

Forecasting Daily Confirmed Cases of Covid-19 in Nigeria as a Public Health Tool Using Arima Model

F. C. Nkeshita*, A. A. Adekunle, S. S. Owosho and A. A. Adewale

Department of Civil Engineering, Federal University of Agriculture, Abeokuta, Nigeria

*Corresponding Author: nkeshitafc@funaab.edu.ng

ARTICLE INFO

Received: January, 2021

Accepted: May, 2021

Published: July, 2021

Keywords:

Arima model
Coronavirus
Covid-19
Forecast
Pandemic.

ABSTRACT

The coronavirus virus scourge, also known as COVID-19 has continued to record an increase in the number of confirmed cases in Nigeria with numbers exceeding over a hundred thousand and more than a thousand deaths by the middle of January, 2021. An ARIMA model was developed to forecast the confirmed cases of covid-19 in Nigeria using a dataset of daily confirmed cases from 27th February, 2020 to 12th January 2021 sourced from the Nigeria Centre for Disease Control (NCDC). The Box-Jenkins method for analyzing ARIMA models was adopted for this research using the EVIEWS version 10 and SPSS version 21.0 software, with the former used for checking stationarity, seasonality, and trends as well as developing a time series chart and model identification, while the latter was used for model selection and forecast. The most preferred model was a non-seasonal ARIMA model (4,1,8) which was further used to forecast future confirmed cases in Nigeria up to 30 days ahead. The result of the forecast shows an exponential increase in confirmed cases with grave implications for the country if stringent measures are not taken.

1. INTRODUCTION

Communicable diseases have continued to portend danger to public health due to the transmission of such diseases from person to person. This category of diseases had been previously reported to have caused about 16% of total deaths worldwide in the year 2004 (Shim, 2004). There has been a growing concern among public health experts on the outbreak of the novel coronavirus SARS-CoV-2 also known as COVID-19 worldwide. This has resulted in the declaration by the World Health Organization (WHO) that the outbreak requires a Public Health Emergency at the international level (WHO, 2020). The novel virus was first reported in China in early December 2019 and has evolved into an epidemic that is now ravaging several countries (Li *et al.*, 2020). The first case of covid-19 was reported in Nigeria on 27th February 2020, when an expatriate worker in Ewekoro, Ogun State, South-western Nigeria tested positive for the virus, caused by SARS-CoV-2 (NCDC, 2020) and it has resulted in a rapid spread to many states in the country.

Since COVID-19 was first detected in Nigeria, there have been over 100,000 confirmed cases and over a thousand deaths, even after the Federal Government had imposed several restrictions including a restriction of movement and social distancing. However, the gains began to yield results when the number of confirmed cases began to drop by the third quarter of the year (2020). By early December 2020, there were records of a spike in confirmed cases which the Nigeria Centre for Disease Control (NCDC) termed “second wave” in which some daily records show confirmed cases reaching over a thousand. This has resulted in a Nkeshita *et al.*: Forecasting Daily Confirmed Cases of Covid-19 in Nigeria as a Public Health Tool Using Arima Model

renewed call for people to act more responsibly by obeying several public health guidelines including handwashing, use of hand sanitizers, social distancing, wearing of nose masks, etc. There is a need for COVID-19 cases need to be forecasted as this enhances the ability of health officials to make important decisions and also enables policymakers to take important decisions that can aid the prevention of pandemics (Singh *et al.*, 2020).

Many researchers have adopted several methods aimed at predicting the occurrence of pandemic and infectious diseases (Papastefanopoulos *et al.*, 2020; Tran *et al.*, 2020) including the SIR model, SEIR model, Artificial Neural Networks (ANN) and ARIMA time series model, etc. Time series models had previously been used by researchers to model infectious diseases including covid-19 (Kotwa *et al.*, 2020; Wang *et al.*, 2018; Wang *et al.*, 2020). ARIMA models are reliable when used to forecast disease novelty or in the absence of adequate data (Singh *et al.*, 2020). This study aims to develop stochastic models using ARIMA TO forecast the daily confirmed cases of covid-19 confirmed cases in Nigeria for the next thirty days using data sourced from the NCDC website. This will enable policymakers and health workers to formulate public health-related decisions that can aid the slowing down of the spread of the pandemic.

2. METHODOLOGY

Description of Study Area

Nigeria is located in the western part of Africa. It shares its boundary with Niger Republic in the North, Benin republic in the west, and both Chad and Cameroun in the east. Its southern boundary is delimited by the Atlantic Ocean and occupies a landmass of about 900,000 sq. Km. It occupies latitudinal and longitudinal coordinates extending from 4° to 14°N and 2° to 15°E respectively and has a population of about 206,000,000 people (NBS, 2020).

Source of Data and Preliminary Analysis

The data for confirmed cases were retrieved from the Nigeria Centre for Disease Control (NCDC) website (www.ncdc.gov.ng) for daily confirmed cases reports from February 27th, 2020 to 12th January, 2021. The Box-Jenkins method for analyzing ARIMA models was adopted for this research with the aid of the following software: The software Statistical Package for the Social Sciences (SPSS) version 21.0 which was used to develop the time-series database and prediction models and EVIEWS 10, was used to check for stationarity, seasonality and trends by visual inspection of the autocorrelation (ACF) and partial autocorrelation (PACF) functions charts, as well as for the Augmented Dickey-Fuller (ADF) test. The ADF test was also used to confirm the stationarity of the series. When found to exhibit non-stationary and seasonality features, the time series was subjected to differencing or seasonal differencing as appropriate.

Model Identification and Selection

To determine an appropriate ARIMA model, the autocorrelation and partial autocorrelation plots of the time series were investigated for significant spikes, and the corresponding autoregressive (AR) and moving average (MA) components were derived. The best-integrated model (ARIMA) was selected by applying the criteria of choosing the model with the highest Correlation coefficient (R^2), lowest Schwartz Bayesian information criterion (BIC), and lowest root mean square error (RMSE). The rule of parsimony was also adopted where applicable.

Model Diagnostics and Forecast

A visual inspection of the autocorrelation and partial autocorrelation charts of the residuals of the selected model was carried out to ensure that there were no spikes that extend beyond the continuous lines below which imply the approximate 95% confidence limits. The best model was adopted for the prediction of future results after diagnostic checks were carried out on its residual. This model was finally used to forecast future confirmed cases of COVID-19.

2. RESULTS AND DISCUSSION

Test for Stationarity, Trend, and Seasonality

Figure 1 depicts the time series plot of covid-19 cases recorded in Nigeria which was implemented in SPSS version 21. From visual inspection, it could be observed that there appears to be an upward trend but no seasonality. Further inspection of the autocorrelation and partial autocorrelation charts were carried out visually to confirm the preliminary observations. From the ACF plot shown in Figure 2 below, the slow but consistent decline towards zero with each additional lag further lends credit to the trending nature of the raw data. In terms of stationarity, a visual inspection of figure 1 shows that the time series is non-stationary while the slowly decreasing lags in the ACF plot also point to non-stationarity. An Augmented Dickey-Fuller (ADF) test or the unit root test was also conducted to determine the stationarity of the time series as shown below in Figure 3. The p-value of 0.9756 being above 0.05 indicates that the time series data of confirmed cases in Nigeria has a unit root and thus confirms the non-stationarity of the time series data. From the foregoing, this time series data set displayed trend, is non-seasonal but also displays non-stationarity, therefore, there is a need to subject the time series data to a non-seasonal differencing process in order to remove the trend and also to make the raw data stationary.

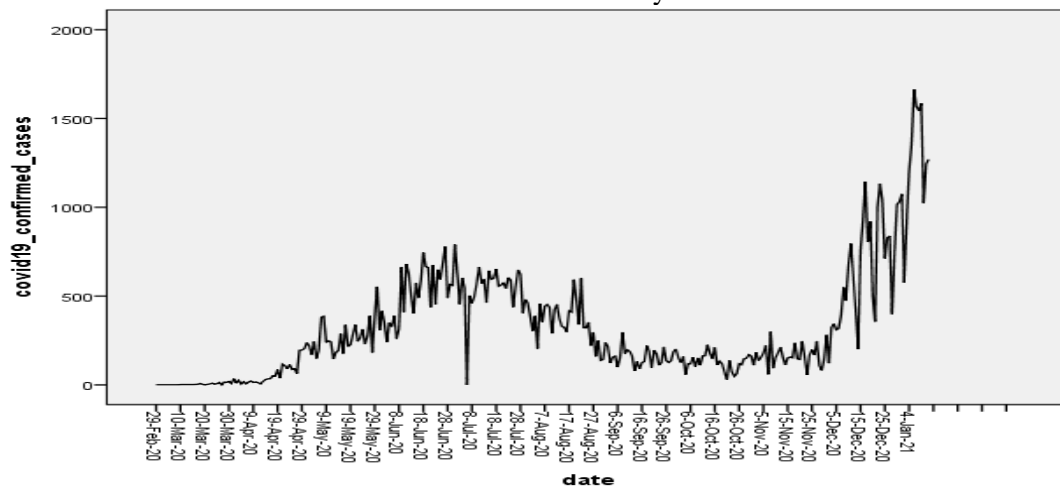


Figure 1: Time series plot of raw data of covid-19 confirmed cases in Nigeria

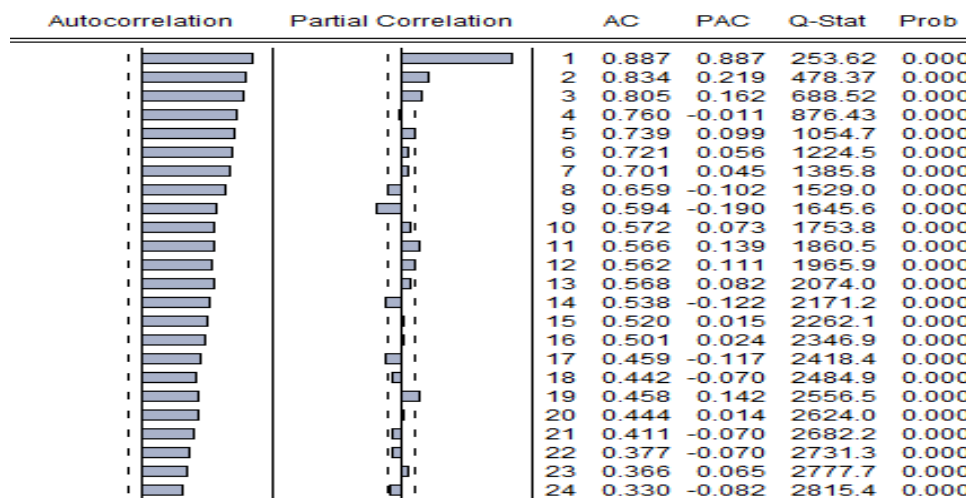


Figure 2: Autocorrelation and Partial autocorrelation plots of raw data of Nigeria covid-19 cases

Null Hypothesis: NIGERIA_COVID19_CASES has a unit root
Exogenous: Constant
Lag Length: 8 (Automatic - based on SIC, maxlag=16)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.255718	0.9756
Test critical values:		
1% level	-3.451283	
5% level	-2.870651	
10% level	-2.571695	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(NIGERIA_COVID19_CASES)
Method: Least Squares
Date: 01/13/21 Time: 09:39
Sample (adjusted): 3/09/2020 1/12/2021
Included observations: 310 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NIGERIA_COVID19_CASES(-1)	0.007145	0.027940	0.255718	0.7983
D(NIGERIA_COVID19_CASES(-...)	-0.509920	0.065009	-7.843891	0.0000
D(NIGERIA_COVID19_CASES(-...)	-0.358307	0.071985	-4.977501	0.0000
D(NIGERIA_COVID19_CASES(-...)	-0.262041	0.076965	-3.404697	0.0008
D(NIGERIA_COVID19_CASES(-...)	-0.293683	0.077104	-3.808909	0.0002
D(NIGERIA_COVID19_CASES(-...)	-0.209137	0.076409	-2.737093	0.0066
D(NIGERIA_COVID19_CASES(-...)	-0.094860	0.074625	-1.271143	0.2047
D(NIGERIA_COVID19_CASES(-...)	0.156410	0.069374	2.254606	0.0249
D(NIGERIA_COVID19_CASES(-...)	0.247365	0.060445	4.092381	0.0001
C	7.716324	10.49723	0.735082	0.4629
R-squared	0.289995	Mean dependent var		4.093548
Adjusted R-squared	0.268695	S.D. dependent var		137.2461
S.E. of regression	117.3679	Akaike info criterion		12.40023
Sum squared resid	4132566.	Schwarz criterion		12.52076
Log likelihood	-1912.036	Hannan-Quinn criter.		12.44841
F-statistic	13.61470	Durbin-Watson stat		2.005623
Prob(F-statistic)	0.000000			

Figure 3: Augmented Dickey-Fuller test for stationarity of raw data

The time series raw data was subjected to a first order differencing and the preliminary observation processes were carried out again and a look at the ACF and PACF plots in Figure 4 shows that the differenced time series look stationary. This was confirmed by applying the Augmented Dickey-Fuller (ADF) test as shown in Figure 5 with a p-value less than 0.05 thus indicating that the null hypothesis that confirmed cases in Nigeria is rejected, and therefore has no unit root, and confirming the stationarity of the time series data. The first order differenced time series plot of confirmed cases in Nigeria is as shown in Figure 6 where the series appear to display a constant mean and variance.

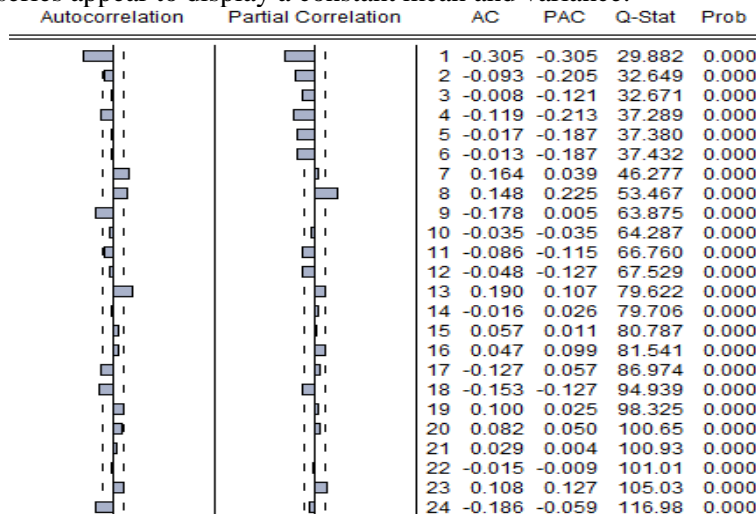


Figure 4: Autocorrelation and Partial autocorrelation plots of differenced data of covid-19 confirmed cases

Null Hypothesis: D(NIGERIA_COVID19_CASES) has a unit root
Exogenous: Constant
Lag Length: 7 (Automatic - based on SIC, maxlag=16)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.619137	0.0000
Test critical values:		
1% level	-3.451283	
5% level	-2.870651	
10% level	-2.571695	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(NIGERIA_COVID19_CASES,2)
Method: Least Squares
Date: 01/13/21 Time: 09:40
Sample (adjusted): 3/09/2020 1/12/2021
Included observations: 310 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(NIGERIA_COVID19_CASES(-1))	-2.260026	0.341438	-6.619137	0.0000
D(NIGERIA_COVID19_CASES(-1),2)	0.758525	0.317813	2.386703	0.0176
D(NIGERIA_COVID19_CASES(-2),2)	0.409389	0.281637	1.453608	0.1471
D(NIGERIA_COVID19_CASES(-3),2)	0.157179	0.239906	0.655166	0.5129
D(NIGERIA_COVID19_CASES(-4),2)	-0.127019	0.195599	-0.649385	0.5166
D(NIGERIA_COVID19_CASES(-5),2)	-0.327324	0.150762	-2.171128	0.0307
D(NIGERIA_COVID19_CASES(-6),2)	-0.414127	0.104452	-3.964742	0.0001
D(NIGERIA_COVID19_CASES(-7),2)	-0.251386	0.058273	-4.313954	0.0000
C	9.750239	6.839897	1.425495	0.1551
R-squared	0.727927	Mean dependent var	0.083871	
Adjusted R-squared	0.720696	S.D. dependent var	221.7355	
S.E. of regression	117.1855	Akaike info criterion	12.39400	
Sum squared resid	4133467.	Schwarz criterion	12.50248	
Log likelihood	-1912.069	Hannan-Quinn criter.	12.43736	
F-statistic	100.6650	Durbin-Watson stat	2.007776	
Prob(F-statistic)	0.000000			

Figure 5: Augmented Dickey-Fuller test for stationarity of differenced data

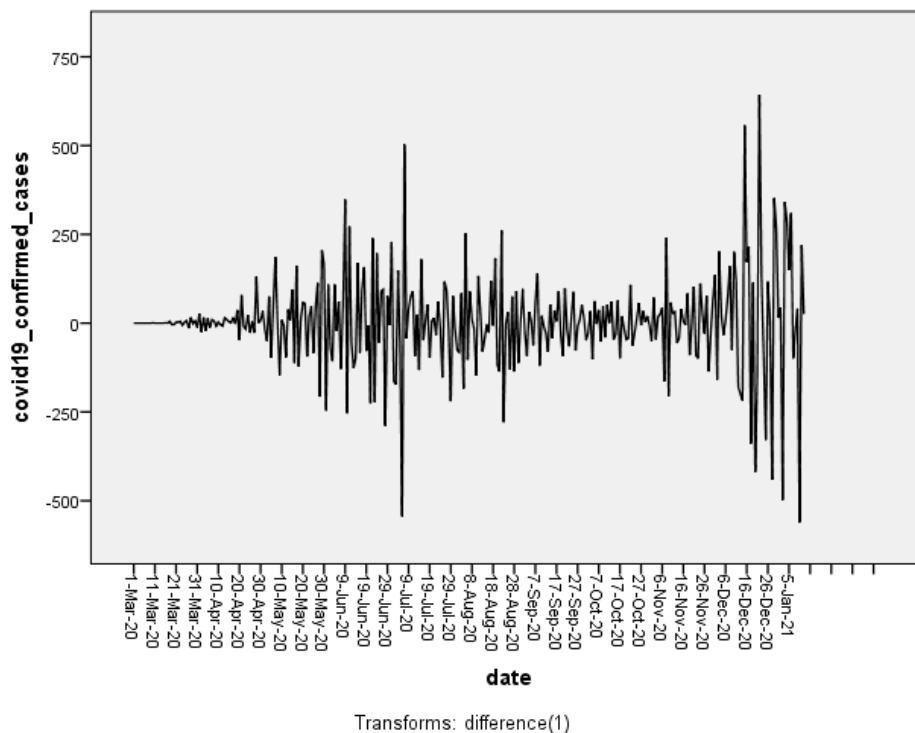


Figure 6: Time series plot of first order differenced data of covid-19 confirmed cases in Nigeria.

Model Identification and Selection

The first order differenced ACF and PACF plots were used to identify and select appropriate non- seasonal ARIMA models that could be used for the COVID-19 forecast. This was done by identifying the autoregressive (AR) component of the models from the PACF plot and the moving average (MA) component from the ACF plot. From the ACF, it would be observed that there were significant spikes at Nkeshita *et al.*: Forecasting Daily Confirmed Cases of Covid-19 in Nigeria as a Public Health Tool Using Arima Model

lags 1, 4, 7 and 8, while for the PACF plots, there were significant spikes at lags 1, 2, 3, 4, 5, 6, and 8. Thus, the tentative ARIMA models to be considered include; (1,1,1), (1,1,4), (1,1,7), (1,1,8), (2,1,1), (2,1,4), (2,1,7), (2,1,8), (3,1,1), (3,1,4), (3,1,7), (3,1,8), (4,1,1), (4,1,4), (4,1,7), (4,1,8), (5,1,1), (5,1, 4), (5,1,7), (5,1,8), (6,1,1), (6,1,4), (6,1,7) and (6,1,8).

Model Diagnostic check

In order to pick the best model, a selection criterion was adopted using the highest coefficient of determination, R², lowest RMSE, and lowest BIC to filter the ARIMA model. The rule of parsimony was also adopted where applicable. Each model was further subjected to diagnostic check of its autocorrelation and partial autocorrelation plots of the residuals to determine that there were no spikes that extend beyond the continuous lines which imply the approximate 95% confidence limits. The candidate models that passed the diagnostic checks are shown below in Table 1 with their corresponding result of parameters. Based on the criteria for selection, the ARIMA model **(4,1,8)** was adopted as the best candidate for the subsequent forecast and its ACF and PACF plots for residuals is shown below in Figure 7. In order to validate the model, a diagnostic check was carried out on the normality of noise residuals as can be seen in Figure 8 shown below; the bell shaped distribution of the histogram indicates that the noise residual was normal. The Q-Q plot shown in Figure 9 also indicates normality of residuals since the points lie on or around the straight line, thus validating the model for use in forecasting future values of confirmed cases of covid-19 in Nigeria.

Table 1: Tentative non-seasonal ARIMA Models selected

S/No.	MODEL	R ²	RMSE	BIC
1	(4,1,7)	0.87	111.92	9.65
2	(4,1,8)	0.88	110.99	9.65
3	(5,1,4)	0.87	114.44	9.66
4	(5,1,8)	0.88	111.55	9.68
5	(8,1,8)	0.87	112.91	9.76

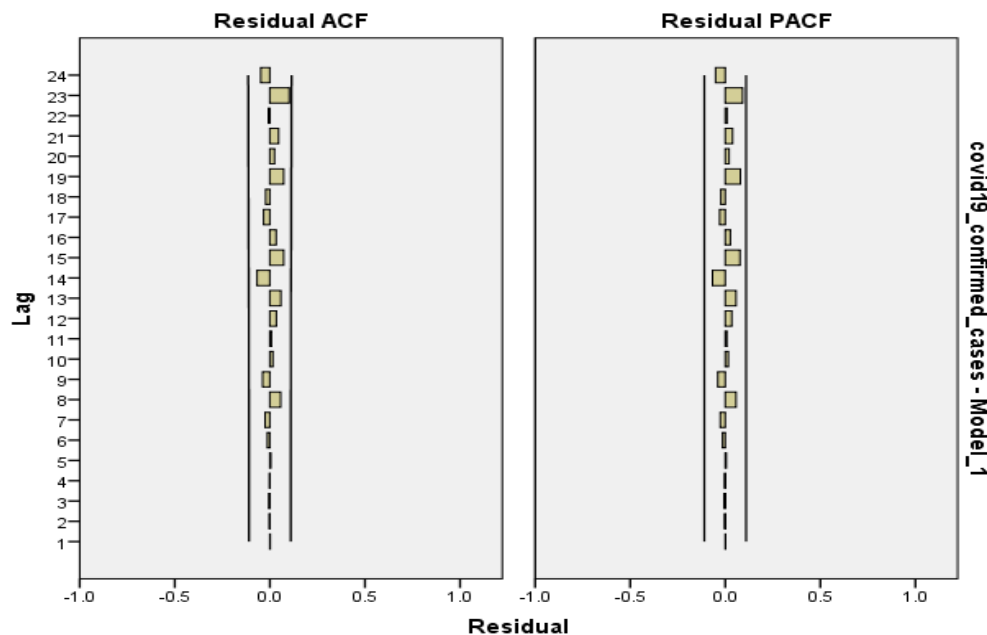


Figure 7: Autocorrelation and Partial autocorrelation plots of ARIMA model (4,1,8) residuals

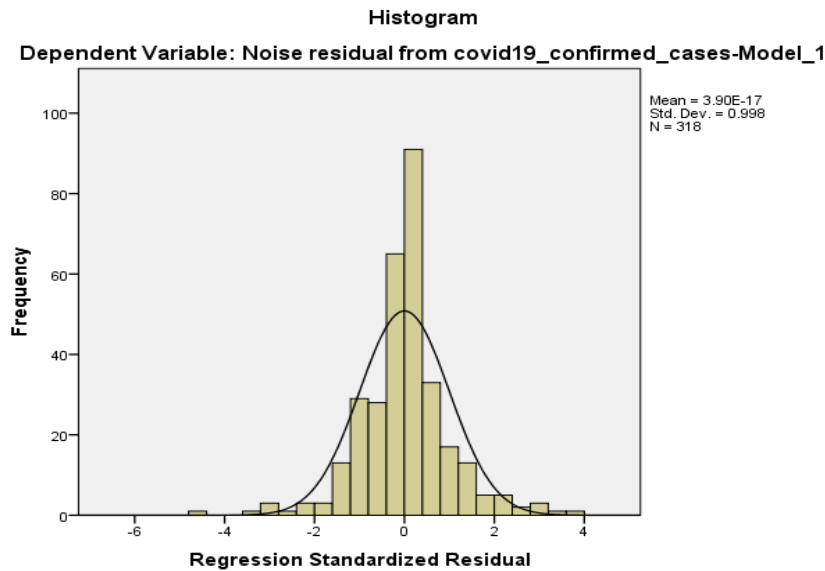


Figure 8: Histogram of Noise residuals distribution of confirmed cases

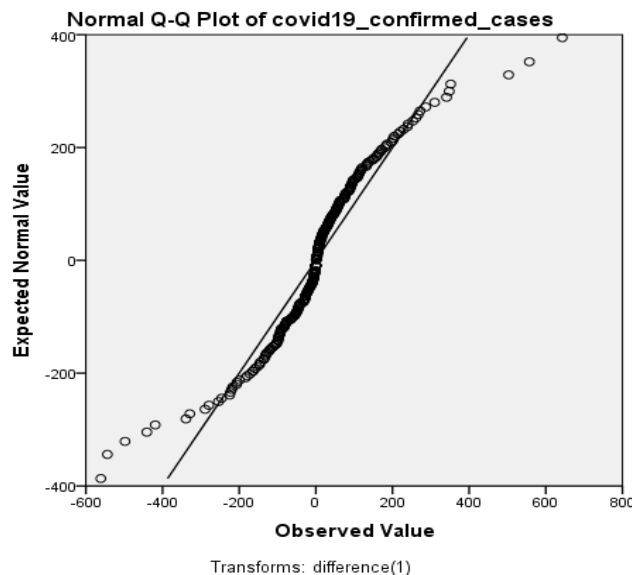


Figure 9: Normality diagnostic plot of the residuals of confirmed cases of covid-19 in Nigeria

Model Forecast

The ARIMA (4,1,8) was subsequently used to forecast for 30 days and the plot of the forecast is as shown in Figure 10 while the predicted values are accompanied by the lower class limits (LCL) and upper-class limits (UCL) displayed in Table 2. A look at the time series forecast plot shows that the predicted values assume an upward undulating trend which is similar to the previously recorded values of confirmed cases. However, the news of a covid-19 variant and the report of the daily confirmed cases increasing exponentially with an upward trend appears to illustrate the growing cases of infections in the “second wave” and lends credence to the fact that public health measures put in place by the Government at all

levels are failing either due to non-compliance of covid-19 protocols by citizens or inadequate enforcement of preventive measures put in place by the presidential task force (PTF).

With the variant of the covid-19 strain in existence and race against time to produce vaccines for the world's population, there is a need for the Nigerian government to tighten its grip in areas where the scourge has continued to grow unabated. This could be in the area of subjecting hotspots to quarantine measures, reduction of overseas travels to high-risk countries such as England and Russia, compulsory wearing of nose masks, subjecting defaulters of Government's prevention protocol measures to heavy fines and looking inward to provide a homegrown vaccine by fully supporting researchers just as it is done in developed countries. This would no doubt help to stem the tide of the covid-19 scourge in the country. The continued confirmed cases of this pandemic in the last ten months appear to show that the trend is a non-linear occurrence, hence, the higher values of autoregressive and moving average components of the model. There are also grave implications if public health measures put in place end up not being able to slow down the scourge. The forecasted values for the next one month are an indication that the pandemic in the country is not leaving any time soon.

Table 2: Predicted values for covid-19 confirmed cases in Nigeria

S/No	Predicted			
	Date	Confirmed cases	LCL Predicted	UCL Predicted
1	13-Jan-21	1498	1280	1716
2	14-Jan-21	1506	1264	1747
3	15-Jan-21	1339	1080	1599
4	16-Jan-21	1278	1004	1551
5	17-Jan-21	1150	864	1437
6	18-Jan-21	1172	873	1472
7	19-Jan-21	1383	1074	1691
8	20-Jan-21	1504	1170	1838
9	21-Jan-21	1504	1148	1861
10	22-Jan-21	1444	1079	1809
11	23-Jan-21	1280	906	1654
12	24-Jan-21	1173	791	1555
13	25-Jan-21	1257	869	1644
14	26-Jan-21	1395	994	1796
15	27-Jan-21	1505	1086	1925
16	28-Jan-21	1552	1121	1984
17	29-Jan-21	1453	1010	1895
18	30-Jan-21	1294	844	1745
19	31-Jan-21	1240	785	1694
20	1-Feb-21	1291	829	1753
21	2-Feb-21	1411	938	1885
22	3- Feb-21	1545	1060	2030
23	4- Feb-21	1566	1068	2063
24	5- Feb-21	1460	953	1968
25	6- Feb-21	1341	828	1854
26	7- Feb-21	1280	762	1798
27	8- Feb-21	1319	795	1844
28	9-Feb-21	1453	919	1986
29	10-Feb-21	1567	1022	2112
30	11-Feb-21	1573	1017	2130

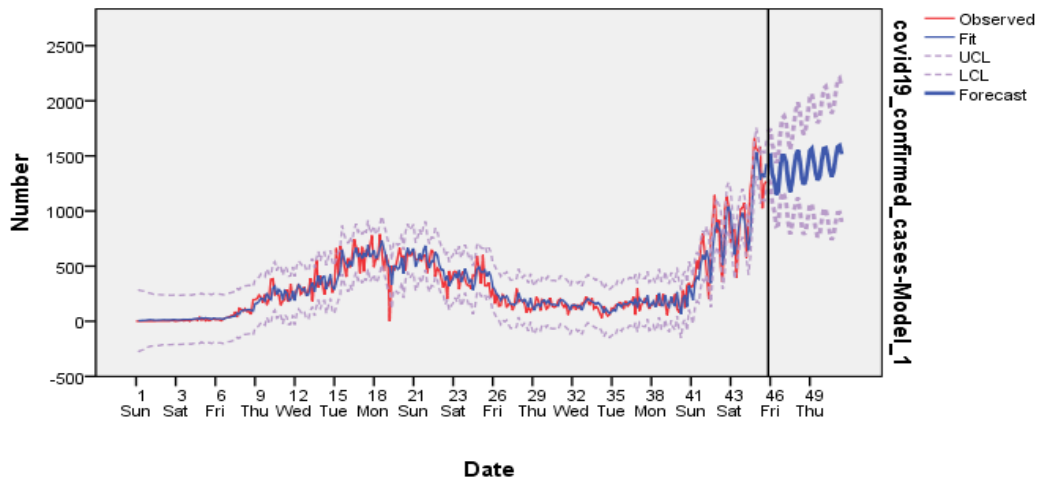


Figure 10: Forecast of ARIMA (4,1,8)model

4. CONCLUSIONS

The coronavirus pandemic “second wave” has continued to result in increased confirmed cases in Nigeria. An attempt was made to forecast the occurrence of the scourge for the next month and the result of the forecast shows an exponentially increasing trend in the coming weeks ahead. It is strongly suggested that the government at all levels, should consider adopting more stringent measures to slow down the scourge.

References

- Kotwal, A., Yadav, A. K., Yadav, J., Kotwal, J. and Khune, S. (2020). Predictive models of COVID19 in India: a rapid review. *Medical Journal of Armed Forces*, 76(4): 377–386.
- Li, Q., Guan, X. and Wu, P. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia; 2020.
- NBS, 2020. National Bureau of Statistics website. Retrieved from <https://www.nigerianstat.gov.ng/>.
- NCDC, 2020. Nigeria Centre for Disease Control. Available from: <https://ncdc.gov.ng>.
- Papastefanopoulos, V., Linardatos, P. and Kotsiantis, S. (2020). COVID-19: A comparison of time series methods to forecast percentage of active cases per population. *Applied Sciences*, 10(11): 3880
- Perone G. (2020). An ARIMA model to forecast the spread of COVID-2019 epidemic in Italy. *SSRN Electron Journal*.
- Shim, E. (2004). An Epidemic Model with Immigration of Infectives and Vaccination. B.Sc. thesis, The University of British Columbia, Canada.
- Singh, S., Sundram, B. M., Rajendran, K., Law, K. B., Aris, T., Ibrahim, H., Dass, S. C. and Gill, B.S. (2020). Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models. *The Journal of Infections in Developing Countries*, 14(9): 971-976.
- Tran, T. T., Pham, L. T. and Ngo, Q. X. (2020). Forecasting epidemic spread of SARS-CoV-2 using ARIMA model (Case study: Iran). *Global Journal of Environmental Science and Management*, 6(SI): 1-10 Autumn.
- Wang Y, Shen, Z. and Jiang, Y. (2018). Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China. *PloS One*, 13(9):1-11
- Wang, L. Y., Peng, B., Zhou, R., Zhan, C.Y. and Liu, Z. (2020). Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures. *Annals of Infectious Diseases and Epidemiology*, 5(1):1052.