Journal of Science and Technology Research

Journal homepage: www.nipesjournals.org.ng

Modelling Nigeria Crude Oil Prices using ARIMA Time Series Models

Mujtaba Suleiman [1], Isah Muhammad [2*], Adamu Zakar Adamu [2], Yahaya Zakari [3], Rufai Iliyasu [2], Ali Muhammad [2], Ibrahim Adamu [4] and Mannir Abdu [2]

1 Department of Health Information Management, Federal Polytechnic, Daura, Nigeria.
2 Department of Statistics, Binyaminu Usman Polytechnic, Hadejia, Nigeria.
3 Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.
4 Department of Statistics, Federal University of Technology, Owerri, Nigeria.
Corresponding Author: Isah Muhammad, Email: isahsta@gmail.com, Tel: +2348065071770

Article Info

Keywords:
Box-Jenkins methods, ARIMA model, crude oil prices, information selection criteria

Received 15 February 2023
Revised 14 March 2023
Accepted 15 March 2023
Available online 17 March 2023

https://doi.org/10.5281/zenodo.7744937

ISSN-2682-5821/

Abstract

This paper identified the best ARIMA time series model for monthly crude oil price in Nigeria spanning from 2006 to 2020. At first, the stationary condition of the data series are observed by time plot, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, and then confirmed using Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron (PP) test statistic, which has been found that the crude oil price series is non-stationary. After taking first difference of logarithmic values of data series, the crude oil prices data become stationary. Box-Jenkins four-step iterative methodology comprising of model identification, model fitting, diagnostic and forecasting is also applied to the crude oil prices data. Two optimal time series models were selected namely; ARIMA (2,1,1) and ARIMA (3,1,1) based on the three information criteria AIC, BIC and HQC. Thus, based on the criteria of mean square error; root mean square error; mean absolute error; the ARIMA (3,1,1) model best fits the data with minimum values of predictive measures.

1.0 Introduction

The crude oil industry is the largest and most significant sector in Nigeria, contributing significantly to the country's Gross Domestic Product (GDP) and providing a major source of energy for both Nigeria and the world [1]. Due to its importance, the oil industry has a significant influence on Nigeria's economic and political landscape. While the industry was established at the start of the 20th century, it wasn't until the end of the Nigerian civil war in 1970 that the sector began to take center stage in the country's economic activities [2]. Nigeria relies heavily on the export of oil, which accounts for more than 95% of its export earnings, making it a mono-product economy. The government also heavily relies on oil revenue, which contributes about 70% of its revenue. Furthermore, the majority of new investments in Nigeria are related to oil and its products. As an oil-exporting developing country, Nigeria's economic growth has been unstable due to its dependence on the fluctuating international oil market. The nation's over-reliance on oil exports has made it vulnerable to oil price shocks, which is evident from the significant decline in non-oil exports over the past three decades. This information was reported in a study by [3].

Over the last twenty years, the oil market has experienced predictable seasonal fluctuations. From 1999 to early 2004, oil prices were relatively stable, averaging between USD20 and USD30 per barrel, although it reached as high as USD147 per barrel in July 1998. However, the global financial crisis in September 2008 led to a sharp drop in oil prices, with an average of around USD53 per barrel by the end of 2008. From 2014 to 2016, there was another sharp drop in oil prices, from an average of USD110 in June 2014 to a low of USD29 per barrel in January 2016, due to increased oil production in the United States. The current situation in the global market, including the Saudi Arabia-Russia price war and the COVID-19 pandemic, has resulted in a significant downward trend in the market, creating uncertainty for oil producers. This information was reported in studies by [4], [5], [6], [1] and [7].

The study of the relationship between oil prices and macroeconomics has rapidly grown in the literature, to the point where "oil price" has become a commonly used term. While the initial wave of research on this topic was centered on developed economies, subsequent studies have expanded to include a wider range of countries and regions. Hamilton's seminal study in 1983 served as the basis for extensive research on the relationship between oil prices and macroeconomic variables, particularly in the context of the United States. The main goal of this research has been to examine how oil price shocks impact the cyclical fluctuations of the US economy [8]. Hamilton discovered a negative correlation between oil prices and economic growth, suggesting that the US economy tended to experience a recession following an increase in oil prices. In fact, Hamilton's findings indicate that almost all recessions in the US since World War II were preceded by oil price shocks. [9] conducted a study in Nigeria to investigate the effects of oil price volatility on economic growth using time-series data from 1970 to 2010. They employed Granger causality testing and vector autoregressive modeling in their analysis. Their results showed that while oil prices had a significant impact on real GDP, this effect was only observed through two main variables: the exchange rate and government expenditure. [10] forecasted crude oil prices by applying the ARIMA model to West Texas Intermediate spot prices from January 1970 to December 2003. They compared the result with those of support vector machine and artificial neural networks methods. Once again, the out of sample forecasts indicate that the ARIMA model provides the poorest forecasting performance among the methods considered. [11] performed an out of sample forecast for short and long term horizons employing daily natural gas and Dubai crude oil price from 1994- 2005 using an ARIMA model. The results indicate that for very short horizon forecasts, the ARIMA model outperforms the artificial neural networks and the support vector machine approaches, however, for long horizon forecasts, the ARIMA model underperforms the other approach. [12] conducted a panel study of 72 developed and developing countries to investigate the impact of oil price changes on domestic inflation from 1970 to 2015. Their findings indicated that a 10% increase in international oil prices led to a 0.4% increase in domestic inflation. However, this effect tended to diminish over time, which was consistent with Hooker's findings. They also discovered that oil price changes had an asymmetric impact, with the effect being stronger for increases in oil prices than decreases. The vanishing effect observed was attributed to sound monetary policy frameworks in countries.

In a study conducted by [13] on eleven European countries using wavelet-based Markov switching methodology, the researchers examined the impact of oil price changes on inflation. They discovered that an increase of 100% in oil prices led to an inflation increase ranging from 1% to 6% units. However, the effect of oil price changes on inflation was found to be stronger over a longer time frame, indicating that short-term oil price increases had a relatively low impact on inflation in the countries studied. These varying results can be attributed to differences in sample period, methodology, and how the crude oil indicator was measured. The inconclusive results from previous studies highlight the need for further investigation in this area, which justifies the focus of this study. It is important to have a proper understanding of crude oil prices in Nigeria for both the oil industry and policymakers. This study is unique in several ways, including its use of the Autoregressive Integrated Moving Average (ARIMA) model with the best model selection criteria to fit Nigeria's crude oil price. The study also covers a wide range of data, using monthly data from 2006 to 2020, which includes recent economic events such as Nigeria's economic recession in 2016. Overall, this study adds to the existing literature on the topic.

2.0 General Theories and Methodology

To evaluate and determine the best model for the monthly Nigerian crude oil price, this study utilized Box-Jenkin's iterative four-step method, which involves model identification, model fitting, model diagnostic, and forecasting. The study drew inspiration from the methods of [10], [11] and [6] and used data primarily sourced from the Central Bank of Nigeria's (CBN) 2020 statistical bulletin. The sample period for the study covers the years between 2006 and 2020.

2.1 Autoregressive Integrated Moving Average Process ARIMA (p, d, q)

This process was developed to help remove trends and uncover hidden patterns in non-stationary data because; ARMA process can only model stationary data. Although the theory behind ARIMA time series model was developed much earlier, the systematic procedure for applying the technique was documented in the landmark book by [14]. Since the ARIMA forecasting and box and Jenkins forecasting usually refer to the same set of techniques. Knowing that stationary time series is integrated of the order d and if by differencing the terms it becomes an ARIMA (p, d, q) process, then the difference process can have an ARIMA (p, d, q) representation. In this case the time series $\{X_t\}$ can be expressed as

$$\phi(L)\Delta^d X_t = \theta(L)e_t \qquad (1)$$

where $\Delta^d = (1-L)^d$

2.2 Unit Root Test

Unit root testing has become a standard practice in time series analysis, and tests such as Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) are frequently used to determine the presence of a unit root. Stationarity is a crucial concept, and the absence of a unit root implies that the series does not have time-dependent variances, and any effects of shocks will dissipate over time. Therefore, it is important to apply unit root tests to determine the presence or absence of a unit root in a time series. [15] proposed a test of the null hypothesis that an observable series is trend stationary (stationary around a deterministic trend). [15] proposed the following test statistic which is given below;

$$\text{T-statistic} = \frac{1}{T^2}\sum_{t=1}^{r}\frac{S_t^{\,2}}{\sigma_\infty^{\,2}} \qquad (2)$$

where $S_t = \sum_{j=1}^{t} w_j$

The Augmented Dickey-Fuller (ADF) regression equation due to [16] is given by:

$$\nabla y_t = \mu_0 + \mu_1 + \emptyset y_{t-1} + \sum_{j=1}^{p} \alpha_j \nabla y_{t-1} + e_t \qquad (3)$$

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

where μ_0 is the intercept, $\mu_1 t$ represents the trend in case it is present, ϕ is the coefficient of the lagged dependent variable. y_{t-1} And p lags of Δy_{t-j} with coefficients a_j are added to account for series correlation in the residuals the null hypothesis $H_0 : \phi = 0$ is that the series has unit root while the alternative hypothesis $H_1 : \phi \neq 0$ is that the series is stationary. The ADF test statistics is given by:-

$$\text{ADF} = \frac{\hat{\phi}}{SE(\hat{\phi})} \qquad (4)$$

where $SE(\hat{\phi})$ is the standard error for $\hat{\phi}$, and $\hat{}$ denotes estimate. The null hypothesis of unit root is accepted if the test statistic is greater than the critical values.

2.3 Model Identification

Once the necessary statistical computations and data visualizations are completed for crude oil prices, the next step is to choose an appropriate model based on the research topic. To determine the best model, we use model selection criteria, which help in finding a model that strikes a balance between fit and simplicity. The criteria assist in identifying candidate models that are not complex enough to capture the data or overly complicated. The popular model selection criteria are AIC due to [17], HQC due to [18] and SIC due to [19].

Let $L_n(k)$ be the maximum likelihood of a model with k parameters based on a sample of size n. The information criteria for selecting the most parsimonious correct model proposed by [17] is given by:

$$Akaike: \qquad C_n(k) = -\frac{2\ln\left(L_n(k)\right)}{\frac{n+2k}{n}} \tag{5}$$

Hannan-Quinn information criterion (HQC) is an alternative to Akaike information criterion (AIC) and Bayesian information criterion (BIC) given as;

$$Hannan-Quinn: \qquad C_n(k) = -\frac{2\ln\left(L_n(k)\right)}{\frac{n+2kln(\ln(n))}{n}} \tag{6}$$

Schwarz information is derived using Bayesian arguments, this criterion is also known as the Bayesian Information Criterion (BIC). These criteria take the general form;

$$C_n(k) = -\frac{2\ln\left(L_n(k)\right)}{\frac{n+k\varphi(n)}{n}} \tag{7}$$

where $\varphi(n) = 2$ in Akaike case, $\varphi(n) = 2\ln(\ln(n))$ in Hannan – Quinn case $\varphi(n) = \ln(n)$ in the Schwarz case. Using these criteria, a model is selected that corresponds to:

$$\hat{k} = \arg min_{k\leq m} C_n(k) \tag{8}$$

where, the parameters bear the usual meaning. Schwartz also shows that this criterion is better than AIC. The model with minimum SIC assumes to describe the data series adequately. The minimum value of this criterion is desirable for the adequacy of a model.

2.4 Model Estimation

To estimate the unknown parameters in the preliminary ARIMA (p, d, q) model, various criteria are employed, such as the least square method and maximum likelihood approach, to obtain the best possible estimates of the parameters for a given model. In the case of an independent and identically distributed sample, the joint density function is used.

$$f(x_1, x_2, ..., x_n \mid \theta) = f(x_1 \mid \theta) \times f(x_2 \mid \theta) \times \cdots \times f(x_n \mid \theta). \tag{9}$$

The likelihood function of the independent and identically distributed sample is given as:

$$\mathcal{L}(\theta; x_1, \cdots, x_n) = f(x_1, x_2 \cdots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta). \tag{10}$$

where ";" denote a separation between the two input arguments: θ and the observations $x_1, x_2 \cdots, x_n$. In practice it is often more convenient to work with the natural logarithm of the likelihood function, called the log likelihood to estimate the unknown parameters:

$$\ln\mathcal{L}(\theta; x_1, \cdots, x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta), \quad (11)$$

2.5 Diagnostic Checking

Once the model is estimated using the Box-Jenkins approach, the next step is to diagnose the adequacy of the model. It is essential to identify how the model is adequate and how it is inadequate. To assess the adequacy of the overall Box-Jenkins model, it is necessary to analyze the residuals obtained from the model. The following methods can be used to evaluate the model adequacy based on the residuals:

2.5.1 Jarque-Bera Test

[20], proposed test for normality based on skewness and kurtosis of a distribution. The Jarque-Bera test is a two-sided goodness of fit test suitable when a fully-specified null distribution is unknown and its parameters must be estimated. The test statistic is given by

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right) \quad (12)$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis. The test check the null hypothesis that, the distribution is symmetry and hence normal.

2.5.2 Box-Ljung Test

Box-Ljung test proposed by [14], is a diagnostic tool used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an ARIMA (p, d, q) models to the data, the test examines correlations of the residuals. If the autocorrelations are very small, we conclude that the model does not exhibit significant lack of fit.

2.6 Forecasting

Forecasting as described by [14] provides basis for economic and business planning, inventory and production control and optimization of industrial processes. The efficiency validation of the considered models was evaluated by means of the following measures;

2.6.1 Root Mean Square Error (RMSE)

The RMSE is a measure of how well the model fits the data. It is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(yi-\hat{y}i)^2}{n}} \quad (13)$$

where the $\hat{y}_i$ are the values of the predicted variable when all samples are include in the model formation, and n is the number of observations. RMSE

2.6.2 Mean Absolute Error (MAE)

The MAE is a quantity used to measure how close predictions are to the eventual outcomes.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (14)$$

It is an average of the absolute errors. i.e.$|e_i| = |f_i - y_i|$, where f_i is the prediction and y_i is the true value.

2.6.3 Mean Absolute Percentage Error (MAPE)

The MAPE is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| * 100 \quad (15)$$

Where, A_t is the actual value and F_t is the forecast value.
The difference between A_t and F_t is divided by the Actual value A_t again.

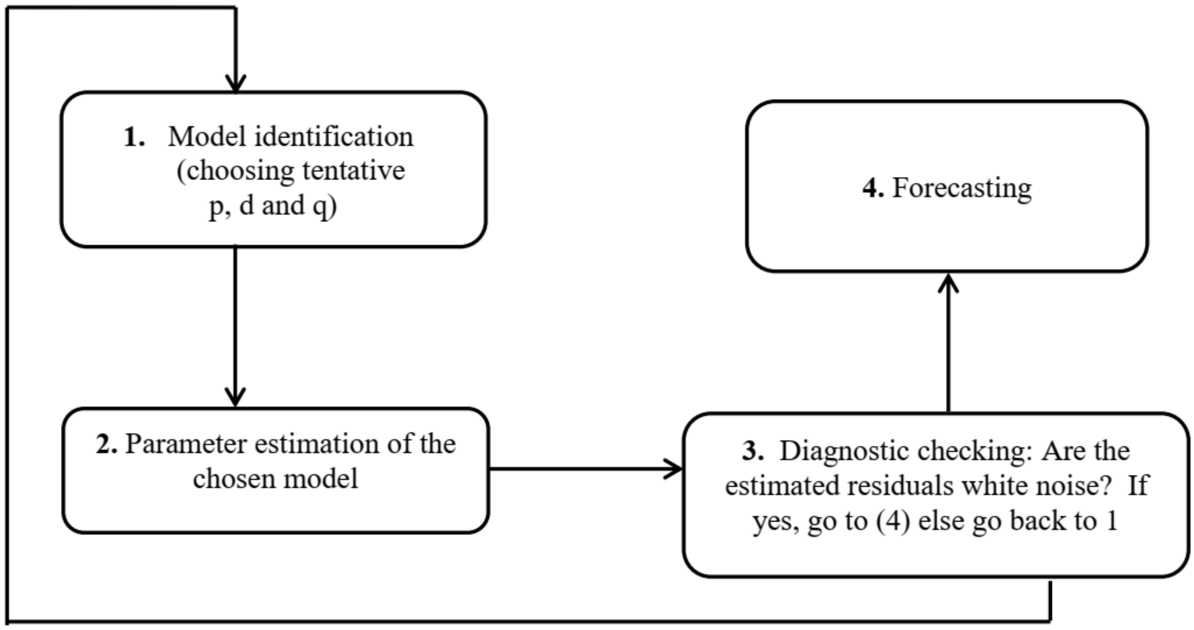

Fig 3.1: Box–Jenkins Methodology

3.0 Results and Discussion

3.1 Explore the Time Series

To identify the model of any time series data, it is necessary to identify the process of data generation and to achieve this, a study of the pattern and behavior was done by plotting the time plot, ACF and PACF of the series.

3.1.1 Time Plot of the Crude Oil Prices

Fig 1: Time plot of the Crude Oil Prices Data

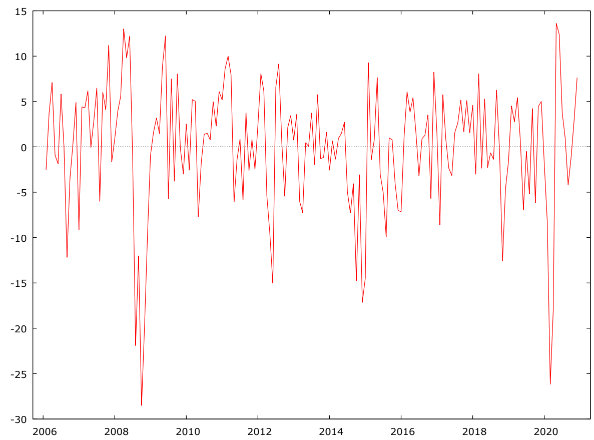

Fig 2: Time plot of the first difference of the data

From Fig 1, it could be seen that the year 2008 recorded the highest oil price while the year 2020 recorded the lowest oil prices for the period of study. The oil prices showed a fluctuationnal pattern, suggesting that the mean and variance of the crude oil price in nigeria have been changing over time. This means we have two sources of non-stationary variables, the variance and the mean. The former can be removed by taking the log of the data while the later can be removed by differencing.

However, Fig 2 shows the time plot of the first difference of the original series. 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 for both mean and variance suggesting that the time series is stationary.

3.1.2 Unit Root Tests of the Crude Oil Prices Data

Table 1: ADF and KPSS tests of the data

	Unit Root Tests of the Original Data		Unit Root Tests of the First Difference of the Data	
Test	T-Statistic	Critical/P-Values	T-Statistic	Critical/P-Values
ADF without constant	-0.850571	0.3477	-9.10733	6.182e-016
ADF with constant	-2.54409	0.105	-9.08094	8.18e-014
ADF with constant and trend	-2.96882	0.1411	-9.07561	8.482e-013
KPSS without trend	0.964772	10% 5% 1% 0.349 0.462 0.739	0.0757909	1% 5% 10% 0.349 0.462 0.739
KPSS with trend	0.388718	10% 5% 1% 0.120 0.148 0.216	0.0375596	1% 5% 10% 0.120 0.148 0.216

From Table 1, the KPSS test result showed that the data is not stationary before first differencing since the KPSS t-statistic is greater than the critical value at 1%, 5% and 10% levels of significance, implying that the null hypothesis (H_0) was rejected. However, the data is stationary at first differencing since the KPSS t-statistic is less than the critical value at 1%, 5% and 10% level of significance, implying that we fail to reject H_0, which claims stationary on the data at first difference. Similarly, from Table 1, we have strong evidence to fail to reject H_0, which says there is presence of unit root in the data before first difference since the p-values are greater than the value of alpha at 1%, 5% and 10% level of significance. However, we have strong evidence to reject H_0, which state that there is no presence of unit root in the data at first difference since the p-values are less than the value of alpha at 1%, 5% and 10% level of significance.

3.1.3 ACF and PACF of the Crude Oil Prices Data

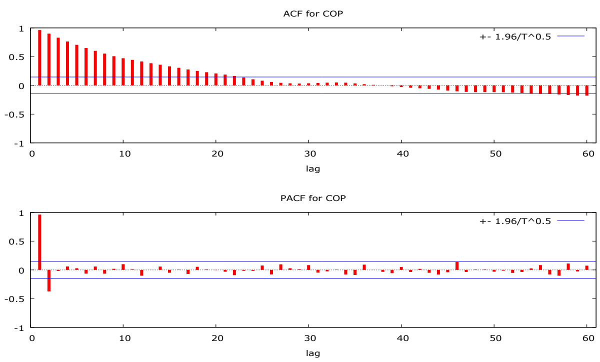

Fig 3: Correlogram (ACF and PACF) Plots of the Series

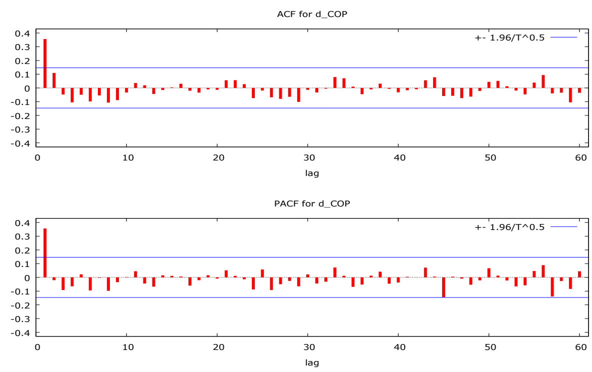

Fig 4: Correlogram (ACF and PACF) Plots of the 1st difference

After the preliminary analysis it was observed that the data can be modelled using Box-Jenkins procedure. Hence in this section, the correlation structure of the data (crude oil price series) using ACF and PACF are then examined as shown above. Fig 3 shows that ACF is exponentially decaying and the PACF is significant at lag 25 before dropping to zero. So, this shows AR feature. Similarly, Fig 4 shows the estimated ACF and PACF of series' first difference. Thus, the ACFs at lag 1 is statistically different from zero (at the 95% confidence limit), but at all other lags is not statistically different from zero i.e. the ACF decayed after lag 1. Also, the PACF is statistically different from zero only at lag 1, 45, and 57; all other lags are not different from zero (at the 95% confidence limits). It could be seen that the data is now stationary and that there is no existence of seasonality. This also suggest that AR and MA exists in the time series data. Hence the series is non-seasonal but integrated and therefore the best model that will fit the data is ARIMA model.

3.2 Model Identification

By applying the principle of parsimony, and since after first difference, the data is said to be stable. Thus, d = 1, p,q = (0, 1, 2, 3, 4). 25 models were generated as shown below;

Table 2: Result of ARIMA model identification and selection

MODEL	AIC	BIC	HQC
ARIMA(0,1,0)	1205.648	1212.023	1208.233
ARIMA(0,1,1)	1185.986	1199.548	1189.863
ARIMA(0,1,2)	1184.392	1199.142	1189.562
ARIMA(0,1,3)	1186.351	1202.287	1192.813
ARIMA(0,1,4)	1186.723	1205.847	1194.478
ARIMA(1,1,0)	1185.288	1197.850	1197.165
ARIMA(1,1,1)	1185.237	1197.987	1190.404
ARIMA(1,1,2)	1186.375	1202.312	1192.837
ARIMA(1,1,3)	1185.873	1202.997	1191.627
ARIMA(1,1,4)	1185.329	1207.641	1194.376
ARIMA(2,1,0)	1185.212	1212.961	1190.382
ARIMA(2,1,1)+	**1182.255**	**1191.191**	**1188.717**
ARIMA(2,1,2)	1184..212	1203.337	1191.967
ARIMA(2,1,3)	1185.554	1207.865	1194.601
ARIMA(2,1,4)	1185.883	1211.383	1196.223
ARIMA(3,1,0)	1185.635	1201.572	1192.097
ARIMA(3,1,1)+	**1184.199**	**1193.323**	**1191.953**
ARIMA(3,1,2)	1186.105	1208.417	1195.153
ARIMA(3,1,3)	1190.364	1215.863	1200.704
ARIMA(3,1,4)	1187.785	1216..472	1199.418
ARIMA(4,1,0)	1186.864	1205.988	1194.619
ARIMA(4,1,1)	1187.605	1209.917	1196.652
ARIMA(4,1,2)	1187.650	1213.149	1197.990
ARIMA(4,1,3)	1188.022	1216.708	1199.654
ARIMA(4,1,4)	1189.665	1221.539	1202.589

Table 2 shows the results of model selection, 25 models were tested based on the Alkaike information criteria (AIC); Bayesian information criteria (BIC); Hannan-Quinn information criteria (HQC), and two models were selected for further examination namely; ARIMA(2,1,1) and ARIMA (3,1,1) models since they have the minimum values of AIC, BIC and HQC.

3.3 Model Estimation

Table 3: Result of ARIMA(2,1,1) model estimation for the Crude Oil Price

Parameter R	Coefficient	Std. Error	Z-Statistics	P-Values
Constant	−0.152056	0.156619	−0.9709	0.3316
phi_1	1.33248	0.0688246	19.36	1.66e-083 ***
phi_2	−0.381153	0.0692061	−5.508	3.64e-08 ***
theta_1	−1.00000	0.0186233	−53.70	0.0000 ***

The model for ARIMA (2,1,1) is fitted and shown below;

$$y_t = -0.152056 + 1.33248y_{t-1} - 0.381153y_{t-2} + \varepsilon_t - 1.00000\varepsilon_{t-1} \quad (16)$$

From Table 3, the p-value for the constant term is grater than 5% level of significance, therfore we failed to reject the null hypothesis and conclude that its statistically insignificant to the model, while the parameters phi_1, phi_2 and theta_1 are statistically significant to the model since their p-values are less than 5% level of significance there by rejecting the null hypothesis.

Table 4: Result of ARIMA(3,1,1) model estimation for the Crude Oil Price

Parameter	Coefficient	Std. Error	Z-Statistics	P-Values
Constant	−0.152843	0.154444	−0.9896	0.3224
phi_1	1.32559	0.0745426	17.78	9.57e-071 ***
phi_2	−0.357607	0.120061	−2.979	0.0029 ***
phi_3	−0.0177411	0.0739566	−0.2399	0.8104
theta_1	−1.00000	0.0189446	−52.79	0.0000 ***

The model for ARIMA (3,1,1) is fitted and shown below;

$$y_t = -0.152843 + 1.32559y_{t-1} - 0.357607y_{t-2} - 0.01774116y_{t-3} + \varepsilon_t - 1.00000\varepsilon_{t-1} \quad (17)$$

From Table 4, the p-values for the constant term and the parameter phi_3 are greater than 5% level of significance, therefore we failed to reject the null hypothesis and conclude that the parameters are statistically insignificant to the model, while the parameters phi_1, phi_2 and theta_1 are statistically significant to the model since their p-values are less than 5% level of significance there by rejecting null hypothesis.

3.4 Model Checking

3.4.1 ACFs and PACFs of the Residuals Using the Selected Models

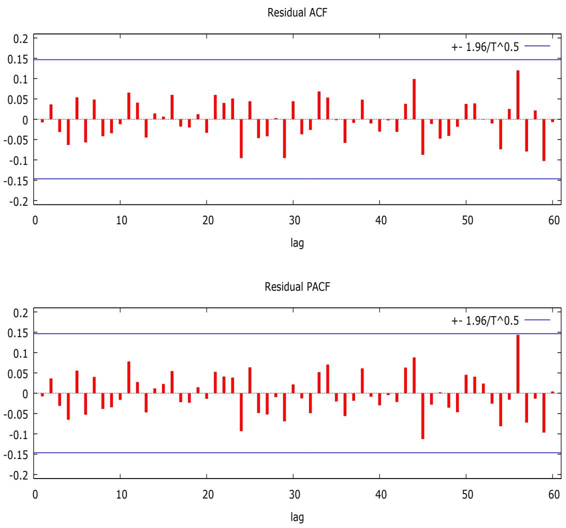

Fig 5: ACF and PACF of the residual from ARIMA (2,1,1) model

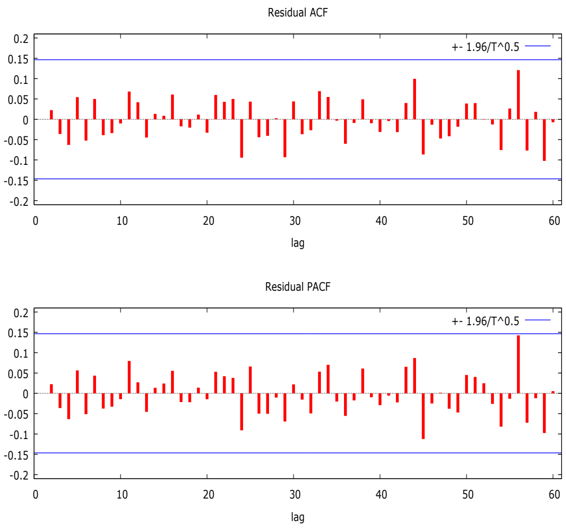

Fig 6: ACF and PACF of the residual from ARIMA (3,1,1) model

Fig 4 and Fig 6 confirmed that there is no form of correlation amongst the residuals, this means that the ARIMA (2,1,1) and ARIMA (3,1,1) models have passed the standard test criteria of being white noise, since the residuals are uncorrelated and stationary.

3.4.2 Normality Test of the Residuals

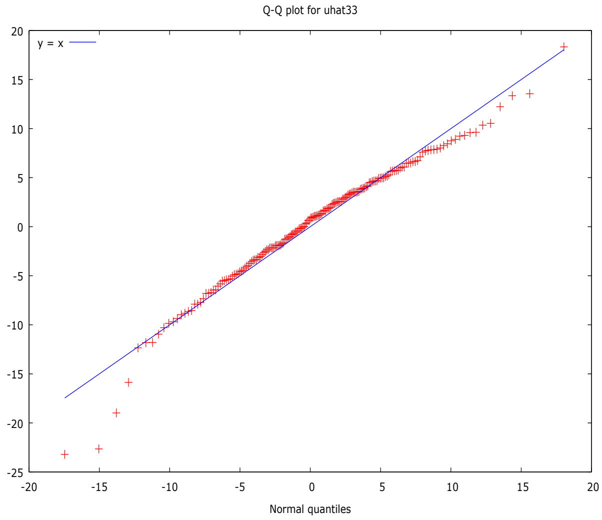

Fig 7: Normal probability plot from ARIMA (2,1,1) model

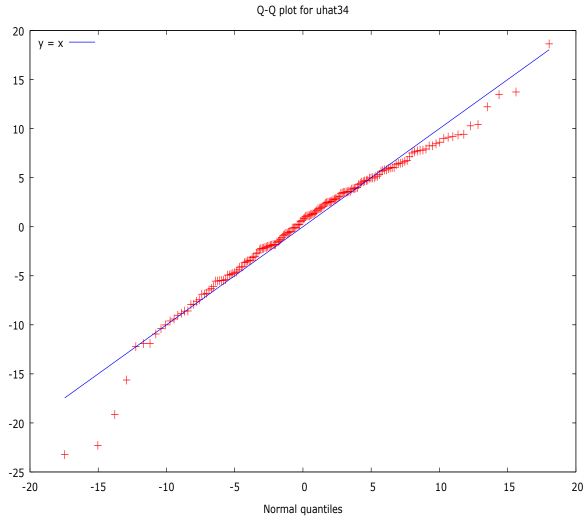

Fig 8: Normal probability plot from ARIMA (3,1,1) model

From Fig 7 and Fig 8 above, it is observed that the relationship between the theoretical percentiles and the sample percentiles is approximately linear. Therefore, the Normal Probability Plot of the residuals of the data using the two models suggests that the error terms are indeed normally distributed.

3.4.3 Jarque-Bera and Ljung-Box Test

Table 5: Result of Jarque-Bera and Ljung-Box Tests of the squared residuals

Test	Model	Test Statistic	P-Value
Jarque-Bera	ARIMA(2,1,1) model	1.113455	0.110233
	ARIMA(3,1,1) model	1.596970	0.150980
Ljung-Box	ARIMA(2,1,1) model	1.5437	0.8189
	ARIMA(3,1,1) model	1.8911	0.2105

Table 5 shows the result of Jarque-Bera and Ljung-Box tests, it could be seen that the p-values of the two selected models are all greater than the value of alpha at 5% level of significance, we therefore failed to reject the null hypothesis and conclude that the models selected are normally distributed and does not exhibit significant lack of fit, respectively.

3.5 Forecasting

Table 6: Result of ARIMA (2,1,1) and ARIMA (3,1,1) Forecasted Values for the Crude Oil Price

	ARIMA (2,1,1)			ARIMA (3,1,1)		
Year/Month	Prediction	std. error	95% interval	Prediction	std. error	95% interval
2021:1	53.67	6.355	41.21 - 66.12	53.69	6.353	41.24 -66.14
2021:2	55.19	10.587	34.44 - 75.95	55.35	10.549	34.68 -76.03
2021:3	55.95	13.806	28.89 - 83.01	56.21	13.797	29.17 -83.25
2021:4	56.37	16.255	24.52 - 88.23	56.69	16.291	24.76 -88.62

2021:5	56.64	18.141	21.08 - 92.19	56.98	18.218	21.28 -92.69
2021:6	56.82	19.616	18.38 - 95.27	57.17	19.721	18.52 -95.83
2021:7	56.96	20.785	16.22 - 97.70	57.31	20.908	16.33 -98.28
2021:8	57.07	21.724	14.49 - 99.64	57.40	21.853	14.57 - 100.23
2021:9	57.15	22.483	13.08 - 01.21	57.47	22.612	13.15 - 101.79
2021:10	57.21	23.103	11.93 - 102.49	57.51	23.226	11.99 - 103.04
2021:11	57.25	23.611	10.97 - 103.52	57.54	23.726	11.04 - 104.04
2021:12	57.27	24.030	10.18 - 104.37	57.55	24.134	10.25 104.85

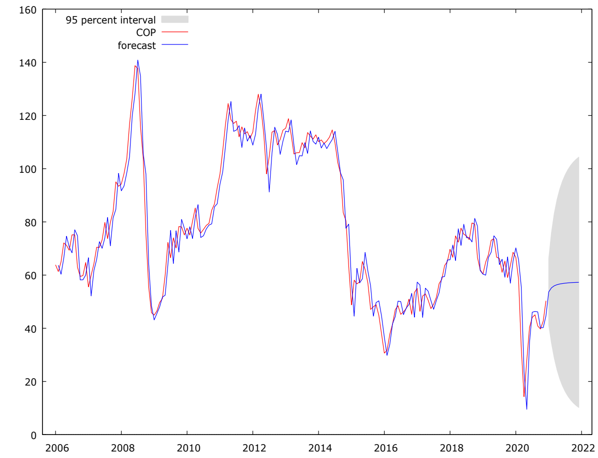

Fig 9: Plot of actual and forecasted Series of ARIMA (2,1,1)

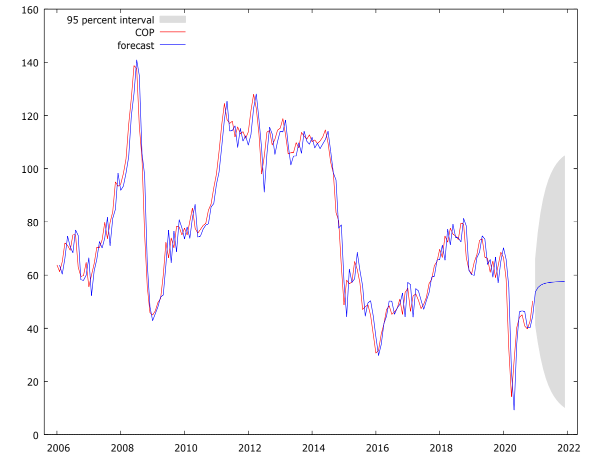

Fig 10: Plot of actual and forecasted Series of ARIMA (3,1,1)

Fig 9 and Fig 10 revealed the original and the forecasted values of the crude oil price series in Nigeria produced by ARIMA (2,1,1) and ARIMA (3,1,1) models, respectively. The figures signifies that the models fitted the data well since there were no much difference in the pattern of the forecasted values and the actual values.

3.5.1 Model Validation Based on Forecasting Power

Table 7: Forecast Evaluation Statistics

Predictive Measures	ARIMA (2,1,1)	ARIMA (3,1,1)
Mean Squared Error	40.88196	40.86789
Root Mean Squared Error	6.3939	6.3928
Mean Absolute Error	4.9971	4.9942
Mean Percentage Error	-0.45729	-0.45829
Mean Absolute Percentage Error	7.7621	7.7226

The smaller the value of the error, the better the forecasting performance of the model. From Table 7, it could be seen that both the two selected models have shown good result (minimum predictive measures). But forecast result of ARIMA (3,1,1) model is more closer to the actual series. Therefore, the prediction power of ARIMA (3,1,1) model is better and suitable for monthly periods forecasting, as such the model best fit the data.

4.0 Conclusion

Box-Jenkins four-step iterative methodology comprising of model identification, model fitting, diagnostic and forecasting is also applied to the Nigerian monthly crude oil prices data between the periods of January 2006 to December 2020. Three optimal time series models were selected namely; ARIMA (1,1,0), ARIMA (2,1,1) and ARIMA (3,1,1) based on the three information criteria AIC, BIC and HQC. Moreover, based on the criteria of mean square error; root mean square error; mean absolute error; the ARIMA (3,1,1) model best fits the data with minimum values of predictive measures. Therefore, the study findings could be helpful for monitoring oil markets and developing policies for stabilizing oil price. With the knowledge of other factors the forecast result can be used to achieve viable and workable framework on future price pattern.

References

[1] Ramat, M. H., Zakari, Y., Usman, M., Muhammad, I., Falgore, J.Y. & Dikko, H. G. (2022). The Effect of Exchange Rate on Gross Domestic Product (GDP) on the Nigerian Economy using ARDL-ECM approach. Communication in Physical Sciences, 8(2): 211-221.

[2] Central Bank of Nigeria (CBN), Annual Reports and Statement of Accounts – Various Years. (9.9 Central Bank of Nigeria, CBN (2014)): "The Structure of the Nigerian Economy and Implications for Development", Realm Communication Ltd, Lagos.

[3] Huntington, H. G. (2014). Oil Price Forecasting in the 1980s: What Went Wrong? The Energy Journal, 15 (2), 1-22.

[4] Bolton, P. (2019) oil price. Brie.

[5] Maugeri, L. (2009). Understanding oil price behavior through an analysis of a crisis. Review of Environmental Economics and Policy, 3(2), 147-166.

[6] Adamu, I., Ogbonna, C.J., Adamu, Y., & Zakari, Y. (2021). Modeling Coronavirus Pandemic Using Univariate and Multivariate Models. The Nigerian Perspective. Asian Journal of Probability and Statistics, 15(4), 134-143.

[7] Ellwanger, R., Sawatzky, B. and Zmitrowicz, K. (2017). Factors behind the 2014 oil price decline. URL: http://www.bankofcanada.ca/wp-content/uploads/2017/11/boc-review-autumn 2017.

[8] Hamilton, J.D. (1983). Oil and Macro Economy since World War II. The Journal of Political Economy, 91: 228-248.

[9] Oriakhi, D. E., & Osaze, I. D. (2013). Oil price volatility and its consequences on the growth of the Nigerian economy: An examination (1970-2010). Asian economic and financial review, 3(5), 683.

[10] Xie, W., Yu, L. Xu, S. & Wang, S. (2006). "A new method for Crude Oil Price Forecasting Based on Support Vector Machines." International Conference on Computation Science (Part IV), pp. 444–451.

[11] Farnandez, C. (2010). A sample Approach to Short-term and Long-term Oil Price Forecasting. Energy Economics, 23, 325-338.

[12] Choi, S., Furceri, D., Loungani, P., Mishra, S., & Poplawski-Ribeiro, M. (2018). Oil prices and inflation dynamics: Evidence from advanced and developing economies. Journal of International Money and Finance, 82, 71-96.

[13] Živkov, D., Đurašković, J., & Manić, S. (2019). How do oil price changes affect inflation in Central and Eastern European countries? A wavelet-based Markov switching approach. Baltic Journal of Economics, 19 (1), 84-104.

[14] Box, G. E. P., & Jenkins, G. M. (1976).Time Series Analysis: Forecasting and Control. San Franciso: Holden- Day.

[15] Kwiatkowski, P .C. B. Phillips, P. Schmidt, & Yongoheol, S. (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometric, 54: 159-178.

[16] Dickey, D. A. & Fuller, W. A. (1979). Distributions of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association, 74: 427-431.

[17] Akaike, H. (1974). A New Look at the Statistical Model Identification. I.E.E.E. Transaction on Automatic Control, 19, 716-723.

[18] Hannan, E. J. & Quinn, B. G. (1979).The Determination of the Order of an Autoregression. Journal of the Royal Statistical Society, 41, 190-195.

[19] Schwarz, K. J. (1978).The Determination of the Order of an Autoregression. Journal of the Royal Statistical Society, 41, 190-195.

[20] Jarque, C. M. & Bera, A. K. (1987). A Tests for Normality of Observations and Regression Residuals. International Statistical Review, 55 (2), 163-172.