International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 1, Jan-Feb 2022
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Monte Carlo Simulation to Autoregressive Integrated Moving Average (MS-ARIMA) Model for Time Series Modelling and Forecasting: Case Study of Nigerian Forcados Price

Djoï N. André[1], Dr. Nwosu I. Joseph[2], Prof. Ikiensikimama S.Sunday[3]

[1]World Bank PhD Scholar, Department of Petroleum and Gas Engineering, World Bank Africa Centre of Excellence in Oil Field Chemicals Research (ACE-CEFOR), University of Port Harcourt, Nigeria

Email: djoi.andrew@gmail.com

[2]Department of Geology, Facultyof Sciences, University of Port Harcourt, Nigeria

[3]Shell Professorial Chair,Department of Petroleum and Gas Engineering,World Bank Africa Centre of Excellence in Oil Field Chemicals Research (ACE-CEFOR), University of Port Harcourt, Nigeria

--************************----------------------------------

Abstract:

Time series modelling is a valuable tool in commodity price forecasting, especially crude oil and natural gas price forecasting. The random nature of crude oil and natural gas prices has been proven and is taken into account in the forecasting process. The Autoregressive Moving Average (ARIMA) and its multiple variants are the common techniques used for this purpose. This research proposes a stochastic approach that applies the Monte Carlo Simulation to Autoregressive Moving Average (MS-ARIMA) in forecasting the random behavior of oil price into the future. This is an improvement over the ordinary ARIMA model for oil price forecasting. The result of oil price forecasting using the MS-ARIMA and ordinary ARIMA model for Nigerian forcado crude oil between 2000-2019 shows that the oil price forecasted using the MS-ARIMA model was closer to the actual oil price within the same period than those forecasted using the ordinary ARIMA model. In this research the MS-ARIMA model has been used to forecast the yearly price of Nigerian forcados crude oil for a period of 10 years between 2020 and 2029. The result shows that the oil prices have high chances of being between $US 64.32 and $US 69.57 with10, 50 and 90 percent of chance beingrespectively higher than $US81.28, $US 66.75 and $US43.25 from 2020 to 2029.

Keywords **—Time series, ARIMA, Monte Carlo simulation, MS-ARIMA, P10, P50, P90, Nigeria.**

--************************----------------------------------

1. INTRODUCTION

Commodity price prediction is vital in any economic activity. Crude oil price forecasting has been one of the key inputs of exploration and/or field development projectseconomic models. Unlike other commodities, crude oil and natural gas prices fluctuation is not only caused by the demand and supply chain imbalances and shocks in the energy market(Arnold and Rume, 2020)[4]. The ever-changing futures-price and chaotic volatilities result from a series of dynamic events such as change in regime, politics, economic crises, formation or breakdown in trade agreements, war or conflict in regions, unexpected weather patterns etc. This shows the stochastic nature of crude oil and natural

 ©IJSRED: All Rights are Reserved

gas price and that must be taken into account in crude oil price forecasting.

In Nigeria, oil and gas exports accounted for more than 98% of export earnings and about 83% of the total revenue in 2000 (Omekara, C. O. et al., 2015)[14]. Kayode and Habib (2013)[12]have shown that about 70% of government revenue is derived from oil and over 90% of new investments are associated with oil and allied products. Several times series forecasting methods have been used for Nigeria's crude oil/natural gas price as well as for crude oil production and export. Arnold and Rume (2020)[4] as well as Balogun and Ogunleye (2015)[7] have modelled and forecasted Bony Light (a Nigerian crude oil) price and monthly crude oil export with Autoregressive Integrated Moving Average (ARIMA) model. Garzali and Bashir (2018)[11] have shown that crude oil forecasting using seasonal Conditional Heteroscedasticity (GARCH) produced a result that matched with crude oil production data from April 2013 to December 2015. Kayode and Habib (2013)[12] as well as Omekara et al. (2015)[14] have respectively used the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Multiplicative SARIMA model for Nigerian monthly crude oil production and export. These methods are only based on deterministic approach with probabilistic assumptions at model validation stage.

In this work we propose an approach which takes into account the randomness of oil price by combining the Monte Carlo Simulation with the Auto-regressive Integrated Moving Average (ARIMA) Model and compare the results of this model with the ordinary ARIMA.

2. MATEIALS AND METHODS

2.1. Materials

The proposedapproach is being tried using Nigerian forcados crude oil price. The yearly oil prices were collected from BP Statistical Review of Global Energy website[8]. It spans a period of time of twenty (20) years (2000 to 2019). The computation was made using Microsoft Excel and R Studio (a statistical analysis software). The models were implementedvia R-programming for the following reasons: (a) high execution speed, (b) the availability of necessary libraries and tools for their open-source nature, (c) ease of implementation, and (d) flexibility of the system via t-series lib, forecast lib and ggplot2 lib (Arnold and Rume, 2020)[4].

2.2. Methods

The essence of our approach, called Monte Carlo Simulation to Autoregressive Integrated Moving Average (MS-ARIMA)Model is the combination of the Monte Carlo Simulation to ARIMA model. We will first model the time series with the ordinary Box-Jenkins (ARIMA) method and simulate the oil price around the forecasted value on the basis of the probability distribution pattern of the data set. The work will be performed in four steps: (1) determination of ARIMA model, (2) forecasting with the ARIMA model, (3) Monte Carlo Simulation of oil price and (4) comparison of MS-ARIMA and ordinary ARIMA results.

2.2.1. ARIMA Model

ARIMA modelling of a time series consists of identifying the model, selecting the model, estimating model parameters and proceeding to its validation.

Time series is a series of data points indexed or graphed in the order of time (Selviet al., 2018)[17].There are many different notations used for time-series analysis. $(X_t)_{0\leq t\leq T}$ or $(X_t)_{t\epsilon T}$ or $X=\{X_1,X_2,\cdots,X_T\}$are notations used for discrete parameter processes and $X(t)$ is a notation for continuous parameter processes. For times series analysis with Box-Jenkins method, the ordinary index is used. 0 is affixed to the first value of the time series, 1 to the next one, and so forth.

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; that is, change in time series does not affect the trend line, no increase or decrease(Balogun and Ogunleye, 2015)[7]. A stationary time series has a constant mean, a constant variance and the covariance is independent of time.**Trend** also called**secular** or **long-term**

 ©IJSRED: All Rights are Reserved

trend, is the basic tendency of a series to grow or decline over a long period of time.

Box-Jenkins methods states that for a time series $(X_t)_{0 \le t \le T}$, an observation X_t is related to a certain number of the previous values X_{t-j}, j>0 and/or errors terms through a linear relationship (Anderson, 1977)[3].

Mathematically, a stationary time series $(X_t)_{0 \le t \le T}$ is said to follow the Autoregressive Moving Average model of order p and q, that is, ARMA (p, q) when for a given t, X_t can be expressed from Equation 1 (Nashirah and Sofian, 2017)[13]:

$$X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} + \mu + \varepsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \quad (1)$$

Where: φ_0 and μ are constants, X_{t-i} are called the autoregressive terms, ϵ_{t-j} the moving average terms, φ_i the parameters of the autoregressive term, θ_j the parameters of the moving average term of the model**;** p and q are respectively the number of the autoregressive and moving average terms.

When μ and all θ_j are equal to zero, the model is called Autoregressive model of order p, that is, AR (p) and is expressed in (Equation 2):

$$X_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i X_{t-i} \quad (2)$$

When all $\boldsymbol{\varphi_i}$ are equal to zero, the model is called Moving Average model of order q, that is, MA(q) and expressed as in (Equation 3):

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{p} \theta_j \epsilon_{t-j} \quad (3)$$

The ARMA (p, q) model expression can be reduced to the following (Equation 4) by adding $\varphi_0$and μ.

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{p} \theta_j \epsilon_{t-j} \quad (4)$$

The error terms ε_t are generally assumed to be independent identically distributed(i.i.d.) random variables sampled from a normal distribution with zero mean, $\epsilon_i \sim N(0, \sigma^2)$, where $\sigma^2$is the variance. In this case,ϵ_i are said to be white noises.

When the time series X_t is not stationary, it can be converted into a stationary time series Y_t by differencing (Reza and Ahmad, 2012)[16].

If the differencing of order d is performed before making the time series stationary, the model becomes ARIMA (p, d, q) and is expressed from Equation 5 (Nashirah and Sofian, 2017)[13]:

$$X_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{p} \theta_j \epsilon_{t-j} \quad (5)$$

Where Y_t is the stationary time series obtained after integration (differencing) and d, the order of integration. Y_t is defined as in Equation 6, (Rangsan and Titid, 2006)[15]:

$$Y_t = \Delta^d X_t = (1 - B)^d X_t \quad (6)$$

Where B is the lag or backward linear operator defined by Equation 7:

$$B^i X_t = X_{t-i} \quad (7)$$

$(1 - B)^d$ is expanded with polynomial development process. One has for instance the expressions of Equations 8 and 9:

$$\Delta X_t = \Delta^1 X_t = (1 - B) X_t = X_t - X_{t-1} \quad (8)$$

$$\Delta^2 X_t = (1 - B)^2 X_t = X_t - 2X_{t-1} + X_{t-2} \quad (9)$$

A. *Model Identification*

For the ARIMA model identification, two main tasks are performed: the time series stationarity test and the determination of the model orders p, d, and q.

- ***Stationarity Test***

The stationarity of the times series can be checked from three common techniques: the graphical analysis of the time series plot, the correlograms analysis and a statistical test called Unit Root Test. The first two are the most used; a unit root test comes when there is doubt about the graphical analyses results.

Graphical Analysis of the Time Series Plot: A time series $(X_t)_{0 \le t \le T}$ plot is the graph of X_t versus t. A quick way of telling whether a process is stationary is to plot the series against time. If the graph crosses the mean of the sample many times, then the variable is stationary; otherwise that is an

 ©IJSRED: All Rights are Reserved

indication of persistent trends away from the mean of the series (Balogun and Ogunleye, 2015)[7].

Correlograms Analysis: The correlograms of a time series are the plots of its autocorrelation function (ACF) and partial autocorrelation function (PACF) against lag length (Adolphus, 2016)[1]. The autocorrelation and partial autocorrelation functions are respectively the set of the autocorrelation and partial autocorrelation coefficients as function the lags.

The autocorrelation coefficient expresses the correlation between a variable lagged one or more periods and itself. The formula for computing the lag k autocorrelation coefficient between X_t and X_{t-k}which are k periods apart, is given byEquation 10(Selvi et al., 2018)[17]:

$$\boldsymbol{r_k} = \frac{\sum_{t=k+1}^{T}(X_t-\bar{X})(X_{t-k}-\bar{X})}{\sum_{t=1}^{T}\left((X_t-\bar{X})\right)^2} \quad (\mathbf{10}) \text{for } k = 0, 1, 2, \ldots.$$

Equations 11 to 13help in getting a good estimate of the partial autocorrelation coefficient ρ_k(Adolphus, 2016) [1].

$$\boldsymbol{\rho_{11} = r_1} \quad (\mathbf{11})$$

$$\boldsymbol{\rho_{kk}} = \frac{\boldsymbol{r_k - \sum_{j=1}^{k-1}\rho_{k-1,j}.r_{k-j}}}{\boldsymbol{1 - \sum_{j=1}^{k-1}\rho_{k-1,j}.r_j}} \quad (\mathbf{12})$$

For k = 2, 3, ….

$$\boldsymbol{\rho_{kj} = \rho_{k-1,j} - \rho_{kk}\cdot\rho_{k-1,k-j}} \quad (\mathbf{13})$$

k = 3, 4, …, j = 1, 2, …, k-1.

The tips for checking whether a time series is stationary or not is as follow. If the graph of a time series ACF values either cuts off fairly quickly or dies down fairly quickly, then the time series values should be considered stationary. If a graph of ACF dies down extremely slowly, then the time series values should be considered non-stationary (Rangsan and Titida, 2006)[15].

A unit root test is used to find whether or not a time series is stationary using an autoregressive model (Balogun and Ogunleye, 2015)[7]. Because the unit root test is a statistical test for the proposition that for an autoregressive statistical model of a time series, the autoregressive parameter is one (Wolters and Hassler, 2005)[19]. The most common unit root tests are Augmented Dickey – Fuller (ADF) test, Kwiatkowski Phillip Schmidt Shin (KPSS) test, Phillips – Perron (PP) test and DF – GLS test (Kayode and Habib, 2013)[12]. In our study we are going to test Nigerian Forcados crude oil price time series stationarity with the Augmented Dickey – Fuller (ADF) test, if needed.

Said and Dickey (1984) augmented the basic autoregressive unit root test to accommodate general ARMA (p, q) models with unknown orders and their test is referred to as the Augmented Dickey- Fuller (ADF) test (Wolters and Hassler, 2005)[19]. The ADF test is used to test the null hypothesis H_0: the time series is non-stationary against the alternative H_1: the time series is stationary.

The ADF test statistic is given by Equation 14 (Suleiman et al., 2015)[18]:

$$\boldsymbol{ADF_t = tau = \frac{\hat{\theta} - 1}{SE(\varphi)}} \quad (\mathbf{14})$$

Where SE ($\boldsymbol{\varphi}$) is the standard error for $\boldsymbol{\varphi}$ and $\boldsymbol{\hat{\theta}}$ the estimated $\boldsymbol{\theta}$ for the ARMA (p, q) model.

The useful rule of thumb is that the null hypothesis is accepted if the test statistic is greater than the critical value (or the p-value is less than the confidence threshold) (Suleiman et al., 2015)[18].

- ***Model Orders Determination***

The order of integration (differencing) d is set at zero when the initial time series is stationary. If not, d is the number of times the series is integrated (differenced) to get a stationary time series. The stationarity is tested for each differenced time series with the technique indicated above.

When the series is identified stationary, the next step is to determine the different values of p and q. The correlograms (ACF and PACF) analysis is the common techniques used for this purpose. Usually, PACF is used for the AR order p identification and ACF for the MA order q (Nashirahand Sofian, 2017)[13]. The rule is that the potential p and q are respectively the first p and q at which the PACF and ACF show significance spikes. Anderson (1977)[3],Rangsan and Titida (2006)[15] and Adolphus (2016)[1] have also highlighted a rule of determining p and q from ACF and PACF patterns, as showed in Table 1.

 ©IJSRED: All Rights are Reserved

TABLE I
MODEL ORDER DETERMINATION USING THE ACF AND PACF PATTERNS

Model	ACF	PACF
AR(p)	Dies down	Cut off after lag p
MA(q)	Cut off after lag q	Dies down
ARMA(p, q)	Dies down	Dies down

B. Model selection

The model selection consists of the choice of the best ARIMA model. The rule is as follows. When p, d and q are determined, the different ARIMA (i, d, j) are set for $0 \leq i \leq p$ and $0 \leq j \leq q$. A total number of $p \cdot q$ models is got. The selection of the best model is based on some statistics characterizing these models. The common statistics used for model selection are the Akaike Information Criteria (AIC), Mean Squared Error (MSE), Shwartz Bayesian Criteria (SBC) (Adolphus, W., 2016)[1], Hannan Quinn Information Criterion (HQC) (Kayode and Habib, 2013)[12] and mean absolute percentage error (MAPE) (Nashirahand Sofian, 2017)[13]. For our work the Akaike Information Criteria (AIC) and Mean Squared Error (MSE) will be used.

The AIC takes into account both how well the model fits the observed series and the number of parameters to be used in this model. Akaike (1969) has defined AIC as in Equation 15 (Reza and Ahmad, 2018)[16]:

$$AIC = T \cdot log(\hat{\sigma}^2) + 2k \qquad (15)$$

Where T is the length of the time series, $\hat{\sigma}^2$the variance of the model residual and k the number of the model parameters.

The Mean Squared Error (MSE) of the model is computed from Equation 16 (Adolphus, 2016) [1]:

$$MSE = \frac{\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}{T} \qquad (Eq. 16)$$

Where x_t is the sample observed value and $\hat{x}_t$ the estimated value.

Rule for best model selection: The model that gives the minimum AIC or MSE is considered to be the best one (Kayodeand Habib, 2013)[12].

C. Model Estimation

The model estimation consists of computing the parameters φ_i and $\theta_j$of the model. Two techniques are commonly used: the Ordinary Least Square (OLS) and Maximum Likelihood (ML) methods. Alexanderand Sebastian (2014)[2] have stated the fact that OLS method helps in getting the parameters that minimize the square distance between the estimated points and the observed ones, called Sum of Squared Error (SSE). As far as the estimator of the exact maximum likelihood is concerned, it gives the parameters which maximize the log-likelihood function (Kayode and Habib, 2013)[12]. Our model estimation will rely on the OLS method. Some statistical tests on the significance of each parameters can also be performed.

D. Model Validation

After the modelestimation, the Box-Jenkins model building strategy entails a test of the adequacy of the model (Kayode and Habib, 2013)[12].In the test for residual, there are various test including Portmanteus test, Lagrange multiplier, autocorrelation and partial autocorrelations of residual analysis, normality test, Jarque-Bera test, and Ljung–Box test, etc. (Suleiman et al., 2015)[18]. Our model validation will rely on the analysisof residual autocorrelation and partial autocorrelation functions.

The residual autocorrelation analysis consists of assessing the sample autocorrelation function (ACF) and partial autocorrelation function(PACF) of the residual.One can conclude that the model is adequate if there are no significant spikes in the ACF and PACF (Kayode and Habib, 2013)[12].

2.2.2. *Forecasting with ARIMA Model*

This stage involves getting the precise estimates of oil prices from the validated ARIMA model from the time range of the data set. These estimates are obtained by computing the oil price from the model equation. For a good assessment of the forecasting, the forecasts and the observed values of the times series are plotted on the same graph and the

 ©IJSRED: All Rights are Reserved

absolute error on each estimation are computed in a table.

2.2.3. Monte Carlo Simulation

Developed in 1949 by the American mathematicians John Von Neumann and Stanislav Ulam, Monte Carlo Simulation is a series of techniques used to solve complex problems but very often deterministic by introducing random sampling (Baissa, K., 2012)[6]. The term ''Simulation'' is related to random variables simulation, that is, the generation a huge number thereof.

The use of Monte Carlo Simulation in our case study rely on the mathematical theorem called **Huge Numbers Law**. This law states that for an independent sample X_1, X_2, …, X_n of a random variable X, the empirical average $\bar{X}$ tends to the mathematic expectancy $E(X)$ of X when n is huge (Christian, 2015)[9]. That is, (Equation 17):

$$\boldsymbol{E(X) = \lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n}} \qquad (17)$$

In case of Nigerian forcados crude oil price forecasting, the application of Monte Carlo Simulation will take into account the variability of the price and make the results much more accurate than those of an ordinary ARIMA. This will be performed in two steps: (1) probability distribution modelling and (2) oil price simulation and forecasting.

A. Probability Distribution Modelling

For the simulation of a random time series X, the determination of its probability distribution function (PDF) and cumulative density function (CDF) is needed. The first task is to determine these statistical functions of the time series. A time series can follow one of existing distribution functions. Frequency analysis, CDF analysis coupled with a statistical test of distribution (Chi-square test for instance) is used for probability distribution determination. Frequency analysis and CDF analysis consist respectively of plotting the empirical frequency and CDF and checking which theoretical PDF and CDF comes closest the empirical one. Thereafter, Chi-square test can be carried out to check whether the frequency or CDF analysis results are good, if needed.

For the purposes of our study, the triangular distribution will be assumed for Nigerian forcados crude oil price. According to Wikipedia, a random variable X is said to follow a triangular law of parameters a, b, c $\left(noted\ X\sim T(a,b,c)\right)$, with $a \le c \le b$, if its probability distribution function f is the one of the Equation 18:

$$\boldsymbol{f(t)} = \begin{cases} \dfrac{\boldsymbol{2(t-a)}}{\boldsymbol{(b-a)(c-a)}}, & a \le t \le c \\ \dfrac{\boldsymbol{2(b-t)}}{\boldsymbol{(b-a)(b-c)}}, & c \le t \le b \end{cases} \qquad (\mathbf{18})$$

a and b are respectively the minimum and maximum of the possible values of X and c the most expected value.

In case of our study a and b are the minimum and maximum of the time series $\left(a = min(X_t),\ b = max(X_t)\right)$ and the most expected value will be the forecasted value from the ARIMA model, $\hat{x}_{ARIMA}(t)$. For a given forecast $\hat{x}_{ARIMA}(t)$, the corresponding triangular random variable will be noted $T_{\hat{x}_{ARIMA,t}}$.

The cumulative distribution function F of a triangular law of parameters a, b, c is defined by Equation 19 (Fred, 2018)[10]:

$$\boldsymbol{F(t)} = \begin{cases} \dfrac{\boldsymbol{(t-a)^2}}{\boldsymbol{(b-a)(c-a)}}, & a \le t \le c \\ \boldsymbol{1 - \dfrac{(b-t)^2}{(b-a)(b-c)}}, & c \le t \le b \end{cases} \qquad (\mathbf{19})$$

For a given forecast $\hat{x}_{ARIMA}(t)$, the CDF of the random variable $T_{\hat{x}_{ARIMA,t}}$ will be noted $F_{\hat{x}_{ARIMA,t}}$.

B. Oil Price Simulation and Forecasting

- ***Oil Price Simulation***

The simulation of oil price will result in generating a huge number (thousands or tens of thousands even hundreds of millions) of its random variable T. Since some random number generators are not able to generate other than uniform law, the simulation of T will be carried out by CDF inversion method. This relies on the following theorem: If U is a uniform random variable on $[0, 1]$, and F an inversible continuous function with

 ©IJSRED: All Rights are Reserved

value in $[0, 1]$, then $F^{-1}(U)$ is the random variable whose CDF is F (Aude, 2015)[5].

On the basis of this theorem, for a given forecasted value $\hat{x}_{ARIMA}(t)$, simulating an n-size sample consists of generating randomly n values $u_1, u_2, \cdots, u_n$of a uniform law U on $[0, 1]$.
Then: ${F_{\hat{x}_{ARIMA,t}}}^{-1}(u_1)$, ${F_{\hat{x}_{ARIMA,t}}}^{-1}(u_2)$, …, ${F_{\hat{x}_{ARIMA,t}}}^{-1}(u_n)$ is the simulated sample of $T_{\hat{x}_{ARIMA,t}}$.

- ***Oil Price Forecasting for Model Validation***

By application of the Huge Number Law, the forecasted value $\hat{x}_{MS-ARIMA}(t)$ of oil price at t with our proposed model, the Monte Carlo Simulation to Autoregressive Moving Average (MS-ARIMA) model, is given by Equation 20 as follows:

$$\hat{x}_{MS-ARIMA}(t) = \frac{1}{n}\left[{F_{\hat{x}_{ARIMA,t}}}^{-1}(u_1) + {F_{\hat{x}_{ARIMA,t}}}^{-1}(u_2) + \cdots + {F_{\hat{x}_{ARIMA,t}}}^{-1}(u_n)\right] \quad (20)$$

The algorithm of the forecasting process from MS-ARIMA (p, d, q) model is as follow:

Algorithm:

Step 1: For a given t, forecast the value of $\hat{x}_{ARIMA,t}$ from the ARIMA (p, d, q).

Step 2: Generate a huge number of random values $u_1, u_2, \cdots, u_n$of a uniform law U on $[0, 1]$.

Step 3: Calculate the images ${F_{\hat{x}_{ARIMA,t}}}^{-1}(u_1)$, ${F_{\hat{x}_{ARIMA,t}}}^{-1}(u_2)$, …, ${F_{\hat{x}_{ARIMA,t}}}^{-1}(u_n)$ of $u_1, u_2, \cdots, u_n$ from the CDF inverse of the variable $T(\min(X), \max(X), \hat{x}_{ARIMA,t})$.

Step 4: Calculate$\hat{x}_{MS-ARIMA}(t)$ as follows.

$$\hat{x}_{MS-ARIMA}(t) = \frac{1}{n}\left[{F_{\hat{x}_{ARIMA,t}}}^{-1}(u_1) + + {F_{\hat{x}_{ARIMA,t}}}^{-1}(u_2) + \cdots + {F_{\hat{x}_{ARIMA,t}}}^{-1}(u_n)\right]$$

$\hat{x}_{MS-ARIMA}(t)$ is the estimated oil price from MS-ARIMA model.

Some statistics will also be computed: P10, P50 (median) and P90 for the forecast $\hat{x}_{MS-ARIMA}(t)$. By definition, P10, P50 and P90 are the values of $T_{\hat{x}_{ARIMA,t}}$ for which any other value has respectively 10, 50 and 90% of chance to be higher than it, that is:

$P(T_{\hat{x}_{ARIMA,t}} \geq P10) = 0.1$;
$P(T_{\hat{x}_{ARIMA,t}} \geq P50) = 0.5$ *and*
$P(T_{\hat{x}_{ARIMA,t}} \geq P90) = 0.9$.

Therefore, one has: $P10 = G^{-1}(0.1)$; $P50 = G^{-1}(0.5)$ *and* $P90 = G^{-1}(0.9)$, with $G(x) = 1 - F_{\hat{x}_{ARIMA,t}}(x)$.

2.2.4. Comparison of MS-ARIMA and Ordinary ARIMA

At this stage the oil prices will be forecasted for t between 2 and T with both models (ARIMA and MS-ARIMA). The forecasts and the observed values of the times series as well as the relative errors will be presented in a table and plotted on the same graph.

3. RESULTS AND DICUSSION

3.1. ARIMA Model

The case study is made on the time series of Nigerian forcados crude oil annum average price from 2000 to 2019. The boxplot of Fig. 1 shows the main statistics of the data set. Its minimum value, average and maximum value are respectively US$ 24.23, US$ 65.97 and US$ 114.21. The first quartile, the median and third quartile are respectively US$ 42.94, US$ 64.20 and US$ 86.12 while the standard deviation is US$ 30.26.

A. Stationarity Test

The pattern of Nigerian forcados crude oil price is shown in Fig. 1. It can be clearly seen that the oil price does not really oscillate around the average of the data set, represented by the horizontal blue straight line. As a result, the time series can be said non-stationary. This statement needs to be confirmed by the correlogram analysis and, if needed, the Unit Root test (ADF test).

The autocorrelation function (ACF) graph of the time series is shown in Fig. 2. The ACF goes down fairly quickly when the lag increases, showing that the time series is non-stationary.

We need to integrate (difference) the time series to get a stationarity. Fig. 2 shows the configuration of the Autocorrelation function of the time series first (1st) difference. It fairly dies down from the lag 1. Therefore, this time series is stationary. However, it is necessary to confirm this through an Augmented Ducky-Fuller (ADF) test.

 ©IJSRED: All Rights are Reserved

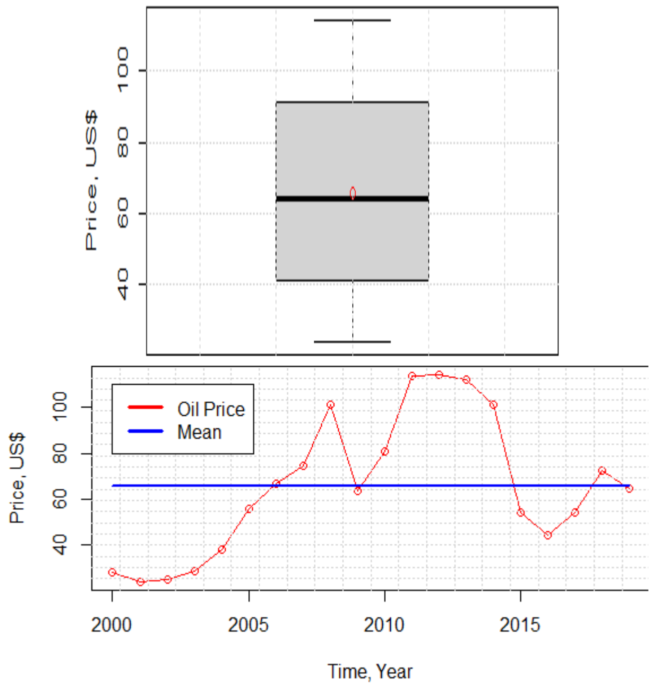

Fig. 1 Top: Nigerian forcados crude oil price boxplot. Bottomt: Nigerian forcados crude oilprice pattern

The results of the Augmented Ducky-Fuller test performed on the time series 1st difference are summarized in Table 2. The p-value of the test statistic is higher than the confidence threshold of 5%. The null hypothesis is then rejected and the alternative one is accepted. Therefore, the 1st difference time series is stationary.

TABLE III
ADF TEST RESULTS FOR THE 1ST DIFFERENCE TIME SERIES

Null Hypothesis H0	Alternative Hypothesis H1	ADF Test Statistic	Lag Order	p-Value
The time series is not stationary	The time series is stationary	-2.679	2	0.3137

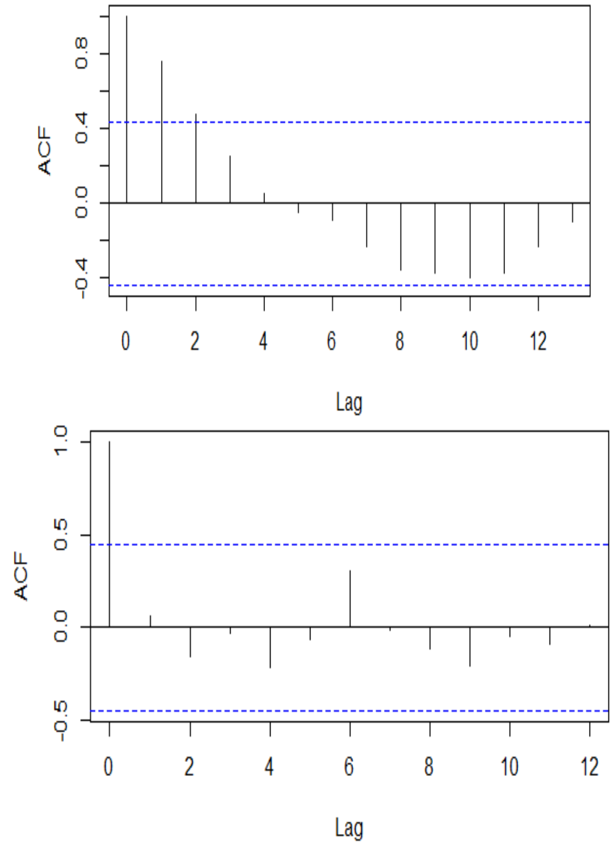

Fig. 2 Top: Nigerian forcados crude oilprice ACF plot. Bottom: ACF plot of 1st difference

B. Model Orders Determination

The first difference of the Nigerian forcados crude oil price time series is stationary. As a result, the integration (differencing) order of the ARIMA model is d = 1.

As highlighted in the methodology section, the Autoregressive (AR) order p will come from the first difference Partial Autocorrelation Function graph interpretation (Fig. 4). The first lag at which a significant spike is gotten on the PACF plot is the lag 2, which sets the AR order at p = 2.

On the first difference ACF graph, the significant spikes are observed at the lag 0 (Fig. 2). The Moving Average (MA) order q must be equal to 0. The potential ARIMA models for Nigerian forcados crude oil price forecasting are ARIMA (1, 1, 0) and ARIMA (2, 1, 0).

 ©IJSRED: All Rights are Reserved

C. *Model Selection*

From the previous section results, two models are candidates to a selection: ARIMA (1, 1, 0), and ARIMA (2, 1, 0). Table 3 shows the Akaike Information Criterion (AIC) of these models.
ARIMA (1, 1, 0) is the best model since it has the lowest AIC.

TABLE IIIII
AIC OF THE MODELS FOR SELECTION

Model ARIMA	(1, 1, 0)	(2, 1, 0)
AIC	170.25	171.83

D. *Model Estimation*

Table 4 summarizes the results of the model parameters estimation. The model equation is the Equation 21:

$$\Delta^1 X_t = 0.0696\, \Delta^1 X_{t-1} + \varepsilon_t \quad (21)$$

After transformation ($\Delta^1 X_t = X_{t-1} - X_{t-2}$), we got the Equation 22:

$$X_t = (1.0696) X_{t-1} - (0.0696) X_{t-2} + \varepsilon_t \quad (22)$$

TABLE IVV
AIRMA (1, 1, 0) MODEL CHARACTERISTICS

AIRMA (1, 1, 0) model characteristics			
Coefficients		**Log Likelihood**	**AIC**
Phi 0	Phi 1	-83.12	170.25
0	0.0696		

E. *Model Validation*

The Autocorrelation function of the model residuals shows no significant spike (Fig. 3). Also the shape of the model histogram indicates a centered normal law. We can then conclude that the model error is a white noise. Therefore, the model satisfies the validation requirements.

F. *Forecasting for ARIMA Model Efficiency*

Table 5 shows the forecasted values of the model and the relative errors compared to the observed values of the time series. The mean square error of the model is 388.92. Fig. 4 shows a graphical comparison of forecasted values to the observed ones.

3.2. *Monte Carlo Simulation*

A. *Probability Distribution Modelling*

For our case, as stated in the methodology about forecasting with our proposed model called MS-ARIMA (1, 1, 0), a triangular distribution is used for the simulation. For each forecasted value $\hat{x}_{ARIMA,t}$ from ARIMA (1, 1, 0) the parameters of the associated triangular distribution $T_{\hat{x}_{ARIMA,t}}$ are a a = 24.23, b = 114.21 and model c = $\hat{x}_{ARIMA,t}$, that is, $T_{\hat{x}_{ARIMA,t}} \sim T\left(24.23,\ 114.21,\ \hat{x}_{ARIMA,t}\right)$.

TABLE V
AIRMA (1, 1, 0) MODEL EFFICIENCY

Years	2000	2001	2002	2003	2004
Observed (US$)	28.41	24.23	25.04	28.66	38.13
Forecasts (US$)			23.94	25.10	28.91
Relative Error (%)			4.39	12.43	24.18
Years	**2005**	**2006**	**2007**	**2008**	**2009**
Observed (US$)	55.69	67.07	74.48	101.43	63.45
Forecasts (US$)	38.79	56.91	67.86	75.00	103.31
Relative Error (%)	30.35	15.15	8.89	26.06	62.81
Years	**2010**	**2011**	**2012**	**2013**	**2014**
Observed (US$)	81.05	113.65	114.21	111.95	101.35
Forecasts (US$)	60.81	82.27	115.92	114.25	111.79
Relative Error (%)	24.98	27.61	1.50	2.05	10.30
Years	**2015**	**2016**	**2017**	**2018**	**2019**
Observed (US$)	54.41	44.54	54.31	72.47	64.95
Forecasts (US$)	100.61	51.14	43.85	54.99	73.73
Relative Error (%)	84.91	14.82	19.25	24.12	13.52

B. *Oil Price Forecasting for MS-ARIMA Model Efficiency*

As stipulated in the forecasting algorithm of the methodology section, the estimated values of oil price from ARIMA (1, 1, 0) model are used as most expected value (parameter c) of the triangular distributions. 100,000,000 random numbers from a uniform law on [0, 1] are generated. This huge number is used for convergence purposes. The

 ©IJSRED: All Rights are Reserved

average of their images from the CDF inverse of the triangular random variable are computed.

The results are summarized in Table 6. The Mean Squared Error of the model is 187.79.

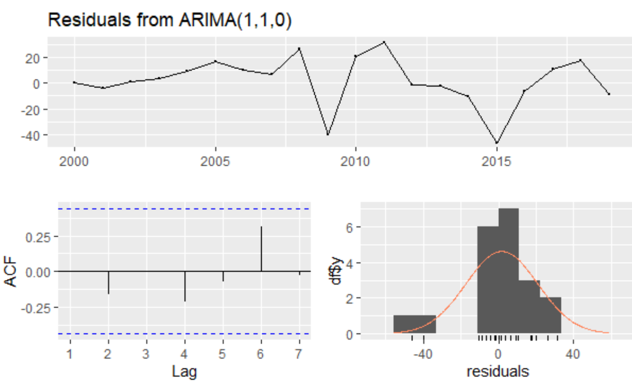

Fig. 3Model validationgraphics

3.3. Comparison of MS-ARIMA and Ordinary ARIMA

Table 7 shows the relative errors of both models. The Mean Square Errors (MSE) of ARIMA (1, 1, 0) and MS-ARIMA (1, 1, 0) are respectively 388.92 and 187.79. The MSE of ARIMA (1, 1, 0) is more than twice the one ofMS-ARIMA (1, 1, 0). Therefore, MS-ARIMA (1, 1, 0) is much more efficient than ARIMA (1, 1, 0). Fig. 5 shows graphically this difference of efficiency.

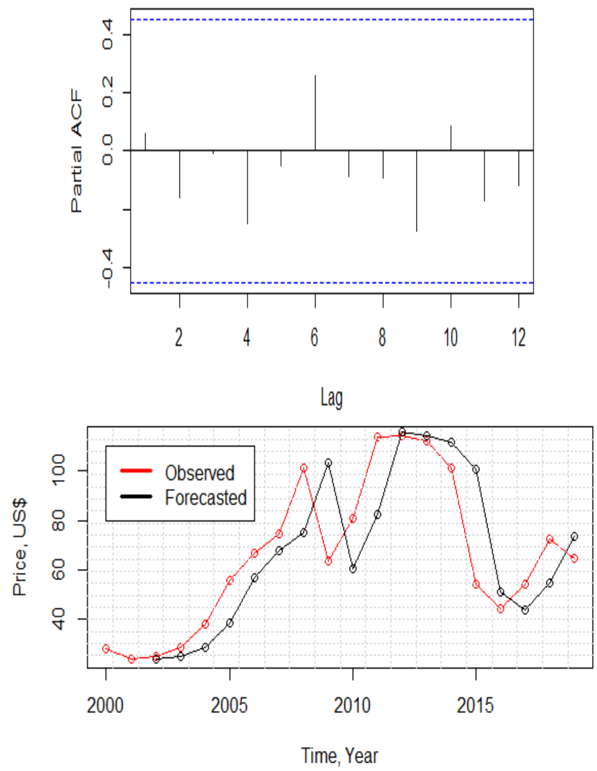

Fig. 4 Top: PACF plot of the first difference. Bottom: Observed and forecasted oil prices

3.4. Oil Future Price Forecasting with MS-ARIMA

Contrarily to the ordinary ARIMA model which gives single values, MS-ARIMA provides a range of values for oil price gotten from 100,000,000-size simulated samples. These ranges are calculated for a 10-year period of time (2020 to 2029). The oil price values for which the expected values have respectively 10, 50 and 90% chance of being higher than the expected (herein called P10, P50 and P90) are also computed for each year.

The results of the forecasting from MS-ARIMA show that the oil prices have high chances to be between $US 64.32 and $US 69.57 from 2020 to 2029. The P10, P50 and P90 are respectively $US 81.28, $US 66.75 and $US 43.25.

 ©IJSRED: All Rights are Reserved

TABLE VI
MS-AIRMA (1, 1, 0) MODEL EFFICIENCY

Years	**2000**	**2001**	**2002**	**2003**	**2004**
Observed (US$)	28.41	24.23	25.04	28.66	38.13
Forecasts (US$)			27.64	30.11	37.91
Relative Error (%)			10.38	5.06	0.58
Years	**2005**	**2006**	**2007**	**2008**	**2009**
Observed (US$)	55.69	67.07	74.48	101.43	63.45
Forecasts (US$)	59.28	66.34	71.16	74.61	80.46
Relative Error (%)	6.45	1.09	4.46	26.44	26.81
Years	**2010**	**2011**	**2012**	**2013**	**2014**
Observed (US$)	81.05	113.65	114.21	111.95	101.35
Forecasts (US$)	67.99	78.47	105.94	112.96	104.37
Relative Error (%)	16.11	30.95	7.24	0.90	2.98
Years	**2015**	**2016**	**2017**	**2018**	**2019**
Observed (US$)	54.41	44.54	54.31	72.47	64.95
Forecasts (US$)	79.49	57.03	52.24	65.54	68.72
Relative Error (%)	46.09	28.04	5.01	9.56	5.80

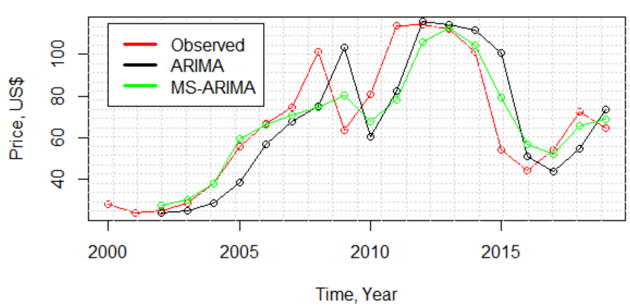

Fig. 5Comparison of forecasted oil prices for both models

TABLE VII
ARIMA (1, 1, 0) AND MS-AIRMA (1, 1, 0) RELATIVE ERRORS

Years	**2000**	**2001**	**2002**	**2003**	**2004**
ARIMA Error (%)			4,39	12,43	24,18
MS-ARIMA Error (%)			10,38	5,06	0,58
Years	**2005**	**2006**	**2007**	**2008**	**2009**
ARIMA Error (%)	30,35	15,15	8,89	26,06	62,81
MS-ARIMA Error (%)	6,45	1,09	4,46	26,44	26,81
Years	**2010**	**2011**	**2012**	**2013**	**2014**
ARIMA Error (%)	24,98	27,61	1,50	2,05	10,30
MS-ARIMA Error (%)	16,11	30,95	7,24	0,90	2,98
Years	**2015**	**2016**	**2017**	**2018**	**2019**
ARIMA Error (%)	84,91	14,82	19,25	24,12	13,52
MS-ARIMA Error (%)	46,09	28,04	5,01	9,56	5,80

4. CONCLUSIONS AND RECOMMANDATIONS

The application of Monte Carlo simulation to the Autoregressive Integrated Moving Average (MS-ARIMA) model developed in this work has proved that when the randomness of time series is taken into account in modelling and forecasting process, the model accuracy rises. The search for adequate probability distributions of the time series is the most difficult task.

Based on the MS-ARIMA case study results, it is more likely that Nigerian forcados crude oil price will be between $US64.32 and $US69.57 from 2020 to 2029. The P10, P50 and P90 are respectively $US 81.28, $US66.75 and $US43.25. That is, in the next ten years Nigerian forcados crude oil price has 10% of chance being higher than $US81.28, 50% of chance being greater than $US 66.75 and 90% of chance being over $US43.25. The MS-ARIMA ismore efficient result than the ordinary ARIMA because it permits us to take into account the randomness aspect of oil price. Monte

Carlo Simulation is also recommended for ordinary time series forecasting.

REFERENCES

[1] AdolphusWagala. ''Non Linear Time Series Modelling Of the Diesel Prices in Kenya'', International Journal of Academic Research in Economics and Management Sciences, Vol. 5, No. 4, pp. 105-125, 2016.

[2] Alexander Gustaffson and Sebastian Wogenius. Modelling Apartment Prices with the Multiple Linear Regression Model. Thesis, KTH Royal Institute of Technology, SCI School of Engineering Sciences, Sweden, 2014.

[3] Anderson, O. D. ''The Box-Jenkins approach to time series analysis''. RAIRO, Recherche Opérationnelle, tome 11 no 1, pp. 3-29, 1977.

[4] Arnold Adimabua Ojugo and Rume Elizabeth Yoro. ''Predicting Futures Price and Contract Portfolios Using the ARIMA Model: A Case of Nigeria's Bonny Light and Forcados'', Quantitative Economics and Management Studies (QEMS), Vol. 1 No. 4, pp. 237-248,2020.

[5] Aude Illig. TP 4: Simulation avec R. Université Versailles Saint-Quentin, 2015.

[6] Baissa Kenza. Simulation Statistique. Faculté Des Sciences Exactes et Sciences de la Nature et la Vie. Thesis, Université Mohamed Khider, Biskra, Algérie, 2012.

[7] Balogun Oluwafemi Samson and Ogunleye Opeoluwa Mercy. ''On the Time Series Modelling of Crude Oil Exportation in Nigeria'', European Journal of Academic Essays 2(3), pp. 1-11, 2015.

[8] (2020).British Petroleum homepage on Statistical Review of World Energy. [Online]. Available:https://ourworldindata.org/grapher/crude-oil-spot-prices?time=1972..latest.

[9] Christian Léonard. Notes de cours de statistique mathématique élémentaire. Département de mathématiques et informatique. Université Paris Ouest Nanterre, 2015.

[10] (2021) Fred Schenkelberg Website. [Online]. Available:https://accendoreliability.com.

[11] GarzaliMuhammed and Bashir Umar Faruk. (2018): The relevance of GARCH-family models in forecasting Nigerian oil price volatility. ResearchGate, CBN Bullion, Vol 42, no 2.

[12] Kayode Ayinde and Habib Abdulwahab. ''Modeling and Forecasting Nigerian Crude Oil Exportation: Seasonal Autoregressive Integrated Moving Average Approach'', International Journal of Science and Research (IJSR), Volume 2 Issue 12, pp. 245-249, Dec 2013.

[13] NashirahAbu Bakar and SofianRosbi. ''Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction'', International Journal of Advanced Engineering Research and Science (IJAERS), Vol-4, Issue-11, pp. 130-137, Nov 2017.

[14] Omekara C.O., Okereke O. E, Ire K. I. and OkamgbaC. O. ''ARIMA Modeling of Nigeria Crude Oil Production'', Journal of Energy Technologies and Policy, Vol.5, No.10, pp. 1-5.

[15] RangsanNochai and Titida Nochai. ''Arima Model for Forecasting Oil Palm Price'' in 2nd IMT-GT Regional Conference on Mathematics, Statistics and Application, 2006.

[16] Reza Narimani and Ahmad Narimani. ''A New Hybrid Model for Improvement of ARIMA by DEA''. Decision Science Letter 1, pp. 59-68, 2012.

[17] Selvi J. Jaya, R. Kaviya Shree and J.Krishnan. ''Forecasting Crude Oil Price Using ARIMA Models'', International Journal of Advanced Research in Science and Engineering, Volume No 07, Special Issue No 05, pp. 334-343, Mar 2018.

[18] Suleiman S., Alabi M. A., Issa Suleman, Usman U. and Umar Adamu. ''Modeling and Forecasting the Crude Oil Price in Nigeria'', International Journal of Novel Research in Marketing Management and Economics, Vol. 2, Issue 1, pp. 1-13, April 2015.

[19] Wolters Jurgen and Hassler Uwe. Unit root testing. School of Business & Economics, Free University Berlin, 2005.