A Time Series Model on the Occurrence of COVID-19 Pandemic in Nigeria

ABSTRACT

Coronaviruses belong to a large family of viruses which affect the hepatic, gastrointestinal, neurological and respiratory systems. The increase in the daily number of COVID-19 confirmed and deaths cases from different countries of the world has brought social, economic and political activities to a standstill, affecting individuals, government, public and private sectors. In this study, autoregressive integrated moving average (ARIMA) time series model for modeling and forecasting daily confirmed, recovered, and deaths cases of COVID-19 in Nigeria was used with data on daily cases of confirmed, recovered and deaths due to COVID-19 in Nigeria from 27/02/2020-31/07/2020 obtained from Nigeria Centre for Disease Control (NCDC) website. The data from 27/02/2020-16/07/2020 were used for model building while 15 observations from 17/07/2020-31/07/2020 were used for training and forecast evaluations. Time plots and Dickey-Fuller Generalized Least Squares unit root test were used to investigate the stationarity properties of the data. Schwarz Information Criterion (SIC) in conjunction with log likelihood were used to search for optimal ARIMA models while Mean Absolute Percentage Error (MAPE) was used for forecast evaluation. Results showed that all the study variables were differenced stationary and hence integrated of order one, I(1). ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models were selected as the best candidates for modeling and forecasting the confirmed, recovered and deaths cases of COVID-19 in Nigeria respectively. The study found an approximate COVID-19 life cycle of 12 days among the infected population. The 15 days forecasts from ARIMA (2,1,4) and ARIMA (2,1,2) models showed increases in the daily number of confirmed and recovered cases of COVID-19 in Nigeria. The forecasts from ARIMA (2,1,3) model however showed fluctuating trend with decline in the number of deaths cases due to the disease. The result of the study further showed that improving on the present approach to treatment will further decrease the number of casualties due to COVID-19 in Nigeria.

1. INTRODUCTION

The family of Coronaviridae comprised two other sub-families of Coronavirinae and Torovirinae. The five members of the genera are the Alpha-coronavirus, Beta-coronavirus, Gamma-coronavirus, Torovirus and Bafinivirus. These large families of viruses affect the gastrointestinal, hepatic and neurological respiratory systems and can easily be found among animals like bats, mice, livestock, birds and other animals as well as humans [1 & 2]. A famous type of virus in the Coronaviridae family is the Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) which is distributed from one animal to another and to humans. Middle East Respiratory Syndrome coronavirus (MERS-CoV) is another type of coronavirus that is only distributed from one human being to another [3].

In December, 2019 many cases of respiratory diseases originating from a seafood market in Wuhan, China were reported to World Health Organization (WHO). The virus responsible for these respiratory problems was later termed coronavirus while the disease was coined Coronavirus Disease 2019 (COVID-19). The coronavirus belongs to the Beta-coronavirus genera of the Coronaviridae family and spread from human to human [4]. The Centre for Disease Control and Prevention (CDC) has verified that COVID-19 virus is distributed from human to human and can spread by air, close contact, by touching surfaces, or

through objects that are contaminated with viral particles. COVID-19 has an incubation period of at least 14 days. An infected person can transmit the disease to others during the incubation period [5].

People that are infected with COVID-19 experience mild to moderate to severe respiratory problems and most of them regain their normal health without requiring special medical attention. The older age group and people with underlying illnesses such as diabetes, asthma cardiovascular diseases, tuberculosis, chronic respiratory diseases as well as cancer are more prone to develop serious medical complications due to COVID-19 infection.

The mode of transmission of COVID-19 virus is primarily through the discharges from the nose of an infected individual when he/she coughs or sneezes or through the droplets of saliva of an infected person. Practicing respiratory etiquette like coughing into a flexed elbow, washing hands with running water and soap/detergent, frequently usage of alcohol based sanitizer, not rubbing/touching your hands on the face as well as practicing social distancing are some of the ways to prevent the transmission of the virus. Having informed knowledge about COVID-19 virus and understanding how the disease is caused and spread will also prevent one from getting infected with the virus. Currently, there is no specific vaccine or medical treatment for COVID-19 infection. However, many clinical and traditional trials evaluation potential treatments are ongoing around the world.

According to European Centre for Disease Prevention and Control (ECDC) COVID-19 situation update worldwide, as of 31st July, 2020, there were over 17 million confirmed cases of COVID-19 and over 600,000 deaths due to coronavirus disease globally. The disease has spread to about 214 countries and two international conveyances in the world and there is every tendency that the disease will reach all regions and countries of the world. In Africa alone, as at 31st July, 2020, over 900,000 confirmed cases and over 19 000 deaths due to COVID-19 were reported. The five countries in Africa with high reported cases of COVID-19 were South Africa with over 500,000 confirmed cases, Egypt with over 90,000 confirmed cases, Nigeria with over 43,000 confirmed cases, Ghana with over 37,000 confirmed cases and Algeria with over 31,000 confirmed cases of coronavirus disease [6].

In Nigeria, the first confirmed case of Coronavirus was reported on the 27th of February, 2020 by the Virology Laboratory of the Lagos University Teaching Hospital. The case was from an Italian citizen working in Nigeria who travelled and returned from Milan, Italy to Lagos. The second case of the virus was announced on the 9th of March, 2020 in Ewekoro of Ogun State from a Nigerian who had contact with the Italian citizen. Since then there has been a steady increase in the number of confirmed cases and deaths due to COVID-19 pandemic in Nigeria. As at 31st July, 2020, Nigeria had 43,151 confirmed cases 19,565 recovered cases and 879 deaths due to coronavirus disease [7]. The summary of COVID-19 cases across the globe, regions, and countries of the world including Nigeria are reported in Table 1.

Table 1: Summary of Confirmed and Death Cases due to COVID-19 across the Globe as at 31/07/2020

Country/Region	Total Confirmed Cases	Total Deaths
Nigeria	43,151	879
Africa	926,917	19,645
Asia	4,226,508	95,490
America	9,501,405	359,738
Europe	2,904,970	204,397
Oceania	18,701	228
Others	696	7
Worldwide	**17,579,197**	**679,505**

Source: www.ecdc.europa.eu

Modeling and forecasting the confirmed, recovered and death cases of COVID-19 pandemic will help the government, Nigeria Centre for Disease Control (NCDC) and other health authorities in planning to control the coronavirus disease and the utilization of health care resources. Statistical methodologies like time series models are useful statistical tools for modeling and forecasting infectious diseases and time-indexed data. Employing time series models to predict infectious diseases like COVID-19 will aid the

Nigerian government and medical professionals to prepare for the upcoming coronavirus cases and be more ready in the healthcare systems. Time series models are useful in producing forecasts that present early warning signals and case detections pointing to areas of unstable transmission and in areas where infection is likely to spread before large number of infection cases are detected. The models are also helpful in establishing evidence that significant amounts of transmission will occur before the onset of disease symptoms as well as determining if the approach utilized in treating COVID-19 victims is successful or not. Thus, mathematical and statistical models are proudly useful tools for making public health decisions and ensuring optimal utilization of resources in order to reduce morbidity and mortality incidences associated with infectious diseases like COVID-19 pandemic when the estimations and projections of such models are robust, accurate and reliable.

The aim of this study is to search for optimal time series models for modeling and forecasting the confirmed, recovered and deaths cases of COVID-19 in Nigeria. Specific objectives are (i) to examine the stationarity properties of the series (ii) to fit appropriate time series models to the confirmed, recovered and deaths cases of coronavirus disease data (iii) to estimate the coronavirus disease life cycle among the infected population and (iv) to produce short-term in-sample forecasts and estimate models accuracies. The rest of the paper is organize as follows: Section 2 dwells on the review of related literature, Section 3 presents materials and methods, Section 4 hinges on presentation and discussion of results while Section 5 presents the concluding remarks.

2. EMPIRICAL LITERATURE REVIEW

The outbreak of Coronavirus disease has prompted attention from health professionals, academic researchers and all sundry all over the world to put all hands on desk in conducting researches whose findings are well documented in literature. For example, [8] proposed a new model to predict COVID-19 epidemic in Italy and to plan an effective control strategy. About eight compartments were considered in the model: susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E), collectively termed SIDARTHE. The SIDARTHE model discriminated between infected individuals depending on whether they had been diagnosed and on the severity of their symptoms. The difference between the diagnosed and non-diagnosed individuals was important since the former was typically isolated and hence less likely to spread the virus. The delineation also helped in explaining the misperceptions of the case fatality rate and of the epidemic spread. The simulated results were compared with real data on the COVID-19 epidemic in Italy. The results demonstrated that restrictive social-distancing measures was needed to be combined with widespread testing and contact tracing to end the ongoing COVID-19 pandemic in Italy.

Yonar *et al*. [9] modeled and forecasted the confirmed number of COVID-19 epidemic cases of the G8 Countries: Germany, United Kingdom, France, Italy, Russian, Canada, Japan, and Turkey between 1/22/2020 and 3/22/2020 using curve estimation models, Box-Jenkins (ARIMA) and Brown/Holt linear exponential smoothing methods. Results showed that Holt linear exponential smoothing model was more suitable in modeling COVID-19 infection cases in Japan, UK, Canada and Italy while ARIMA (1,4,0), ARIMA (0,1,3) and ARIMA (1,4,0) models were more suitable in modeling and forecasting COVID-19 index cases in Germany, France and Turkey respectively.

Liu *et al*. [10] extracted real-time data on the susceptible, the exposed, the infectious and the recovered of COVID-19 and the population mobility data and employed ARIMA time series model with neural network models (NNs) to model the disease trends in Wuhan, Beijing, Shanghai and Guangzhou in China. The result showed that the number of infected persons would increase by 45% while the number of deaths would increase by 56% in the study areas. Pontoh *et al*. [11] modeled the index cases of COVID-19 in South Korea using nonparametric time series Neural Networks models. Among the neural networks models, Multilayer Perception (MLP) with two hidden layers outperformed with the lowest RMSE and MAE values and was the best in forecasting the confirmed, recovered, and death cases due to COVID-19 in South Korea.

Dehesh *et al*.[12] employed autoregressive integrated moving average (ARIMA) time series models to study and predict the daily confirmed cases of COVID-19 in countries with high number of confirmed

cases in the world. The daily data on confirmed cases of COVID-19 were collected from the official website of Johns Hopkins University for the period of January, 22nd, 2020 to March, 1st, 2020. The study found ARIMA (2,1,0) suitable for Mainland China, ARIMA (2,2,2) for Italy, ARIMA (1,0,0) for South Korea, ARIMA (2,3,0) for Iran, and ARIMA (3,1,0) for Thailand. Mainland China and Thailand had stable trends while Iran and Italy had unstable trends. The trend of South Korea was decreasing with tendency of becoming stable in the near future. Tran *et al*. [13] utilized ARIMA time series prediction model to study the daily total confirmed cases, total deaths, growth rate in confirmed cases and growth rate in deaths. The model used SARS-CoV-2 daily data collected from the official website of the European Centre for Disease Prevention and Control for the period of February 20 to May 04, 2020. The result of the ARIMA model showed that Iran exhibited an increase in the daily total confirmed cases and the total deaths, while the daily total confirmed new cases, total new deaths, and growth rate in confirmed cases/deaths became stable. The study predicted that Iran could control the SARS-CoV-2 disease in the near future. The ARIMA model could aid the forecasting of patients and render a better preparedness plan in Iran.

Chintalapudi *et al*. [14] conducted a study in Italy using COVID-19 infected patient data collected from the Italian Health Ministry website on registered and recovered cases from mid-February to end of March, 2020. The study employed SARIMA time series model for forecasting. The predictions from the SARIMA model showed that the infected patients could reach the value of 182,757 with 93.75% accuracy while the recovered cases could reach the value of 81,635 with 84.4% accuracy at end of May, 2020. The findings of the study further suggested that about 35% decrement of registered cases and 66% growth of recovered cases would be possible in the near future. Benvenuto *et al*. [15] utilized a simple econometric ARIMA model to predict the spread of COVID-2019 in Italy using daily data extracted from Johns Hopkins epidemiological data to predict the epidemiological trend of the prevalence and incidence of COVID-2019. Tandon *et al*. [16] developed a model to forecast future COVID-19 cases in India. The result of the study indicated an ascending trend for new COVID-19 cases in India. The time series model also presented an exponential increase in the number of new cases. Malekia *et al*. [17] used autoregressive time series forecasting models based on two-piece scale mixture normal distributions called TP-SMN-AR models to analyze the real world time series data of confirmed and recovered COVID-19 cases.

In Nigeria, [18] presented a stochastic model that captured the random behaviour of daily reported cases of COVID-19 infections in Nigeria for the period of March 17th, 2020 - May 17th, 2020. The new stochastic model was expressed in form of a distribution function with four parameters. The results obtained from the stochastic model established the model adequacy in explaining the random behaviour of the daily reported COVID-19 infections in Nigeria. The fitted model was also applied in describing the number of infections exceeding certain thresholds. Ibrahim and Oladipo [19] used daily data on the spread of COVID-19 from February, 27 to April, 26, 2020 and utilized ARIMA time series model to provide a ten-day forecast of the daily occurrence of COVID-19 cases in Nigeria. Result showed that ARIMA (1,1,0) model was suitable in providing the forecast. The forecast from the model showed a steep upward trend of the spread of the COVID-19 in Nigeria within the selected time period.

The above literature review points to the fact that ARIMA time series models are good statistical tools for modeling and forecasting infectious diseases and time indexed data. Autoregressive Integrated Moving Average time series models are particularly important in modeling infectious diseases because they can be used to estimate the disease life cycle among the infected individuals when the model is correctly specified.

3. MATERIALS AND METHODS

3.1 Data Source

The data utilized in this research are daily secondary data on confirmed, recovered and deaths cases of COVID-19 infection for the period of 29th February, 2020 to 16th July, 2020 making a total of 139 observations. The data from 27/02/2020-16/07/2020 were used for model building while 15 observations from 17/07/2020-31/07/2020 were used for training and forecast comparisons. The data was obtained from Nigeria Centre for Disease Control (NCDC) website: https://ncdc.gov.ng/diseases/sitreps. The daily

cases of confirmed, recovered and deaths are transformed to natural logarithms using the following formula:

$$X_t = \ln X_t \tag{3.1}$$

where X_t is the time series observation indexed by time t and ln is the natural logarithm.

3.2 Dickey-Fuller Generalized Least Squares Unit Root Test

To investigate the stationarity and unit root properties of the series under investigation, Dickey-Fuller generalized least squares unit root has been employed. Elliott et al. [20] have modified the popular Dickey-Fuller unit root test to Dickey-Fuller Generalized Least Squares (DF GLS) unit root test with efficient and improved statistical features. In terms of small sample size and power properties, the DF GLS modified unit root test always perform better. The DF GLS testing procedure is as follows:
Let X_t be a time series and let $z_t = (1, t)$, we regress $[X_1, (1 - \alpha L)X_2, ..., (1 - \alpha L)X_T]$ on $[z_1, (1 - \alpha L)z_2, ..., (1 - \alpha L)z_T]$ giving $\hat{\beta}_{GLS}$ where $\alpha = 1 + \bar{c}/T$, $u_0 = 0$ and $\bar{c} = -13.5$ in case of detrended statistic. Detrended $\hat{X}_t = X_t - z_t'\hat{\beta}_{GLS}$ is then employed in the Augmented Dickey-Fuller regression without intercept and time trend. The t −statistic on $\hat{X}_{t-1}$ is the DF-GLS statistic. For the demeaned case, the t is omitted from z_t and $\bar{c} = -7.0$.
In the GLS detrending, the series under consideration is regressed on a constant and linear time trend while utilizing the residual series in a standard Dickey-Fuller regression. In the GLS demeaning, only an intercept appears in the first stage regression while the residual series is used as the regressand in a Dickey-Fuller regression.

3.3 Time Series Models Specifications

To specify an ARIMA model which is the model framework use in this study, we first specify autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model before specifying autoregressive integrated moving average (ARIMA) model. These models are specified as follows.

3.3.1 The Autoregressive Model

A stochastic time series process $\{X_t\}$ is an autoregressive process of order p, denoted AR(p) if it satisfied the difference equation

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \tag{3.2}$$

where ε_t is a white noise and $\phi_1, \phi_2, ..., \phi_p$ are constants to be determined. The autocorrelation function of an AR(p) process can be found by multiplying equation (3.2) by X_{t-k}, taking expectation and dividing the resultant expression by γ_0, this gives the Yule-Walker equations:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, \qquad k = 1, 2, ... \tag{3.3}$$

The general solution of the above recurrence relations (3.3) is of the form

$$\rho_k = C_1 G_1^{|k|} + C_2 G_2^{|k|} + \cdots + C_p G_p^{|k|} \tag{3.4}$$

where $G_1, G_2, ..., G_p$ are the roots of $G^p - \phi_1 G^{p-1} - \phi_2 G^{p-2} - \cdots - \phi_p = 0$.
By setting $\rho_0 = 1$ for $k = 1, 2, ..., p - 1$, then $C_1, C_2, ..., C_p$ are determined. It is required that $\gamma_k \rightarrow 0$ which exhibit exponential decay as $k \rightarrow \infty$ and the roots lie inside a unit circle. That is $|G_i| < 1$ and the process is said to be stationary.

3.3.2 Moving Average Model

A time series $\{X_t\}$ which satisfies the difference equation

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{3.5}$$

where $\theta_1, \theta_2, ..., \theta_q$ are fixed constants with ε_t as white noise is called a moving average process of order q, denoted MA(q).

3.3.3 Autoregressive Moving Average Model

A stochastic time series process $\{X_t\}$ which results from a linear combination of autoregressive and moving average processes is called an Autoregressive Moving Average (ARMA) process of order p, q, denoted ARMA (p, q) if it satisfies the following difference equation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \qquad (3.6)$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are fixed constants associated with the AR terms and $\theta_1, \theta_2, \ldots, \theta_q$ are fixed constants associated with the MA terms with ε_t being a white noise. The stationarity of an ARMA (p, q) process is guaranteed if the roots of the polynomial
$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ lie outside the unit circle. Equation (3.6) can also be written as

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) X_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right) \varepsilon_t \qquad (3.7)$$

where L is the lag operator, the ϕ_i are the parameters of the autoregressive part of the model, the θ_j are the parameters of the moving average part.

3.3.4 Autoregressive Integrated Moving Average Model

If the original time series process $\{X_t\}$ is not stationary, then the first order difference process

$$X_t = \nabla X_t = X_{t-1} - X_t, \qquad \nabla = 1 - L \Rightarrow \nabla X_t = (1 - L)X_t \qquad (3.8)$$

Or the second-order difference process

$$X_t = \nabla^2 X_t = \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} \Rightarrow \nabla^2 X_t = (1 - L)(1 - L)X_t = (1 - L)^2 X_t$$
$$= (1 - L - L + L^2)X_t = (1 - 2L + L^2)X_t = X_t - 2LX_t + L^2 X_t \qquad (3.9)$$

will be stationary noting that ∇ is the difference operator and L is the lag operator. A stochastic time series process $\{X_t\}$ is said to be an autoregressive integrated moving average process, ARIMA (p, d, q) if $X_t = \nabla^d X_t$ is an ARMA (p, q) process.

Suppose that the polynomial $\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)$ in equation (3.7) has a unitary root of multiplicity d, then it can be written as:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)(1 - L)^d \qquad (3.10)$$

An ARIMA (p,d,q) process expresses this polynomial factorization property, and is given by:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \qquad (3.11)$$

3.4 Model Order Selection

We employed Schwarz information Criterion (SIC) due to [21] to select an appropriate model order that is both parsimonious and accurately model the data. The SIC is given as:

$$SIC(P) = -2\ln(L) + P\ln(T) \qquad (3.12)$$

where T is the number of observations, P is the number of free parameters to be estimated in the model and L is the likelihood function given by:

$$L = \prod_{i=0}^{n} \sqrt{\frac{1}{2\pi\sigma_i^2}} \cdot exp\left[-\sum_{i=1}^{n} \frac{(y_i - f(x))^2}{2\sigma_i^2}\right] \qquad (3.13)$$

$$\ln(L) = ln\left[\prod_{i=1}^{n} \sqrt{\frac{1}{2\pi\sigma_i^2}}\right] - \frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - f(x))^2}{\sigma_i^2} \qquad (3.14)$$

Given a range of ARIMA model orders, a model with the least SIC value and the largest log likelihood value is the best fitting model.

3.5 Breusch-Godfrey Serial Correlation Lagrange Multiplier Test

The Breusch-Godfrey serial correlation Lagrange Multiplier test is a test for autocorrelation of residuals in time series and econometric regression models. The null hypothesis of the test is that there is no serial correlation of any order up to order p in the residuals of an estimated model. The test proceeds as follows [22 & 23]
Consider a linear regression of the form

$$Y_t = \beta_1 + \beta_2 X_{t,1} + \beta_3 X_{t,2} + u_t \tag{3.15}$$

where the errors are assumed to follow an autoregressive process of order p given as

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t \tag{3.16}$$

If $\hat{u}_t$ is a set of sample residuals obtained from the fitted regression model using OLS, then

$$\hat{u}_t = \alpha_0 + \alpha_1 X_{t,1} + \alpha_2 X_{t,2} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_p \hat{u}_{t-p} + \varepsilon_t \tag{3.17}$$

If R^2 statistic is computed for the above model, then the following asymptotic approximation can be used for the distribution of the test statistic $nR^2 \sim \chi_p^2$ when the null hypothesis:

$H_0: \rho_i = 0$ for all i holds. That is, there is no serial correlation of any order up to order p. Here n is the number of data points available for the second regression for $\hat{u}_t$. $n = T - p$ where T is the total number of observations.

3.6 Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) is a statistical tool for measuring the accuracy of a forecast method. This accuracy is measured as a percentage and can be computed as the average absolute percent error for each observation minus the actual value divided by the actual value. It is given as:

$$MAPE = \frac{1}{n}\left(\sum_{t=1}^{n} \left|\frac{X_t - \hat{X}_t}{X_t}\right|\right) \tag{3.18}$$

where X_t is the actual value of the series at time t, $\hat{X}_t$ is the forecasted value of the series and n is the number of observations. The lower the value of MAPE, the better the model is able to forecast future values.

4. RESULTS AND DISCUSSION

4.1 Descriptive Statistics and Normality Measures of Time Series Variables

The summary statistics and normality measures for the confirmed, recovered and deaths cases of COVID-19 in Nigeria for the period 29th February to 31st July, 2020 are summarized in Table 2.

Table 2: Descriptive Statistics and Normality Measures of Time Series Variables

Statistic	Confirmed Cases	Recovered Cases	Death Cases
Mean	280.1364	127.3896	5.785714
Median	241.5000	79.50000	5.000000
Maximum	790.0000	829.0000	31.00000
Minimum	0.000000	0.000000	0.000000
Standard Deviation	245.8505	150.6242	5.772330
Skewness	0.301715	1.520762	1.061659
Kurtosis	1.654017	5.838761	4.276795
Jarque-Bera Statistic	13.96137	111.0691	39.38988
JB P-value	0.000930	0.000000	0.000000
No. of Observations	154	154	154

The result of summary statistics presented in Table 2 reveals that the daily means of the confirmed, recovered and deaths cases of COVID-19 in Nigeria are all positive indicating increases in the number of these cases. The results reported an average of 280 confirmed cases per day, 127 recovered cases per day and approximately 6 deaths per day. All the cases are positively skewed with high kurtosis values indicating that the occurrence amongst others were high in the first place but continuously decline over time. The non-normality of the three series is confirmed by Jarque-Bera statistics which are very high with marginal probability values. The implication is that the COVID-19 confirmed, recovered and deaths cases in Nigeria are non-Gaussian and do not follow normal distributions as evidence by asymmetry earlier observed.

4.2 Graphical Examination of the Series

The time plots of the level, transformed and first difference series for the confirmed, recovered and deaths cases of COVID-19 in Nigeria from 29th February to 31st July, 2020 are presented in Figure 1.
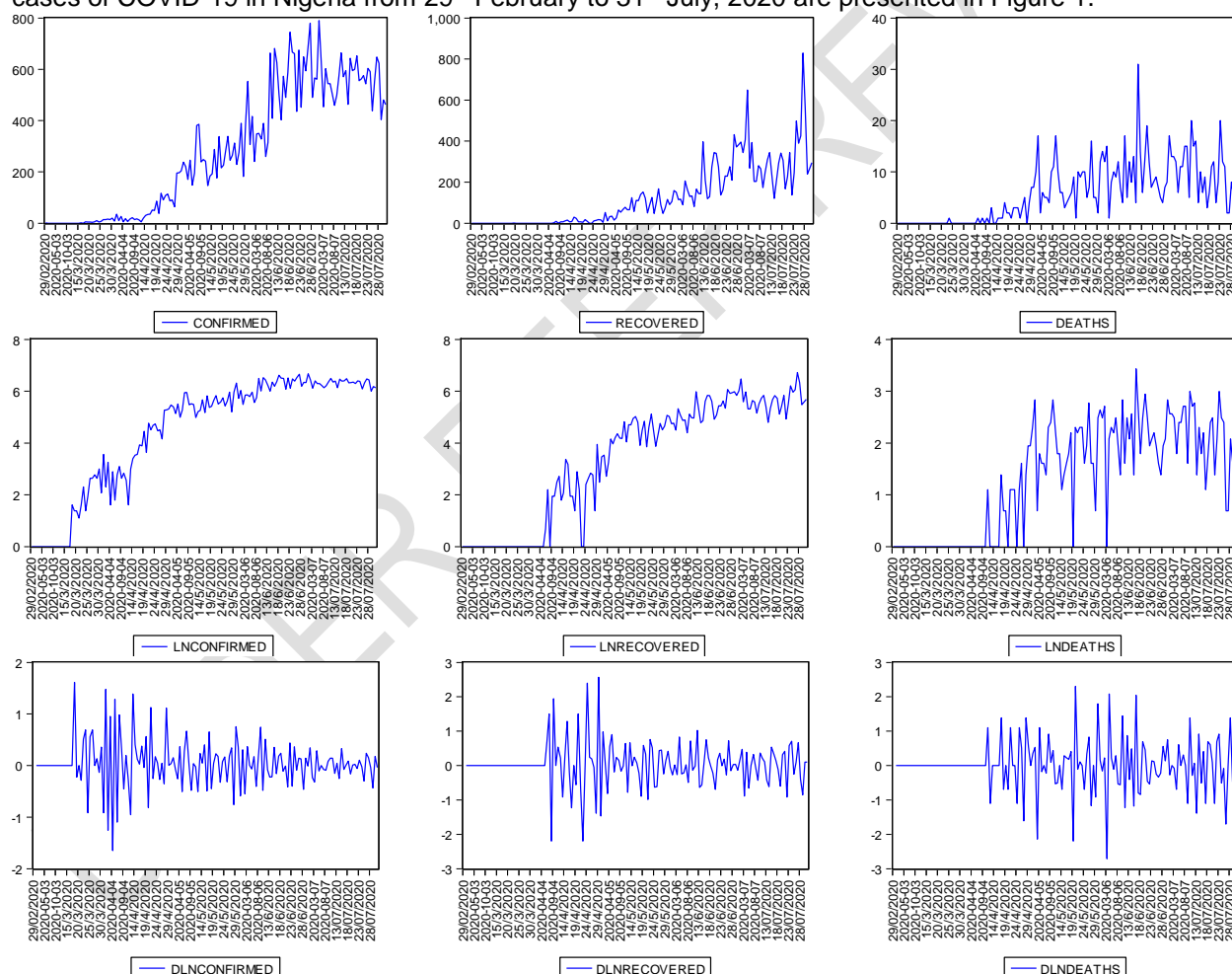
Figure 1: Time Plots of Level, Transformed and First Difference of Confirmed, Recovered and Deaths Cases of COVID-19 in Nigeria from 29/02/2020-31/07/2020.

The time plots of the confirmed, recovered and deaths cases of COVID-19 in Nigeria both in levels and transformed indicate increasing trends in the Data Generating Process (DGP) typical of non-stationary series. The trends are however eliminated from the three series by taking the first differences of the series

which showed some evidence of differenced stationary processes. The stationarity of the series under investigation are also checked using unit root test presented in Table 3.

Table 3: Dickey-Fuller Generalized Least Squares (ERS) Unit Root Test Result

Variable	Option	Level Series		First Difference Series	
		Test Statistic	5% critical value	Test Statistic	5% critical value
$\ln confirmed$	Intercept only	0.6554	-1.9432	-22.2204	-1.9432
	Intercept & trend	-1.3689	-2.9930	-22.4278	-2.9930
$\ln recovered$	Intercept only	0.1630	-1.9433	-10.0086	-1.9432
	Intercept & trend	-1.2152	-2.9980	-4.1957	-2.9980
$\ln deaths$	Intercept only	-0.1451	-1.9432	-9.2977	-1.9432
	Intercept & trend	-1.4349	-2.9930	-9.6395	-2.9960

The DF-GLS unit root test result reported in Table 3 showed that the confirmed, recovered and deaths cases of COVID-19 in Nigeria are all non-stationary in levels (contaminated with unit root). The DF-GLS result of the first differences of the series showed that the three series are covariance-stationary indicating that the confirmed, recovered and deaths cases of COVID-19 in Nigeria are all integrated of order one, I(1). All the subsequent analysis on the three series are based on the stationary series.

4.3 Results of Models Order Selection

Having determined the correct order of integration of the three series, we now search for optimal ARIMA models for fitting the confirmed, recovered and deaths cases of COVID-19 data in Nigeria. The model search is presented in Table 4. The ARIMA model with the highest log likelihood (LogL) value and the least Schwartz Information Criterion (SIC) value is the best fitting model.

Table 4: Model Order Selection

Model	Confirmed		Recovered		Deaths	
	LogL	SIC	LogL	SIC	LogL	SIC
ARIMA(1,1,0)	-70.8773	1.1716	-126.5512	1.8939	-132.2588	1.9773
ARIMA(0,1,1)	-74.5348	1.1539	-117.1345	1.7663	-114.7320	1.7090
ARIMA(1,1,1)	-70.4323	1.1900	-116.2124	1.7770	-114.3690	1.7394
ARIMA(1,1,2)	-70.0842	1.1593	-115.3520	1.7757	-114.1810	1.7599
ARIMA(2,1,1)	-70.2878	1.1542	-114.3790	1.7565	-114.0077	1.7702
ARIMA(2,1,2)	-67.7923	1.1544	**-102.4825**	**1.7244**	-113.7968	1.7905
ARIMA(2,1,3)	-65.0549	1.1673	-109.7525	1.7756	**-91.3107**	**1.4833**
ARIMA(3,1,2)	-63.2247	1.1813	-110.3100	1.7964	-112.7873	1.8123
ARIMA(3,1,1)	-70.0477	1.1831	-113.3063	1.7736	-113.1408	1.7939
ARIMA(1,1,3)	-70.0342	1.1660	-113.5273	1.7736	-106.3636	1.6691
ARIMA(3,1,3)	-63.9856	1.1659	-111.1214	1.8112	-112.6368	1.8336
ARIMA(3,1,4)	-66.2496	1.1658	-109.2366	1.8068	-114.5166	1.9872
ARIMA(1,1,4)	-70.8277	1.2019	-114.3057	1.8483	-113.4314	1.7955
ARIMA(2,1,4)	**-61.5947**	**1.1518**	-109.0631	1.7677	-106.5820	1.7313
ARIMA(4,1,2)	-65.2617	1.1934	-106.0785	1.7493	-112.5455	1.9356
ARIMA(4,1,3)	-62.8942	1.1946	-106.0712	1.7729	-108.6290	1.9137
ARIMA(4,1,4)	-63.1249	1.2047	-108.7529	1.7470	-107.4682	1.9329

From the list of 17 ARIMA models presented in Table 4, the best candidate to model the confirmed cases of COVID-19 in Nigeria is the ARIMA (2,1,4) model with the highest LogL value and least SIC value. Similarly, ARIMA (2,1,2) and ARIMA (2,1,3) models are the best candidates for fitting the recovered and deaths cases of COVID-19 in Nigeria.

4.4 Parameter Estimates of ARIMA Models

Having selected ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models as the best candidates for fitting the confirmed, the recovered and deaths cases of COVID-19 in Nigeria, the parameter estimates of the models are presented in Table 5.

Table 5: Parameter Estimates of the ARIMA Models

Variable	Coefficient	Std. Error	t-Statistic	P-value
ARIMA (2,1,4) Model for the Confirmed Cases				
AR(1)	1.467979	0.079253	18.52259	0.0000
AR(2)	-0.812012	0.069281	-11.72059	0.0000
MA(1)	-2.193614	0.112629	-19.47638	0.0000
MA(2)	2.080965	0.225543	9.226461	0.0000
MA(3)	-0.950848	0.211976	-4.485637	0.0000
MA(4)	0.269227	0.097438	2.763060	0.0066
F-statistic	**P-value**	**LogL**	**SIC**	**Durbin Watson**
14.24217	0.00000	-61.594671	1.151762	2.036334
Inverted AR Roots	0.73+0.52i	0.73-0.52i		
Inverted MA Roots	0.85-0.49i	0.85+0.49i	0.25+0.47i	0.25-0.47i
ARIMA (2,1,2) Model for the Recovered Cases				
AR(1)	0.619465	0.192348	3.220551	0.0016
AR(2)	-0.470273	0.116916	-4.022305	0.0001
MA(1)	-1.159735	0.189433	-6.122138	0.0000
MA(2)	0.620885	0.138205	4.492495	0.0000
F-statistic	**P-value**	**LogL**	**SIC**	**Durbin Watson**
11.452163	0.000000	-102.48253	1.72436	2.039478
ARIMA (2,1,3) Model for the Deaths Cases				
AR(1)	0.124641	0.053072	2.348512	0.0204
AR(2)	-0.803687	0.052873	-15.20030	0.0000
MA(1)	-0.869524	0.065934	-13.18775	0.0000
MA(2)	0.246890	0.042468	5.813554	0.0000
MA(3)	-0.937970	0.081685	-11.48272	0.0000
F-statistic	**P-value**	**LogL**	**SIC**	**Durbin Watson**
36.61138	0.000000	-91.31069	1.48326	2.211201

From the results of ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models reported in Table 5, all the parameters of the three models are statistically significant at 1% marginal significance levels. The sum of AR and MA terms are less than unity in all the estimated models (i.e., $\sum(\phi_i + \theta_i) < 1$). This show that the estimated ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models for confirmed, recovered and deaths COVID-19 cases in Nigeria are dynamically stable, stationary and predictable.

The F-statistic measures the overall fitness of a time series model. From the results of model estimation presented in Table 5, the p-values of the F-statistics for the three models are all statistically significant at the 1% marginal significance levels indicating that the estimated models are good fits for the data. The Durbin Watson statistics are less than 3 in the estimated models. This means that there are no positive serial correlations in the ARIMA models estimated and that the estimated models are not spurious and hence not biased.

4.4.1 The Stability and Invertibility Analysis of ARIMA (2,1,2) Model

The AR/MA inverted roots of polynomials of the estimated ARIMA model help in judging whether the estimated model is stable, stationary and invertible. The roots are also used in estimating the disease life cycle among the infected population if the model is accurate and correctly specified. The AR/MA inverted roots of polynomials are presented in Table 6.

Table 6: AR/MA Roots of Polynomial for ARIMA (2,1,4) Model for Confirmed Cases

Root	Real	Imaginary	Modulus	Total

AR	Root 1	0.73	0.52	0.8993	
	Root 2	0.73	-0.52	0.8993	**1.7986**
MA	Root 1	0.85	-0.49	0.9811	
	Root 2	0.85	0.49	0.9811	
	Root 3	0.25	0.47	0.5324	
	Root 4	0.25	-0.47	0.5324	**3.0270**

From the results of the AR/MA polynomial roots of the estimated ARIMA (2,1,4) model for the confirmed infection cases, observe that all the roots lied inside a unit circle indicating that the model is dynamically stable and satisfied the stability and invertibility conditions. The sum of $AR\ roots = 1.7986$ and sum of $MA\ roots = 3.0270$ and it is estimated that $\tan\theta = y/x = 1.7986/3.0270 = 0.5942$ implying $\theta = 30.72°$. Thus, the COVID-19 infection life cycle among the infected population is computed as $360°/30.72° = 11.71875 \approx 12$ days. Hence the COVID-19 infection life cycle among the infected people is approximately 12 days.

4.4.2 Models Diagnoses

The Breusch-Godfrey serial correlation Lagrange Multiplier test has been employed to diagnose the estimated ARIMA models for serial correlation/autocorrelation, result is presented in Table 7.

Table 7: Models Diagnosis using Breusch-Godfrey Serial Correlation LM Test

	Model	F-statistic	P-value	nR^2	P-value
Confirmed	ARIMA(2,1,4)	0.596872	0.5521	1.265940	0.5310
Recovered	ARIMA(2,1,2)	0.319469	0.7271	1.243324	0.5534
Deaths	ARIMA(2,1,3)	0.526953	0.5792	1.300962	0.6273

From the LM serial correlation test result reported in Table 7, the null hypotheses of no serial correlation/autocorrelation in the residuals of the fitted ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models for confirmed, recovered and deaths COVID-19 cases in Nigeria have been accepted. The acceptance of the null hypotheses is justified by the p-values of the F-statistics and nR^2 tests being greater than the alpha value of 0.05. The implication is that the estimated ARIMA models are free from serial correlation problem and that the estimates of the models are not bias and can be used to make predictions of future values that can be relied upon for policy implementation.

4.5 Forecast and Forecast Evaluation

Since the estimated ARIMA models have passed the diagnostic tests, the models are then used to provide future in-sample forecasts for 15 days starting from 17th July, 2020 to 31st July, 2020. The forecasts for the confirmed, recovered and deaths cases of COVID-19 in Nigeria are all presented in Table 8. We also evaluated the accuracy of the predictions by employing mean absolute percentage error (MAPE) separately for the three estimated models which are also presented in Table 8.

Table 8: Forecast Evaluation and Computation of MAPE for Estimated Models

Period	Confirmed			Recovered			Deaths		
	Actual	Forecast	APE (%)	Actual	Forecast	APE (%)	Actual	Forecast	APE (%)
17/07/2020	600	603	0.50	341	248	27.27	3	3	0.00
18/07/2020	653	605	7.35	305	210	31.15	6	5	16.67
19/07/2020	556	608	9.35	167	214	28.14	11	10	9.09
20/07/2020	562	611	8.72	228	240	5.26	12	11	8.33
21/07/2020	576	615	6.77	344	264	23.26	4	4	0.00
22/07/2020	543	619	13.99	138	274	98.55	8	9	12.50
23/07/2020	604	623	3.15	246	277	12.60	20	15	25.00

24/07/2020	591	628	6.26	498	294	40.96	12	11	8.33
25/07/2020	438	632	44.29	389	309	20.57	11	10	9.09
26/07/2020	555	637	14.77	426	326	23.47	2	3	50.00
27/07/2020	648	638	1.54	829	345	58.38	2	3	50.00
28/07/2020	624	641	2.72	561	357	36.36	8	7	12.50
29/07/2020	404	645	59.65	240	359	49.58	5	5	0.00
30/07/2020	481	649	34.93	266	363	36.47	5	5	0.00
31/07/2020	462	653	41.34	295	368	24.75	1	3	200.00
Total	**8297**	**9407**	**255.33**	**5273**	**4448**	**516.77**	**110**	**104**	**401.51**
Average	**553**	**627**		**352**	**297**		**7**	**7**	
MAPE			**17.02%**			**34.45%**			**26.77%**

From the forecast results reported in Table 8, the ARIMA (2,1,4) and ARIMA (2,1,2) models for predicting the confirmed and recovered cases of COVID-19 in Nigeria presented ascending trends with exponential increases. This means that the confirmed cases and recovered cases of COVID-19 in Nigeria will continue to increase over time. This result should usher in a better preparedness plan by the Nigerian government, NCDC, individuals and other relevant stakeholders in curbing the menace of Coronavirus disease in the country. Whereas, ARIMA (2,1,3) model for predicting deaths cases due to COVID-19 in Nigeria presented a fluctuating trend showing decline in daily number of deaths cases due to the disease. The forecasts reported an average number of 627 confirmed cases of COVID-19 per day in Nigeria, 297 recovered cases per day and an average of 7 deaths due to coronavirus disease per day in Nigeria.

The mean absolute percentage error (MAPE) value of the model for the confirmed cases is 17.02%. The MAPE value of 17.02% indicates the average difference between the forecasted values and the actual values. This means that the accuracy of the estimated ARIMA (2,1,4) model for predicting COVID-19 confirmed cases is 82.80%. The MAPE value of ARIMA (2,1,2) for recovered cases is 34.45% giving accuracy rate of 65.55% whereas the MAPE value of ARIMA (2,1,3) for forecasting deaths cases is 26.77% with accuracy rate of 73.23%. The MAPE values for the three models are low with high accuracy demonstrating the acceptability and suitability of ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models for predicting the confirmed, recovered and deaths cases due to coronavirus disease in Nigeria.

5. CONCLUSION

The increase in the daily number of confirmed cases and deaths due to COVID-19 pandemic from different parts of the world has made social, economic and political activities to come to a standstill, affecting individuals, government, public and private sectors. Governments look out for the time when the curves will be flattened to un-lock the lockdown for normal activities. Due to the severe health impact of the Coronavirus disease, there is an urgent need for methods that will allow forecasting and early warning with timely case detection in areas of unstable transmission, so that preventive and control measures can be implemented effectively. The autoregressive integrated moving average (ARIMA) time series model utilized in this study is a useful tool for modeling and predicting infectious diseases with time indexed. This study attempts to search for optimal ARIMA models that will fit and predict daily cases of confirmed, recovered and deaths of COVID-19 in Nigeria. The study utilized daily data on the confirmed, recovered and deaths due to COVID-19 in Nigeria from 27/02/2020-31/07/2020 obtained from NCDC website. The data from 27/02/2020-16/07/2020 were used for model building while 15 observations from 17/07/2020-31/07/2020 were used for training and forecast evaluations. Time series plots and Dickey-Fuller Generalized Least Squares unit root test were used to investigate the stationarity properties of the data. Schwarz Information Criterion (SIC) in conjunction with log likelihood was used to search for optimal ARIMA models while Mean Absolute Percentage Error (MAPE) was used for forecast evaluation. The results of the study showed that all the study variables were differenced stationary and hence integrated of order one, I(1). The ARIMA (2,1,4), ARIMA (2,1,2) and ARIMA (2,1,3) models were selected as the best candidates for modeling and forecasting the confirmed, recovered and deaths cases of COVID-19 in Nigeria respectively. The study found an approximate COVID-19 life cycle of 12 days among the infected population. The 15 days forecast evaluation showed the MAPE value of 17.02% for ARIMA (2,1,4) model for predicting COVID-19 confirmed cases with 82.80% accuracy. The MAPE value of ARIMA (2,1,2) for recovered cases was found to be 34.45% with 65.55% forecast accuracy while the MAPE value of ARIMA

(2,1,3) for forecasting deaths cases was 26.77% with 73.23% forecast accuracy. The forecasts from ARIMA (2,1,4) and ARIMA (2,1,2) models showed ascending trends with stable increases indicating significant increases in the daily number of confirmed and recovered cases of COVID-19 in Nigeria. The forecasts from ARIMA (2,1,3) model showed fluctuating trend with decline in the number of deaths cases due to the disease. The result of this study has ushered in a better preparedness plan by the Nigerian government, NCDC, individuals and other relevant stakeholders in curbing the menace of Coronavirus disease in the country.

Although, Nigeria has made some tremendous achievements toward the control and prevention of COVID-19, Nigeria Centre for Disease Control (NCDC) and relevant health agencies and personnel need to take more serious actions and decisions to control and prevent the spread of the disease thereby reducing the number of new cases and deaths due to this pandemic in the future. As the predictions of this study suggest rising confirmed cases of the disease, the Nigerian government and relevant stakeholders should focus more on formulating policies that hinge on the control and prevention of the virus. To achieve success in combating COVID-19 pandemic, rapid and severe infection control in healthcare systems should be a top priority.

REFERENCES

[1] Wang LF, Shi Z, Zhang S, Field H, Daszak P, Eaton BT. Review of bats and SARS. Emerg. Infect. Dis., 2006; 12:18-34.
[2] Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. J. Med. Viro., 2020. Aailable at https://doi.org/10.1002/jmv.25681.
[3] Cauchemez S, Van Kerkhove MD, Riley S, Donnelly CA, Fraser C, Ferguson NM. Transmission scenarios for Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and how to tell them apart. Euro Surve., 2013; 18(24): 20503.
[4] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet, 2020; 395:565-574.
[5] Cheng ZJ, Shan J. Novel coronavirus: where we are and what we know. Infection 2019; 2020. Available at https://doi.org/10.1007/s15010-020-01401-y.
[6] European Centre for Disease Prevention and Control (ECDC) COVID-19 Situation Update Worldwide, 23rd July, 2020. Retrieved from https://www.ecdc.europa.eu
[7] Nigeria Centre for Disease Control (NCDC) COVID-19 Situation Update, July, 2020: Website:https://ncdc.gov.ng/diseases/sitreps/?cat=14&name=An%20update%20of%20COVID-19%20outbreak%20in%20Nigeria.
[8] Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nature Med., 2020; 1-32.
[9] Yonar H, Yonar A, Tekindal MA, Tekindal M. Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods. Eurasian J. Med. Oncol., 2020; 4(2):160-165.
[10] Liu Z, Huang S, Lu W, Su Z,Yin X, Liang H, Zhang H. Modeling the trend of coronavirus disease 2019 and restoration of operational capability of metropolitan medical service in China: A machine learning and mathematical model-based analysis. Glo. Hea. Res. Poli., 2020; 5(20): 1-11.
[11] Pontoh RS, Solichatus Z, Yuyun H, Aldella R, Mahardika N, Sukono J. COVID-19 Modelling in South Korea using a Time Series Approach. Inter. J. Adv. Sci. Tech., 2020; 29(7): 1620-1632.
[12] Dehesh T, Mardani-Fard HA, Dehesh P. Forecasting of COVID-19 Confirmed Cases in Different Countries with ARIMA Models. 2020; Retrieved from https://doi.org/10.1101/2020.03.13.20035345.

[13] Tran TT, Pham LT, Ngo QX. Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (Case study: Iran). Glo. J. Environ. Sci. Mgt., 2020; 6: 1-10.

[14] Chintalapudi N, Battineni G, Amenta F. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty-day lockdown in Italy: A data driven model approach. J. Microbio., Immuno. Infect., 2020; 53, 396-403.

[15] Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in Brief, 2020; 29, 1-4.

[16] Tandon H, Ranjan P, Chakraborty T, Suhag V. Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. Inter. J. Appl. Sci., 2020; 12(4): 67-79.

[17] Malekia M, Mahmoudib MR, Wraithd D, Phoe KH. Time series modelling to forecast the confirmed and recovered cases of COVID-19. Trav. Med. Infect. Dis., Retrieved from https://doi.org/10.1016/j.tmaid.2020.101742

[18] Osatohanmwen P. Modeling the Daily Number of Reported Cases of Infection from the COVID-19 Pandemic in Nigeria: A Stochastic Approach. Retrieved on 29th May, 2020 from https://www.researchgate.net/publication/341482901

[19] Ibrahim RR, Oladipo HO. Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins Modeling Procedure. Retrieved on 17th June 2020 from https://doi.org/10.1101/2020.05.05.20091686.

[20] Elliott G, Rothenberg TJ, Stock JH. Efficient Tests for an Autoregressive Unit Root. Economet., 1996; 64(4), 813-836.

[21] Schwarz G. Estimating the dimension of a model. The Ann. Stats., 1978; 6(2), 461-464

[22] Breusch TS. Testing for Autocorrelation in Dynamic Linear Models. Austra. Econ. Papers, 1978; 17: 334-355.

[23] Godfrey LG. Testing against General Autoregressive and Moving Average Error Models when the Regression include Lagged Dependent Variables. Economet., 1978; 46: 1293-1301.