

Original Research Article

MODELING CORONAVIRUS PANDEMIC USING UNIVARIATE AND MULTIVARIATE MODELS: THE NIGERIAN PERSPECTIVE

Abstract

Corona virus Disease (COVID-19) is a disease which was discovered in December, 2019 has been spreading worldwide like wildfire. In view of this, there is need of continuous findings on the impact, consequence and possible medications of the pandemic in Nigeria and the world at large. Therefore, this research is aimed at Analyzing the spread of Coronavirus pandemic in Nigeria, using univariate and multivariate models namely;(ARIMA) and (ARIMAX). The daily data used in this research was obtained from the NCDC official website dated from 19th April, 2020 to 20th April, 2021 with total of 384 observations using R and Eview10 software for the analysis. Three different variables were examined. The variables are; total confirmed, discharged and death cases for the purpose of establishing reliable forecast, for better decision making and a helping technique for drastic action in reducing the day to day spread of the pandemic. Summary statistics and stationary test were checked with the data being stationary at the first difference and design technique was conducted as well. Also, best fitted model was selected using Akaike Information Criteria (AIC). The ARIMA (1,1,3) model with an exogenous variable was chosen from the ARIMA models with minimum AIC. From the model, a prediction of sixty-days forecast showed the upward trend of the total confirmed cases of the pandemic in the country. The government on its part via its task force can use the predicted line to take much necessary measures and emphases on taking COVI-19 vaccines so as to prevent further spread of the virus.

Keywords: COVID-19, ARIMA, ARIMAX, NCDC, Nigeria .

1.0 Introduction

Coronavirus Disease was discovered in December, 2019 and has been spreading like wildfire. Its initial outbreak was in the Wuhan city, Hubei, China showing a significant influence on universal incidence, death and health precautions. Symptoms of the pandemic include; fever, dry cough, fatigue, and occasional gastrointestinal symptoms Ayinde *et al.*, (2020). The government of peoples republic of China announced the detachment of the disease on January 7, 2020, which brought about the name Coronavirus (COVID-19) by the World Health Organization on January 12, 2020. Coronavirus is a Zoonosis pathogen, discovered in human and animals with the ability to show clinical features in human. There are four types of prototypic in human being that lead to respiratory diseases and endemic. They include the human beta coronaviruses OC43, human alpha coronaviruses 229E, HKU1 and the NL63 (Drexler *et al.*, 2010). Persons who are infected may be asymptomatic or symptomatic as the case may be and this affects the gastrointestinal, neurological system, hepatic and respiratory. This may lead to admitting the affected person in an emergency care unit of a hospital or any health care Centre. (Drexler *et al.*, 2010; Woo *et al.*, 2010; Yin & Wunderink, 2018). Pathogenicity of CoV could earlier be dated back to the years 2002 and 2003 when they were considered to be highly pathogenic due to intense respiratory syndrome (SARS) in Guangdong State of China for the first time. Prior to outburst of the pandemic, the famous CoV was presumed to bring about slight infections in humans with sufficient immune system which include; CoV OC43 and CoV 229E (Peiris *et al.*, 2003; Yin & Wunderink, 2018). Barely a decade of post SARS, the exalted pathogenic CoV, Middle East Respiratory Syndrome Coronavirus (MERS-CoV) transpired in the Middle East countries (Zaki, Van Boheemen, Bestebroer, Osterhaus, & Fouchier, 2012) and lately in the month of December 2019, a novel Coronavirus (nCoV) transpired at a livestock/Seafood Market in the city of Wuhan State of Hubei Province in China and has metamorphosed into a global pandemic (“Seven days in medicine,” 2020). The Chinese authorities ensured the isolation of the pandemic on January 7 2020, the virus was labeled COVID-19 by the World Health Organization on January 12, 2020, and as at February 12, 2020, a total of 43,103 confirmed cases and 1,018 deaths had been reported (Organization, 2020). Ever since the outbreak, there have been a good number of research and predictions about the pandemic employing various statistical techniques worldwide. This research focused more on application of ARIMA (univariate) and ARIMAX (multivariate) models using time series data from Nigeria Center for Disease Control (NCDC).

Andam et al, (2020), analyzed the economic significance of the pandemic and conduct employed to curb the spread of the pandemic in Nigeria where they carried out simulations using the multiplier model based on the 2018 Social Accounting Matrix (SAM) for Nigeria. Their research focused on the five week lockdown implemented by the

Nigerian Government across two states; Lagos and Ogun and the federal capital territory. Abuja. Hence their findings had implications for understanding the direct and indirect of the pandemic during the period understudy.

Piccolomini & Zama (2020), also proposed an adjustment of the Susceptible- Infected-Exposed-Recovered-Dead (SEIRD) differential approach for the analysis and forecast of the coronavirus spread in some regions of Italy. They recommended a time-dependent transmitting rate and reported the highest rate of infection spread for the three Italian regions firstly affected by the coronavirus outburst (Lombardia, Veneto and Emilia Romagna). Jit *et al.*, (2020,) employed the exponential growth approach to fit several care admissions from multiple surveillance to study probable coronavirus case numbers and progress in the United Kingdom from the 16th of February to March 23rd, 2020. They projected that on 23rd March, there were 102,000 (median; 95% credible interval 54,000 -155,000) new cases and 320 (211 - 412) new critical care reports, with 464,000 (266,000 – 628,000) cumulative cases since February 16, 2020.

Rauf and Olukemi (2020), estimated the spread of coronavirus in Nigeria using the Box-Jenkins Modeling method using daily spread data from February 27 to April 26, 2020, and modelled the autoregressive integrated moving average (ARIMA) model using the R software. Stability analysis and stationary test, parameter test, and model diagnostic were also carried out. Finally, the fitting, selection and prediction accuracy of the ARIMA model was assessed with the aid of the AIC model selection criteria. The ARIMA (1,1,0) model was finally selected as it was presumed to be the best among ARIMA models based on the parameter test and Box–Ljung test. Furthermore, a ten-day forecast was also carried out from the model and this shows an absolute upward trend of the spread of the COVID-19 pandemic in Nigeria within the period understudy.

Malavika et al., (2020), used the Susceptible-Infected-Recovered (SIR) model for spread of disease to make a short term prediction in India and high incidence states, where they forecasted the maximum number of cases and peak time in order to evaluate the impact of lockdown and other measures in containing the pandemic.

Another way of reducing the spread of Covid-19 is the ban on social gathering. Such substantial reduction is feasible in the jurisdictions (and the entire Nigerian nation) if these protocols are strictly adhered to.

Hussain et al (2020), studied a stochastic model on the spread and containing of the pandemic in USA, where the population of infected persons is divided into susceptible, infected and recovered classes. They began by studying the stability of the corresponding model. They also calculated the unique non-negative solution and investigated the inequality managing that led to the control of the virus.

As at 27th of February, 2020, Nigeria recorded its first case of Covid-19 with the index case being an Italian who arrived the country via the Murtala Muhammed International Airport, Lagos at 10pm aboard a Turkish airline from Milan in Italy. On 9th of March, 2020, the second case of the pandemic was recorded in Nigeria and it was a contact of the index case. The number of persons who have contracted the pandemic kept increasing day by day. As at March 29 2020, the total confirmed cases in Nigeria had risen to 97 (ninety-seven) with the first death from complications of the pandemic recorded on March 23, 2020. The federal government of Nigeria on 29th of March 2020, declared a lockdown on two states; Lagos and Ogun and the Federal Capital Territory, Abuja with effect from 11 pm of March 30, 2020. In Nigeria, Lagos State became the epicentre of the pandemic.

The government extended the restrictions on March 21, 2020 by closing two main international airports in the country. That is, The Murtala Muhammed Airport in Lagos and The Nnamdi Azikiwe Airport in Abuja. The country also went further by suspending rail services on March 23, 2020, due to the spread of the disease as a result of local transmission.

In this work, we decide to employ two models; the univariate (ARIMA) model and the multivariate (ARIMAX) models to model of the spread of the pandemic in Nigeria. The data used for the analysis comprise of three variables which are; total confirmed, recovered and death cases.

2.0 Methodology for Analysis

Information on the variables understudy were extracted from the official website of the Nigeria Centre for Disease Control (NCDC), (<https://covid19.ncdc.gov.ng/>). The data is on daily cases and covered the period from April 19, 2020, to 20th April 2021 with total 384 observations. The total confirmed cases variable shows an exponential surge, notably in March, 2020. Nevertheless, the curved line of recovered cases increased but not as high as that of total confirmed cases within the stipulated period. The considered variables are ordered at time of daily interval. The Univariate (ARIMA) model and Multivariate (ARIMAX) model were considered.

2.1 Unit roots Test

Empirically using time series data is essential to establish the presence or absence of unit root in the series understudy. This is because, modern econometrics has shown that, regression analysis using non-stationary time series variables produce spurious regression since standard results of OLS do not hold. A variable is said to be

stationary if the mean, the variance and the covariance of the series are constant over time. A stationarized series is easier to predict than the non-stationary series. Where;

$$E(Y_t) = E(Y_{t-1}) = \mu$$

Which is a constant and

$$Cov(Y_t, Y_{t-1}) = \gamma l$$

This depends only on the lag l and not on time t . If there is no unit root, the time series fluctuates around a constant long-run mean with finite variance which does not depend on time. There are several proposed quantitative methods of testing for stationary of a time series variable.

2.2 Augmented Dickey Fuller (ADF) Unit Root Test

This research adopted the Augmented Dickey-Fuller (ADF) test to ascertain if the coronavirus series data is stationary. The ADF test suggested by Dickey and Fuller (1979) is an upgrade of the Dickey-Fuller (DF) test. The ADF handles bigger and more complex models. This test is based on the assumption that the series follow a random walk with model;

$$\Delta R_t = \Phi Y_{t-1} + u_t \quad 2.1$$

And tests the hypothesis: $H_0 : \Phi = 0$ (Non-stationary) against $H_1 : \Phi \neq 0$ (Stationary) where ϕ is the characteristic root of an AR polynomial and u_t is an uncorrelated white noise series with zero mean and constant variance σ^2 . When $\Phi = 1$, equation (1) does not satisfy the weakly stationary condition of an AR (1) model hence the series becomes a random walk model known as a unit root non-stationary time series. Subtracting Y_{t-1} from both sides of equation (1) we get;

$$\Delta R_t = \phi Y_{t-1} + u_t, t = (1, \dots, T) \quad 2.2$$

Where $\phi = \Phi - 1$ and $\Delta R_t = Y_t - Y_{t-1}$. For estimating the existence of unit roots using equation (2), we test hypothesis $H_0 : \phi = 0$ against $H_1 : \phi \neq 0$ under H_0 , if $\phi = 0$, then $\Phi = 1$, thus the series has a unit root hence is non-stationary. The rejection or otherwise of the null hypothesis, H_0 is based on the t-statistic critical values of the Dickey Fuller statistic. The Dickey Fuller test assumes that the error terms are serially uncorrelated; however, the errors terms of the Dickey Fuller test do show evidence of serial correlation. Therefore, the proposed ADF test includes the lags of the first difference series in the regression equation to make u_t a white noise. The Dickey and Fuller (1979) new regression equation is given by;

$$\Delta R_t = \phi Y_{t-1} + \sum_{j=1}^p \gamma \Delta r_{t-j} + u_t, t = (1, \dots, T) \quad 2.3$$

If the intercept and time trend $\beta + \alpha t$ are included, then equation (3) is written as;

$$\Delta R_t = \beta + \alpha t + \phi Y_{t-1} + \sum_{j=1}^p \gamma \Delta r_{t-j} + u_t, t = (1, \dots, T) \quad 2.4$$

Where β is an intercept, α defines the coefficient of the time trend factor, $\sum_{j=1}^p \gamma \Delta r_{t-j}$ defines the sum of the

lagged values of the response variable ΔR_t and p is the order of the autoregressive process. If ' of the Augmented Dickey Fuller model is zero , then there exist a unit root in the time series variable considered, hence the series is not covariance stationary. The choice of the starting augmentation order depends on the periodicity of the data, the significance of i estimates and the white noise residuals series u_t . The ADF test statistic is given by;

$$F_r = \frac{\hat{\phi}}{SE(\hat{\phi})} \quad 2.5$$

Where is the estimate of ϕ and $SE(\hat{\phi})$ is the standard error of the least square estimate of $\hat{\phi}$. The null hypothesis H_0 is rejected if, the p -value $< \alpha$ (significance level). If the series is not stationary, it is transformed

by differencing to make it stationary and stationary tested again. If the time series is not stationary but its first difference is stationary, then the series is said to be an integrated at order one (1) or simply an I (1) process.

2.3. Univariate Ljung-Box Test

The research employed the univariate Ljung and Box (1978) test to jointly check whether or not autocorrelations r_l of the residuals of the individual ARIMA models fitted were zero. It is based on the assumption that the residuals contain no serial correlation (no autocorrelation) up to a given lag m . The Univariate Ljung-Box statistic is given by:

$$Q(m) = T(T + 2) \sum_{l=1}^m \frac{r_l^2}{T - l} \quad 2.6$$

2.4 Model Specification

The models specifications proceed with the use of Univariate (ARIMA) and the multivariate (ARIMAX) models.

2.5 The univariate (ARIMA) Model

The univariate (ARIMA) model is ARMA modeled via differencing series which is a method for removing non-constant trend by making the mean stationary but not the variance. The model was suggested by G.E.P.BOX and G.M.Jenkins. Practically, majority of the series don't conform to the requirements of stationarity. Although some of the non-stationary series have a particular trend; hence, a differencing operator could help in ensuring stationarity of the series. For instance, first-order differencing can change a time series with a constant slope into a new series with constant mean. The univariate (ARIMA) model shows a time series using appropriate differencing and an Autoregressive Moving Average model, which includes moving an average (MA) process and autoregressive (AR) model. This model is indicated by ARIMA (p, d, q) and is expressed as;

$$\Phi(B)\nabla^d y_t = \Theta(B)\varepsilon_t \quad 2.7$$

Where B is the backward shift operator

$$\nabla^d = (1 - B)^d$$

$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ Is a moving average polynomial of order q

$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ is an autoregressive polynomial of order p

A Univariate technique (ARIMA) is an integral tool when dealing with time series analysis and forecast making. Outline for analysis and prediction as proposed by Box and Jenkins are as follows.

- i. Check the series reliability. Where stationarity isn't ascertained, we may achieve stationarity after first lag of the series.
- ii. Evaluate some invincible parameters in the series and also check their significance.
- iii. Create an analysis of diagnosis to analysis in order to ensure the series equates the attributes of the observation.

2.6 The Multivariate (ARIMAX) Model

From the univariate (ARIMA) model, the multivariate (ARIMAX) approach may consider covariates into account by including the covariate to the right side of the univariate (ARIMA) mathematical expression.

The mathematical expression of the multivariate (ARIMAX) technique is given as;

$$\Phi(B)\nabla^d y_t = \mu + \Theta(B)X_t + \Theta(B)\varepsilon_t \quad 2.8$$

- i. Assess the stationarity of the series, where stationarity is achieved take a difference at first lag.
- ii. Evaluate result indicating attributes of response series in the model, for instance, Autocorrelation function (ACF) and Partial autocorrelation function (PACF) in order to estimate the parameters p and q .
- iii. Estimate the hidden parameters of the series as well as ascertain their importance and residuals of the model.
- iv. Estimate the cross correlation coefficient between the response series and the input series to determine the configuration of the ARIMAX model.

- v. Ensure diagnostic analysis to verify whether the model corresponds to the characteristics of the data.

4.0 Result and Discussion

Daily COVID-19 data for the period 19/04/2020 to 14/04/2021 were used with total of 384 observations. The data was obtained via the website of the Nigeria Center for Disease Control (NCDC). Table 4.1 depicts output summary of the series.

Table 4.1: Summary Statistics

	Total Confirmed	Discharged	Death
Mean	66877.34	54052.45	966.2599
Median	60216.00	39967.50	952.0000
Maximum	164633.0	154643.0	2061.0000
Minimum	0.000000	0.000000	0.000000
Std. Deviation	59303.63	53167.03	722.2681
Skewness	0.447961	0.652527	0.121207
Kurtosis	1.758397	2.025285	1.698051
Jarque-Bera	34.57773	39.13520	25.86906
Probability	0.000000	0.000000	0.000002

The table 4.1 shows the summary statistics of the daily COVI-19 data with the average of total confirmed, discharged and death cases as 66877.34, 54052.45 and 966.2599, the range of the data is self explanatory, the standard deviations of all the variables 59303.63, 53167.03 and 722.2681 which show variation from the average and extent of the variability of the variables understudy. All the variables were positively asymmetry and platykurtic based on the nature of kurtosis. The distribution isn't normal as the likelihood of the Jarque-Bera is below 0.05.

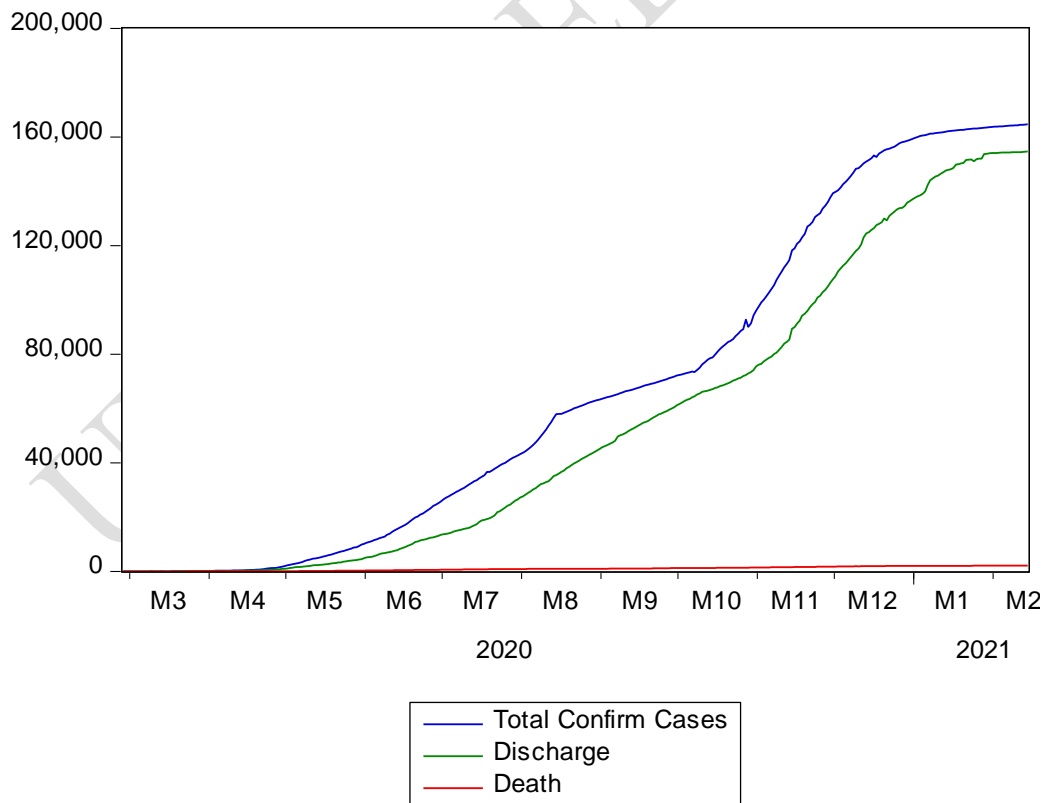


Figure 4.1: The time plot of total confirmed, discharged and Death cases.

The total confirmed and discharged series depicts an exponential surge, from April 2020. Moreover, the plot of recovered series increased but not higher as total confirmed cases and the death cases trend increasing slowly compared with total confirmed and discharged cases throughout the study period.

Table: 4.2 Unit Root Test

Series	Probability	Lag	Max Lag
Total Confirmed	0.7461*	7	12
Discharged	0.7361*	11	12
Death	0.8829*	6	12

(*) Denote non acceptance of null hypothesis at 1%, 5% and 10% significant level and conclude non-stationarity of the series based on the result of Augmented Dicky Foller (ADF) Test.

After taken the 1st Difference of the series

Series	Probability	Lag	Max Lag
Total Confirmed	0.03839**	6	12
Discharged	0.03993**	9	12
Death	0.00197**	5	12

(**) Denote the rejection of null hypothesis at 1, 5 and 10 percent level of significance and concluded that the series is stationary based on the result of Augmented Dicky Fuller (ADF) Test.

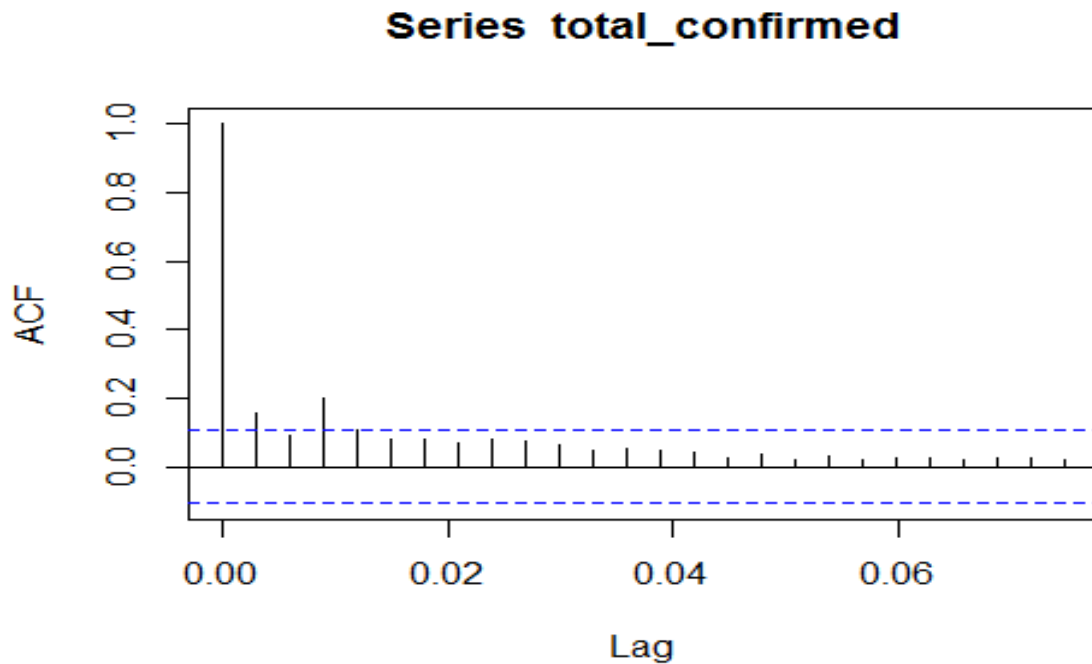


Figure 4.2: Autocorrelation Function (ACF) of the differenced total confirmed cases

All the lags are significant but precisely, the first four lags as most significant lags and suggested MA (1) and MA (3) terms.

Series total_confirmed

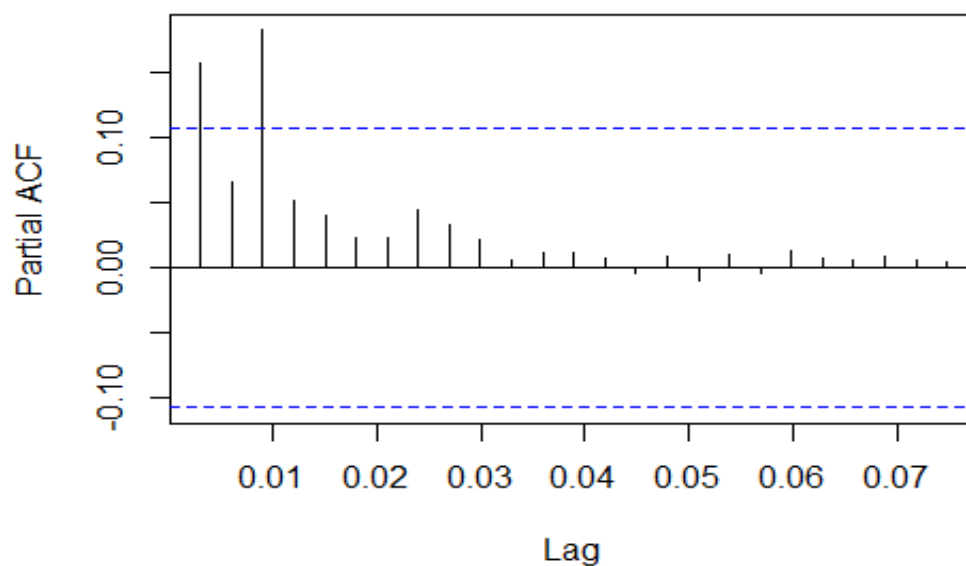


Figure 4.3: Partial Autocorrelation Function (PACF) of the differenced total confirmed cases. Lag one to lag four are the most significant but for convenience, AR(1) and AR(3) terms were suggested.

Table 4.3: ARIMA Model Selection

Model	Log likelihood	AIC
ARIMA (1, 1, 1)	-2594.69	5195.39
Model Terms	Coefficient	Std Error
AR(1)	-0.2405	0.0572
MA(1)	-0.7825	0.0313
Model	Log likelihood	AIC
ARIMA (1, 1, 3)***	-2589.93	5189.87***
Model Terms	Coefficient	Std Error
AR(1)	-0.1812	0.5502
MA(1)	-0.8790	0.5470
MA(2)	-0.0323	0.5778
MA(3)	0.1542	0.1381
Model	Log likelihood	AIC
ARIMA (3, 1, 1)	-2590.77	5190.77
Model Terms	Coefficient	Std Error
AR(1)	-0.3437	0.0815
AR(2)	-0.1890	0.0809
AR(3)	-0.0016	0.0683
MA(1)	-0.7090	0.0619
Model	Log likelihood	AIC
ARIMA (3,1,3)	-2589.49	5192.99
Model Terms	Coefficient	Std Error
AR(1)	0.1101	0.4979
AR(2)	-0.1189	0.1903

AR(3)	0.0871	0.1066
MA(1)	-1.1708	0.4950
MA(2)	0.1653	0.0874
MA(3)	0.1527	0.0374

Table 4.4: ARIMAX Model

Model	Log likelihood	AIC
ARIMA (1,1,3)**xreg(death)	297.09**	-582.19**
Model Terms	Coefficient	Std Error
AR(1)	-0.0466	0.1106
MA(1)	-0.2656	0.0310
MA(2)	-0.0544	0.0294
MA(3)	0.8539	0.0265
Xreg	-0.0012	0.0001

The table 4.3 and table 4.4 are the ARIMA and ARIMA with an exogenous variables (ARIMAX), based on the result ARIMA (1,1,3) with exogenous is the best model with minimum Akaike Information Criteria and highest log likelihood.

Table 5: Box-Ljung Test

Model	Chi-Square	d.f	P-Value
ARIMA	39.303	20	0.100960**
ARIMAX	37.238	20	0.006112***

(**) Denote the acceptance of null hypothesis and conclusion that there correlation in the residual while (***) Denote the rejection of null hypothesis and conclusion that there is no correlation in the residuals; therefore the model is good for forecast.

Conclusion

This research aimed at Analyzing the wide spread of coronavirus in Nigeria, using Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Integrated Moving Average with an Exogenous Variable (ARIMAX) models with three different variables were examined; total confirmed, discharged and death cases to generate a reliable forecast that may be employed for proper planning in future. The best fitted model was selected using Akaike Information Criteria (AIC). The ARIMA (1,1,3) model with an exogenous variable was finally selected among ARIMA models with minimum AIC. A sixty-days prediction was made and the result depicts the surge of the total confirmed cases of the coronavirus in the country. The Nigerian government is therefore advised to evaluate this prediction and adopt it in tackling the menace.

Forecasts from ARIMA(1,1,3)

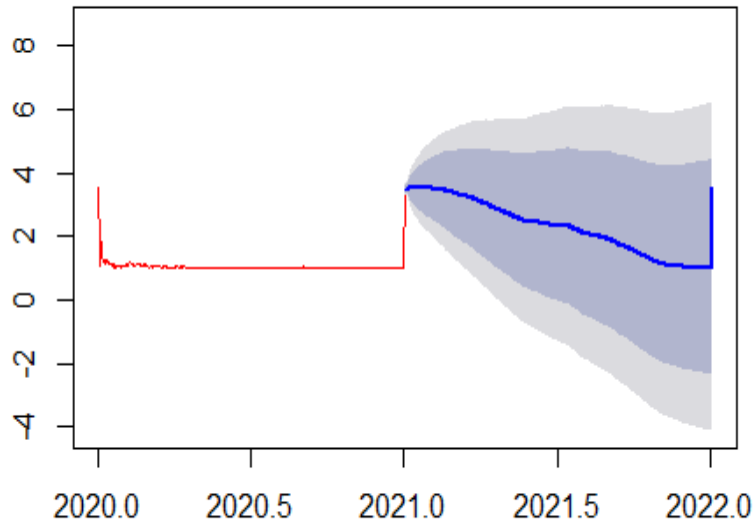


Figure 4.4: Time Plot for the two month forecast of ARIMA with an exogenous variable

COMPETING INTERESTS DISCLAIMER:

Authors have declared that no competing interests exist. The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

REFERENCES

- Aas, K. and Dimakos, X. K. (2004). Statistical Modeling of Financial Time Series. *Applied Research and Development*. 1-37
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366), 427–431.
- Drexler, J. F., Gloza-Rausch, F., Glende, J., Corman, V. M., Muth, D., Goettsche, M., ... Drosten, C. (2010). Genomic Characterization of Severe Acute Respiratory Syndrome-Related Coronavirus in European Bats and Classification of Coronaviruses Based on Partial RNA-Dependent RNA Polymerase Gene Sequences. *Journal of Virology*, 84(21), 11336–11349. <https://doi.org/10.1128/jvi.00650-10>
- Hussain, S., Zeb, Anwar., Rasheed, A., and Saeed, T., (2020). Stochastic Mathematical Model for the spread and control of coronavirus. *Springer open Journal of Advances in differential Equations*. 574(2020).
- Kayode, A., Adewale, F.L., Opeyemi, E, A. (2020). Modeling Nigerian Covid-19 cases: A comparative analysis of model and estimators. *Elsevier Chaos Solitons Fractals*.
- Ljung, G. M., & Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.
- Organization, W. H. (2020). Coronavirus disease 2019. *World Health Organization*, (April), 2633.
- Malavika, B., Marimuthu, S., Joy, M., Nadaraj, A., Sam, A, E., Jeyaseelan, L. Forecasting Covid-19 epidemic in India and high incidence states using SIR and logistic growth model . *Elsevier Science Direct*, 9(2021), 26-33.
- <https://doi.org/10.1001/jama.2020.2633>
- Rauf, I.R and Oluwakemi, H.O.(2020). Forecasting the spread of COVID-19 in Nigeria using Box-Jenkins Modeling Procedure. [CC-BY-NC-ND 4.0 International license](#) .
- Sahin, A. R. (2020). 2019 Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature. *Eurasia Journal of Medical Investigation*, 4(1), 1 <https://doi.org/10.14744/ejmo.2020.12220>
- Woo, P. C. Y., Huang, Y., Lau, S. K. P., & Yuen, K. Y. (2010). Corona virus genomics and bioinformatics analysis. *Viruses*, Vol. 2, pp. 1805–1820. <https://doi.org/10.3390/v2081803>
- Yin, Y., & Wunderink, R. G. (2018, February). MERS, SARS and other corona viruses as causes of pneumonia. *Respirology*, Vol. 23, pp. 130–137. <https://doi.org/10.1111/resp.13196>