A Time Series Model to Forecast COVID-19 Infection rate in Nigeria Using Box-Jenkins Method

[1] Ortese, C.A. [2]Ieren, T.G. and [1]Tamber, A.J.
[1]Department of Mathematics and Computer Science, Benue State University, Makurdi.
[2]Department of Statistics and Operations Research, Madibbo Adama University of Technology, Yola, Nigeria.
[a]oaondona@bsum.edu.ng, ternagodfrey@gmail.com[b], atamber@bsum.edu.ng[c]

doi: https//doi.org/10.46912/napas.232

Abstract
Coronavirus declared as a global pandemic by WHO has emerged as the most aggressive disease negatively affecting over 90% of the countries in the world. Nigeria, one of the most populated countries in Africa is not an exception. This study focuses on analyzing the intrinsic patterns in the COVID-19 spread in Nigeria using the Box-Jenkins procedure. Data of daily confirmed cases of COVID-19 in Nigeria was retrieved from Nigeria Centre for Disease Control (NCDC) official website from February 27, 2020 to October 31, 2020 to identify the series components, estimate parameters, develop an appropriate stochastic predictive model and use the model to forecast future trend of the deadly virus. The Autoregressive Integrated Moving Average (ARIMA) of order (0,1,1) was identified as the most suitable model based on the analysis of the autocorrelation (ACF), partial autocorrelation functions (PACF) and Akaike Information Correction (AICc) value. R software version 4.0.3 was used to analyze the trend which moothen the series by using 8-point moving average to extract the irregular component as wellas differencing the series one step further to obtain a stationary series. We performed the Augmented Dickey-Fuller Unit root test, parameter estimation and Ljung-Box test to check the proposed model's conformity to the stationary univariate process. A 85 – day (1st Oct., 2020 – 24th Jan., 2020)forecast shows a gradual decline in the successive number of confirmed cases of infection indicating the effectiveness of the intervention strategies employed by the Task Force to contain the virus. The concerned authorities can apply the forecasted trend to make further informed decisions on the measures to be put in place to reduce diffusion of the deadly virus into the country.

Keywords: *COVID-19, ARIMA, forecast, Autocorrelation, Partial Autocorrelation*

Introduction

Coronavirus popularly known as COVID-19 is a severe viral disease caused by a contagious acute Respiratory Syndrome Coronavirus 2 (SARS-CoV 2). It belongs to the genus 'coronavirus' of the Coronaviridae family (Sahin, 2020). It is characterized by crown-shape (the name "coronavirus" is derived from the Greek κορώνα, meaning crown.) peplomers with 80-160 nM in size. The genome of CoV contains a linear, single-stranded RNA molecule of positive (mRNA) polarity and about 28-32Kb in length (woo *et al*., 2020)

It was first discovered in Wuhen, Hubei District of China in December, 2019(WHO, 2020) since then, it has spread across over 200 countries of the world. On March 11th, 2020, the World Health Organization declared the outbreak a pandemic. Covid-19 is currently a major worldwide threat to human existence and has caused the largest global recession. It has been spreading rapidly globally, with a considerable impact on global morbidity, mortality and healthcare utilization (Rauf and Oladipo, 2020). As of 31st October, 2020, the world has registered over 46.4 million confirmed cases of the deadly virus from which 1,200,565 and 33,493,349 are the recorded deaths and recoveries respectively.

On February 27, 2020, Nigeria recorded its first case of Covid-19. The index case was an Italian citizen who arrived Nigeria via the Murtala Mohammed International Airport, Lagos at 10pm aboard Turkish airline from Milan, Italy. Since then, there has been an exponential rise in the number of confirmed cases of the virus. As of 31st October, 2020, Nigeria has 62,853 confirmed cases, 58,675 discharged and 1144 unfortunate deaths recorded (NCDC, 2020)

Curtailing infection rate, preventing transmission and reducing death is the goal of every society. How many persons will be infected on daily bases, how to manage them and future occurrence is stochastic (uncertain) and the effect of the intervention strategies employed by the government greatly rely on past and future trends of the pandemic. Due to the varying trend, it is therefore pertinent to construct a realistic model that will competently help policy makers, medical field, government and other relevant authorities to understand the components of the series to control the global epidemic threat and provide future forecast of possible number of daily infections. These will prepare healthcare for the upcoming cases. Using statistical models to study of the trend of the Covid-19pandemic in Nigeria can provide critical information for responding to outbreaks and understanding the impact of strategies employed by the government in containing the spread of the disease.

Time series modeling is a dynamic area that carefully collects and rigorously studies past observations to develop an appropriate model which describes the inherent structure of the series and also used to generate future values (Cochrane, 1997). Time series forecasting is the act of predicting the future by understanding the past (Raicharoen *et al*., 2001). One of the most popular and frequently used stochastic time series model is the Autoregressive Integrated Moving Average (ARIMA) Model (Zhang, 2003)

During the ongoing pandemic, some research publications have focused on the epidemiology, trend analysis and forecasting for different cities and countries. These studies presented long-term and short-term trend using time series data from relevant database and offered forecasting applications using models such as ARIMA model, Exponential Smoothing methods, SEIR model and Regression Model.

Applying purely data-driven statistical method, Yang *et al*. (2020) estimated the case fatality rate (CFR) for COVID-19 in three clusters: Wuhan city, other cities of Hubei province, and other provinces of mainland China. A simple linear regression model was applied to estimate the CFR from each cluster. The result obtained showed that CFR during the first weeks of the epidemic ranges from 0.15% (95% CI: 0.12-0.18%) in mainland China excluding Hubei through 1.41% (95% CI: 1.38-1.45%) in Hubei province excluding the city of Wuhan to 5.25% (95% CI: 4.98-5.51%) in Wuhan. Their results conclusively indicate CFR of COVID-19 was lower than the previous coronavirus epidemics caused by SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV).

To study the epidemic trend of COVID-19 in mainland China, Hubei province, Wuhan city and other provinces outside Hubei from

January 16 to February 14, 2020, Zhu et al. (2020) generated the epidemic curve of the new confirmed cases, multiple of the new confirmed cases for period-over-period, multiple of the new confirmed cases for fixed-base, and the period-over-period growth rate of the new confirmed cases using data from National Health Commission. From January 16 to February 14, 2020, the cumulative number of new confirmed cases of COVID-19 in mainland China was 50 031, including 37 930 in Hubei province, 22 883 in Wuhan city and 12,101 in other provinces outside Hubei.

Fanelli and Piazza, (2020) analyzed the temporal dynamics of COVID-19 outbreak in China, Italy and France with the timeframe of January 22 to March 15 2020. A first analysis of simple day-lag maps points to some universality in the epidemic spreading and the analysis of the same data within a simple susceptible-infected-recovered-deaths model indicated that the kinetic parameter that described the rate of recovery appeared to be the same, regardless of the country, while the infection and death rates appeared to be more variable.

Piccolomini and Zama, (2020) also proposed the modification of the Susceptible-Infected-Exposed-Recovered-Dead (SEIRD) differential model for the analysis and forecast of the COVID-19 spread in some regions of Italy. They introduced a time-dependent transmitting rate and reported the maximum infection spread for the three Italian regions firstly affected by the COVID-19 outbreak (Lombardia, Veneto and Emilia Romagna).

Danon *et al*. (2020), applied an existing national-scale metapopulation model to capture the spread of CoVID-19 in England and Wales. They captured data from population sizes and population movement, together with parameter estimates from the current outbreak in China and were able to predict the peak of the outbreak after person-person transmission was established in England and Wales.

Jit *et al*. (2020) applied exponential growth model to fit critical care admissions from multiple surveillance to study likely COVID-19 case numbers and progress in the United Kingdom from February 16 – March 23, 2020. They estimated that on 23 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.

Prashant *et al.* (2020) applied the ARIMA and Fuzzy Models in Forecasting COVID-19 Outbreak in India. Both models suggested an exponential uplift in COVID-19 cases in the near future.

Rauf and Oladipo (2020) applied the Box-Jenkins procedure in forecasting the spread of COVID-19 in Nigeria. The ARIMA (1, 1, 0) was selected as the best model fit for the dataset. The limitation to this study was 10-day forecast

The main aim of the study is to employ the Box-Jenkins modeling approach to develop a model and apply it to forecast future incidences of COVID-19 disease in Nigeria using a more robust dataset and projections of future occurrences.

The specific objectives are

i. Develop a time series model that will identify the trend of COVID-19 occurrence in Nigeria
ii. Estimate parameters of the developed model
iii. Diagnosing the model
iv. Predict the future incidence of COVID-19 disease in Nigeria.

Materials and Method

Data and Source

Confirmed cases of COVID-19 infections are collected for Nigeria by the Nigeria Centre for Disease Control, NCDC. Data was therefore extracted from the official website of NCDC (http://www.ncnc.org) from February 11, 2020 to October 31, 2020 (7 months) to build a predictive model.

Notations

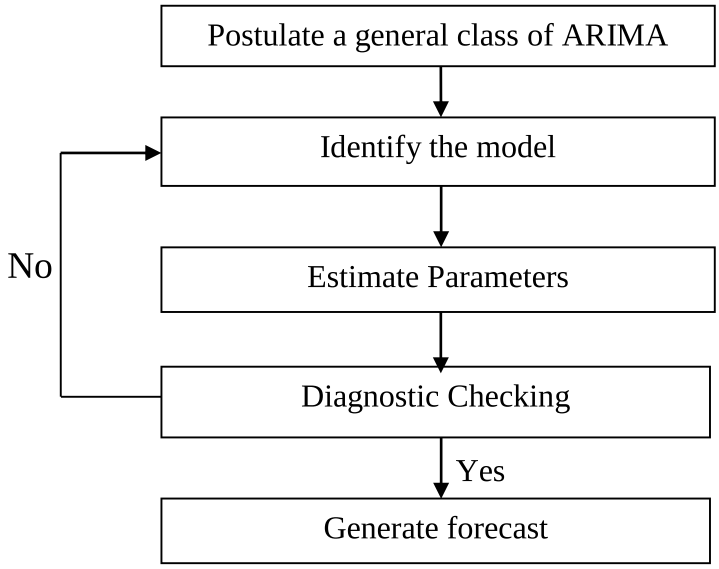
y_t : $COVID-19\ series$ ε_t : $random\ error\ (shocks)$
$\varphi_i (i = 1,2,3 \ldots p)$: $autoregressive\ model\ parameters$
c : $constant$
p : $order of the\ AR\ component\ of\ the\ model$
q : $order\ of the\ MA\ component\ of\ the\ model$
$\theta_j (j = 1,2, \ldots q)$: $moving\ average\ model\ parameters$
μ : $mean$
$\varphi(L)$: $polynomial\ of order\ p\ in\ the\ lag\ operator$
$\theta(L)$: $polynomialof order\ q\ in\ the\ lag\ operator$
γ_o : $autocovarianceatlagzero$ (0)
γ_k : $autocovariance\ at\ lag\ k$

Procedures

The Box-Jenkins method was employed in building the Autoregressive Integrated

Moving Average (ARIMA) model. This is an iterative three-stage approach to modeling as shown in the diagram

Figure 1: The Box-Jenkins methodology for optimal model selection



Source: G.E.P Box and Jenkins (1970)

Postulating a general class of ARIMA

The selection of a proper model is extremely important as it reflects the underlying structure of the series and this fitted model in turn is used for future forecasting.

A linear time series model was considered as the current value of the observed series is a linear function. Different univariate time series models are used in literature such as the Autoregressive (p) and Moving Average (q) Models (Hipel and McLeod, 1994). The combination of these two models forms the Autoregressive Moving Average (ARMA) models. However, in this study, the Autoregressive Integrated Moving Average (ARIMA) Model is considered.

This ARIMA model is a transformed ARMA models which means it combine the Autoregressive (p) and Moving Average(q) and transforms the trend from a non-stationary to a stationary one (constant mean and variance)

In an Autoregressive (p) model, the future value of a variable is assumed to be a linear combination of p past observations and a random error together with a constant term.

Mathematically, the Autoregressive (p) model can be expressed as (Lee, 2010)

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t \quad (1)$$

In a Moving Average (q) model, the model regress against past values of the series, it used past errors as the explanatory variables. The MA (q) model is given by (Lee, 2010),

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

The random shocks are assumed to be a white noise process.

As stated earlier, Autoregressive (AR) and Moving Average (MA) models can be effectively combined together to form a general more useful model known as ARMA model Mathematically, an ARMA (p,q) model is represented as (Lee, 2010)

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (3)$$

With $\varphi_2 = 0,\ \theta_q \neq 0\ and \sigma^2{}_{\varepsilon} > 0$

Usually, ARMA models are manipulated using the lag operator notation (Lee, 2010), the lag or backshift operator is defined as

$$Ly_t = y_{t-1} \quad (4)$$

Polynomials of lag operator are used to represent ARIMA models as follows

Autoregressive (p) model: $\varepsilon_t = \varphi(L) y_t$ (5)

Moving Average (q), model: $y_t = \theta(L) \varepsilon_t$ (6)

Autoregressive Moving Average (p, q)

$$\varphi(L)y_t = \theta(L)\varepsilon_t \quad (7)$$

Where

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p$$
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$$

In practice, ARMA (p, q) models can only be used for stationary time series data. However, many time series show non-stationary behavior in such situations therefore, the ARIMA Model is implemented instead (Hipel and McLeod, 1994) This study is not an exception in this scenario as the data contain a trend and non-stationary behavior, therefore, it is inadequate to implement the ARMA model in this situation, the research propose an ARIMA model which is a generalization of an ARMA model to include the case of non-stationarity. Here, we apply finite differencing of the data points so as to transform the non-stationary data to stationary.

The mathematical formulation of the ARIMA (p,d,q) model using lag polynomials is given by

$$\varphi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t \quad (8)$$

That is

$$1 - \sum \varphi_i L^i](1 - L)^d y_t = \left[1 + \sum_{j=1}^{q} \theta_j L^j\right]\varepsilon_t \quad (9)$$

where p, d and q refer to the order of autoregressive, integrated and moving average parts of the model respectively.

Model Identification

To determine a proper model and the order of the autoregressive and moving average term for a given time series data, the autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis was carried out to decide which autoregressive or moving average component to be used in the model. The plot of the autocorrelation function (ACF) and partial autocorrelation function (PACF) against consecutive time lags was done in order to determine this.

The autocorrelation coefficient at lag k is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_o} \quad (10)$$

$$-1 < \rho_k < 1$$

where

$\gamma_o = autocovarianceatlagzero$ (0)

γ_k is the autocovariance at lag k defined as

$$\gamma_k = cov(y_t, y_{t+k}) \quad (11)$$

Autocorrelation function (ACF) of an Autoregressive (p) is given as

$$\rho(h) - \varphi_1\rho(h-1) - \cdots - \varphi_p\rho(h-p) = 0 \quad (12)$$

where $h \geq p$

Autocorrelation function (ACF) of an Autoregressive Moving Average (p, q) is given as:

$$\gamma(h) - \varphi_1\gamma(h-1) - \cdots - \varphi_p\gamma(h-p) = 0 \quad (13)$$

Where $h \geq \max(p, q+1)$

Order Determination

The order of a time series model was determined by defining the criteria for choosing the order of a model or by testing hypothesis $\varphi_{kk} = 0, \ \rho_k = 0$

Parameter Estimation

To obtain the best estimates for $\varphi and \theta$ parameters for Autoregressive moving average(p, q)

$$y_t - \varphi_1 y_{t-1} - \cdots - \varphi_p y_{t-p} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \quad (14)$$

Recall that

$$e_k - \varphi_1 e_{k-1} - \cdots - \varphi_p e_{k-p} = 0 \quad fork > q$$

Setting $k = q+1, \ldots q+p$ can be solved simultaneously to obtain the estimates of $\varphi_1, \varphi_2, \ldots, \varphi_p$

Similarly, we can show that

$$e(k) = \theta_1 p(k-1) - \cdots - \theta_q p_k(k-q) = 0, \quad k > q$$

Setting $k = p+1, \ldots p+q$ can be solved simultaneously to obtain the estimates of $\theta_1, \theta_2, \ldots, \theta_q$

calculate

$$S(\varepsilon) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} \left(y_t - \hat{\varphi}_1 y_{t-1} - \cdots - \hat{\varphi}_p y_{t-p}\right) + (\hat{\theta}_1\varepsilon_{t-1} + \cdots + \hat{\theta}_q\varepsilon_{t-q}))^2$$

where $\xi = (\hat{\varphi}_1, \ldots, \hat{\varphi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$

where the grid search is used to obtain the value of ξ that maximizes S(ξ)

Diagnostic checking

After fitting the model, the estimated model is tested to determine whether the estimated model conform to the specification of a stationary univariate process. The Ljung – Box test is performed to test the model adequacy and the Autocorrelation Function of the residuals plotted. The steps are reiterated until the required adequacy is achieved.

Results and Discussion

The overall distribution of daily COVID-19 confirmed number of infection from February 27, 2020 to September 10, 2020 was retrieved from the Nigerian Centre for Disease Control (NCDC) official website (http://covid19.ncdc.gov.ng/). Analysis was conducted with the use of R and Python statistical software.

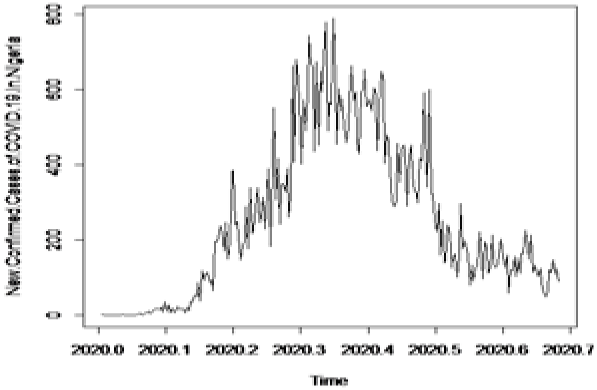


Figure 2: Time Series Plot of confirmed COVID-19 cases in Nigeria

Figure 2 have plotted a dataset for daily confirmed cases of COVID-19 infection cases in Nigeria. From the plot above, it can be deduced that the series there is random fluctuations in the data which is roughly constant over time.

Decomposing the COVID-19 Series

In separating the series into its constituent components which are mainly the trend and the irregular components in the case of this series, the trend is estimated using the additive model to compute the simple moving average. In this case, the 8-point moving average (n=8) is used to obtain the smoothened series to estimate the trend.

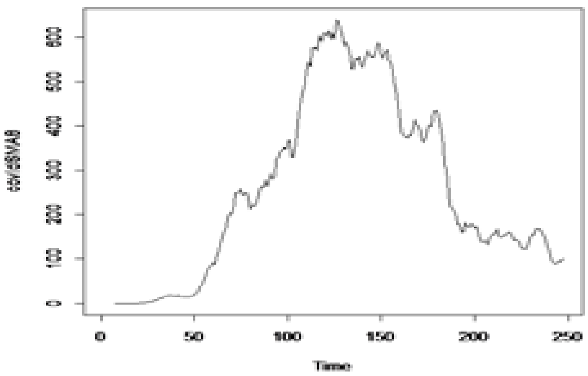


Figure 3: Simple Moving Average to estimate trend using 8-point moving average (n=8)

Figure 3 showed the smoothened series which estimates the trend for the series. We have hence removed the trend component and we are left with the irregular component.

Autoregressive Integrated Moving Average (ARIMA) Model

The research considered the ARIMA (p,d,q) model for the analysis as it allows for non-zero autocorrelation in the irregular component and also it makes assumptions about correlations between successive values of the series.

ARIMA Models are defined for stationary series. In our time series plot on figure 2, the plot is a non-stationary in nature, we hence need to difference the series d times to obtain a stationary series.

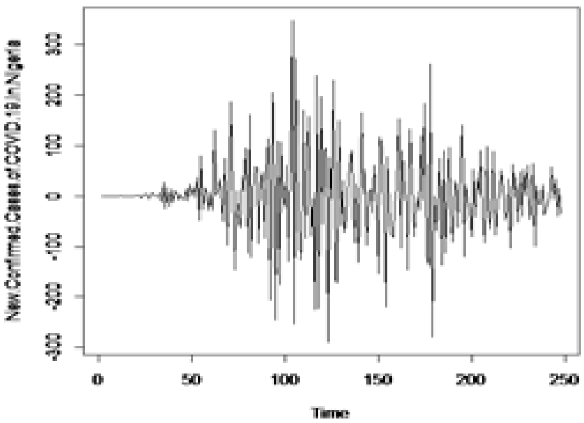


Figure 4: COVID-19 confirmed cases in Nigeria at first differenced (d=1)

Table 1: ARIMA Models and corresponding AICc

S/N	Models	AICs	Type
1	ARIMA(2,1,2)	2337.928	with drift
2	ARIMA(0,1,0)	2396.869	with drift
3	ARIMA(1,1,0)	2352.64	with drift
4	ARIMA(0,1,1)	2329.819	with drift
5	ARIMA(0,1,0)	2394.837	
6	ARIMA(1,1,1)	2332.861	with drift
7	ARIMA(0,1,2)	2331.864	
8	ARIMA(1,1,2)	Inf	
9	ARIMA(0,1,1)	2327.878	
10	ARIMA(1,1,1)	2330.905	
11	ARIMA(0,1,2)	2329.906	
12	ARIMA(1,1,0)	2350.597	
13	ARIMA(1,1,2)	Inf	

From the table above, the best model is the one with the lowest Akaike Information Criterion correction (AICc) value which is ARIMA (0, 1, 1) model.

The best order for the series is ARIMA (0,1,1) is defined as

ARIMA (0,1,1) = $y_t - \mu = \varepsilon_t - (\theta * \varepsilon_{t-1})$

Model Parameter Estimation

As observed from the above analysis, the best model is the ARIMA (0,1,1) base on the AICc criterion and ACF and PACF graphs. The model is then estimated with its parameter estimates for forecasting the daily spread series of COVID-19 in Nigeria. From the output, the estimated value of ϴ is -0.6543 (see Appendix IV). Therefore, the workable predictive model obtained after the substitution of estimated parameters is represented as

ARIMA (0,1,1) = $y_t = 0.6543\varepsilon_{t-1}$

Table 2: **ARIMA Model Results**

	Coef	Std. err	Z	P>|z|	0.025	0.975
const	0.6539	1.819	0.360	0.719	-2.910	4.218
ma.L1.D.New Cases	-0.6543	0.050	-12.965	0.000	-0.753	-0.555

Diagnostic Checking

Box-Ljung test

The Ljung-Box test is a diagnostic tool used to test for lack of fit of a time series model (Box and Jenkins, 1976). The value for the Ljung-Box test statistic (X-Squared) is 20.688 with a p-value as 0.5157. These has hence provided relevant validation (p>0.5) in favor of the null hypothesis at 5% level of significance thereby establishing the suitability of the model. The results of the and obtain the results.

Residual ACF:

The correlogram plot for the forecast errors (residuals) to measure the goodness of fit as shown below.

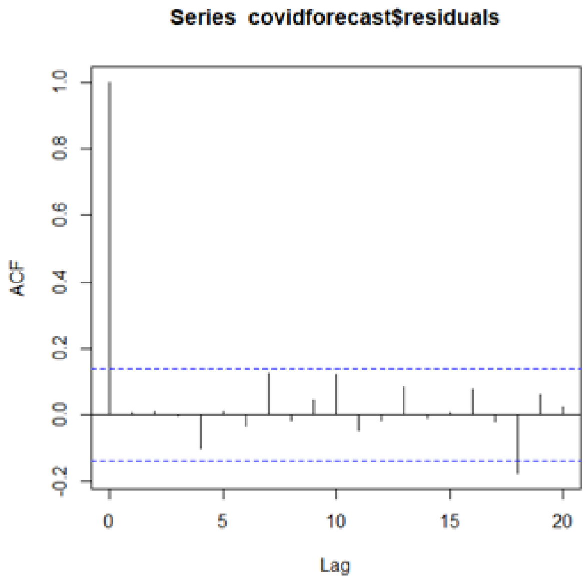


Figure 7: Residual ACF Plot

From Figure 7 above, it can be deduced that, the spikes in between the horizontal dotted lines are random and gradually decreasing to zero. This implies that the ARIMA model is fitted appropriately. The Residual Error density plot is shown below.

From figure 8, We get a density plot of the residual error values suggests that the residual errors are Gaussian.

Forecast ARIMA Model

The ARIMA model is used to forecast future time steps of the Covid-19 confirmed cases in Nigeria. A one-step forecast using the ARIMA model is used. It accepts the index of the time steps to make predictions as arguments, 228 observations are used in the training dataset to fit the model. Therefore, the index of the next time step for making prediction start at 229. The training dataset is splitted into train and test sets, use the train set to fit the model and generate a prediction for each element on the test set. The forecast is performed by re-creating the ARIMA model after each new observation is received. All observations are tracked in the history that is seeded with the training data and to which new observations are appended each iteration. These procedure prints the prediction and expected value of each iteration. The results of the iteration is as tabulated below.

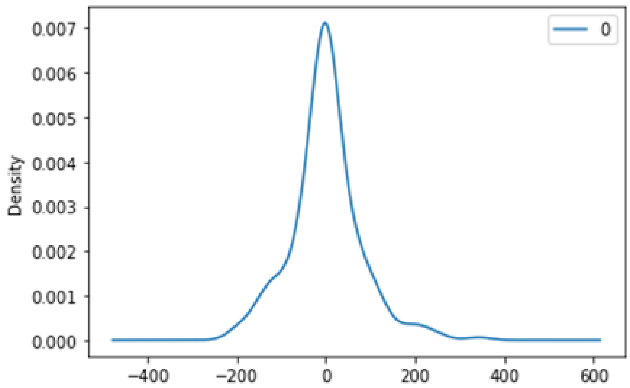


Figure 8: ARIMA fit Residual Error Density Plot

From figure 8, We get a density plot of the residual error values suggests that the residual errors are Gaussian.

Table 4: predicted and expected values with lower (Lo) and Upper (Hi) prediction intervals

Date	predicted	expected	Lo 80	Hi 80	Lo 95	Hi 95
11/1/2020	404.626224	453	7.837458	203.8709	-44.0495	255.7578
11/2/2020	422.480671	437	1.978449	209.7299	-53.01	264.7184
11/3/2020	429.693861	290	-3.56729	215.2756	-61.4915	273.1998
11/4/2020	388.316548	423	-8.8452	220.5535	-69.5634	281.2717
11/5/2020	401.693248	453	-13.8907	225.599	-77.2798	288.9881
11/6/2020	420.15666	373	-18.732	230.4404	-84.684	296.3923
11/7/2020	408.010894	329	-23.3922	235.1005	-91.811	303.5193
11/8/2020	385.741091	325	-27.89	239.5983	-98.6899	310.3982
11/9/2020	368.689339	298	-32.2414	243.9497	-105.345	317.0531
11/10/2020	347.905158	417	-36.4599	248.1682	-111.796	323.5047
11/11/2020	372.237431	410	-40.5568	252.2651	-118.062	329.7704
11/12/2020	386.278927	593	-44.5422	256.2505	-124.157	335.8655
11/13/2020	453.500732	476	-48.4246	260.133	-130.095	341.8032
11/14/2020	463.572931	340	-52.2118	263.9201	-135.887	347.5951
11/15/2020	427.18518	601	-55.9103	267.6186	-141.543	353.2515
11/16/2020	482.948192	322	-59.5261	271.2344	-147.073	358.7814
11/17/2020	436.752043	321	-63.0645	274.7728	-152.485	364.1929
11/18/2020	404.76621	252	-66.5303	278.2386	-157.785	369.4934
11/19/2020	359.325866	221	-69.9278	281.6361	-162.981	374.6894
11/20/2020	313.974353	296	-73.2608	284.9691	-168.079	379.7869
11/21/2020	309.339215	160	-76.533	288.2413	-173.083	384.7912
11/22/2020	257.925158	250	-79.7474	291.4558	-177.999	389.7073
11/23/2020	256.43393	138	-82.9072	294.6155	-182.831	394.5397
11/24/2020	214.810812	143	-86.0149	297.7232	-187.584	399.2925
11/25/2020	188.976924	239	-89.0731	300.7814	-192.261	403.9696
11/26/2020	208.46445	216	-92.084	303.7923	-196.866	408.5744
11/27/2020	212.275224	125	-95.0498	306.7581	-201.402	413.1102
11/28/2020	182.216053	156	-97.9724	309.6808	-205.872	417.5801
11/29/2020	173.691462	162	-100.854	312.5621	-210.278	421.9867
11/30/2020	170.382929	100	-103.696	315.4038	-214.624	426.3327
12/1/2020	145.765492	155	-106.499	318.2075	-218.912	430.6206
12/2/2020	149.891109	296	-109.266	320.9747	-223.144	434.8526
12/3/2020	202.854307	176	-111.998	323.7067	-227.323	439.0309
12/4/2020	194.305219	197	-114.697	326.4049	-231.449	443.1574
12/5/2020	196.260374	188	-117.362	329.0705	-235.526	447.234
12/6/2020	194.334343	160	-119.996	331.7046	-239.554	451.2625
12/7/2020	183.157883	79	-122.6	334.3083	-243.536	455.2446
12/8/2020	147.029231	132	-125.174	336.8827	-247.473	459.1818

12/9/2020	142.353037	90	-127.72	339.4287	-251.367	463.0756
12/10/2020	124.17469	126	-130.239	341.9473	-255.219	466.9274
12/11/2020	125.454778	131	-132.731	344.4393	-259.03	470.7386
12/12/2020	128.070628	221	-135.197	346.9055	-262.802	474.5103
12/13/2020	161.863899	189	-137.638	349.3467	-266.536	478.2439
12/14/2020	172.382289	97	-140.055	351.7637	-270.232	481.9404
12/15/2020	146.37461	195	-142.449	354.1572	-273.893	485.6009
12/16/2020	164.320497	176	-144.82	356.5278	-277.518	489.2265
12/17/2020	169.256	111	-147.168	358.8763	-281.11	492.8181
12/18/2020	149.463059	125	-149.495	361.2031	-284.668	496.3766
12/19/2020	141.513695	213	-151.801	363.5089	-288.195	499.9031
12/20/2020	167.42442	136	-154.086	365.7943	-291.69	503.3982
12/21/2020	157.151182	126	-156.351	368.0597	-295.155	506.8629
12/22/2020	146.933421	136	-158.597	370.3057	-298.59	510.2979
12/23/2020	143.773049	187	-160.825	372.5328	-301.996	513.704
12/24/2020	159.646077	201	-163.033	374.7415	-305.374	517.0819
12/25/2020	174.937086	153	-165.224	376.9322	-308.724	520.4322
12/26/2020	168.027762	126	-167.397	379.1053	-312.047	523.7557
12/27/2020	154.056373	160	-169.553	381.2613	-315.345	527.053
12/28/2020	156.850951	58	-171.692	383.4005	-318.616	530.3247
12/29/2020	122.831372	120	-173.815	385.5234	-321.863	533.5713
12/30/2020	122.387127	118	-175.922	387.6302	-325.085	536.7935
12/31/2020	121.39363	155	-178.013	389.7214	-328.283	539.9917
1/1/2021	133.75478	103	-180.089	391.7974	-331.458	543.1666
1/2/2021	123.568027	151	-182.15	393.8584	-334.61	546.3186
1/3/2021	133.739113	111	-184.196	395.9047	-337.74	549.4482
1/4/2021	126.369074	163	-186.228	397.9367	-340.847	552.5558
1/5/2021	139.752077	164	-188.246	399.9546	-343.934	555.642
1/6/2021	148.863468	225	-190.251	401.9588	-346.999	558.7072
1/7/2021	176.306239	179	-192.241	403.9496	-350.043	561.7517
1/8/2021	178.031097	148	-194.219	405.9271	-353.068	564.7761
1/9/2021	168.237925	212	-196.183	407.8916	-356.072	567.7806
1/10/2021	184.324745	113	-198.135	409.8435	-359.057	570.7658
1/11/2021	160.153245	133	-200.075	411.7829	-362.024	573.7319
1/12/2021	151.338805	118	-202.002	413.7101	-364.971	576.6793
1/13/2021	140.285388	102	-203.917	415.6254	-367.9	579.6083
1/14/2021	127.387873	124	-205.82	417.5288	-370.811	582.5194
1/15/2021	126.73247	86	-207.712	419.4207	-373.704	585.4128
1/16/2021	112.882077	60	-209.593	421.3013	-376.581	588.2889
1/17/2021	94.553449	48	-211.462	423.1707	-379.44	591.1479
1/18/2021	78.248996	62	-213.321	425.0291	-382.282	593.9901
1/19/2021	72.725279	119	-215.168	426.8768	-385.108	596.8159
1/20/2021	89.549701	113	-217.006	428.7139	-387.917	599.6256
1/21/2021	98.240136	147	-218.832	430.5406	-390.711	602.4193
1/22/2021	115.960578	108	-220.649	432.3572	-393.489	605.1974
1/23/2021	113.603393	123	-222.455	434.1636	-396.252	607.9601
1/24/2021	117.404559	87	-224.252	435.9602	-398.999	610.7077

A time series plot showing the expected value (blue) and the forecast prediction (red) is as shown below.

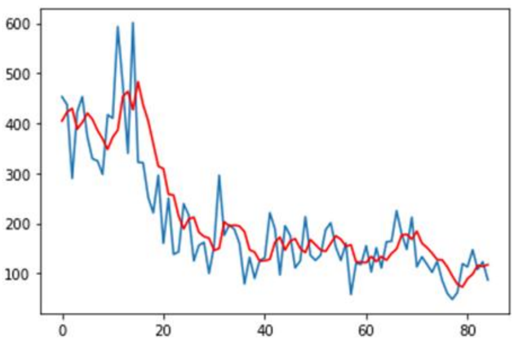


Figure 9: Graphical representation of forecasted (red) and expected (blue) number of COVID-19 Cases in Nigeria.

The forecast for the next 85 days (November 1, 2020 – January 24, 2020) as well as the lower (Lo) and upper (Ho) predictive intervals 80% and 95% respectively as shown in the table below.

The figure below shows the plot of the confirmed cases of COVID-19 for the first 235 days and for the next 85 days using the estimated ARIMA (0,1,1) model.

Conclusion

There was no case of COVID-19 in Nigeria until February, 2020. Since then, the deadly virus has been reported on daily basis by the NCDC and showed an upward trend, special precautionary measures were taken such as total lockdowns, use of facemasks, social distancing, personal hygiene and other strategies to curtail the virus. This yielded positive results which is evident in the downward trend observed on the reported cases in recent days. The research have used stochastic time series model called ARIMA to estimate all aspects of the looming COVID-19 infections in Nigeria based on the contemporary statistical data obtained from Nigeria Centre for Disease control (NCDC). In this rationale, The research have modeled the non-stationary number of infections in Nigeria, transformed the series from non-stationary to stationary by differencing the series once thus obtaining the value of d as 1. Therefore the resultant class of ARIMA model is ARIMA (p,1,q). The Autocorrelation function (ACF), Partial Autocorrelation function plot (PACF) and Akaike Information Criterion (AICc) was used to decide the best model, from our analysis, ARIMA (0,1,1) was chosen as the most suitable model for our dataset.

A workable predictive model was obtained and estimated the parameter (Θ) from our model using maximum Likelihood estimation in python 3.8 software. The predictive model is used to forecast new infections from September, 2020 to November 30, 2020 (85 days). From the forecast trend a significant decline in the number of infections in Nigeria over time was observed. This implies that the virus may not be much effective in the long run as the rate of transmission is reduced. These permit the lifting of lockdowns and resumption of economic activities.

The model may assist them to estimate new infections even after the last prediction date (30th November, 2020) and to plan the required strategy for supplying potential health facilities to all individuals. This preparatory preparation will heal the current adversity, protect human lives and prevent heavy economic losses. Realizing the crucial importance of such modeling, the work can be extended to use multivariate models for forecasting which will provide a further precise description of the current pandemic situation.

Competing Interest: The authors declare that they have no competing interest

References

Cochrane H.J. (1997), "Time Series for Microeconomics and Finance", Graduate School of Business, University of Chicago, Spring 1997.

Danon L., Brooks-Pollock E., Bailey M., Keeling M., (2020) A spatial COVID-19 transmission in England and Wales: early spread and peak timing. *The Preprint server for health sciences.* 3(1), 167-178

Fanelli D and Piazza F., (2020), Analysis and forecast of COVID-19 spreading in China, Italy and France. *Elsevier*. 123

George E., Box P (1970) Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics. 5th Edition. ISBN: 978-1118675021

Hipel K.W., McLeod A.I. (1994), "Time Series Modelling for Water resources and Enveronmental Systems", Amsterdam, Elsevier 1994

Jit M., Jombart T., Nightingale E., Endo A., Abbot S., Edmunds J (2020), Estimating number of cases and spread of coronavirus disease (COVID-19) using critical care admissions, United Kingdom. *Euro Surveill.* 2(5), 18

Lee, J., "Univariate Time Series Modeling and Forecasting (Box-Jenkins Method)", ECO 413. Lecture 4. 2010 (Unpublished)

Lombardo R, Flaherty J., " Modelling Private New Housing Starts in Australia", Pacific – Rim Real Estate Society Conference, University of Technology Sydney (UTS) Jan 24 – 27, 2000

NCDC, "First Coronavirus disease confirmed in Nigeria", http://ncdc.gov.ng/news/227/first_case_of_coronavirus_disease_confirmed_in_Nigeria. (accessed September 7, 2020)

Piccolomini E.L., and Zama F., (2020), Monitoring Italian COVID-19 spread by a forced SEIRD model. *PLoS ONE* 2020 Aug; 15(8): e0237417 Doi: 10.1371/journal.pone.0237417

Prashant V, Mukti K, Shikha D, Shweta D (2020), Forecasting the COVID-19 Outbreak: An Application of ARIMA and Fuzzy Time Series Models. :

Raicharoen, C., Lursinsap, C., Saguanbhoki, P "Application of Critical Support Vector Machines to Time Series Prediction", Circuits and Systems, 2003. ISCAS'03 *Proceedings of the 2003 International Symposium*. Volume 5, 25 – 28 May, 2003 pages 741 – 744.

Rauf R.I, Oladipo H.O. (2020), Forecasting the spread of COVID-19 in Nigeria Using Box-Jenkins Modeling Procedure. *MedRxiv preprint*.

Sahin, A.R. (2020) Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature. *Eurasian Journal of Medical Investigation*, 4(1), 1 – 7.

Tandon H, Ranjan P., Chakraborty T., Suhag V., (2020), Coronavirus (COVID-19): ARIMA based Time Series Analysis to Forecast Near Future.

Website: www.ncdc.gov.ng

Website: www.worldometers.info/coronavirus

WHO(2020),https://www.who.int/emergencies/diseases/novel_coronavirus_2019/evens_as_they_happen (accessed September 8, 2020)

WHO, Coronavirus disease (COVID-19) outbreak. https://www.who.int/emergencies/diseases/Novel-corona virus-2019 (accessed April 14, 2020)

Woo, P.C., Huang, Y., Lau, S.K. and Yuen, K.Y. (2010). Coronavirus Genomic and bioinformatics Analysis. *Viruses*, Vol. 2, pp. 1805 – 1820.

www. wikipedia.org http://en.m.wikipedia.org/wiki/COVID-19_Pandemic

www.bbc.com>world-51839944. https://www.bbc.com/news/world-51839944

Yang S., Cao P., Du P., Wu Z., Zhuang Z., Lin Y., Zhou Q., Feng X., Wang X., Li W., Liu E., Chen J. (2020), Early Estimation of the case faatality rate of COVID-19 in Mainland China: A data-driven analysis. *Annals of Translational Medicine*, 8(4): 128.

Zhang G.P. (2003), “Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model”, Neurocomputing 50 (2003). Pages: 159 – 175

Zhu, T. T., Wu, P., Lin, Y., Lau, E., Leung, G.,& Cowling, B. (2020). Effect of changing case definition for COVID-19 on the epidemic curve abd transmission parameters in Mainland China: A modellling study. *The Lancet Public Health*, 5(5), 289-296.