TIME SERIES ANALYSIS OF BRENT CRUDE OIL PRICES PER BARREL: A BOX-JENKINS APPROACH

Adebowale Olusola Adejumo [1], James Daniel [2]

[1] **University of Ilorin, Ilorin, Nigeria**
[2] **National Bureau of Statistics, Abuja, Nigeria**

Corresponding Author: Adebowale Olusola Adejumo, aodejumo@unilorin.edu.ng

ABSTRACT: This study was aimed at analyzing the Brent Crude Oil Price Per Barrel. The data was extracted from the Organisation of Petroleum Exporting Countries (OPEC) Bulletin and it covered the period of January 1985 to September 2016. The Box and Jenkins approach of model identification, parameter estimation and diagnostic checking was adopted in the analysis with the aid of S-plus Package. In view of the above problem, this research was set to: explore and explain the behavior of the series; determine the best model; and forecast future values of the series. With the aid of S-Plus programming ware, R language ware; ARIMA (2,1,2) (2,0,0) [12] was reached as an optimum and parsimonious model for the series. An arithmetic increase in the price was forecast.
KEYWORDS: Autocorrelation, Partial Autocorrelation, Autoregressive Integrated Moving Average (ARIMA), Autoregressive Moving Average (ARMA), Autoregressive (AR), Moving Average (MA), Differencing, Brent Crude Oil, Stationary Series.

1. INTRODUCTION

Oil is one of the most important commodities in the world. The fluctuation of crude oil price affects global economy, and also affects our daily lives. The oil market is quite complex. Crude oil is a naturally occurring, yellow-to-black liquid found in geologic formations beneath the Earth's surface. It is a fossil fuel which is commonly refined into various types of fuels. Crude oil is distinguishing from petroleum that includes both naturally occurring unprocessed crude oil and petroleum products. Crude oil is a mixture of a very large number of different hydrocarbons including alkanes (paraffins), cycloalkanes (naphthenes), aromatic hydrocarbons, or more complicated chemicals like asphaltenes and sulfur. Each crude oil variety has a unique mix of hydrocarbons, which define its physical and chemical properties, like colour and viscosity.

Figure 1: Barrels of Crude Oil

Figure 1 shows barrels of crude oil. The oil is one of the most important sources used in our daily lives. We often think that petroleum is mostly used to power internal combustion engines in the form of gasoline or petrol. In fact, petroleum is not only used for transportation and producing electricity, but also used for producing clothes, plastics, beauty products and so on. Apart from the common petroleum products such as LPG, Gasoil, Gasoline, Naphtha, Bitumen, Fuel Oil, Plastics, etc. There are lots of incredible petroleum products such as bicycle tires, fishing lures, perfumes, food preservatives, dentures, lipstick, vitamin capsules, petroleum jelly and so on.
The oil price or the price of oil, generally refers to the spot price of a barrel of benchmark crude oil. The major benchmark oil prices in the world contain Brent crude oil price, WTI (West Texas Intermediate) crude oil price and Dubai/Oman crude oil price. The different types of oil are with different density and sulfur content, that leads to the oil price difference. Crude oil prices are commonly measured in USD per barrel. The price of oil is affected by global economic conditions and supply and demand as well as market speculation. The International Energy Agency reported that high oil prices generally have a large negative impact on global economic growth. In the United States and Canada, the oil barrel (abbreviated as bbl) is a volume unit for crude oil, it is defined as 42 US gallons, which is equal to 159 liters or 35 imperial gallons. However, Outside the above two countries, volumes of oil are usually reported in cubic meters (m^3) instead of oil barrels [A+17].

The aim of this research work is to find model that best fit time series of Brent crude oil price per barrel. The specific objectives are to:

i. explore and explain the behavior of the series;
ii. determine the best model;
iii. forecast future values of the series; and
iv. cross-validate suggested model in (ii) through traditional methods.

2. LITERATURE REVIEW

[DW51] tested whether or not residuals are autocorrelated. The differences between successive residuals were used to calculate the test statistics through this equation:

$$DW = \frac{(\sum_{t-1}^{n} \varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t-1}^{n} \varepsilon_t^2} \qquad (1)$$

It was discovered that they were correlated. Thus, they concluded that there was presence of serial correlation.

[DM06] proposed a weighted bootstrap method, also known in the literature as the wild bootstrap, which results in consistent variance of test statistics even in the presence of heteroscedacity. In this procedure, each observation of the original series is weighted, resampled with reposition from a standard normal distribution. Neumann and [KP03] tested the validity of this method, in the contest of time series.

[HJ96] investigated the accuracy of the subsampling technique for estimating one-sided and asymmetrical distribution functions of a studentized, asymptotically normal statistics. Hall and [HJ96] found that when l is chosen optimally, the rate of convergence of the root mean square error are $n^{-\frac{1}{4}}$ respectively, for one-sided and symmetrical distribution functions.

[Kob14] applied a simple neural-network method of seasonal adjustment and data filtering to estimate broiler cycles. He employed the Bureau of census method for seasonal adjustment and found what he called '4 well defined cycles'. For broiler chick placements a cycle of approximately 27 months is needed and for broiler prices a cycle of approximately 30 months was suggested for peaks in broiler prices by about 21 months at these cycles.

[H+06] in their book titled "Time Series Analysis, Forecasting and Control" emphasized that for an autoregressive process of order ρ, the Partial Autocorrelation Function (PAF) ϕ_{kk} will be non-zero for k less than or equals to ρ and zero for k greater than ρ. In another words, the Partial Autocorrelation Function (PAF) of ρ^{th} order autoregressive process has a cut-off after lag k. Generally, the autoregressive model is given as:

$$X_t = \emptyset_{11}X_{t-1} + \emptyset_{22}X_{t-2} + \emptyset_{33}X_{t-3} + \cdots + \emptyset_{pp}X_{t-p} + \varepsilon_t \qquad (2)$$

The complexity of such a long model was given relieve by their work which provided a cut off at a point where an autocorrelation coefficient breaks out of their confidence interval. The Box and Jenkins ARIMA techniques are based on the idea that a time series in which successive values are highly dependent can be regarded as being generated from series of independent shocks.

3. METHODOLOGY

3.1 Introduction to Time Series Analysis

A time series is a stochastic process in discrete time with a continuous state space.

Notation: $\{X_1, X_2, \ldots, X_n\}$ denotes a time series process, whereas $\{x_1, x_2, \ldots, x_n\}$ denotes a univariate time series, i.e. a sequence of realizations of the time series process. $X_1, X_2, \ldots, X_n, X_{n+1}$.

3.2 Autoregressive (AR) Model

An Autoregressive (AR) model of order ρ, or an AR(ρ) model, satisfied the equation:

$$X_t = \mu + \sum_{j=1}^{p} \emptyset_{jj} X_{t-j} + \varepsilon_t$$

$$X_t = \mu + \emptyset_{11}X_{t-1} + \emptyset_{22}X_{t-2} + \emptyset_{33}X_{t-3} + \cdots + \emptyset_{pp}X_{t-p} + \varepsilon_t$$

$$\text{For} \begin{cases} j: (1)p; \\ t \geq 0; \\ \sum_{k=1}^{p} \varepsilon_t = 0. \end{cases} \qquad (3)$$

Where AR is a constant. The p denotes the order of autoregressive model, defining how many previous values the current value is related to. The model is called autoregressive because the series is regressed on to past values of itself. The error term ε_t in equation 3 refers to the noise in the time series. The error is said to be independently and identically distributed (iid). Commonly, they are also assumed to have a normal distribution. $\varepsilon_t \sim \mathrm{N}(\mu, \sigma^2)$. For the model in equation 3 to be of use in practice, the estimator must be able to estimate the value of ϕ_{kk} and μ. Note the subscripts are defined so that the first value of the series to appear on the left of the equation is always one.

3.3 Moving Average (MA) Models

The moving Average (MA) model of order p or MA(q) model is of form:

$$X_t = \mu + \sum_{k=1}^{q} \theta_{kk}\, \varepsilon_{t-k} + \varepsilon_t$$

$$X_t = \mu + \theta_{11}\varepsilon_{t-1} + \theta_{22}\varepsilon_{t-2} + \theta_{33}\varepsilon_{t-3} + \cdots + \theta_{pp}\varepsilon_{t-p} + \varepsilon_t$$

$$\text{For} \begin{cases} k{:}\,(1)p; \\ t \geq 0; \\ \sum_{k=1}^{q} \varepsilon_t = 0. \end{cases} \quad (4)$$

MA models imply the time series signal can be expressed as a linear function of previous value on the time series. The error (or noise) term in the equation 4. is the one step and ahead forecasting error. In contrast, MA(q) models imply the signal can be expressed as function of previous forecasting errors. It suggests MA(q) models make forecast based on the error made in the past, and so one can learn from the error made in the past to improve current forecast.

3.4 Autoregressive Moving Average (ARMA) Models

The best model is the simplest model that captures the important features of the data (parsimonious model).
Sometimes, however, neither a simple AR(p) model nor simple MA(q) model exists. In these case, a combination of AR(q) any MA(q) models will almost produce a simple model. These models are called Autoregressive Moving Average Models, or ARIMA (p, q) models. Once again p is used for number of Autoregressive components, and q is for the number of Moving Average Component.
Definition: The Form ARMA(p, q) model is given by the equation:

$$X_t = \mu + \sum_{j=1}^{p} \emptyset_{jj}\, X_{t-j} + \varepsilon_t = \mu + \sum_{k=1}^{q} \theta_{kk}\varepsilon_{t-k} + \varepsilon_t$$

$$X_t - \sum_{j=1}^{p} \emptyset_{jj}\, X_{t-j} + \varepsilon_t = \mu + \sum_{k=1}^{q} \theta_{kk}\varepsilon_{t-k} + \varepsilon_t$$

$$X_t - \mu + \emptyset_{11}X_{t-1} + \emptyset_{22}X_{t-2} + \emptyset_{33}X_{t-3} + \cdots + \emptyset_{pp}X_{t-p} + \varepsilon_t = \mu + \theta_{11}\varepsilon_{t-1} + \theta_{22}\varepsilon_{t-2} + \cdots + \theta_{qq}\varepsilon_{t-q} + \varepsilon_t$$

$$\text{For} \begin{cases} j{:}\,(1)p; \\ k{:}\,1(p); \\ t \geq 0; \\ \sum_{k=1}^{p} \varepsilon_t = 0; \\ \sum_{k=1}^{q} \varepsilon_t = 0. \end{cases} \quad (5)$$

Where X_t is the observed data point. μ is some constants and ϕ_{jj}, θ_{kk} are defined as for AR and MA model respectively.

3.5 Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA model, sometimes called the Box-Jenkins models; named after the authors of the iterative Box-Jenkins methodology typically applied to time series data for forecasting consists of three parts: An Autoregressive (AR) part, a Moving Average (MA) part and the difference part. The model is usually then referred to as the ARIMA models where p is the order of the Autoregressive part, d is the order of difference and q is the order of the moving Average part. For example, a model is referred to have ARIMA(1,1,1) when it has only one Autoregressive parameter called order one and only one moving Average parameter also called order one for the time series data after it was differenced once to attain stationary.
If the model becomes ARMA, which is linearly stationary on its self without being differenced. ARIMA is a linear non-stationary model. If the underlying time series is non-stationary, taking the difference of the series with itself some d-times makes it stationary, and then ARIMA model is applied onto the difference series.

3.6 Differencing and Unit Root

In the full class of ARIMA(p,d,q) models, the 'I' stand for integrated. The idea is that one might have a model whose terms are the partial sum up to time t, of some ARIMA models.
Thus X consist of accumulated past shock, that is, shocks to the system, have a permanent effect. Note also that the variance of X_t increase without bound as time passes. A series in which the variance and mean are constant and the covariance between X_t and X_s is a function only of the time difference (t-s) is called a stationary series. Clearly these integrated series are non-stationary.
Another way to approach this issue is through the model.
Again let us consider an autoregressive order 1 model, AR(1):

$X_t - \mu = \rho(X_{t-1} + \mu) + \varepsilon_t$

$X_{t-1} - \mu = \rho(X_{t-1} + \mu) + \varepsilon_{t-1}$ if $\begin{cases} |\rho| < 1 \\ t = (t-1) \end{cases}$

$X_t - \mu = \rho\varepsilon_{t-} + \rho^2(X_{t-1} - \mu) + \varepsilon_t$

$$X_t - \mu = \sum_{t=1}^{n}\rho^i + \varepsilon_{t-1} + \varepsilon_t \qquad (6)$$

Which is convergent expression satisfying the stationary conditions. However, if $|\rho| = 1$ the infinite sum does not converge so one require a starting value say $X(0) = \mu$ in which case $X_t = \mu + \sum_{t=1}^{n}\varepsilon_{t-1} + \varepsilon_t$ is the initial value plus an unweighted sum of or shocks they are permanent and the variance of X grows without bound over time. As a result, there is no tendency of the series to return to any particular value and the use of a symbol for the starting value is perhaps a poor choice of symbols in that this is does not really represent a mean of any sort.

This type of series is called a random walk. Many stock series are believed to follow this or some closely related model. The failure of forecasts to return to mean in such a model implies that the best forecast of the future is current value and hence the strategy of buying low and selling high is pretty much eliminated in such a model whether high or low is unknown. Any Autoregressive model like $X_t = \phi_{11}X_{t-1} + X_{t-1}$ has associated with a 'characteristic equation' whose root determine the stationary or non-stationary of the series.

3.7 Autocorrelation

The correlation between a variable, lagged one or more periods, and itself is called **authocorrelation**. While the graphical tool that displays the correlations for various lags of time series is called **correlogram**.

3.8 Hypothesis Testing for Trend in the Series

Actually, the autocorrelation coefficients for all time lags can be tested simultaneously. If the series is truly random, most of the series autocorrelation coefficients should lie within the specified by 0 plus or minus a certain number of standard errors, arranged as the bellow expression:

$$0 \pm \left(\frac{1}{\sqrt{t}}\right) Z_{\frac{\alpha}{2}}$$

Where z is the standard normal value for a given confidence level and n is the number of of observation in the series.

3.9 Steps in Hypothesis Testing

In order to test for the presence of stationarity, randomness or seasonality, the following steps are taken:

i. Test Statistic
ii. Hypothesis setting
iii. Decision Rule
iv. Decision
v. Conclusion

3.10 Partial Autocorrelation Function (PACF)

In the previous section we have seen how the ACF can be used to identify MA processes, clearly indicating the order q of the process by the number of non-zero terms in the ACF. It would be great if we could similarly identify the order of an AR process. There is another correlation function which allows us to do precisely that: The Partial Autocorrelation Function or PACF.

The PACF computes the correlation between two variables Y_t and Y_{t-k} after removing the effect of all intervening variables $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k}$. We can think of the PACF as a conditional correlation.

Another tool that will be needed is the partial autocorrelation function denoted by: ϕ_{kk} where $k = 1,2,\ldots,k$ is the set of partial autocorrelation at different or various lag k and it is defined by $\phi_{kk} = \frac{\rho^*{}_k}{\rho_k}$, and k by k autocorrelation matrix and

$[\rho_k{}^*]$ with the last column represented by $\begin{bmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_k \end{bmatrix}$.

$[\rho_k]$ is a general form of Yule Walker's equation written as below

$$[\rho_k] = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & . & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & . & . & . & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & . & . & . & \rho_{n-3} \\ . & . & . & . & & & \\ . & . & . & & . & & \\ . & . & . & & & . & \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & . & . & . & 1 \end{bmatrix}$$

Through which the Yule Walker's equation is formed as shown in the below subsection.

3.11 Parameter Estimation for Autoregressive Process

How does PACF behave? Well, by convention we set

$$\phi_{11} - \rho_1 \tag{7}$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \tag{8}$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \tag{9}$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2^2 - \rho_1^2\rho_3 - 2\rho_1\rho_2 - \rho_3}{2\rho_1^2\rho_2 - 2\rho_2^2 - \rho_1^2 + 1}$$

Etc.

3.12 Parameter Estimation for Moving Average Process

$$\phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4} \tag{10}$$

$$\phi_{kk} = \frac{-\theta^k\left(1 - \theta^2\right)}{1 - \theta^{2(k+1)}} \tag{11}$$

3.12 Measure of Forecasting Accuracy

The forecast error is the difference between the actual value and the forecast value for the corresponding period. $\varepsilon_t = A_t - F_t$ Where ε_t is the forecasting error at period t, A_t is the Actual Value at period t and F_t is the forecast for period t . These are as shown in Table 1 below.

Table 1: Formulas for Measuring Accuracy

Error measurement	Formula
Mean Absolute Error (MAE)	$MAE = \frac{1}{T}\sum_{t=1}^{T}\lambda_t$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\lambda_t}{A_t}\right)$
Mean Squared Error (MES)	$MSE = \frac{1}{T}\sum_{t=1}^{T}\lambda_t^2$

3.13 Model Building Strategy

The Figure 2 below is the flow chart that depicts steps taken in the series model building.

4. RESULTS AND DISCUSSION

4.0 Data Presentation

381 univariate data set was collected every first trading day of the month from January 1985 to September 2016 chronologically; which I adopted from the Organization of Petroleum Exporting Countries (OPEC) Statistical Bulletin shown in Table 2 below.

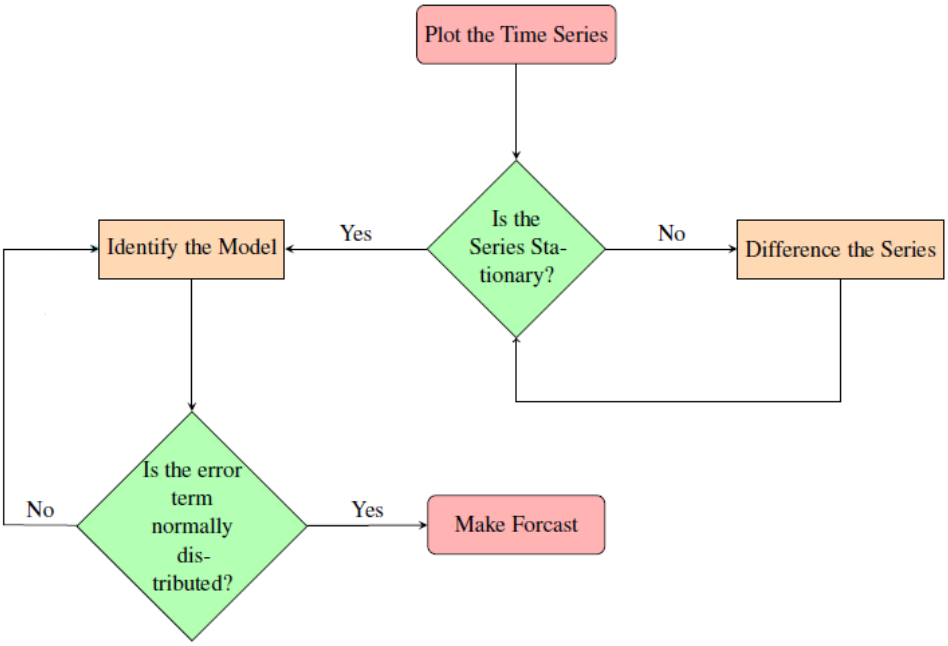

Figure 2: ARIMA Model Building Flowchart

Table 2: Brent Crude Oil Prices Per Barrel ($)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1985	26.6	27.6	27.9	27.8	26.7	26.4	26.6	27.1	27.5	28.3	29.2	26.8
1986	22.3	16.4	12.8	11.9	13.5	12.0	9.9	13.4	14.0	14.0	14.5	15.4
1987	18.1	17.3	17.7	18.1	18.4	18.7	19.6	18.9	18.3	18.6	17.9	16.8
1988	16.5	15.9	14.9	16.4	16.4	15.5	14.5	14.6	13.2	12.2	12.5	14.7
1989	16.5	16.5	18.1	19.4	18.2	17.8	17.7	16.9	17.7	18.4	18.4	19.4
1990	20.6	19.7	18.1	16.3	16.2	14.9	16.8	26.5	33.6	34.9	31.5	26.6
1991	22.8	18.5	18.2	18.5	18.7	17.8	19.0	19.3	20.0	21.6	20.4	17.6
1992	17.5	17.7	17.4	18.7	19.5	20.9	20.2	19.6	20.2	20.0	18.9	17.9
1993	17.2	18.2	18.5	18.4	18.2	17.4	16.4	16.4	15.8	16.4	15.1	13.4
1994	14.2	13.8	13.7	15.2	16.4	17.2	18.0	17.0	16.1	16.5	17.2	16.1
1995	16.9	17.4	17.4	18.8	18.4	17.3	16.1	16.5	16.8	16.2	16.8	17.9
1996	17.8	17.7	19.5	20.8	19.1	18.6	19.6	20.2	22.1	23.4	22.3	23.5
1997	23.3	20.5	19.4	18.0	19.5	18.0	18.5	18.8	18.7	20.1	19.2	17.2
1998	15.1	14.2	13.2	13.4	14.0	12.5	12.7	12.5	13.8	13.3	11.9	10.4
1999	11.3	10.8	12.9	15.7	16.1	16.2	18.8	20.2	22.4	22.2	24.2	25.0
2000	25.2	27.2	27.5	23.5	27.2	29.6	28.2	29.4	32.1	31.4	32.3	25.3
2001	26.0	27.2	25.0	25.7	27.6	27.0	24.8	25.8	25.0	20.7	18.7	18.5
2002	19.2	20.0	23.6	25.4	25.7	24.5	25.8	26.8	28.3	27.5	24.8	27.9
2003	30.8	32.9	30.4	25.5	26.1	27.9	28.6	29.7	26.9	29.0	29.1	30.0
2004	31.4	31.3	33.7	33.7	37.6	35.5	37.9	42.1	41.7	46.9	42.2	39.1
2005	42.9	44.6	50.9	50.6	47.8	53.9	56.4	61.9	61.7	58.2	55.0	56.5
2006	62.4	59.7	60.9	68.0	68.6	68.3	72.5	71.8	62.0	58.0	58.1	61.0
2007	53.4	57.6	60.6	65.1	65.1	68.2	73.7	70.1	76.9	82.2	91.3	89.4
2008	90.8	93.8	101.8	109.1	122.8	131.5	132.6	114.6	99.3	72.7	54.0	41.5
2009	43.9	41.8	47.0	50.3	58.1	69.1	64.7	71.6	68.4	74.1	77.6	74.9
2010	77.1	74.7	79.3	84.1	75.5	74.7	74.5	75.9	76.1	81.7	84.5	90.1
2011	92.7	97.7	108.7	116.3	108.2	105.9	107.9	100.5	100.8	99.9	105.4	104.3
2012	106.9	112.7	117.8	113.8	104.2	90.7	96.8	105.3	106.3	103.4	101.2	101.2
2013	105.0	107.7	102.6	98.9	99.4	99.7	105.2	108.1	108.8	105.5	102.6	105.5
2014	102.3	104.8	104.0	104.9	105.7	108.4	105.2	100.1	95.9	86.1	77.0	60.6
2015	47.5	54.9	53.8	57.4	62.5	61.3	54.4	45.7	46.3	47.0	43.1	36.6
2016	29.9	31.1	37.3	40.8	46.0	47.7	44.2	40.0	46.2			

Source: Organisation of oil Exporting Countries (OPEC)

4.1 Data Exploration

The most obvious reasons for analyzing an important series like this, is to find a way to accurately forecast its future values. However, the analysis process itself sometimes reveals important insight into the series that will help make better decisions.

4.1.1 Histogram

Histogram of Crude Oil Per Barrel is seen in Figure 3.

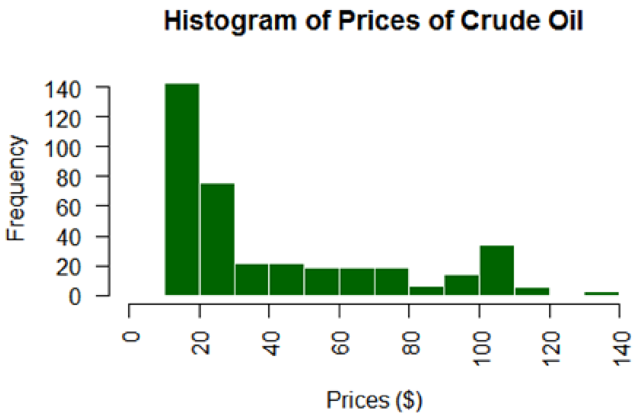

Figure 3: Histogram of Crude Oil Per Barrel

4.1.2 Whisker-Box Plot

The oversight way to check for seasonality in time series is through the box plot as shown in Figure 4.

4.1.3 pareto Diagram

The Pareto theorem of (80/20)% holds for the price regime of $20 interval. Figure 5 shows that 20% {($1 - $21) and ($41 - $41)} of price subgroup experienced 80% of price reign of the crude oil prices.

4.1.4 Time Plot

The first step in any time series analysis is to plot the observations against time.
The time plot was achieved through the use of R software.
The time plot in Figure 6 revealed that the monthly crude oil price is non-stationary. The focal assumption of time series analysis is stationarity of the series out of other assumptions like randomness of the series and normality of the series. To be critical in examining the presence trend (non-stationary series) a hypothesis is set up. The time plot is as shown as in Figure 6.
Time series decomposition diagram is shown in Figure 7.

4.2 Hypothesis Test for First-Order Stationary

To start the process of hypothesis testing for test of stationarity, a 95% confidence interval is computed as follows:

$$0 \pm \left(\frac{1}{\sqrt{t}}\right) Z_{\frac{\alpha}{2}} = \pm 0.1318382182 = \pm 0.13 \begin{cases} t = 380 \\ Z_{\frac{\alpha}{2}} = 2.57 \end{cases} \qquad (12)$$

Next is the computation of autocorrelation coefficients through R programme with the out-put in Table 3.

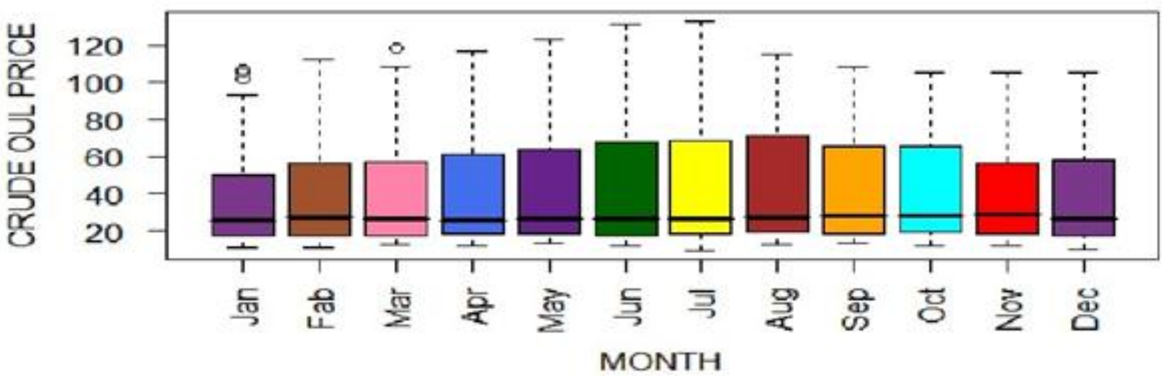

Figure 4: Whisker Box plot of Crude Oil Per Barrel

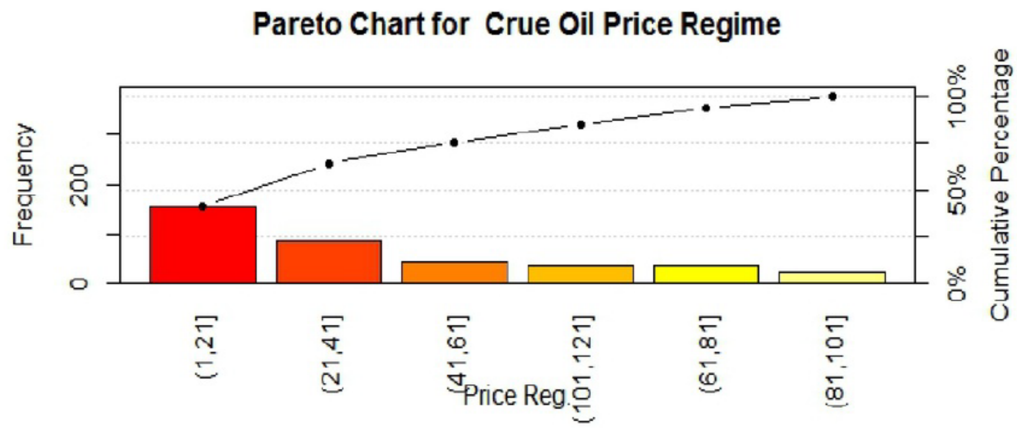

Figure 5: Wisker-Box Plot of Crude Oil Prices Per Barrel ($)

Figure 6: Time plot of Crude Oil Prices Per Barrel

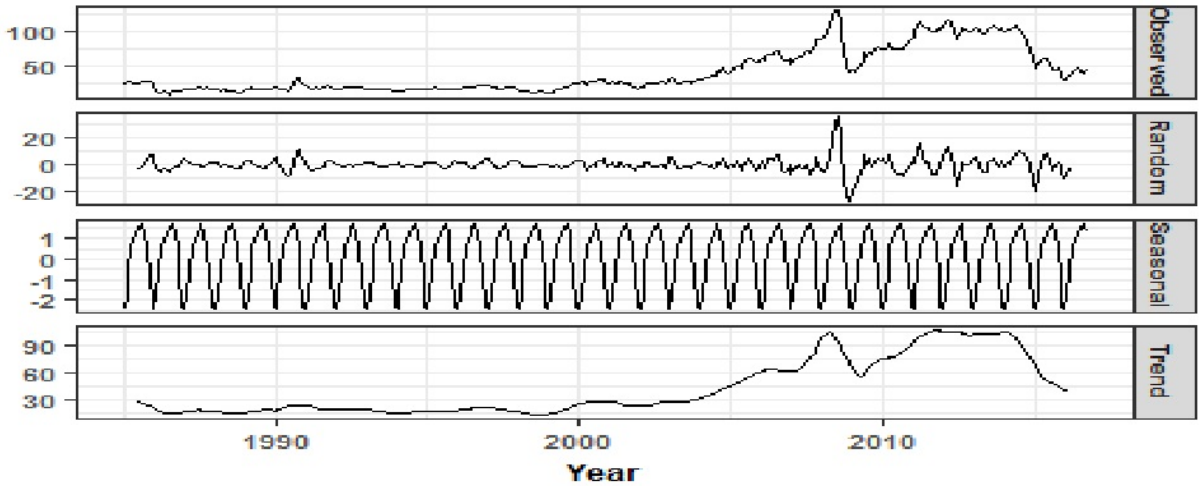

Figure 7: Time Series Decomposition Diagram

Table 3: Autocorrelation Coefficients of First Few Lags of Crude Oil Prices

Lag	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
ACF	1.00	0.99	0.97	0.95	0.93	0.91	0.90	0.88	0.87	0.87	0.86	0.85	0.84

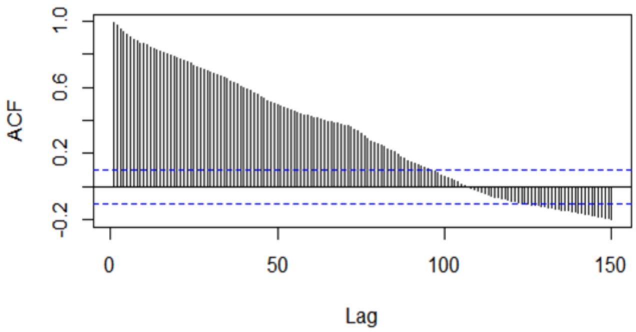

Figure 8: Correlelogram of the Monthly Crude Oil Prices

4.2.1 Hypothesis Setting for First-Order Stationarity

H_0: The series is stationary.
H_0: The series is not stationary.
At 95% confidence level.

4.2.2 Decision Rule for First-Order Stationarity

Reject H_0: The series is stationary, if the first several lag value of autocorrelation coefficients are significant and gradually decreases until zero, accept $H_{0:}$ The series is stationary, otherwise.
At 95% confidence level.

4.2.3 Decision for First-Order Stationarity

I reject H_0: The series is stationary, since the first several lag values of autocorrelation coefficients are significant and gradually decreases until zero.

4.2.4 Conclusion for First-Order Stationarity

At 95% significant level, the data do provide sufficient evidence (as seen in Table 3 and Figure 9) to conclude that the series is not stationary.

4.3 Hypothesis Test for Randomness

4.3.1 Hypothesis Setting for Randomness

H_0: The series is random.
H_0: The series is not random.
At 95% confidence level.

4.3.2 Decision Rule for Randomness

Reject H_0: The series is random, if Autocorrelation Coefficients Lag 1 is close to unity. The series is random, otherwise.
At 95% confidence level.

4.3.3 Decision for Randomness

The series is random, since Autocorrelation Coefficients Lag 1 = 0.9913 and is close to unity.

4.3.4 Conclusion for Randomness

At 95% significant level, the data do provide sufficient evidence (as seen in Table 3 and Figure 9) to conclude that the series is not random.
It is noted that the autocorrelation coefficients of the first several lags are significantly different from zero for instance, autocorrelation coefficient of lag 1, lag 2, lag 3, …, lag 20 = 0.9913, 0.9913, 0.9754, …, 0.7820 respectively which individually less than the ± 0.13 (the confidence bound for autocorrelation) and the autocorrelation coefficients gradually regressed towards zero rather than dropping exponentially. This shows that a trend exists in the series (that is the series is non stationary) or there is presence of serial correlation in the series. The ACF for the series decays very slowly indicating that it is non-stationary. Non-stationary stochastic processes tend to generate series whose estimated autocorrelation function fail to die out rapidly; that is, the estimated autocorrelations for non-stationary processes tend to persist for a large number of lags. Persistently large values of them indicate that the time series is non-stationary, or serially correlated and that transformation of the series through at least one difference is needed. Since the theoretical autocorrelations and partial autocorrelations are only independent of time for stationary processes, it is necessary to difference the original series until it can be assumed to be a realization of a stationary process.
The differencing is achieved through R programme and the output is presented in the Table 4.

Table 4: Display of First Difference of Crude Oil Prices

	0.9	0.4	-0.1	-1.2	-0.3	0.3	0.5	0.4	0.8	0.9	-2.4
-4.4	-6.0	-3.6	-0.9	1.6	-1.5	-2.1	3.5	0.6	-0.0	0.5	1.0
2.7	-0.8	0.5	0.3	0.4	0.3	0.9	-0.7	-0.6	0.3	-0.8	-1.1
-0.3	-0.6	-1.0	1.6	-0.1	-0.7	-1.0	0.1	-1.4	-1.0	0.3	2.2
1.8	0.1	1.6	1.3	-1.3	-0.4	-0.1	-0.8	0.8	0.7	-0.0	1.0
1.2	-0.9	-1.6	-1.8	-0.1	-1.3	1.9	9.7	7.1	1.2	-3.3	-4.9
-3.8	-4.3	-0.3	0.3	0.2	-0.9	1.2	0.3	0.7	1.6	-1.2	-2.8
-0.11	0.1	-0.3	1.3	0.9	1.4	-0.7	-0.6	0.6	-0.2	-1.1	1.0
-0.7	1.0	0.3	-0.1	-0.3	-0.8	-1.0	0.1	-0.6	0.6	-1.4	-1.7
0.8	-0.4	-0.1	1.5	1.3	0.8	0.8	-1.1	-0.9	0.4	0.7	-1.1
0.8	0.6	-0.1	1.4	-0.3	-1.1	-1.3	0.4	0.3	-0.6	0.6	1.1
-0.1	-0.1	1.8	1.3	-1.7	-0.6	1.0	0.6	2.0	1.3	-1.2	1.3
-0.2	-2.8	-1.1	-1.4	1.5	-1.5	0.4	0.3	-0.1	1.4	-1.0	-1.9
-2.2	-0.9	-0.9	025	0.6	-1.5	0.2	-0.2	1.3	-0.5	-1.4	-1.5
0.9	-0.6	2.1	2.9	0.4	0.1	2.5	1.5	2.2	-0.2	2.0	0.8
0.2	2.0	0.3	-4.0	3.8	2.4	-1.5	1.3	2.7	-0.7	0.9	-7.1
0.7	1.3	-2.2	0.7	1.9	-0.6	-2.2	1.0	-0.8	-4.3	-2.0	-0.2
0.6	0.8	3.7	1.8	0.3	-1.2	1.3	1.0	1.5	-0.8	-2.7	3.1
2.9	2.1	-2.5	-4.9	0.6	1.9	0.7	1.1	-2.8	2.1	0.1	0.8
1.5	-0.1	2.4	0.0	3.9	-2.1	2.4	4.2	-0.4	5.2	-4.6	-3.1
3.8	1.7	6.4	-0.3	-2.8	6.1	2.5	5.5	-0.2	-3.5	-3.2	1.5
5.9	-2.7	1.2	7.1	0.6	-0.3	4.2	-0.7	-9.8	-4.0	0.2	2.9
-7.6	4.2	3.0	4.5	0.0	3.1	5.5	-3.5	6.8	5.2	9.1	-1.8
1.4	2.9	8.1	7.2	13.7	8.8	1.0	-18.0	-15.3	-26.6	-18.7	-12.5
2.4	-2.2	5.2	3.3	7.8	11.0	-4.5	7.0	-3.3	5.7	3.5	-2.7
2.3	-2.4	4.6	4.8	-8.6	-0.8	-0.2	1.4	0.2	5.6	2.8	5.5
2.6	5.1	10.9	7.7	-8.1	-2.3	2.0	-7.4	0.4	-0.9	5.4	-1.1
2.6	5.8	5.1	-4.0	-9.6	-13.4	6.0	8.5	1.0	-2.9	-2.2	0.0
3.9	2.6	-5.1	-3.8	0.5	0.4	5.5	2.9	0.7	-3.3	-2.9	2.9
-3.2	2.6	-0.8	0.9	0.8	2.6	-3.2	-5.2	-4.2	-9.8	-9.2	-16.4
-13.1	7.5	-1.1	3.6	5.1	-1.2	-6.9	-8.7	0.6	0.7	-3.8	-6.6
-6.6	1.1	6.3	3.4	5.2	1.7	-3.5	-4.2	6.2			

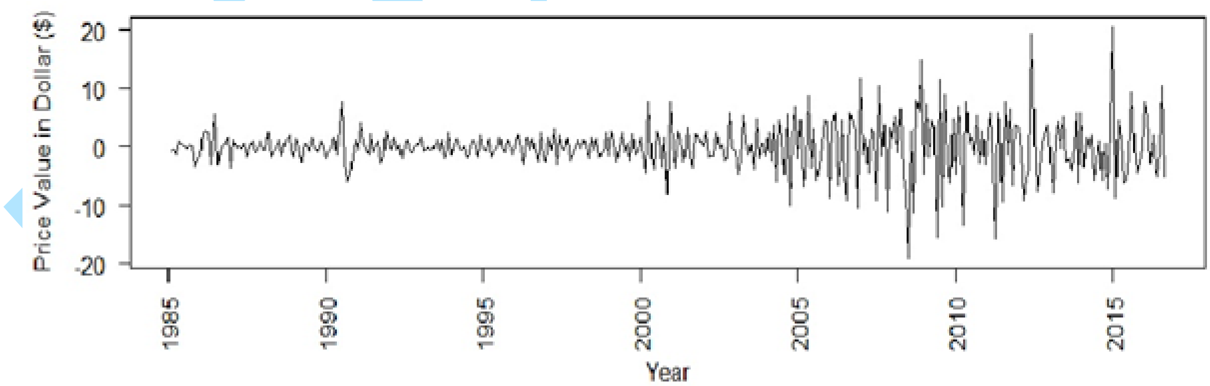

Figure 9: Time Plot of First Difference of the Monthly Crude Oil Prices

Table 5: Autocorrelelogram for the First Difference of the Crude Oil Prices

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
ACF	1.00	0.99	0.97	0.95	0.93	0.91	0.90	0.88	0.87	0.87	0.86	0.85	0.84

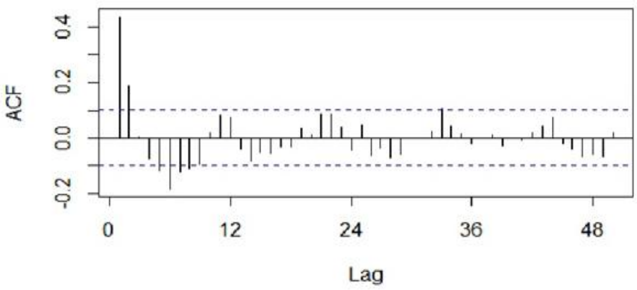

Figure 10: Autocorrelogram of First Difference of the Monthly Crude Oil Prices

4.4 Hypothesis Test for Second-Order Stationary

Having achieved the first differenced series; a test for stationary of the series is necessary and will be conducted again to check if stationarity has been reached. The first step on this is to plot the autocorrelogram of the first differenced series to view how the first differenced time series behaves. The autocorrelogram obtained from Table 5 through S-Plus is presented in Figure 10.

4.4.1 Hypothesis Setting for Second-Order Stationarity

H_0: The Differenced series is stationary.
H_0: The Differenced series is not stationary.
At 95% confidence level.

4.4.2 Decision Rule for Second-Order Stationarity

Reject H_0: The differenced series is stationary, if the first several lag value of autocorrelation coefficients are significant and gradually decreases until zero, accept $H_{0:}$ The series is stationary, otherwise.
At 95% confidence level.

4.4.3 Decision for Second-Order Stationarity

I do not reject H_0: The series is stationary, since the first several lag values of autocorrelation coefficients are not significant and gradually decreases until zero.

4.4.4 Conclusion for Second-Order Stationarity

At 95% significant level, the data do provide sufficient evidence (as seen in Table 5 and Figure 10) to conclude that the series is stationary.
Ordinarily, one would have expected that decision rule for test of seasonality be made on autocorrelation coefficient. The restriction to this is that strength of serial correlation will overpower the autocorrelogram in such a way that spike in it will hardly be noticed. at this stage of the analysis that the trend has been satisfactorily removed, I can now test for seasonality as follows:

4.5 Hypothesis Test for Seasonality

4.5.1 Hypothesis Setting for Seasonality

H_0: The series is seasonal.
H_0: The series is not seasonal.
At 95% confidence level.

4.5.2 Decision Rule for Seasonality

Reject H_0: The series is seasonal, if a spike in Autocorrelation Coefficients after differencing at a regular lag interval. The series is seasonal, otherwise.
At 95% confidence level.

4.5.3 Decision for Seasonality

I do not reject $H_{0:}$ The series is seasonal, since there is are noticeable spikes in Autocorrelation Coefficients 6 lags interval as can be seen in Table 5, and Figure 10.

4.5.4 Conclusion for Seasonality

At 95% significant level, the data do provide sufficient evidence (as seen in Table 5 and Figure 10) to conclude that the series is not seasonal.

Haven achieved detrend, and deseasonalized the series; model identification and parameter estimation is now discussed in the below Subsection 4.6 and 4.8 respectively.

4.6 Model Strategy

The steps demonstrated in Figure 2 are followed appropriately. For easy step-taking in this article, the model strategy involves model identification and parameter estimation as shown in Subsection 4.7 and 4.8 below respectively.

4.7 Model Identification

The model for the monthly crude oil prices is achieved by estimating the autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) of the differenced series. The selection of a time series model is frequently accomplished by matching estimated autocorrelations coefficients with the theoretical autocorrelation. The matching of the first 50 estimated 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 as shown in Table (6a and Table 6b and 6b respectively).

Table 6: ACF and PACF Table of First Differenced Crude Oil Prices

Lag	0	1	2	3	4
ACF	1.00	0.44	0.19	0.01	-0.08
PACF	-0.27	-0.14	-0.16	-0.15	-0.08

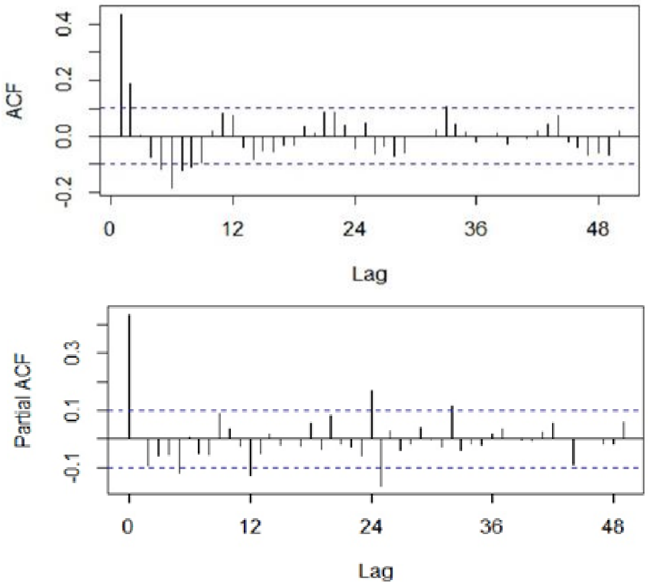

Figure 11: ACF and PACF of First Differenced Crude Oil Prices

With the matching of the ACF with the PACF, the model is identified as an ARIMA (2,1,2) model. Plot of the Autocorrelogram and Partial Autocorrelogram is shown in Figure 11 (11a} and 11b respectively) and their corresponding tables in Table 6 (6a and 6b respectively). Plot of the Autocorrelation Function and Partial Autocorrelation Function greatly assisted in the understanding of the model. From the ACF and PACF, the functions both had a sharp cut off (indicating no additional non-seasonal terms to be included in the model) after lag 2 after the first difference suggest that the model may include in its formation ARIMA (2,1,2). In addition, I had tested the seasonality of the series to be positive in Section 4.5, thus the need to include seasonal component in the series. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA model as demonstrated in the Equation 13 below:

$$\text{ARIMA}\ \underbrace{(p,d,q)}_{\text{Non-Seasonal Part}}\ \underbrace{(P,D,Q)}_{\text{Seasonal Part}}[12] \qquad (13)$$

The "12" in the last bracket specify the periodicity of the seasonal occurrence in month.

4.8 Parameter Estimation

The likely estimates of parameter that will be needed to form appropriate model is summarized in Table 7. The estimates are obtained by R program with the method on Equation 7, 8, 9, and 10.

Table 7: Parameter/Estimate Table

Parameter	ϕ_{11}	ϕ_{22}
Estimate	1.48	-0.59

(a) AR

Parameter	θ_{11}	θ_{22}
Estimate	-1.04	0.49

(b) MA

Parameter	α_{11}	α_{22}
Estimate	0.12	-0.16

(c) SAR

ARIMA(2,1,12) (14)
ARIMA(2,1,2)(1,0,0)[12] (15)
ARIMA(2,1,2)(2,0,0)[12] (16)
and
ARIMA(2,1,2)(3,0,0)[12] (17)

4.9 Diagnostic Check

Equation 14, 15, 16, and 17 will be subjected each through their error terms to different traditional validation methods. Specifically, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mean Absolute Squared Error (MASE), Mean Absolute Error (MAE), Mean Absolute Error (RMAE) Sigma-Square, Log Likelihood method as described in Table 1.

I present a body of conventional method of testing accuracy of time series model in Table forecast-accuracy metric table 8 to measure accuracy of suggested models presented in Equation 14, Equation 15, Equation 16 and Equation 17. Log Likelihood and Mean Error (ME) places a greatest metric value on a best forecast model to be selected, while others like The Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the rest place lowest metric value on best model to be selected.

Obviously, ARIMA (2,1,2)(2,0,0)[12] stands out to be the best model which is in agreement with every accuracy test presented on Table 8.

$$X_t = \emptyset_{11}X_{t-1} + \emptyset_{22}X_{t-2} + \theta_{11}\varepsilon_{t-1} + \theta_{22}\varepsilon_{t-2} + \alpha_{11}X_{t-1} + \alpha_{22}X_{t-2} + \varepsilon_t \qquad (18)$$

Thus, Equation 18 becomes the best, parsimonious, consistent and the most representative model for the series. Also, from the Table of Parameter Estimate in Table 7 the parameters are being fed to Equation 18 which is given as:

$$X_t = 1.48X_{t-1} - 0.59X_{t-2} - 1.04\varepsilon_{t-1} + 0.14\varepsilon_{t-2} + 0.12X_{t-1} - 0.16X_{t-2} + \varepsilon_t \qquad (19)$$

With the Equation 18 and Equation 19 the following are forecast the month of October 2016, November 2016, December 2016 and January 2017 and are presented in Table 9. The forecast is not just on point forecast alone, but also on 80% and 90% interval as can be seen on the same table. To make it eye-friendly, a pictorial diagram of forecast plot is presented in Figure 12.

Figure 12 depicts Table 9 and an In-Series forecast of the series, while Table 10 present the In-Series errors from January 1985 to September 2016.

Table 8: Forecast-Accuracy Metric Table

MODEL	ARIMA (2,1,2)	ARIMA (2,1,2)(1,0,0)[12]	ARIMA (2,1,2)(2,0,0)[12]	ARIMA (2,1,2)(3,0,0)[12]
AIC	2086.870	2085.060	2069.690	2080.190
Sigma^ 2	13.810	13.710	13.080	13.430
Log Likelihood	-1038.430	-1036.530	-1027.840	-1032.090
ME	0.023	0.018	0.073	0.026
RMSE	3.691	3.673	3.588	3.626
BIC	2106.570	2108.700	2097.280	2111.710
MAE	2.494	2.473	2.471	2.490
MASE	0.728	0.722	0.712	0.721

Table 9: Point and Interval Forecast Table

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Oct-16	50.33	45.69	54.96	43.24	57.42
Nov-16	52.81	44.66	60.96	40.34	65.27
Dec-16	55.35	44.14	66.55	38.21	72.49
Jan-17	56.85	43.11	70.59	35.84	77.86
Feb-17	55.75	39.98	71.52	31.63	79.86
Mar-17	56.42	39.04	73.79	29.84	82.99
Apr-17	55.93	37.29	74.57	27.42	84.43
May-17	55.41	35.77	75.05	25.37	85.45
Jun-17	55.48	35.02	75.94	24.20	86.76
Jul-17	55.86	34.74	76.99	23.55	88.18
Aug-17	56.50	34.79	78.21	23.30	89.70
Sep-17	56.99	34.78	79.21	23.02	90.97
Oct-17	57.27	34.47	80.06	22.40	92.13
Nov-17	58.08	34.69	81.48	22.30	93.87
Dec-17	59.37	35.37	83.37	22.67	96.07

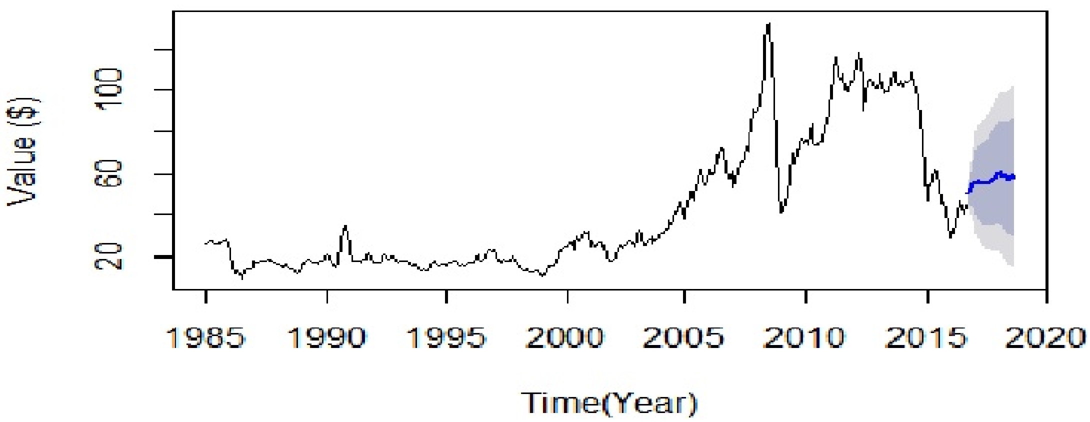

Figure 12: Forecast Plot of ARIMA (2,1,2)(2,0,0)[12]

Table 10: In-forecast Error Table

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1985	0.0	0.8	-0.1	-0.2	-1.1	0.3	0.5	0.4	0.2	0.6	0.6	-2.8
1986	-3.3	-3.9	-0.9	0.6	1.6	-2.9	-2.1	3.9	-1.5	-1.0	0.1	0.8
1987	2.4	-1.6	0.8	0.1	-0.1	0.6	1.2	-1.5	0.0	0.9	-0.7	-1.2
1988	-0.5	-0.8	-1.1	2.0	-0.6	-1.5	-1.1	1.3	-1.7	-0.7	0.8	2.0
1989	0.9	-1.2	1.7	0.4	-1.6	0.4	0.5	-0.8	1.4	0.6	-0.6	0.7
1990	0.8	-1.4	-1.4	-0.7	0.9	-1.4	2.2	9.0	2.2	-2.2	-3.4	-2.5
1991	-0.7	-1.9	2.3	0.7	-0.5	-1.1	1.2	-1.6	0.2	1.5	-1.6	-1.9
1992	-0.5	1.0	0.1	-0.4	-0.4	-1.0	-0.3	0.4	-0.8	1.0	-1.9	-1.6
1993	-0.5	1.0	0.1	-0.4	-0.4	-1.0	-0.3	0.4	-0.8	1.0	-1.9	-1.6
1994	1.7	-1.0	-0.2	1.6	0.5	0.3	0.2	-1.5	-0.1	0.7	0.7	-1.3
1995	1.0	0.6	-0.4	1.3	-1.0	-1.0	-0.8	1.3	0.1	-0.7	0.5	0.8
1996	-0.6	-0.2	1.9	0.6	-2.1	0.4	1.6	-0.1	1.6	0.7	-1.7	1.7
1997	-0.3	-2.4	0.0	-0.6	2.3	-2.3	0.7	0.3	-0.5	1.2	-1.3	-1.7
1998	-1.3	0.4	-0.4	0.5	-0.3	-1.7	0.8	-0.5	1.5	-1.4	-1.4	-0.5
1999	1.5	-1.5	2.3	1.6	-0.8	-0.3	2.6	0.6	1.4	-0.6	2.3	0.1
2000	-0.2	2.4	-0.5	-4.0	6.2	0.9	-2.6	2.0	2.5	-1.7	1.0	-7.2
2001	4.4	1.0	-2.5	2.3	0.8	-1.7	-1.3	1.9	-1.2	-4.0	0.0	1.6
2002	0.0	0.4	3.2	-0.9	-0.1	-1.0	1.8	0.7	1.6	-1.1	-2.0	3.4
2003	2.3	1.0	-4.2	-3.2	3.5	1.8	-0.7	1.0	-3.6	3.0	-0.4	0.3
2004	1.0	-0.7	3.3	-0.4	3.5	-4.0	3.8	3.5	-1.5	5.1	-6.7	-0.0
2005	6.1	0.7	5.0	-3.3	-2.4	8.7	-0.0	4.4	-2.3	-2.9	-0.1	4.0
2006	5.4	-5.1	2.4	7.4	-1.2	-1.8	5.6	-1.9	-9.0	2.4	1.9	2.4
2007	-8.8	8.0	1.8	1.5	-2.4	4.3	3.8	-4.8	9.4	2.3	6.7	-5.1
2008	5.0	2.0	8.2	5.4	11.4	3.5	-1.2	-16.0	-7.3	-17.4	-5.8	-2.5
2009	5.3	-4.5	2.7	-1.5	2.3	5.7	-10.0	9.1	-4.5	9.4	2.8	-4.7
2010	3.3	-1.8	6.7	4.1	-9.3	3.3	1.9	-1.9	-0.3	2.0	-0.6	4.3
2011	1.1	3.9	9.2	2.9	-8.8	3.1	3.1	-6.0	3.6	0.3	6.4	-4.4
2012	3.8	4.1	2.7	-5.8	-7.5	-8.0	12.4	7.3	-4.0	-3.1	-1.5	2.3
2013	3.7	0.9	-5.2	-0.2	1.7	1.6	4.3	-1.6	0.4	-3.1	-0.0	3.9
2014	-4.3	4.6	-0.6	0.4	-1.0	1.0	-2.9	-2.8	-2.0	-8.0	-5.0	-13.0
2015	-5.4	11.9	-7.0	1.1	1.8	-5.6	-6.5	-6.3	3.7	-0.2	-5.2	-4.0
2016	-4.7	2.1	4.9	-1.0	2.1	-0.7	-4.7	-2.8	7.3			

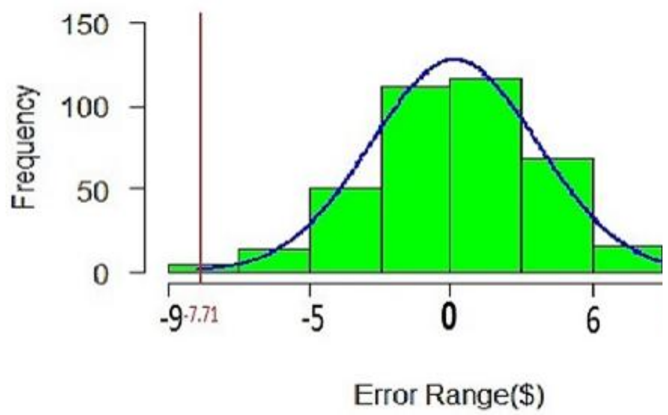

Figure 13: ACF of Error Term

To be double sure, I present a bell-shape-like error term generated from Equation 15 to show that Equation 15 is a good fit of the series.

CONCLUSION

From the investigation performed in this study, the following conclusions can be drawn:

i. Box-Jenkins methodology was used for model identification which indicated that ARIMA (2,1,2) (2,0,0) [12] is best fit for the monthly crude oil prices (January 1985 to September 2016) among set of most fitting model presented in Equation: 14, 15, 16, and 17 in no order.

ii. A traditional method for model validation was used to check the validity of the model also affirmed by forecast-accuracy metrics like Akaike information criterion (AIC), Bayesian information criterion (BIC) among others to be the best model as can be seen in 8 and Figure 13.

iii. The forecast for the price of crude oil per barrel (Brent being the benchmark) for October 2016, November 2016, December

2016 and January 2017 are $50.33, $52.81, $55.35, and $56.85, respectively thus, the crude oil price is forecast to increase up to the December, 2017 to the ton of $59.37 as a point forecast and ($35.37, $96.07) as 95% interval forecast as can be seen in Table 9 and Figure 6.

REFERENCES

[A+17] **O. S. Abdul-Hamid et al.** - *The OPEC annual statistical bulletin*, OPEC, Vienna, 2017.

[B+15] **G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung** - *Time series analysis: forecasting and control*, John Wiley Sons, 2015.

[DM06] **R. Davidson, J. G. MacKinnon** - *Bootstrap methods in econometrics*, 2006.

[DW51] **J. Durbin, G. S. Watson** - *Testing for serial correlation in least squares regression*, ii. Biometrika, 38 (1/2):159 - 177, 1951.

[HJ96] **P. Hall, B. Jing** - *On sample reuse methods for dependent data*, Journal of the Royal Statistical Society. Series B (Methodological), pages 727 - 737, 1996.

[H+06] **J. L. Horowitz, I. Lobato, J. C. Nankervis, N. Savin** - *Bootstrapping the box–pierce q test: a robust test of uncorrelatedness*, Journal of Econometrics, 133(2): 841 - 862, 2006.

[Kob14] **M. Kobayashi** - *Resources for studying statistical analysis of biomedical data and R, Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, Springer, 2014.

[KP03] **J. P. Kreiss, E. Paparoditis** - *Autoregressive-aided periodogram bootstrap for time series*, The Annals of Statistics, 2003.