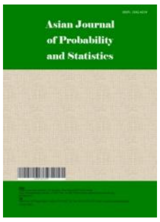


Asian Journal of Probability and Statistics

12(3): 59-80, 2021; Article no.AJPAS.67273
ISSN: 2582-0230

ARIMA Modelling and Forecasting of COVID-19 Daily Confirmed/Death Cases: A Case Study of Nigeria

Essi Isaac Didi[1], Nwuju Kingdom[1*] and Etuk Ette Harrison[1]

[1]Department of Mathematics, Rivers State University, Port Harcourt, Rivers State, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. Author EID designed the study, while author NK formed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author EEH managed the analyses of the study and also the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJPAS/2021/v12i330289
Editor(s):
(1) Prof. Ali Akgül, Siirt University, Turkey.
(2) Dr. Jiteng Jia, Xidian University, China.
(3) Dr. Halim Zeghdoudi, Badji-Mokhtar University, Algeria.
(4) Prof. Mervat Mahdy Ramadan Mahdy, Benha University, Egypt.
Reviewers:
(1) Lamia kalai, Université de Tunis El Manar, Tunisia.
(2) Janith Warnasekara, Rajarata University of Sri Lanka, Sri Lanka.
(3) Matteo Nioi, University of Cagliari, Italy.
(4) Johannes Stübinger, University Erlangen-Nürnberg, Germany.
Complete Peer review History: http://www.sdiarticle4.com/review-history/67273

Original Research Article

Received 22 March 2021
Accepted 21 April 2021
Published 26 April 2021

Abstract

Aims: The aim of this work is to develop suitable ARIMA models which can be sued to forecast daily confirmed/death cases of COVID-19 in Nigeria. This is subject to developing the model, checking them for suitability and carrying out eight months forecast, and making recommendations for the Nigerian Health sector.
Study Design: The study used daily confirmed and death cases of COVID-19 in Nigeria.
Methodology: This work covers times series data on the on the daily confirmed/death cases of COVID-19 in Nigeria, obtained from the Nigerian Centre for Disease Control (NDCD) from 21 March 2020 to 5 May 2020, covering a total of 51 data points. This work is geared towards developing a suitable Autoregressive Integrated Moving Average (ARIMA) models which can be used to forecast total daily confirmed/death cases of COVID-19 in Nigeria. Two adequate subset ARIMA (2, 2, 1) and AR (1) models for the confirmed/death cases, respectively, is fitted and discussed

**Corresponding author: Email: kingdom.nwuju@ust.edu.ng;*

Results: A forecast of 239 days – from 6th May 2020 to 31 December 2020 was conducted using the fitted models and we observed that the COVID19 data has an upward trend and is best forecasted within a short period.
Conclusion: Critical investigation into the rate of spread of COVID-19 pandemic has shown that, that the daily confirmed cases as well as death cases of the disease tends to follow an upward trend. This work aimed at developing a suitable ARIMA models which can be used to fit a most appropriate subsets to statistically forecast the actual number of confirmed cases as well as death cases of COVID-19 recorded in Nigeria for a period of 8 months.

Keywords: ARIMA; COVID-19; forecasting; time series.

1 Introduction

Coronavirus disease 2019 otherwise known as COVID-19, which is a sickness caused by the novel coronavirus which is known as Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which was first discovered amid an outbreak of respiratory sickness cases in Wuhan City, Hubei Provence, China, in December, 2019. This has become an issue of global apprehension. On March 11 2020, the World Health Organisation (WHO) declared the disease a global pandemic, after the virus spread from China to almost every part of the world now, with number of cases as at April 20, 2020, of about two million, four hundred and seventy-two thousand, and sixty-two (2,472,062) with over one hundred and seventy thousand (170,000) deaths recorded worldwide (John Hopkins University, 2020).

In Nigeria, it is not unlike as the index case was chronicled on February 28, 2020, and as at April 20, 2020, we had recorded over 600 cases with about 21 death cases, according to the Nigerian Centre for Disease Control (NCDC). Over the past few months, the federal government under the leadership of His Excellency, President Mohammadu Buhari have taken drastic and containment measure to curb the spread of the Covid-19 disease. Some of this measure include lockdown of the three major places where the Covid-19 has been recorded which includes Lagos State, Ogun State, and the federal capital territory (FCT) Abuja. Other measure being put in place to combat the spread of the virus is the setting up of isolation/quarantine centres in major cities within the country, shutting down of the country's land, air and sea borders for the time being, enacting a 'stay at home' policy, as well as distribution of relief materials to the poor masses especially those been displaced from their homes due to insurgences in the North Eastern part of the country.

In this study, daily data on confirmed cases as well as death cases of COVID-19 in Nigeria shall be used in this study, in other to understand the trend associated with COVID-19 in Nigeria. Also, models will be fitted for future predictions of the state of the daily spread of the disease using its already existing data obtained from the Nigerian Centre for Disease Control (NCDC), with the aim of describing the pattern in which COVID-19 in Nigeria has followed since recording of the index case, to conduct an intervention analysis for the Nigerian health sector and the economy at large. This work will assist in determining if the daily confirmed cases of COVID-19 is growing steadily, slowly or rapidly.

This work covers times series data on the on the daily confirmed/death cases of COVID-19 in Nigeria, obtained from the Nigerian Centre for Disease Control (NDCD) from 21 March 2020 to 5 May 2020, covering a total of 51 data points.

Several types of ARIMA models have been proposed by different scholar over the years. The Autoregressive Moving Average (ARMA) approach was introduced by Box G, et al. [1,2]. Many authors have tried to use the ARIMA model to study diseases in other parts of the world. Some instance, Jing C, et al. [3], Carried out a study on predicting seasonal influenza based on SARIMA model, in Mainland China from 2005 to 2008. They discovered in their study that seasonal influenza is one of the mandatorily monitored infectious diseases, in China. Making full use of the influenza surveillance data helps to predict seasonal influenza. In this study, a seasonal autoregressive integrated moving average (SARIMA) model was used to predict the influenza changes by analyzing monthly data of influenza incidence from January 2005 to December 2018, in China. In another study, Nwuju K & Lekara-bayo I. B. [4] carried out a study on modelling and forecasting daily confirmed cases of COVID-19 in Africa, considering ECOWAS as a case study area. In their work they revealed that a crucial investigation into the rate of spread of Coronavirus Disease 2019 (COVID-19) pandemic has shown that, the daily established cases of the disease have a tendency to to follow an upward trend. Their work which was intended to develop a suitable Autoregressive Integrated Moving Average (ARIMA) model which can be used to statistically forecast the actual number of confirmed cases of COVID-19 recorded in ECOWAS as a whole. They suggested an adequate subset ARIMA (5, 1, 1) model which was fitted and discussed. A forecast of 235 days from 11th May 2020 to 31st December 2020, was conducted using the fitted model, and they discovered that the COVID-19 daily confirmed cases may most likely incline over the next six months.

In the study of Dong et al. [5], carried out a study to understand the epidemiology of COVID-19 in China. In their study, they tried to develop the best trend showing the epidemic curve of the disease in China. They also contrasted onset-to-diagnosis curve for the data they obtained from the Chinese Centre for Disease Control (CCDC), and then fitted a log-normal distribution to it. They discovered amongst other things that more than 90% of all patients investigated showed asymptomatic, mild, or even moderate cases. They also discovered that the number of children infected by the disease were more in Hubei province as compared to other parts of china. They concluded that the number of children living with COVID-19 is not certain as there exists a strong evidence of human-to-human transmission. In another study, Stübinger and Schneider [6], the epidemiology of COVID-19 was studied with the aim of forecasting future occurrences in different countries of the world. They applied the concept of dynamic wrapping to understand the past recorded cases of the disease in the nation sunders study. They obtained from their analysis that China is leading with in the range of 29 days for South Korea and 44 days for the United States. Their model also predicts a future collapse of the UK's healthcare system if not properly managed especially in the face of the pandemic.

Sahai et al. [7], also modelled and forecasted COVID-19 in top five affected countries of the world by Fitting ARIMA model to the data. The countries included in the study includes, Brazil, Russia, Spain, India, and the USA. They estimated the ARIMA model by using Hannan and Rissanen algorithm. A forecast for the next 70 days was carried out using the best fit model. In their result they observed that forecast carried out for the first 18 days of July, indicates that Russia and Spain ay have attained point of inflexion, while USA, India and Brazil is still experiencing exponential increase. Their result tend to help the government of these countries make adequate preparations towards combating the disease. In another study, Qiuying et al. [8], studied the spread of COVID-19 using Hunei, China data to understand the spread in Italy. They understood that since the outbreak of the disease in China around December 2019, many scholars worldwide have tried to study the growing trend of confirmed as well as the death rate that is associated with the disease. They model obtained as best fit was used to forecast the number of confirmed cases as well as death cases in Italy for the next ten days. They discovered that the daily number of case may likely follow an upward trend and the model provides basis for to understanding the development of pandemic in some other countries. In the study of Napoli et al. [9], an investigation of the global spread of COVID-19 and Malaria was conducted as an epidemiological paradox in an juvenile stage of a pandemic. They compared the average rate of spread of COVID-19 and Malaria. They discovered that the epidemic presence of Malaria tend to protect some countries from the risk of contracting the novel Coronavirus disease. They discovered that the protection could be due to the presence of anti-malarial found in the system of these people, while it was also discovered that some part of the world where mosquitoes are in abundance, has overtime developed a natural immunity to such virus and diseases.

Other Applications of The ARIMA model includes the work of Etuk E. H., et al. [10] proposed a seasonal Box-Jenkins Model for Nigerian Inflation rate series. They obtained a seasonal difference as well as a non-sessional difference. The correlogram of the differenced series they obtained, revealed a seasonal nature. It also revealed a seasonal autoregressive component. They fitted a $(5, 1, 0)(0\ 1, 1)_{12}$ seasonal model which was shown to be adequate for the data studied. Nwuju et al. [11], used Box-Jenkins ARMA model in their work on intervention analysis of daily South African Rand/Nigerian Naira exchange rates. In their work they aimed at development a pre-intervention and post-intervention analysis of the exchange rates between the currencies of the two countries involved. They discovered that due to the impact of the economic recession experienced in the country in recent times, a good study of what the exchange rate of foreign currencies is against the Naira is very important. In their study a suitable subset ARIMA (12, 0, 2) was selected to be the best fit, following the results of the residuals and the model selection criteria including AIC and BIC. The data was divided in preintervention part, intervention part, and postintervention part. They suggested that the management of exchange rates should be conducted using the suggested model. The residual diagnostics indicates that the selected model is very adequate and the correlogram of residuals shows no significant lag spike for both PACF and ACF, a clear indication of a white noise process. They discovered that the exchange rate between the Sound African Rand/ Nigerian Naira will gradually increase over the next 12 months.

This study The aim of this work is to develop suitable ARIMA models which can be used to forecast total daily confirmed/death cases of COVID-19 in Nigeria.

The objectives are:

i. to develop a suitable ARIMA models for the Nigerian COVID-19 data
ii. to check the selected model for suitability using residual diagnostic tools
iii. to carry out an eight (8) months future forecast of daily confirmed/death cases of the disease in Nigeria.
iv. to make recommendation to policy makers and the Nigerian Centre for Disease Control (NCDC) on ways to flatten the curve with respect to COVID-19.

2 Materials and Methods

2.1 Data

This work covers times series data on the on the daily confirmed/death cases of COVID-19 in Nigeria as shown on Fig. 1 and Fig. 10. as obtained from Our World in Data https://ourworldindata.org/coronavirus/country/nigeria [12], from 20 March 2020 to 5 May 2020, covering a total of 51 data points. The used data is listed in the appendix A.

2.2 Methodology

The methodology adopted in this work is the Box-Jenkins methodology proposed by Box G. E. P. et al. (1970). They introduced the Autoregressive Moving Average (ARMA) approach, in their work on Time series analysis: forecasting and control. Nonstationary models time series analysis was introduced by them, and methods were discussed for fitting them to data, for forecasting business and economic series, and applying them to feedback and feedforward control schemes.

2.2.1 An AR model

An autoregressive model of order p, AR(P), can be represented as:

$$X_t = c + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + ... + \alpha_p X_{t-p} + \varepsilon_t; \quad t = 1, 2, ..., T, \tag{1}$$

where $\{\varepsilon_t\}$ is the error term in the equation; a white noise process, a sequence of independently and identically distributed (iid) random variables with $E(\varepsilon_t) = 0$ and $var(\varepsilon t) = \sigma^2$; i.e. $\varepsilon_t \sim iid\ N(0, \sigma^2)$, and the α's are the model parameters. All previous values can have additive effects on this level X_t and so on. So, it is a long-term memory model.

2.2.2 A moving average model

A time series { } is said to be a moving-average process of order q, MA (q), if:

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}; t = 1, 2, ..., T, \tag{2}$$

This model is given in terms of past error terms as descriptive variables. Consequently, on q errors will have effects on X_t. However, higher order errors does not have effect on X_t; which means that the Moving-average model, is a short term model.

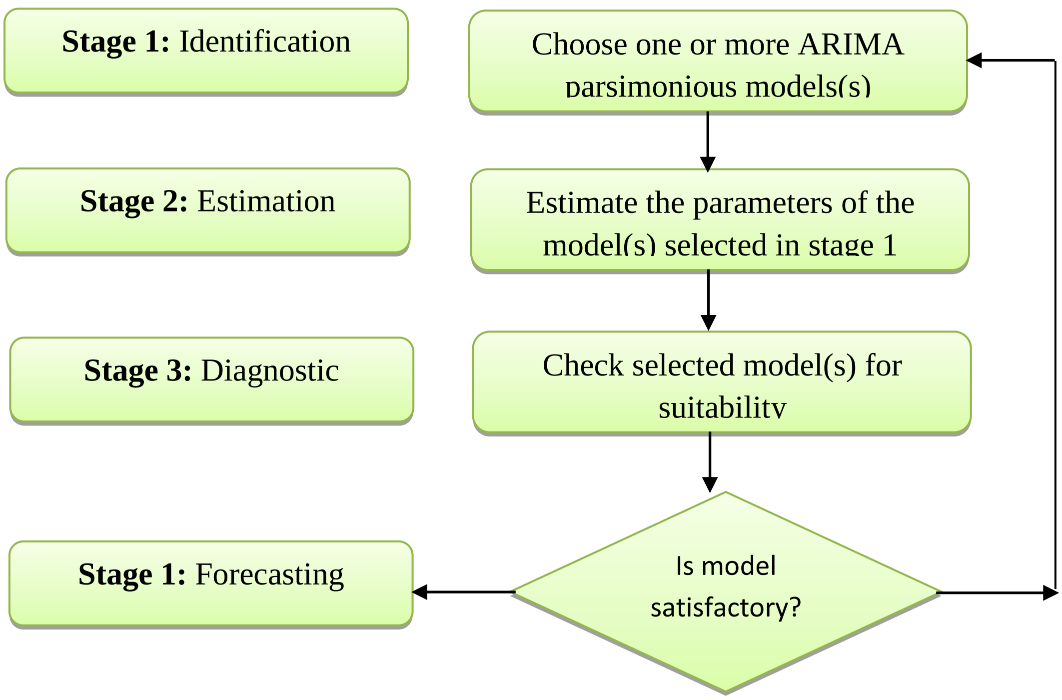


Fig. 1. Flowchart of Stages in the Box-Jenkins' iterative approach

2.2.3 An ARMA model

A time series X_t is said to follow an autoregressive moving-average process of order p and q, i.e. ARMA(p, q), process if:

$$X_t = c + \alpha_1 X_{t-1} + \alpha_1 X_{t-2} + \ldots + \alpha_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} = \varepsilon_t \tag{3}$$

Where

X_t	=	Target variable
c	=	constant term (which allows for a non-zero means)
α_p	=	parameters of the autoregressive part of the model
θ_q	=	parameters of the moving average part of the model
ε_t	=	the stochastic error terms

In summation form we have

$$X_t = \sum_{k=1}^{p} \alpha_k X_{t-k} - \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t + c \tag{4}$$

2.2.4 Autoregressive Integrated Moving-average (ARIMA) Model

The ARMA models can further be extended to non-stationary series by allowing the differencing of the data series resulting to ARIMA models. The general non-seasonal model is known as ARIMA (*p, d, q*): where with three parameters; p is the order of autoregressive, d is the degree of differencing, and q is the order of moving-average.

Thus, a process X_t is said to be an auto regressive integrated moving average (*ARIMA (p,d,q)*) process if:

$$\Delta^d X_t = \sum_{k=1}^{p} \alpha_k \Delta^d X_{t-k} - \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t + c \tag{5}$$

is ARMA (p,q) process. *d* is a non-negative integer representing the differencing order. Where $\Delta X_t = X_t - X_{t-1}$. But if p = q = 0 in equation (3.4), then the model becomes a random walk model which classified as ARIMA (0, 1, 0).

2.3 Box-Jenkins' Approach

In time series analysis, the Box-Jenkins (1970, 1976) approach, named after the statisticians George Box and Gwilym Jenkins, applies ARIMA models to find the best fit of a time series model to past values of a time series.

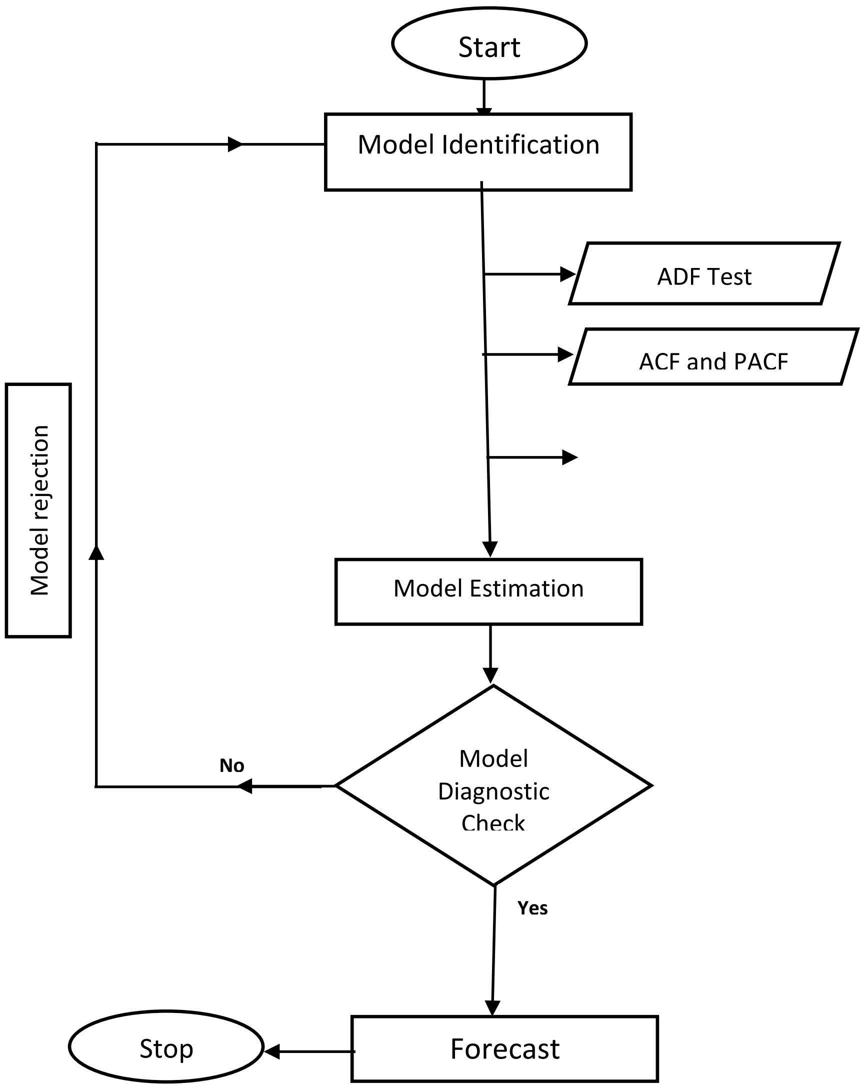


Fig. 2. Detailed flow chart of the box-jenkins' methodology

All plots and numerical computations will be carried out using E-views version 10 on a Windows 10 personal computer with the following specifications.

Processor: Intel® Core i5-3230M CPU @ 2.60GHz
Installed Memory (Ram): 4.00GB
System type: 64-bits Operating System.

3 Results and Discussion

3.1 Results for the Daily Confirmed Cases

3.1.1 Stationarity of the Data.

In (Fig. 3) below we observe that the daily confirmed cases of COVID-19 data in Nigerian follows a fluctuating but upward trend. Thus it was investigated for stationarity using the Augmented Dickey-Fuller (ADF) Test as shown on (Table 1) which shows that indeed there is a trend associated with the original data as the ADF of the actual data shows that the data is not stationary. First difference of the data was taken and the ADF test carried out on it. Also it is seen that the first difference of the data did not also pass the stationarity test as seen (Table 2). Finally the data was made stationary by taking the second difference of the original data and the ADF test of the second difference as shown on (Table 3) is stationary as the p-values are highly significant.

The Correlogram – the plot of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) against corresponding lags of the data as shown on Fig. 4 showed evidence of non-stationarity of the original data as there was a gradual decrease of significant lag spikes. However, the correlogram of the second difference of the data as shown on Fig. 4 shows evidence of a stationary series, and thus was used to selected possible models with parsimony.

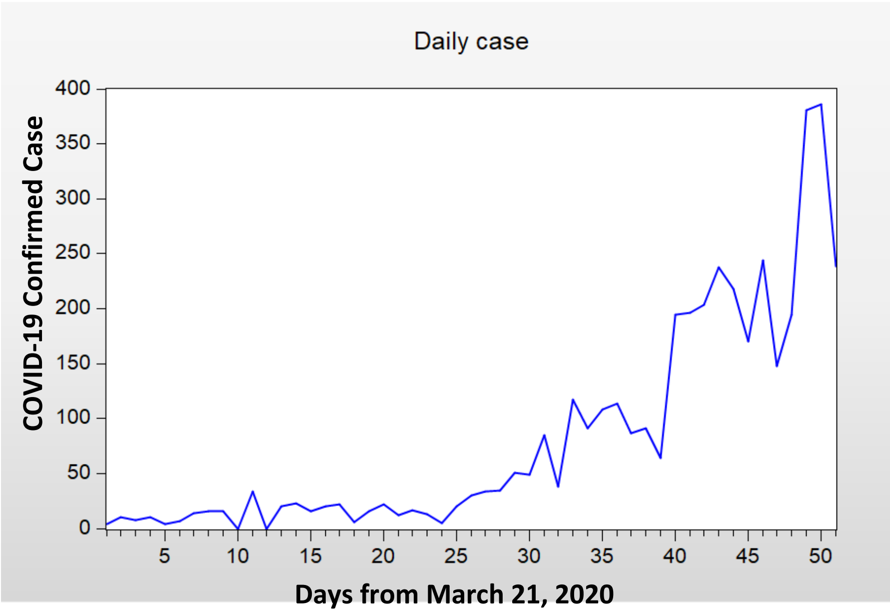
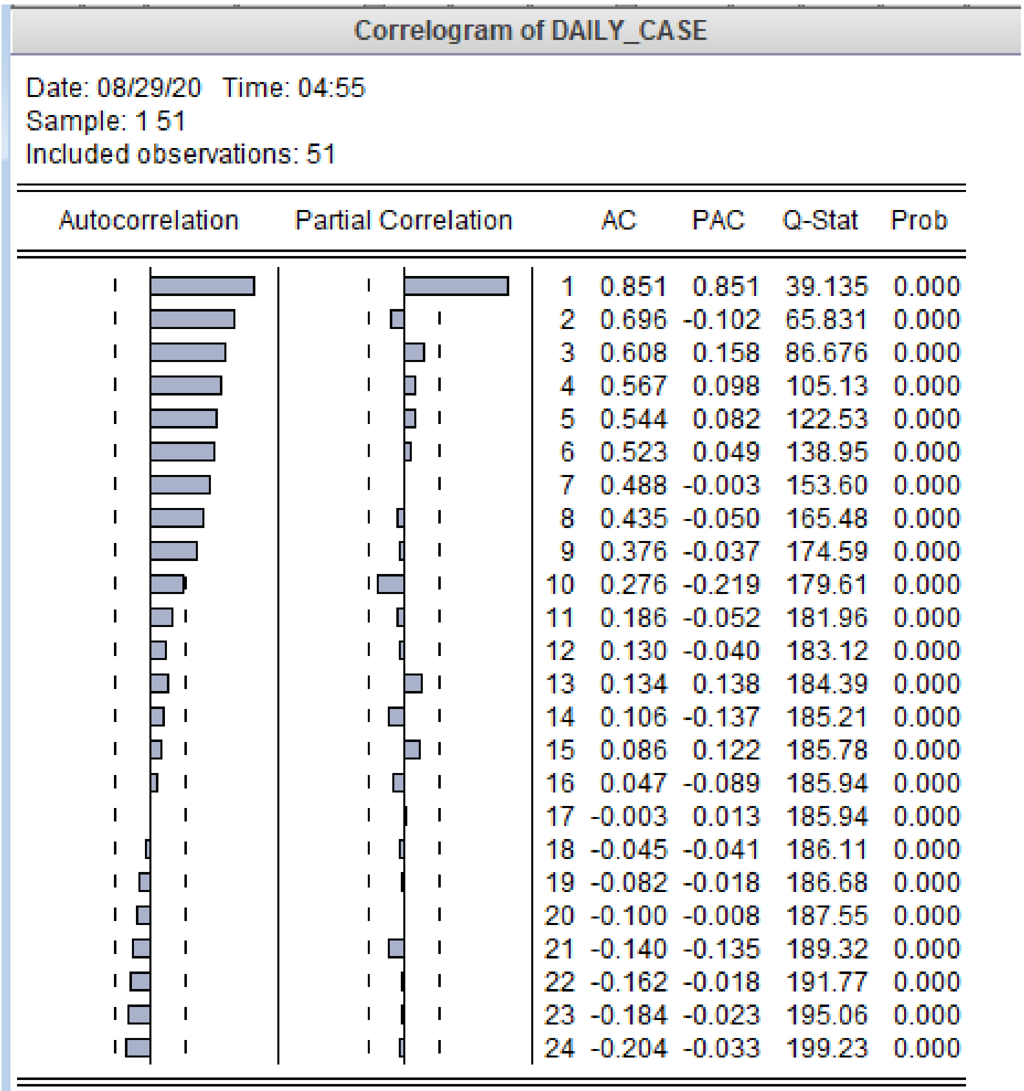


Fig. 3. Time plot of total daily confirmed cases of COVID-19 in Nigeria from 21 March 2020 to 5 May 2020

It can be seen in Table 1 below that the probability of ADF = 0.9750 which is greater than the critical value of the significance level of 0.01, 0.05 and 0.1, that is to say that the original COVID-19 sequence is non-stationary.

A visual examination of Fig. 4 and Table 1 above confirms that the Nigerian COVID-19 series in non-stationary as the lag of the ACF gradually drops to zero. This data will be stationarised by carrying out differencing. Before embarking on further analysis using the Box-Jenkins approach, the data has to be transformed to achieve stationary.

It can be seen in (Table 1) above that the probability of ADF = 0.9750 which is greater than the critical value of the significance level of 0.01, 0.05 and 0.1, that is to say that the original COVID-19 sequence is non-stationary. Also, the ADF test of the second difference as shown on (Table 2) indicates that the first difference data not stationary. However, stationary of the data was attained in the second differencing as shown in (Table 3).


Correlogram of DAILY_CASE

Date: 08/29/20 Time: 04:55
Sample: 1 51
Included observations: 51

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.851	0.851	39.135	0.000
		2	0.696	-0.102	65.831	0.000
		3	0.608	0.158	86.676	0.000
		4	0.567	0.098	105.13	0.000
		5	0.544	0.082	122.53	0.000
		6	0.523	0.049	138.95	0.000
		7	0.488	-0.003	153.60	0.000
		8	0.435	-0.050	165.48	0.000
		9	0.376	-0.037	174.59	0.000
		10	0.276	-0.219	179.61	0.000
		11	0.186	-0.052	181.96	0.000
		12	0.130	-0.040	183.12	0.000
		13	0.134	0.138	184.39	0.000
		14	0.106	-0.137	185.21	0.000
		15	0.086	0.122	185.78	0.000
		16	0.047	-0.089	185.94	0.000
		17	-0.003	0.013	185.94	0.000
		18	-0.045	-0.041	186.11	0.000
		19	-0.082	-0.018	186.68	0.000
		20	-0.100	-0.008	187.55	0.000
		21	-0.140	-0.135	189.32	0.000
		22	-0.162	-0.018	191.77	0.000
		23	-0.184	-0.023	195.06	0.000
		24	-0.204	-0.033	199.23	0.000

Fig. 4. Correlogram of the raw data showing non-stationarity as seen by the gradual exponential decrease of the ACF

Table 1. Augmented Dickey-Fuller (ADF) Test of the Actual Data

Null Hypothesis: DAILY_CASE has a unit root			
Exogenous: Constant			
Lag Length: 10 (Automatic - based on SIC, maxlag=10)			
		t-Statistic	**Prob.***
Augmented Dickey-Fuller test statistic		0.291797	0.9750
Test critical values:	1% level	-3.605593	
	5% level	-2.936942	
	10% level	-2.606857	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DAILY_CASE)
Method: Least Squares
Date: 05/21/20 Time: 07:33
Sample (adjusted): 12 51
Included observations: 40 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DAILY_CASE(-1)	0.080685	0.276510	0.291797	0.7726
D(DAILY_CASE(-1))	-0.896228	0.431920	-2.074985	0.0473
D(DAILY_CASE(-2))	-0.747676	0.506390	-1.476484	0.1510
D(DAILY_CASE(-3))	-0.373319	0.504603	-0.739827	0.4656
D(DAILY_CASE(-4))	-0.357628	0.493737	-0.724330	0.4749
D(DAILY_CASE(-5))	-0.550163	0.565296	-0.973229	0.3388
D(DAILY_CASE(-6))	-0.075801	0.586288	-0.129290	0.8981
D(DAILY_CASE(-7))	-0.056375	0.547344	-0.102997	0.9187
D(DAILY_CASE(-8))	0.200671	0.537355	0.373443	0.7116
D(DAILY_CASE(-9))	1.120813	0.509815	2.198471	0.0363
D(DAILY_CASE(-10))	0.943664	0.356420	2.647619	0.0132
C	8.517553	8.531029	0.998420	0.3266

R-squared	0.644823	Mean dependent var	5.125000
Adjusted R-squared	0.505289	S.D. dependent var	52.67714
S.E. of regression	37.05084	Akaike info criterion	10.30578
Sum squared resid	38437.42	Schwarz criterion	10.81245
Log likelihood	-194.1157	Hannan-Quinn criter.	10.48898
F-statistic	4.621261	Durbin-Watson stat	1.954381
Prob(F-statistic)	0.000509		

Table 2. Augmented Dickey-Fuller (ADF) Test of the 1st difference of the data

Null Hypothesis: DAILY_CASED1 has a unit root
Exogenous: Constant
Lag Length: 9 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.461052	0.8883
Test critical values:	1% level	-3.605593	
	5% level	-2.936942	
	10% level	-2.606857	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DAILY_CASED1)
Method: Least Squares
Date: 09/17/20 Time: 02:45
Sample (adjusted): 12 51
Included observations: 40 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DAILY_CASED1(-1)	-0.521303	1.130682	-0.461052	0.6482
D(DAILY_CASED1(-1))	-1.258992	1.120037	-1.124063	0.2702
D(DAILY_CASED1(-2))	-1.867547	1.070923	-1.743867	0.0918
D(DAILY_CASED1(-3))	-2.110705	0.984975	-2.142903	0.0406
D(DAILY_CASED1(-4))	-2.343811	0.878679	-2.667426	0.0124
D(DAILY_CASED1(-5))	-2.748205	0.769737	-3.570317	0.0013
D(DAILY_CASED1(-6))	-2.672727	0.674411	-3.963052	0.0004
D(DAILY_CASED1(-7))	-2.590498	0.583069	-4.442865	0.0001
D(DAILY_CASED1(-8))	-2.254496	0.469900	-4.797815	0.0000
D(DAILY_CASED1(-9))	-1.008986	0.272937	-3.696765	0.0009
C	8.916292	8.286986	1.075939	0.2908

R-squared	0.836565	Mean dependent var	-4.525000
Adjusted R-squared	0.780208	S.D. dependent var	77.77350
S.E. of regression	36.46174	Akaike info criterion	10.25882
Sum squared resid	38554.31	Schwarz criterion	10.72326
Log likelihood	-194.1764	Hannan-Quinn criter.	10.42675
F-statistic	14.84406	Durbin-Watson stat	2.003264
Prob(F-statistic)	0.000000		

Table 3. Augmented Dickey-Fuller Test of the 2nd differenced data

Null Hypothesis: DAILY_CASED2 has a unit root			
Exogenous: Constant			
Lag Length: 8 (Automatic - based on AIC, maxlag=10)			
		t-Statistic	**Prob.***
Augmented Dickey-Fuller test statistic		-8.190847	0.0000
Test critical values:	1% level	-3.605593	
	5% level	-2.936942	
	10% level	-2.606857	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DAILY_CASED2)
Method: Least Squares
Date: 09/17/20 Time: 02:52
Sample (adjusted): 12 51
Included observations: 40 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DAILY_CASED2(-1)	-22.43785	2.739381	-8.190847	0.0000
D(DAILY_CASED2(-1))	19.66811	2.630080	7.478140	0.0000
D(DAILY_CASED2(-2))	17.32122	2.436569	7.108857	0.0000
D(DAILY_CASED2(-3))	14.78673	2.149689	6.878544	0.0000
D(DAILY_CASED2(-4))	12.08625	1.812190	6.669414	0.0000
D(DAILY_CASED2(-5))	9.044575	1.454472	6.218461	0.0000
D(DAILY_CASED2(-6))	6.146089	1.062300	5.785645	0.0000
D(DAILY_CASED2(-7))	3.395509	0.649895	5.224702	0.0000
D(DAILY_CASED2(-8))	1.044755	0.258222	4.045963	0.0003
C	6.225344	5.805216	1.072371	0.2921

R-squared	0.939614	Mean dependent var	-5.050000
Adjusted R-squared	0.921498	S.D. dependent var	128.4170
S.E. of regression	35.98004	Akaike info criterion	10.21612
Sum squared resid	38836.91	Schwarz criterion	10.63834
Log likelihood	-194.3225	Hannan-Quinn criter.	10.36879
F-statistic	51.86724	Durbin-Watson stat	2.020634
Prob(F-statistic)	0.000000		

3.1.2 Model Estimation for the Daily Case

Following the correlogram on Fig. 4 above, estimation of models with parsimony shall now be estimated using the model estimation command in E-views as the result is presented in Table 4 below.

From the parsimonious models estimation following the significant lags of the PACF for the *AR(p)* and ACF for he *MA(q)* of the correlogram of the differenced data above, only *AR(1) AR(2*) and *MA(1)* are significant at a 0.05 confidence level. These indicates that the selected subset ARIMA (2, 2, 1) is most likely to be adequate.

Therefore, the selected model is ARIMA (2, 2, 1) given by:

$$\Delta^2 X_t = -0.384371\Delta^2 X_{t-2} - 0.477466\Delta^2 X_{t-1} - 0.941409\varepsilon_{t-1} + \varepsilon_t \tag{6}$$

Now, the future forecast of the series will computed using the general form of the model

$$\Delta^2 \hat{X}_t(l) = -0.384371\Delta^2 \hat{X}_{t-1}(l) - 0.477466\Delta^2 \hat{X}_{t-2}(l) - 0.941409\,\varepsilon_{t+l-1} \tag{4.3}$$

For example, the first out of sample forecast is given by

$$\Delta^2 \hat{X}_t(1) = -0.384371\Delta^2 \hat{X}_{t-1}(1) - 0.477466\Delta^2 \hat{X}_{t-2}(1) - 0.941409\,\varepsilon_{t+1-1} \tag{4.4}$$

This shall be discussed in details later in the forecast session.

3.1.3 Residual diagnostics for confirmed cases

We shall now run diagnostics on the residuals obtained from the ARIMA (2, 2, 1) model. This is used to check for possible significant lags in the residue. If there exit one, then that lag in included in the model for adequacy, but where the correlogram of the residual shows no significant spike – a white noise process, then we shall conclude that the selected model is adequate and suitable for forecasting the series under study.

Fig.. 5 below shows the correlogram of residuals of the ARIMA (2, 2, 1) model. We see clearly from above that the correlogram of residuals shows no significant lag spike, that is an evidence of a white noise process. Therefore, the selected model is adequate for forecasting.

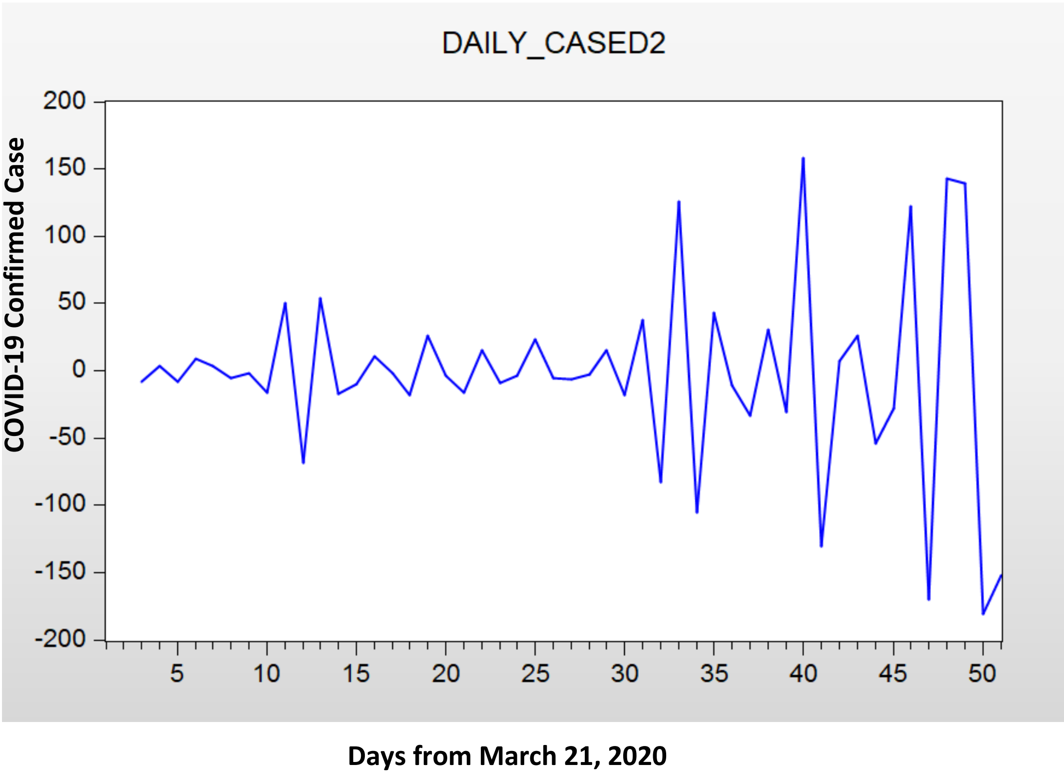
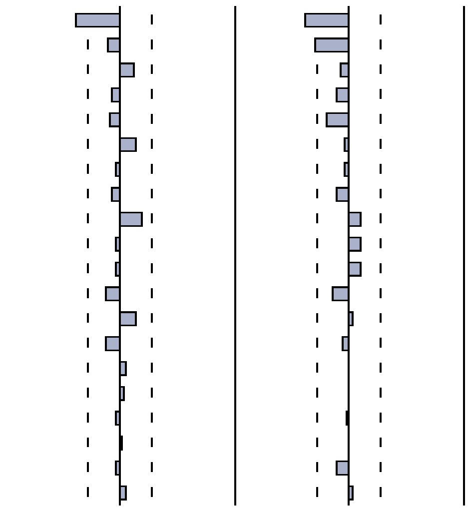


Fig. 5. Time plot of the 2nd differenced data differenced Daily Confirmed Cases data

3.1.4 Forecasting of the daily confirmed cases

Forecasting is carried out by fitting a linear trend equation to the interpolated data and the results are shown on (Table 5) and (Fig. 8) below. The (Fig. 8) below shows also an out-of-sample (future) forecast of the data for a total period of eight(8) months. (between May and December, 2020).

Date: 09/17/20 Time: 03:00
Sample: 1 51
Included observations: 49



	AC	PAC	Q-Stat	Prob
1	-0.403	-0.403	8.4742	0.004
2	-0.101	-0.315	9.0182	0.011
3	0.123	-0.076	9.8455	0.020
4	-0.077	-0.100	10.171	0.038
5	-0.093	-0.190	10.659	0.059
6	0.136	-0.031	11.734	0.068
7	-0.042	-0.033	11.838	0.106
8	-0.081	-0.106	12.237	0.141
9	0.191	0.107	14.519	0.105
10	-0.038	0.112	14.613	0.147
11	-0.037	0.110	14.702	0.197
12	-0.125	-0.154	15.762	0.202
13	0.155	0.040	17.440	0.180
14	-0.123	-0.063	18.511	0.184
15	0.061	-0.000	18.782	0.224
16	0.038	0.002	18.894	0.274
17	-0.047	-0.008	19.068	0.325
18	0.014	0.001	19.083	0.387
19	-0.034	-0.107	19.178	0.445
20	0.061	0.033	19.503	0.489

Fig. 6. Correlogram of the second difference of the daily confirmed cases data

Table 4. Parsimonious models estimation for the confirmed cases

Dependent Variable: DAILY_CASED2
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 09/17/20 Time: 03:03
Sample: 3 51
Included observations: 49
Convergence achieved after 43 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.384371	0.131224	-2.929115	0.0053
AR(2)	-0.477466	0.103240	-4.624818	0.0000
MA(1)	-0.941409	0.082324	-11.43548	0.0000
SIGMASQ	1790.675	253.8917	7.052908	0.0000

R-squared	0.633284	Mean dependent var	-3.122449
Adjusted R-squared	0.608837	S.D. dependent var	70.60265
S.E. of regression	44.15706	Akaike info criterion	10.57144
Sum squared resid	87743.06	Schwarz criterion	10.72587
Log likelihood	-255.0002	Hannan-Quinn criter.	10.63003
Durbin-Watson stat	2.052563		

Inverted AR Roots	-.19+.66i	-.19-.66i
Inverted MA Roots	.94	

Date: 09/17/20 Time: 03:06
Sample: 1 51
Included observations: 49
Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob

Included observations: 51
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.082	-0.082	0.3593	
		2	-0.031	-0.038	0.4132	
		3	0.050	0.044	0.5529	0.457
		4	0.057	0.064	0.7404	0.691
		5	-0.057	-0.045	0.9347	0.817
		6	0.178	0.174	2.8390	0.585
		7	0.087	0.110	3.2997	0.654
		8	0.043	0.077	3.4166	0.755
		9	0.194	0.216	5.8418	0.558
		10	-0.009	0.010	5.8471	0.664
		11	-0.120	-0.111	6.8248	0.655
		12	-0.103	-0.191	7.5627	0.671
		13	0.073	-0.038	7.9408	0.719
		14	-0.087	-0.128	8.4920	0.746
		15	0.028	-0.071	8.5514	0.806
		16	-0.024	-0.089	8.5962	0.856
		17	-0.066	-0.087	8.9383	0.881
		18	-0.089	-0.073	9.5842	0.887
		19	-0.057	-0.052	9.8562	0.910
		20	-0.044	0.054	10.028	0.931
		21	-0.103	-0.014	10.978	0.925
		22	-0.064	-0.059	11.361	0.936
		23	-0.085	-0.069	12.056	0.938
		24	-0.034	-0.020	12.174	0.954

Fig. 7. Correlogram of residuals of the ARIMA(2, 2, 1) model.

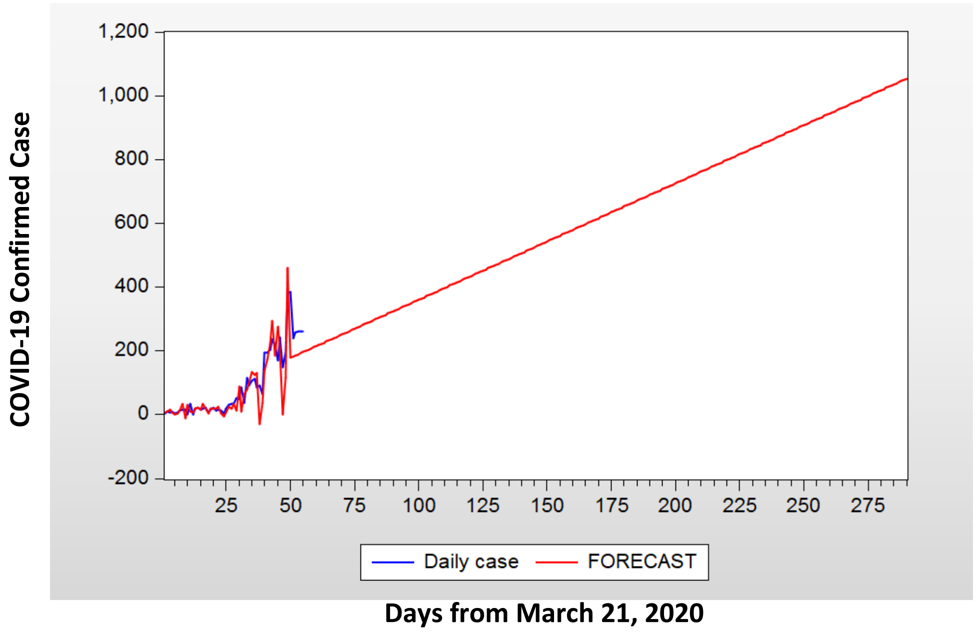


Fig. 8. Actual/Forecasted data time plot following the ARIMA (2, 2, 1) model selected.

Table 5. Forecasting trend model for the confirmed cases

Dependent Variable: DAILY_CASED2
Method: ARMA Maximum Likelihood (BFGS)
Date: 10/12/20 Time: 20:32
Sample: 51
Included observations: 50
Convergence achieved after 26 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@TREND	0.016273	0.006587	2.470648	0.0174
AR(1)	-0.440113	0.134718	-3.266920	0.0021
AR(2)	-0.533603	0.155411	-3.433490	0.0013
MA(1)	-0.999998	3050.301	-0.000328	0.0297
SIGMASQ	1609.048	145202.8	0.011081	0.9912

R-squared	0.670480	Mean dependent var	-3.122449
Adjusted R-squared	0.640524	S.D. dependent var	70.60265
S.E. of regression	42.33077	Akaike info criterion	10.54774
Sum squared resid	78843.34	Schwarz criterion	10.74079
Log likelihood	-253.4197	Hannan-Quinn criter.	10.62098
Durbin-Watson stat	2.084689		

Inverted AR Roots	-.22+.70i	-.22-.70i
Inverted MA Roots	1.00	

3.2 Results for the Daily Death Cases

We shall now put forward the results from the analysis of the daily death cases. This shows the number od daily deaths traceable to COVID-19 disease in Nigerian.

In (Fig. 9) below, we see that there is a trend associated with the recorded death cases of COVID-19 in Nigeria. Thus, the data shall be investigated also for stationarity using the ADF test as like it was carried out for the confirmed cases data.

(Table 6) presents the Augmented Dickey-Fuller Test of the daily death cases data which shows that the data is stationary and hence shall be used for model estimation without any form of differencing.

3.2.1 Stationarity test

This show that the daily death cases of COVID-19 in Nigeria is a stationary data.

3.2.2 Estimation of the daily death cases model

From the parsimonious models estimation following the significant lags of the PACF for the *AR(p)* and ACF for he *MA(q)* of the correlogram of the differenced data above, only *AR(1) is* significant at a 0.05 confidence level. Therefore, the selected model is AR (1) given by:

$$X_t = 0.929997 X_{t-1} + \varepsilon_t \tag{7}$$

3.2.3 Residual diagnostics for the AR(1) Model

Diagnostics on the residuals obtained from the ARIMA (1) model as shown on Figure below shows no significant lag for both the ACF and the PACF. This indicates a white noise process and thus the selected model is adequate for forecasting. The table of Actual, Fitted, and residual as seen in Table 8 below also indicate that the selected model is adequate.

Table 6. Augmented Dickey-Fuller Test of the daily death cases data

Null Hypothesis: DEATHCASE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.902465	0.0322
Test critical values:	1% level	-3.568308	
	5% level	-2.921175	
	10% level	-2.598551	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DEATHCASE)
Method: Least Squares
Date: 10/12/20 Time: 01:07
Sample (adjusted): 2 51
Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DEATHCASE(-1)	-0.351128	0.120976	-2.902465	0.0056
C	1.041641	0.489018	2.130065	0.0383

R-squared	0.149303	Mean dependent var	0.220000
Adjusted R-squared	0.131580	S.D. dependent var	3.025672
S.E. of regression	2.819596	Akaike info criterion	4.950242
Sum squared resid	381.6058	Schwarz criterion	5.026723
Log likelihood	-121.7561	Hannan-Quinn criter.	4.979366
F-statistic	8.424306	Durbin-Watson stat	2.334687
Prob(F-statistic)	0.005577		

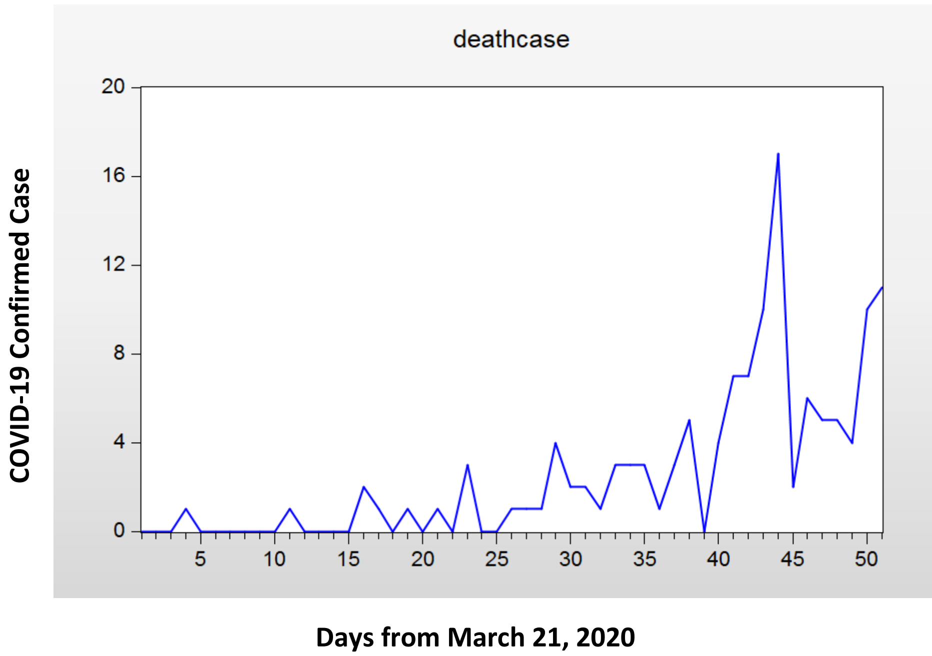


Fig. 9. Time plot of total daily death cases of COVID-19 in Nigeria from 21 March 2020 to 5 May 2020

Table 7. Parsimonious models estimation for the death cases

Dependent Variable: DEATHCASE
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/07/20 Time: 10:34
Sample: 1 51
Included observations: 51
Failure to improve objective (non-zero gradients) after 32 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.929997	0.241685	3.847969	0.0004
AR(6)	0.046859	0.197378	0.237405	0.8135
MA(1)	-0.727897	2.721180	-0.267493	0.7904
MA(2)	0.028728	0.257923	0.111381	0.9119
MA(3)	0.049727	0.537439	0.092526	0.9267
MA(4)	-0.270915	3.825536	-0.070817	0.9439
MA(6)	0.367738	8.041968	0.045727	0.9637
MA(7)	-0.127735	3.048866	-0.041896	0.9668
MA(8)	0.297180	8.345608	0.035609	0.9718
SIGMASQ	4.679112	2.818458	1.660168	0.1045

R-squared	0.613074	Mean dependent var	2.509804
Adjusted R-squared	0.528139	S.D. dependent var	3.512108
S.E. of regression	2.412542	Akaike info criterion	4.917286
Sum squared resid	238.6347	Schwarz criterion	5.296075
Log likelihood	-115.3908	Hannan-Quinn criter.	5.062032
Durbin-Watson stat	1.875541		

Inverted AR Roots	.98	.43-.41i	.43+.41i	-.20+.48i
	-.20-.48i	-.50		
Inverted MA Roots	.95-.31i	.95+.31i	.29+.76i	.29-.76i
	-.08+.76i	-.08-.76i	-.79-.37i	-.79+.37i

Table. 8. Correlogram of residuals of the AR(1) model

Correlogram of Residuals

Date: 10/12/20 Time: 01:20
Sample: 1 51
Included observations: 51
Q-statistic probabilities adjusted for 9 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.006	0.006	0.0019	
		2	-0.032	-0.032	0.0586	
		3	-0.020	-0.020	0.0812	
		4	0.060	0.059	0.2895	
		5	-0.244	-0.247	3.7781	
		6	-0.029	-0.020	3.8294	
		7	-0.012	-0.027	3.8388	
		8	-0.088	-0.112	4.3300	
		9	0.046	0.082	4.4683	
		10	0.161	0.101	6.1738	0.013
		11	0.021	0.005	6.2047	0.045
		12	-0.096	-0.086	6.8430	0.077
		13	0.060	0.018	7.0958	0.131
		14	0.080	0.088	7.5594	0.182
		15	0.059	0.125	7.8240	0.251
		16	-0.139	-0.129	9.3077	0.231
		17	0.014	-0.004	9.3232	0.316
		18	-0.009	0.020	9.3304	0.407
		19	-0.058	-0.066	9.6121	0.475
		20	-0.086	-0.062	10.256	0.508
		21	0.104	0.080	11.230	0.509
		22	-0.054	-0.037	11.506	0.569
		23	0.009	0.007	11.514	0.645
		24	-0.044	-0.143	11.711	0.701

Table 9. Forecasting Trend Model for the Death Cases

Dependent Variable: DEATHCASE
Method: ARMA Maximum Likelihood (BFGS)
Date: 10/07/20 Time: 10:27
Sample: 1 51
Included observations: 51
Convergence achieved after 3 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@TREND	0.120670	0.019300	6.252302	0.0000
AR(1)	0.326112	0.084690	3.850644	0.0003
SIGMASQ	6.030491	0.717991	8.399123	0.0000

R-squared	0.501326	Mean dependent var	2.509804
Adjusted R-squared	0.480547	S.D. dependent var	3.512108
S.E. of regression	2.531284	Akaike info criterion	4.754557
Sum squared resid	307.5550	Schwarz criterion	4.868194
Log likelihood	-118.2412	Hannan-Quinn criter.	4.797981
Durbin-Watson stat	2.028309		

Inverted AR Roots	.33

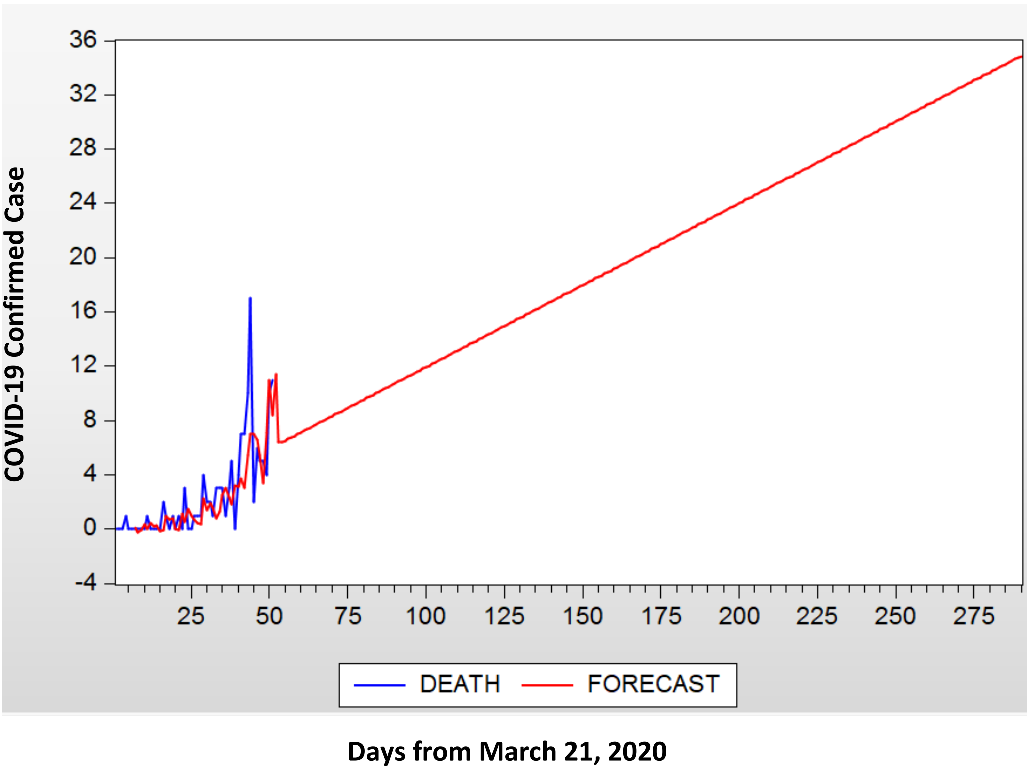


Fig. 10. Actual/Forecasted data time plot following the ARIMA (1) model showing a future forecast

3.2.3 Forecasting the Death Cases using the AR(1) Model

From the selected ARIMA (1) model, we shall make an in data forecast for the first 51 days corresponded to the actual data of confirmed cases of COVID-19 in Nigeria.
using the subsequent forecast of the future values for l >1 shall be computed using the EViews Command below and the result will be integrated twice to obtain the actual forecast of the series.

(Table 9) shows the forecasting equation estimation while (Fig. 10) shows the actual/fitted plot of the daily death cases of COVID-19 in Nigeria, also an out-of-sample (future) forecast of the daily death case data for a total period of eight(8) months. (between May and December, 2020).

"ls daily_cased2 ar(1) @trend @expand(@month, @dropfirst)"

4 Conclusion

Critical investigation into the rate of spread of COVID-19 pandemic has shown that, that the daily confirmed cases as well as death cases of the disease tends to follow an upward trend. This work aimed at developing a suitable ARIMA models which can be used to fit a most appropriate subsets to statistically forecast the actual number of confirmed cases as well as death cases of COVID-19 recorded in Nigeria for a period of 8 months. Data on total daily cases for both confirmed cases and death cases in Nigeria was obtained from the Nigerian Centre for Disease Control (NDCD) online data base on COVID-19, this was used as a secondary data for the work. The data ranged from 21 March, 2020 to 5 May, 2020 covering a total of 51 data points. An adequate subset ARIMA (2, 2, 1) and AR (1) models were fitted to the confirmed cases as well as death cases data respectively. A forecast of 239 days (8 months) from 6th May, 2020 to 31th December, 2020, was carried out and we discovered that, the COVID-19 daily confirmed cases may likely incline over the next eight months.

Following the results obtained and the research findings, we therefore make the following recommendation which shall assist in minimizing the daily confirmed cases as well as mortality rate of COVID-19 in Nigeria.

a. Adequate data collection and presentation

   The statistical process of data identification, collection and presentation has played a very vital role in sufficiently understanding the trend associated with most data sets ranging from health, economic, commerce, agriculture, transport, etc. therefore, an adequate documentation of all data associated with COVID-19 in Nigeria should be collected and properly documented, from the grassroot levels at primary health centres, up to the federal level, this will help in understanding the rate of spread of the disease.

b. Strict adherence to WHO guidelines on COVID-19

   Since the inception of the novel Corona virus disease 2019 (COVID-19) an illness caused by novel coronavirus now called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which was first identified amid an outbreak of respiratory illness cases in Wuhan City, Hubei Provence, China, in December, 2019, has become a thing of global concern, we therefore recommend that further studies should be carried out to understand and model the mortality rate as well as survival rate of COVID-19 in Nigeria, Africa and the world at large, with respect to age, sex, geographical area etc.

c. Further Studies be carried out

   From the results of the forecast, we discovered that the model forecasted with higher precession with the first 3 to 4 months. However, there has been a drastic drop in the number of actual confirmed cases of COVID19 in Nigeria. Therefore an adequate quarterly forecast should be carried out in other to understand the quarterly trend with respect to how fast/slow the virus is spreading in Nigeria and beyond.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Box G, Jenkins G. Time series analysis: Forecasting and control. 2nd ed. San Francisco, CA: Holden-Day; 1970.

[2] Box G, Jenkins G. Time series analysis: Forecasting and control. San Francisco, CA: Holden-Day; 1970.

[3] Jing C, Mengmeng R, Shuyang X, Ingyu W. Predicting seasonal influenza based on SARIMA model, in Mainland China from 2915 to 2018. International Journal of Environmental Research and Public Health. 2019;16(23):47-60.
Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19—studies needed. New England Journal of Medicine. 2020;382(13):1194-1196.

[4] Nwuju, K, Lekara-Bayo IB, Modelling. Forecasting daily confirmed cases of Covid-19 in Africa: A case study of ECOWAS Countries: International Journal of Applied Science and Mathematical Theory. 2020; 6(2):37-54.

[5] Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, Tong S. Epidemiology of COVID-19 among children in China. Pediatrics. 2020;145(6).

[6] Stübinger J, Schneider L. Epidemiology of coronavirus COVID-19: Forecasting the future incidence in different countries. Healthcare. 2020;8(2).

[7] Sahai AK, Rath N, Sood V, Singh MP. ARIMA modelling & forecasting of COVID-19 in top five affected countries. Diabetes & Metabolic Syndrome. 2020;14(5):1419–1427.
DOI: https://doi.org/10.1016/j.dsx.2020.07.042

[8] Qiuying Yang, Jie Wang, Hongli Ma, Xihao Wang. Research on COVID-19 based on ARIMA model—Taking Hubei, China as an example to see the epidemic in Italy. Journal of Infection and Public Health. 2020;13(10):1415 – 1418.

[9] Napoli PE, Nioi M. Global spread of coronavirus disease 2019 and malaria: An epidemiological paradox in the early stage of a pandemic. J Clin Med. 2020 Apr 16;9(4):1138.
DOI: 10.3390/jcm9041138. PMID: 32316118; PMCID: PMC7230338.

[10] Etuk EH, Attoe SA, Essi ID. A seasonal Box-Jenkins model for Nigerian inflation rate series. Journal of Mathematics Research. 2012;4(4):107–113.

[11] Nwuju K, Lekara-Bayo IB, Etuk EH. Intervention analysis of daily South African rand/Nigerian naira exchange rates. International Journal of Management Studies, Business & Entrepreneurship Research. 2019;4:17-26.

[12] Our World in Dat, daily confirmed/death cases of COVID-19 from 20 March 2020 to 5 May; 2020. Available:https://ourworldindata.org/coronavirus/country/Nigeria

APPENDIX

Appendix A

Daily Confirmed/Death Cases of COVID-19 in Nigerian from 20th March 2020 to 5th May 2020.

S/N	Daily Confirmed Cases	Daily Death Cases
1	4	0
2	10	0
3	8	0
4	10	1
5	4	0
6	7	0
7	14	0
8	16	0
9	16	0
10	0	0
11	34	1
12	0	0
13	20	0
14	23	0
15	16	0
16	20	2
17	22	1
18	6	0
19	16	1
20	22	0
21	12	1
22	17	0
23	13	3
24	5	0
25	20	0
26	30	1
27	34	1
28	35	1
29	51	4
30	49	2
31	85	2
32	38	1
33	117	3
34	91	3
35	108	3
36	114	1
37	87	3
38	91	5
39	64	0
40	195	4
41	196	7
42	204	7
43	238	10
44	218	17
45	170	2
46	244	6
47	148	5

S/N	Daily Confirmed Cases	Daily Death Cases
48	195	5
49	381	4
50	386	10
51	239	11

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/67273