to a high quality software program that fits Box-Jenkins models. Fortunately, many commercial statistical software programs now fit Box-Jenkins models. The main approaches to fitting Box-Jenkins models are non-linear least squares and maximum likelihood estimation.

Maximum likelihood estimation is generally the preferred technique. The likelihood equations for the full Box-Jenkins model are complicated and are not included here. See [5] for the mathematical details. At this stage four best models were considered in this analysis.

2.6 Diagnostic Checking (Testing the normality of the residual): Before the interpretation and use of the fitted model, we are to look at some tests to check whether the model specified is suitable for use. The following tests are applied to the residuals: (i) Test for autocorrelation and partial autocorrelation (ii) ARCH test for residual (iii) Jarque-Bera test for non-normality.

2.7 Jarque-Bera test for non-normality: In statistics, the **Jarque–Bera test** is a goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution. The test is named after Carlos Jarque and Anil K. Bera. The test statistic JB is defined as

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

where *n* is the number of observations (or degrees of freedom in general); *S* is the sample skewness, and *K* is the sample kurtosis:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{3/2}}$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{2}},$$

where $\hat{\mu}_3$ and $\hat{\mu}_4$ are the estimates of third and fourth central moments, respectively, $\bar{x}$ is the sample mean, and $\hat{\sigma}^2$ is the estimate of the second central moment, the variance.

If the data come from a normal distribution, the JB statistic asymptotically has a chi-squared distribution with two degrees of freedom, so the statistic can be used to test the hypothesis that the data are from a normal distribution.

Forecast: the best model specified is used to forecast the nearest future. **Time series forecasting** is the use of a derived model to predict future values which can either be short-term or long-term purpose and they are based on the observations on hand.

3. Data Analysis

3.1 Time Plot

The time plot shows that effectively there was an observed increase overtime. Indeed, production was below 60million per month in 2002 while in 2011 no production per month was less than 64million.

Time Plot Of Crude Oil Exportation Between 2002 And 2011

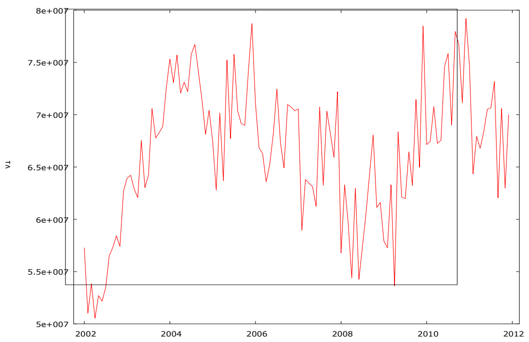

Figure 1: Monthly Time plot of crude oil exportation

3.2 Simple Averages Method For Seasonal Variation

The seasonal index is obtained by

$$Seasonal\ Index = \frac{Monthly\ Average}{Total\ Average} \times 100$$

The seasonal indices from January to December were 100.243, 95.268, 99.05, 93.484, 100.406, 97.816, 102.697, 93.337, 102.075, 104.929, 102.262, 108.435. The minimum was 93.34 while the maximum 104.92 none of which was too far from 100%. The implication is that there were no seasonal variations.

3.2.1 Acf And Pacf Plot (Corellogram) For The Data

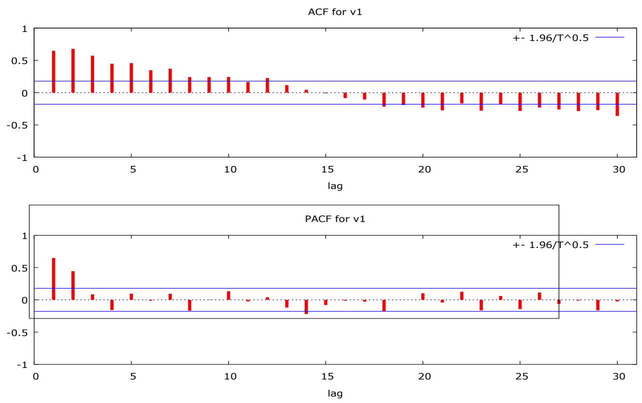

Figure 2: Correllogram Of The Data

The corellogram of figure 2 shows that the data is not stationary and that there is no seasonality in the data. Its also shows that AR and MA also exixts in the time series data

ACF and PACF plot (CORELLOGRAM) is used only for descriptive purpose. Now, the unit root test will be done for inferencial analysis.

3.3 Unit Root Test For The Data

Hypothesis Testing:

H_0: The data is stationary

H_1: The data is not stationary

Below is the Augmented Dickey-Fuller (ADF) test with constant and trend for the series at level.

test statistic: tau = -2.20908

	Level of Significance			
Critical value	10%	5%	2.5%	1%
	-2.64	-2.93	-3.18	-3.46

Table 1: Unit Root Test For The Data

Since the test statistics is greater than all the critical values, we therefore accept the null hypothesis and conclude the time series has unit root, that is our data is not stationary.

The unit root test showed that the data is not stationary, therefore a transformation needs to be done, the transformation can either be log transformation or differencing. In this work, differencing will be carried out and the time plot of the differenced data will be shown in figure 3.

3.4 Time Series Plot After First Differencing

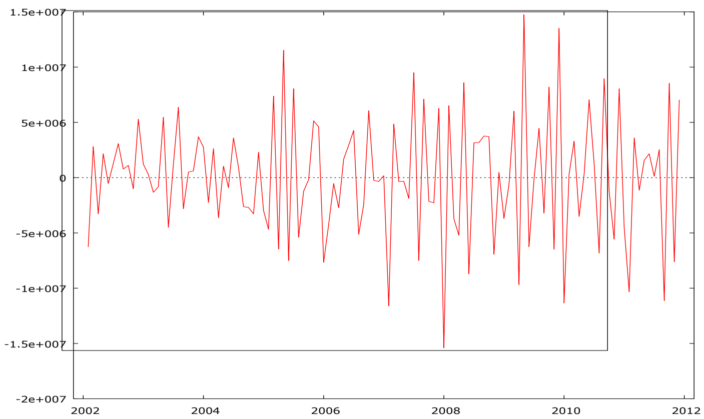

Figure 3: Time Series Plot After First Differencing

The diagram that plots the series after one differencing shows that the variability of the series appears to be stable. The time plot of the series appears to be stationary in both mean and variance suggesting that the time series is stationary.

The ACF, PACF and unit root will be carried out again on the differenced data to ascertain if the data is now stationary.

3.5 ACF AND PACF Plot (Correllogram) At First Diffenrencing

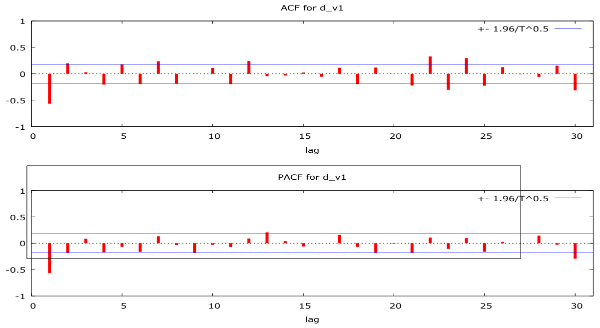

Figure 4: ACF and PACF Plot (Correllogram) At First Diffenrencing

From the correllogram above, its shows that the data is now stationary and that there is no existence of seasonality. Its also shows that the AR and MA exists in the time series data

3.6 Unit Root Test After First Diffenrencing

Hypothesis Testing:

H_0: The data is stationary

H_1: The data is not stationary

Below is the Augmented Dickey-Fuller (ADF) test with constant and trend for the series after first difference.

test statistic: tau = -3.62079

Critical Values			
10%	5%	2.5%	1%
-2.64	-2.93	-3.18	-3.46

Table 2: Unit Root Test After First Differencing

From the unit root test above, the test statisitcs is less than all the critical values which indicates that the series is now stationary after the first differencing and can now be used for analysis purpose.

Since the data has being transformed through differencing, the AR, MA and ARMA models are not efficient for this work, therefore the Autoregressive Integrated Moving Average (ARIMA) model is suitable for this work, AIC and BIC will be used to identify the best model using iterative method.

3.7 Model Identification

At the model identification stage, an iterative method on the number of components for the Moving Average (MA) and Autoregressive (AR) was employed considering the information criteria (i.e. Akaike Information Criterion, Bayesian Information Criterion and Hannan-Quinn Information Criterion).

Information Criteria (IC)

Model	*Information Criteria*		
	AIC	HQC	BIC
0 1 0	4030.989	4032.118	4033.768
1 1 0	3987.925	3991.311	3996.263
0 1 1	3989.354	3992.739	3997.691
1 1 1	3986.528*	3991.042	3997.645
2 1 0	3985.964*	3990.478	3997.667
0 1 2	3988.550	3993.064	3999.667
2 1 1	3984.254*	3989.897	3998.150
1 1 2	3988.658	3994.301	4002.554
2 1 2	3986.365*	3993.136	4003.040

Table 3: Information Criteria

The Asterick in the table 4 indicates models for further study

The models for further study will now be estimated using exact maximum likelihood method.

Model Estimation

This is the section whereby the parameters of the four best models retained were estimated. We estimate the parameters of ARIMA (p,d,q) as the optimal model. Parameters are estimated by exact maximum likelihood method and the order of ARIMA parameters is selected from the lowest Akaike information criteria and is given below.

Model 1: ARIMA (1,1,1)

Parameter	*Coefficient*	*Standard error*	*Z*	*p-value*
Const	110179	208528	0.5284	0.5972
Phi_1	-0.390372	0.149369	-2.613	0.0090
Theta_1	-0.284676	0.160488	-1.774	0.0761

Log-likelihood = -1989.264

Model 2: ARIMA (2,1,0)

Parameter	*Coefficient*	*Standard error*	*Z*	*p-value*
Const	109009	217477	0.5012	0.6162
Phi_1	-0.678974	0.0902506	-7.523	5.34e-014
Phil_2	-.0.182272	0.0907441	-2.009	0.0446

Log-likelihood = -1988.982

Model 3: ARIMA (2,1,1)

Parameter	*Coefficient*	*Standard error*	*Z*	*p-value*
Const	110566	233574	0.4734	0.6360
Phi_1	-1.35603	0.135076	-10.4	1.03e-023
Phi_2	-0.574243	0.0873556	-6.574	4.91e-011
Theta_1	0.727185	0.140660	5.170	2.34e-07

Log-likelihood = -1987.127

Model 4: ARIMA (2,1,2)

Parameter	*Coefficient*	*Standard error*	*Z*	*p-value*
Const	31801.3	230105	0.1382	0.8901
Phil 1	-1.35394	0.152650	-8.870	7.34e-019
Phi_2	-0.570868	0.136538	-4.181	2.90e-05
Theta_1	0.724038	0.188307	3.845	0.0001
Theta 2	-0.00483247	0.168083	-0.02875	0.9771

Log-likelihood = -1987.182

At the stage of model estimation, ARIMA model (2,1,1) and (2,1,2), (2,1,0) were dropped due to high log-likelihood values and ARIMA (1,1,1) was chosen because it lowest log-likelihood value and its easier to use.

Before the interpremation and use of the fitted model, a check will be done to ensure the model selected is suitable for use.

Diagnostic Check

Before the interpretation and use of the fitted model, we are to look at some tests to check whether the model is specified correctly. The following tests are applied to the residuals: (i) Test for autocorrelation and partial autocorrelation, (ii) ARCH test for residual, (iii) Jarque – Bera test for non-normality.

For ARIMA 1,1,1

TESTS	*TEST STATISTIC*	*P-VALUE*
JACKQUE-BERA	1.478	0.4777
ARCH TEST	1.4674	0.48013

Table 4: Diagnostic Check

Since the p-value is greater than 0.05, we therefore conclude that the residual error is normally distributed. The graphical representation will be shown below.

ARIMA (1,1,1)

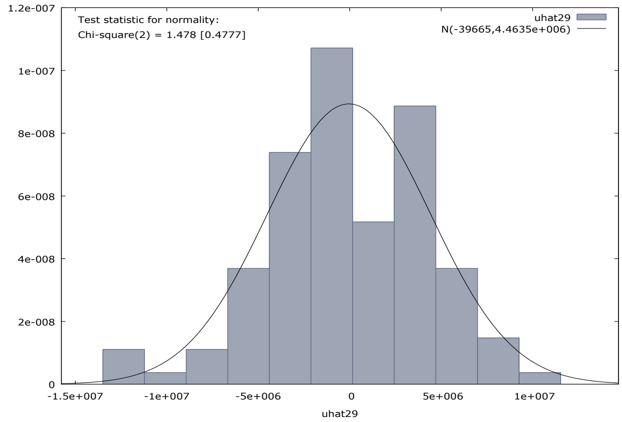

Figure 5: Graphical Representation For Normality

Test for null hypothesis of normal distribution:

Chi-square(2) = 1.478 with p-value 0.47768

Since the p-value is greater than 0.05, we therefore conclude that the residual error is normally distributed

ARIMA (2,1,0)

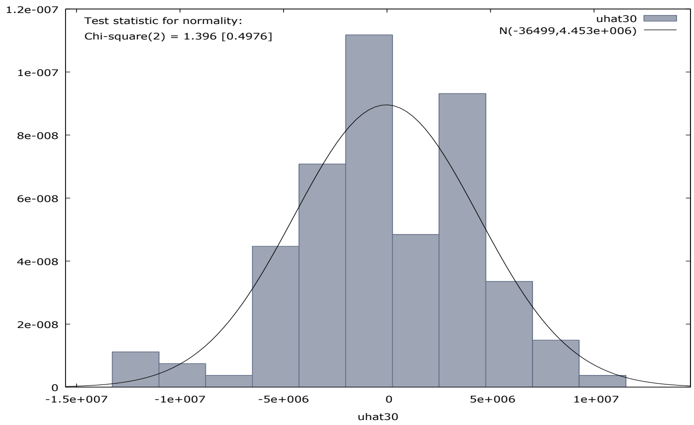

Figure 6: Graphical Representation For Normality

Test for null hypothesis of normal distribution:

Chi-square(2) = 1.396 with p-value 0.49764

Since the p-value is greater than 0.05, we therefore conclude that the residual error is normally distributed.

ARIMA (2,1,0)

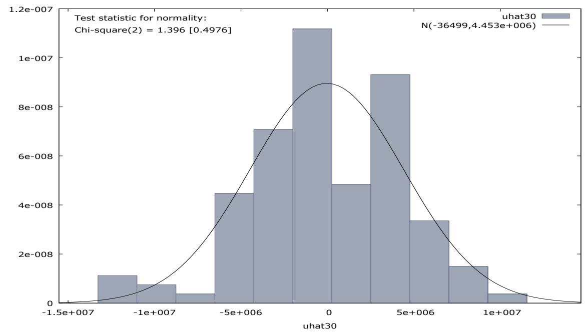

Figure 7: Graphical Representation For Normality

Test for null hypothesis of normal distribution:

Chi-square(2) = 1.396 with p-value 0.49764

Since the p-value is greater than 0.05, we therefore conclude that the residual error is normally distributed.

ARIMA (2,1,1)

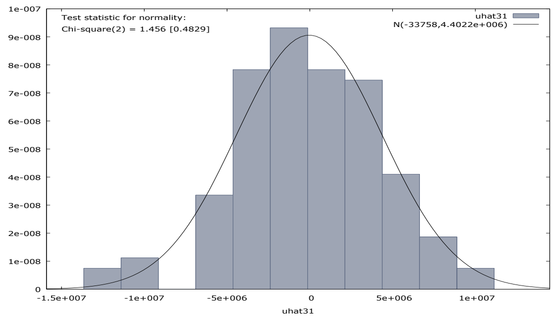

Figure 8: Graphical Representation For Normality Figure

Test for null hypothesis of normal distribution:

Chi-square(2) = 1.456 with p-value 0.48285

Since the p-value is greater than 0.05, we therefore conclude that the residual error is normally distributed

Forecast

After the good ARIMA (p,d,q) model has been derived, the next is to see the ability to forecast. The ability to do so will further testify the validity of the model. What we intend to do is to compare the forecast and the actual values of the three models and choose the one with the minimum error as our best and optimal model. The smaller the value of the error, the better the forecasting performance of the model. Below are the forecasted graphs and forecasted values of the model for a period of 12month.

Graphical Representation For Forecast ARIMA (1,1,1)

Figure 9: Graphical Representation For Forecast ARIMA (1,1,1)

For 95% confidence intervals, z(0.025) = 1.96

Observations	*Prediction*	*std. Error*	*95% interval*
2012:01	66634353.00	4393472.474	58023305.19 - 75245400.82
2012:02	68107150.33	4619614.305	59052872.67 - 77161427.99
2012:03	67685401.90	5293893.261	57309561.77 - 78061242.03
2012:04	68003230.49	5707628.355	56816484.48 - 79189976.50
2012:05	68032349.15	6157328.315	55964207.41 - 80100490.89
2012:06	68174172.00	6552474.398	55331558.17 - 81016785.83
2012:07	68271998.32	6933857.469	54681887.40 - 81862109.23
2012:08	68386999.64	7292073.869	54094797.48 - 82679201.79
2012:09	68495296.32	7634711.415	53531536.92 - 83459055.73
2012:10	68606210.31	7962164.509	53000654.63 - 84211765.98
2012:11	68716102.57	8276843.084	52493788.22 - 84938416.92
2012:12	68826393.69	8579924.084	52010051.49 - 85642735.88

Table 5: Forecast for ARIMA (1,1,1)

To see the efficiancy of the model selected, the ARIMA model (1,1,1) will be used to forecast for the year 2011 and compared with the actual value in the data collected.

Using the ARIMA 1,1,1 Model To Predict for The Year 2011

MONTH	*ACTUAL VALUE*	*PREDICTED VALUE*
JANUARY	74,668,111	75,015,009.58
FEBUARY	64,340,771	76,692,478.44
MARCH	67,931,652	72,041,699.65
APRIL	66,794,717	67,853,096.54
MAY	68,363,645	67,693,029.68
JUNE	70,527,957	67,713,461.63
JULY	70,650,154	69,035,040.87
AUGUST	73,173,567	70,295,857.38
SEPTEMBER	62,053,661	71,522,472.41
OCTOBER	70,586,656	69,243,293.00
NOVEMBER	62,976,593	67,026,382.75
DECEMBER	70,014,740	67,253,414.99

The minimum production in 2011 for actual value is 62.05 million barrel while for the predicted is 67.03 million barrels and the maximum production for actual value is 74.67 million barrels and the predicted is 76.69 miilion barrels, this shows a wide difference in the actual value and predicted value suggesting that the model selected can only be used for short-term supposes.Though the predicted values has a wide difference from the actual values , ARIMA (1,1,1) will still be chosen because of its simplicity to use and with the lowest standard error.

4. Discussion and Conclusion

4.1 Discussion of Results

Data on the exportation of crude oil in Nigeria for 10 years was obtained from the Nigerian National Petroleum Corporation bulletin (NNPC ASB 2011- 1st Edition). From the analyzed data on exportation of crude oil in Nigeria, it was observed in the time plot was non-stationary one due to the presence of an upward trend. Correlogram of the autocorrelation against the l-lag confirmed that there is presence of Trend (non stationary series). There was no seasonality in the data and no clear evidence of cyclic variations. The unit root test was done to determine non-stationary and was not be stationary, therefore a transformation was done . The time plot after one differencing showed that the time plot of the series appears to be stationary in both mean and variance suggesting that the time series is now stationary, the correlogram of the autocorrelation against the lags was done and the unit root was done again and it revealed that the data is now stationary.

Because stationarity was achieved by differencing, any of the models AR, MA and ARMA could not work. The ARIMA was done using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), four best models were retained from 9 iterations and the estimation of the model was carried out. These models are ARIMA (1,1,1), (2,1,0), (2,1,1) and (2,1,2). The simplest among these was ARIMA (1,1,1).

4.2 Conclusion

Based on the analysis carried out and findings made, the result showed there was a general increase in production over time (production was mostly below 60million 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. There was no clear evidence of cyclic variation in the data. All of these are just for descriptive purpose.

The data gave an ARIMA model suggesting that the data set was not stationary at first and a transformation was made through differencing, therefore making the AR, MA and ARMA model not suitable for this data . 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 and its showed that by January the production was 75.02 million barrel and by december production was about 67.25 million barrel, though its a wide difference from the actual value but was still chosen because of its simplicity to use and that its also has the lowest standard error after the forecast.

ARIMA (1,1,1) was used to predict for short-term purpose, therefore the year 2012 was predicted and it revealed that the minimum production for 2012 should be around 66.63 million barrel and the maximum production should be around 68.83 million barrel.

References

[1] Akaike, H. (1973), Information Theory and an Extension of the Maximum Likelihood Principle. In B.N. Petrov and F. Csaki (eds.), 2nd International Theory, PP.267-281. Akademia Kiado, Budapest.

[2] Armstrong, J.S. (1978), “Forecasting with Econometric method: Folklore versus Fact with Discussion”, Journal of Business, 51, 549-600.

[3] Box George and Jenkins, Gwilyn (1970), Times Series Analysis, Forecasting and Control, Holden Day, San Francisco.

[4] Box George & Jenkins, Gwilyn (1976), Time Series Analysis: Forecasting and Control. San Francisco: Holden Day.

[5] Brockwell P. J and Davis, R.A. (1991), Time Series: Theory and Methods, 2nd edition. Springer-Verlag, New York.

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

[7] Chatfield (1987), The Analysis of Time Series: An Introduction. London Chapman and Hall (Third Edition).

[8] Cooper, R.L. (1972), “The Predictive Performance of Quarterly Econometric Models of the United States”, In Econometric Models of Cyclical Behavior (B.G. Hickman, Ed) New York: National Bureau of Economic Research.

[9] Elliot, J.W. (1973), “A Direct Comparison of Short-Run GNP Forecasting Models”, Journal of Business, 46, 33-60.

[10] Frankowne and Ronjones (1982), “Statistics” 2nd Edition, Pitman, Chapter 14 “Sample Design” and Chapter 15, “Planning a Sample Survey”.

[11] Harris, Richard and Robert, Sollis (2003), Applied Time Series Modeling and Forecasting, John Wiley and Sons, Chichester.

[12] Harvey A.C.and R.D. Synder (1990), Structural Time Series Models in Inventory Control, International Journal of Forecasting, 6, 187-198.

[13] Kolmogorov, A.N. (1941), Stationary Sequences in Hilbert space (in Russian). Bull. Math. Univ. Moscow,

2(6), 1-40.
[14] McCleary, R. and Hay, R.A., Jr (1980). Applied Time Series Analysis for the Social Sciences, Beverly Hills, CA: Sage.
[15] McWhhorter, A. Jr. (1975), "Time Series Forecasting using the Kalman Filter: An Empirical Study", Proceeding of American Statistical Association. Business and Economic Section, pp. 436-446.
[16] Narasimham et al (1974), "On the Predictive Performance of the BEA Quarterly Econometric Model and Box-Jenkins Type ARIMA Model". Proceedings of the American Statistical Association: Business and Economics section, pp. 501-504.
[17] Nelson, C.R. (1972), "The Prediction Performance of FRB-MIT-PENN Model of the US Economy", American Economic Review, 5, 902-917.
[18] Newbold, P. (1983), ARIMA Model building and Time Series Analysis approach to Forecasting, Journal of Forecasting, 2, 23-35.
[19] Slutsky, E., (1937), "The Summation of Random Causes as the Source of Cyclic Processes". Econometrica 5, 105-146.
[20] Spiegel M.R, Stephens L.J (1999), Theory and Problems of Statistics (Third edition).
[21] Wold, H. (1938), A study in the Analysis of Stationary Time Series. Almgrist and Wiksell, Stockhom.
[22] Yule, G.U. (1926) "Why Do We Sometimes Get Nonsense-Correlation between Time Series? A study In sampling and the Nature of Time Series, Journal of Royal Statistical Society, 89, 1-64.
[23] Yule, G.U. (1927), On a Method of Investigating Periodicities in Distributed series with Special Reference to Wolfer's Sun spot Numbers. Philosophical Transition Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character, 226: 267-298.

Author Profile

Author Name: Balogun Oluwafemi Samson received the B.Sc. and M.Sc. degrees in Statistics from University of Ilorin in 2007 and 2010, respectively. He currently running his Ph.D. programme in Statistics in the same institution and he is now with Department of Statistics and Operations Research, Modibbo Adama University of Technology, Yola as an Academic Staff.