Pak. j. sci. ind. res. Ser. A: phys. sci. 2024 **67A**(2) 101-112

On the Assessment of Trend and Pattern of COVID-19 Infection in Nigeria: Autoregressive Integrated Moving Average (ARIMA) Approach

Omobolaji Yusuf Halid[a], Bayowa Teniola Babalola*[ab], Joshua Sunday[c], Samuel Oluwaseun Adejuwon[d], Kehinde Abimbola Adigun[d], Ojo Femi Ogunboyo[e], Temitope Olu Ogunlade[f], Anthony Opeyemi Ilesanmi[a] and Sunday Emmanuel Fadugba[f]

[a]Department of Statistics, Ekiti State University, Ado-Ekiti, Nigeria
[b]Department of Mathematics and Statistics, Kampala International University, Uganda
[c]Department of Mathematics, The University of Jos, Plateau, Nigeria
[d]Department of Mathematical and Physical Sciences, Afe Babalola University, Ado-Ekiti-36000, Nigeria
[e]Department of Epidemiology and Biostatistics, University of Medical Sciences, Ondo, Nigeria
[f]Department of Mathematics, Ekiti State University, Ado-Ekiti-360001, Nigeria

(received November 17, 2022; revised September 1, 2023; accepted September 4, 2023)

Abstract. COVID-19 is a deadly infection that causes severe acute respiratory syndrome. Although not particularly spreading rapidly as before due to the introduction of vaccines and other measures, its effect still portends grave danger to human lives in Nigeria and other countries. This study aimed to model and forecast Nigeria's COVID-19 (new) trend of confirmed cases, discharged (recovery) cases and deaths and also to examine the pattern of the infection and survival rate in the face of vaccine introduction. The Box-Jenkins methodology was employed in this study to model and forecast COVID-19 confirmed cases, discharged cases and deaths. The data used for this study was secondary data of weekly confirmed cases, recoveries (discharged) and deaths extracted from the weekly publication of the Nigeria Centre for Disease Control (NCDC). The mean survival rate of COVID-19 was found to be 0.7765 and the three series were found to be stationary after differencing. Also, from an array of candidate models obtained through Autocorrelation Function (ACF) and partial autocorrelation function (PACF) plots, the best-fitted models selected based on minimum Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) were found to be ARIMA (4,1,0), (3,1,0) and (7,3,1) for newly confirmed cases, discharged cases and death, respectively. This implied that these models were adequate for forecasting future rates of infection, recovery and death as further diagnostic tests showed that the ARIMA models were the perfect fit for the three cases (since P>0.05). Finally, a 29-week out of sample forecast showed a steep downward trend in the three cases and in particular a drastic decline to zero of future COVID-19 deaths. Based on these results, it was recommended that existing vaccination strategies should be expanded to achieve near-zero new COVID-19 cases and deaths.

Keywords: COVID-19, infection, recovery, time-series, ARIMA

Introduction

Coronavirus disease (COVID-19) is a respiratory disease characterized by fever, dry cough and fatigue and occasional gastrointestinal among other symptoms. The first human cases of COVID-19, the disease caused by the novel Coronavirus causing COVID-19, subsequently named SARS-CoV-2 were first reported by officials in Wuhan city, China, in December 2019 (Msemburi *et al*., 2023; WMHC, 2020). Nigeria announces the first confirmed case of COVID-19 on 27th February 2020, when an Italian national who entered Lagos tested positive for the virus. The second case of the virus was reported in Ogun State, on 9th March 2020 and this is a Nigerian who was in contact with the Italian man that brought the virus to the country (Betthäuser *et al*., 2023). The disease gradually spread from China to other countries in the world and infected millions of people across all races and age groups (Andersen *et al*., 2020). World Health Organisation (WHO) intensified collaborative works with both animal health and human health experts and other actors to identify gaps and research priorities to combat and control the disease very promptly. Such measures included the eventual tracking of the disease's source (Holmes *et al*., 2021; Zhu *et al*., 2020).

*Author for correspondence;
E-mail: adejuwonso@abuad.edu.ng

The Nigeria Centre for Disease Control (NCDC) (2020) reported its first confirmed COVID-19 case on the 27th of February, 2020. Most countries still battle with thousands of cases and deaths from a few imported cases, as the flow of people through travel has been reported to aid the spread of the disease (Ogundokun *et al*., 2020).

Generally, based on the population growth of Nigerians, the existing health facilities and equipment in the country are inadequate (Ibekwe and Fasunla, 2020). The growing rate of new infection and its consequent effect on the lives of the populace as well as the economic disruptions occasioned by lockdown measures had been a cause of concern. The introduction of the COVID-19 vaccine on the 1st of March, 2021 in Nigeria was reported (NPHCDA, 2021). This uses in new dynamics to this infection, even when the rate of vaccination has been particularly slow as a result of people's poor perception and lack of trust.

This new strategy of combating the spread of the infection and many more necessitated the need for this study, to examine the existing pattern and survival rate of the infection as well as future rates of new infections, recoveries and death in the face of COVID-19 vaccine introduction. The instrumentality of time series approach in the framework of autoregressive integrated moving average was employed for this research. It is intended that the result of this research would reveal the pattern of the new cases periods after the introduction of the vaccine. Although several researchers had worked on COVID-19 data across the globe, the novelty of this article is in the fact that it concentrates on the impact of intervention using Nigerian data. Further, it could allow policymakers to see patterns, with results helping government postulate about a community's health.

Many studies with a primary focus on modelling and forecasting the infection rate had been carried out in several countries. A statistical analysis of the novel corona virus by applying SIR and log-linear models to Italy and Spain's daily and cumulative incidence data was carried out (Chu, 2021). In all the cases, the result shows that the reproductive number was greater than one which means that the infection rate will keep growing. Stochastic modelling of COVID-19 prevalence patterns in four east African nations using ARIMA has been reported (Takele, 2020). The study concluded that COVID-19 cases in Somalia, Djibouti, Sudan and Ethiopia will increase rapidly after four months with an unknown peak.

South Africa's COVID-19 transmission dynamics were investigated using a compartmental model to assess the impact of various control measures (Garba *et al*., 2020). The study observed an extension in the predicted peak time of the pandemic which led to about 10% more cumulative deaths. The study also illustrated the effectiveness of some strategies to reduce the spread.

In different countries, the epidemiology of the infection with focus on forecasting of future incidence using dynamic time warping was reported (Stubinger and Schneider, 2020). The study predicted that among all countries of the world, China will have the highest cases within 29 days, United States will lead after 44 days and there will be a collapse of United Kingdom's healthcare system. Modelling and forecasting of early evolution of COVID-19 in Brazil using two variations of SIR model which takes into account the effect of social distancing which flattened the pattern of infection and the peak of infection will only shift with approximately the same value if social distancing is not sustained long enough was reported in (Bastos and Cajueiro, 2020). Analysis, modelling and forecasting of the infection in India was conducted using a mathematical model that monitors six compartments namely susceptible, asymptomatic, recovered, infected and isolated infected. The authors concluded that a reduction in contact rate between infected and uninfected individuals can only be achieved *via* quarantine of susceptible individuals, while elimination of the pandemic is also possible through the combination of social distancing and contact tracing (Arora *et al*., 2020; Sarkar *et al.*, 2020).

Estimation of the survival rate of COVID-19 in Nigeria using the Autoregressive Integrated Moving Average (ARIMA) was done, which found that the daily average survival rate of COVID-19 patients is 27.5% with a median survival rate of 25.4%. They also identified the ARIMA (0, 1, 1) as a better tool for predicting the survival rate of COVID-19 in Nigeria within the period of review (Aronu *et al.*, 2021). An examination of the transmission rate of COVID-19 in Nigeria using a time series approach was carried out, it was revealed that the daily mean transmission rate of COVID-19 was found to be 7.6% with 95% confidence interval of (0.0554, 0.0971) which shows that for every 100 persons tested, about 8 of them will test positive to COVID-19 in Nigeria (Mmaduakor *et al*., 2022). Also, the ARIMA (0, 1, 1) was identified to be appropriate for predicting the transmission rate of COVID-19 in Nigeria within the observed period. Further, finding showed that a slight variation exists between the forecasted and actual transmission rate of COVID-19 for June 2020 which indicates that the obtained ARIMA (0, 1, 1) model is adequate for estimating the transmission rate of COVID-19 in Nigeria.

In Nigeria, predictive modelling of COVID-19 confirmed cases using linear regression model to measure the impact of travelling history and contact on the spread of the disease before and after travel restriction was reported (Ayinde *et al*., 2020). The result of their prediction shows that travelling history and contacts increase the chances of people being infected by 85% and 88%, respectively. Modelling of Nigerian COVID-19 cases with comparative analysis of models and estimators was also reported (Aronu *et al*., 2021). The study concluded that the quartic linear regression model with an autocorrelated error of order 1 AR(1) was the best model for prediction of the daily COVID-19 data while the ordinary least squares, Cochrane Orcutt, Hildreth-Lu and Prais-Winsten as well as least absolute deviation (LAD) estimators were useful for parameter estimation. Apart from this, the survival rate of novel coronavirus has been forecasted using ARIMA modelling between 28th February and 30th June, 2020. The study found the mean survival rate to be 27.5% with a median survival rate of 25.4%, while it also predicted future survival rates using ARIMA modelling since there were little variations between the actual values and the predicted values. The study recommended the need for effective treatment strategies and local production of personal protective equipment in order to increase the survival rate. Analysis of variants of COVID-19 infections has been reported in some studies. A statistical analysis of COVID-19 initially using first-wave data of the WHO African region was carried out (James *et al*., 2022). The authors found that wealthier African countries with larger tourism industries and older populations have higher peak, cumulative attack rates and lower crude death ratios, while countries with stiffer early control policies experienced greater delay in the detection of first cases. Modelling and forecasting Nigeria's third wave of COVID-19 incidence rate was done (Odekina *et al*., 2022). Their projection showed a constant rise in death cases with a little decline in confirmed cases. Also, an investigation of the fourth wave of infection in India via statistical forecasting was reported (Rajeshbhai *et al*., 2022). A mixture of Gaussian distribution and bootstrap methodology was used to fit the data with the result predicting the peak of the fourth wave to be August 31, 2022. No doubt, the evolution of numerous approaches to evaluate COVID-19 situations since the first case had been reported has created a robust bank of literature for combating the spread of infection at different points. Also, the introduction of the COVID-19 vaccine on the 28th of February, 2021 by the Nigerian government ushers in new dynamics to the infection rate. This raises curiosity about the existing survival rate, current and future infection, recovery and death rates. This curiosity motivates the need for this work.

Materials and Methods

Data collection. The data used in this study was secondary data extracted from the weekly report of the Nigeria Centre for Disease Control (NCDC) website https://ncdc.gov.ng/. It consisted of 58-week data of COVID-19 confirmed (new) cases, discharged cases, and death between 28th February 2020 and 11th April 2021.

Method of data analysis. Numerous studies had been conducted to determine the trend of the COVID-19 pandemic by different methods (Fanelli and Piazza, 2020; Nishiura *et al*., 2020).

Box-Jenkins methodology also known as autoregressive integrated moving average (ARIMA) modelling had been used for time series data encountered in financial volatility and epidemiological studies. Extensive applications of (ARIMA) in modelling and forecasting of epidemic trends of infections and survival rates of the disease have also been documented (Lukman *et al*., 2020; Perone, 2020).

As a result, the methodology was employed in this study to model and forecast COVID-19 confirmed cases, discharged cases and deaths. The four basic steps required for creating a good model, namely identification, estimation, diagnostic checking and forecasting (Box *et al*., 2015; Box and Jenkins, 1976) were carefully followed. Also, the pattern of newly confirmed cases, discharged (recoveries) and deaths were investigated using time plots, while the survival rate is computed as a ratio of total discharged cases to total confirmed (active) cases. STATA (version 17) was used for the analysis.

Model identification, estimation and selection. To identify the best-fitted model for forecasting, the underlying (variation) mechanism in the data were explored using time plots and unit root test for the three series (confirmed, discharged and death). This was to ensure that the data are stationary since Box-Jenkins models are primarily stationary models (Akaike, 1974).

The unit root test by Augmented Dickey-Fuller procedure adjusts appropriately for the occurrence of serial correlation in the data given in the time series model specified by

$$Y_t = b_0 + b_1Y_{t-1} + b_2Y_{t-2} + ... + b_pY_{t-p} + e_t \text{ .... (1)}$$

where:

ε_t is the stochastic (residual) error term. The null hypothesis, that is, Y_t is non-stationary and is rejected if b_i is significantly negative. The number of lag (p) of Y_t is usually chosen to ensure that the regression is approximately white noise. The respective ACF and PACF can be estimated at lag k by

$$\hat{\rho}_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t-k+1}^{T} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{T} (Y_t - \bar{Y})^2} \text{ if } k = 0,1,2 \text{ .... (2)}$$

if k = 1

$$\phi_{kk} = \left[\begin{array}{l} \hat{\rho}_1 \\ \dfrac{\hat{\rho}k - \sum_{j=1}^{k-1} \hat{\rho}_{k-1,j}\, \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\rho}_{k-1,j}\, \hat{\rho}_k} \text{ if } k = 2,3 \text{ .... (3)} \end{array} \right.$$

where:

$\hat{\rho}_{kj} = \hat{\rho}_{k-1,j} - \hat{\rho}_{kk}\hat{\rho}_{k-1,k-j}$ for j = 1,2 ..... k-1. So, that (2) and (3) are fitted using sample values to obtain the ACF and PACF plots, which were used to further obtain the orders p and q. If the ACF and PACF plots slowly (geometrically) decay with significant spikes at some lags, then we will have an array of ARIMA models. If, the AR(p) or MA(q) may be used, the general form of the ARIMA (p,d,q) is given by

$$Y_i = c + \sum_{f=1}^{p} \phi_l Y_{t-l} + \sum_{f-1}^{q} \theta_l \varepsilon_{t-l} + \varepsilon_t \text{ .... (4)}$$

where:

c is any arbitrary constant, ε_t is called, while noise white noise, $E(\varepsilon_t) = 0$ and $V(\varepsilon_t) = \sigma^2$, where *p* is the number of autoregressive terms, *d* is the number of differencing and *q* is the number of moving average terms. In selecting the best ARIMA model from the array of candidate models, AIC (Akaike, 1974) and SIC (Schwarz, 1978) with respective computational formulae:

$$AIC = -2\log_e \hat{L} + 2p \text{ .... (5)}$$

$$SIC = -2\log_e \hat{L} + \log_e n \text{ .... (6)}$$

where:

p is the number of parameters, n is the number of observations in the modeled dataset and $\log_e \hat{L}$ is the log-likelihood of the observed data Y under the model of interest. The model with the least performance score is considered the best (optimal) model which is then estimated by either the ordinary least squares (OLS) or the maximum likelihood (ML) approach. Other methods, models, comparisons and reviews on COVID-19 data can be seen in (Baleanu *et al*., 2022; Mohammadi *et al*., 2022).

Diagnostic checking. The optimal fitted model is tested for lack of fit or otherwise using the Portmanteau Q test reported in (Ljung and Box, 1978) with statistic

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\rho_k^2}{n - k} \text{ .... (7)}$$

where:

n is the length of the time series, ρ_k is the sample autocorrelation at log k and h is the number of logs. The aim is to test whether or not the ARIMA model matches the dataset. Under the null hypothesis, the statistic asymptotically follows a χ^2 (h) at α level of significance.

Results and Discussion

The result in Table 1 shows the descriptive analysis of Nigeria's COVID-19 confirmed cases, discharged and death data extracted from (NCDC, 2020) between 28th February, 2020 and 11th April, 2021. It covers 58 weeks and the tail end of this period coincided with the period of introduction of the COVID-19 vaccine. For effective ARIMA modelling, sample size must be at least 30, and this validates the reason for choosing 58 observations (Box *et al*., 2015).

Table 1 shows that on weekly basis, the highest number of confirmed COVID-19 cases was 11179, the highest number of discharged cases was 9287 and the highest number of COVID-19 deaths was 129. There are weeks with no discharged and death cases and the lowest number of confirmed cases was 15. The result in Table 1 showed that on average 2974 people were infected, 2309 people were discharged and 36 people died of COVID-19 on weekly basis. Skewness values for data of confirmed and discharged cases are higher than 1 indicating that they are highly skewed to the right, while data of death cases is less than 1 indicating that the data is moderately skewed. Kurtosis values for confirmed and discharged cases are leptokurtic since they are higher than 3, indicating the presence of outliers, while data for death cases is platykurtic since it is lower than 3, indicating an absence of outliers. Table 1 also found the coefficient of variation to be 90.9% for confirmed cases, 107.59% for discharged and 89.2% for death, the mean survival rate to be 0.777 and the median survival rate to be 0.804. When the survival rate is 0.5 and above, then it can be said that the survival rate is high. High coefficient of variation values shows high dispersion or variations in the data sets. The high weekly mean and median survival rates observed here may be attributed to the effect of the approval and introduction of COVID-19 vaccines by the government to fight the infection.

The pattern of Nigeria's COVID-19 infection was displayed for confirmed (active), discharged (recoveries) and deaths in Fig. 1(a-e).

Table 1. Descriptive statistics on COVID-19 confirmed cases, discharged cases and death

Statistics	Confirmed cases	Discharged	Deaths
Minimum	15	0	0
Maximum	11179	9287	129
Sum	172479	133931	2068
Mean	2973.776	2309.155	35.655
Median	1924	1546.500	27.500
Standard deviation	2679.037	2484.358	31.740
Coefficient of variation	90.090	107.590	89.020
Skewness	1.514	1.409	0.632
Kurtosis	4.748	3.970	2.563
Mean survival rate		0.777	
Median survival rate		0.804	

Figure 1(a-e), was observed that the three series are generally characterized by steady fluctuations from beginning to end. Fig. 1(a) showed newly confirmed infection figure rose to the highest in the 48th week, which systematically declined from then till the 57th week. Fig. 1(b) like Fig. 1(a) displayed similar movement in the discharged series which peaked in the 47th week but declined systematically from then till the 56th week. Fig. 1(c) showed that the number of deaths peaked in the 19th week, from then on the number of deaths started declining although not sharply. The reason is not far-fetched, since only a small proportion of the population has embraced vaccination. Fig. 1(d) and (e) show patterns of all three cases. From Fig. 1(e), the patterns of the logarithm of all three cases show that the pattern of death was 'expectedly' low indicating low death rates when compared to the infection (confirmed) and recovery (discharged) rates.

Test for stationarity. The Augmented Dickey-Fuller (ADF) test was employed to test if the three series (corresponding to confirmed, discharged and death) in

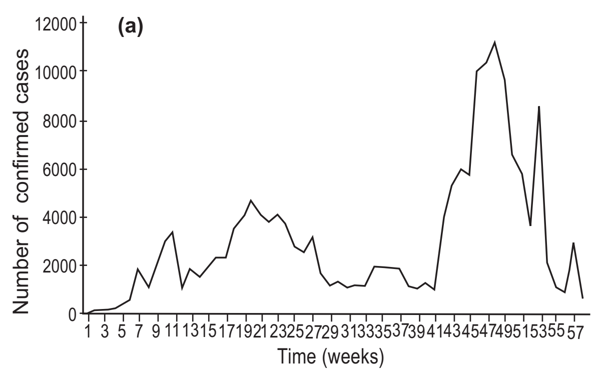

Fig. 1(a). Pattern of confirmed cases.

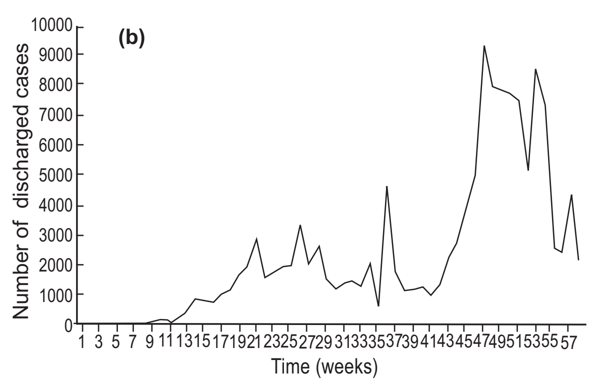

Fig. 1(b). Pattern of discharged cases.

the COVID-19 data are stationary (that is, possess no unit root) or otherwise. The null hypothesis is such that the series is non-stationary. The result of the ADF test is provided in Table 2.

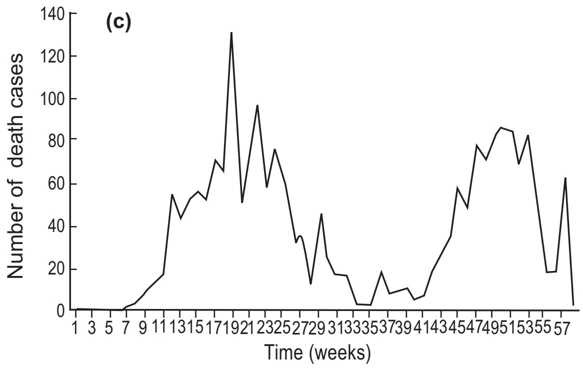

Fig. 1(c). Pattern of deaths.

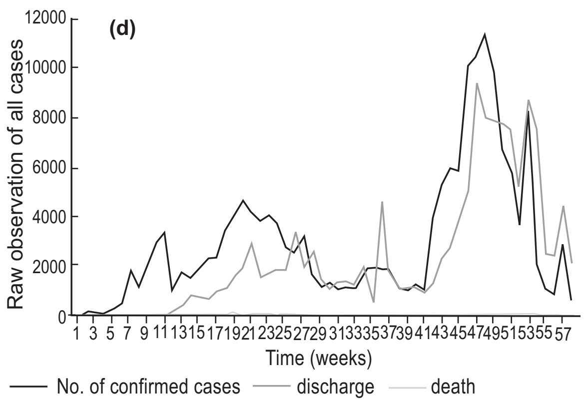

Fig. 1(d). Pattern of confirmed, discharged and death cases.

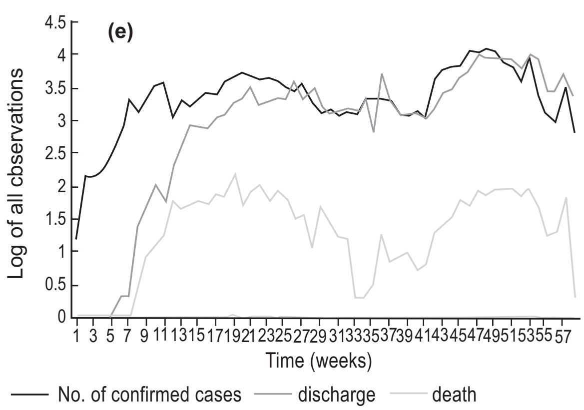

Fig. 1(e). Pattern of log of confirmed, discharged and death cases.

In Table 2, it was observed that the series of confirmed and discharged cases to are I(1), while that of death is I(3). That is confirmed and discharged series are stationary after the first difference while the death series was stationary after the third difference since Mackinnon P-values corresponding to their ADF test statistics was <0.001*, which were less than the chosen critical value of alpha = 5%. This follows that these three data sets have satisfied the condition for ARIMA modelling.

Box-Jenkins modelling. Box method was employed for modelling and forecasting of the COVID-19 data. The ACF and PACF were used to determine the type and order of the models since the three series have become stationary (Yaffe and McGee, 2000). An array of candidate models were obtained using significant

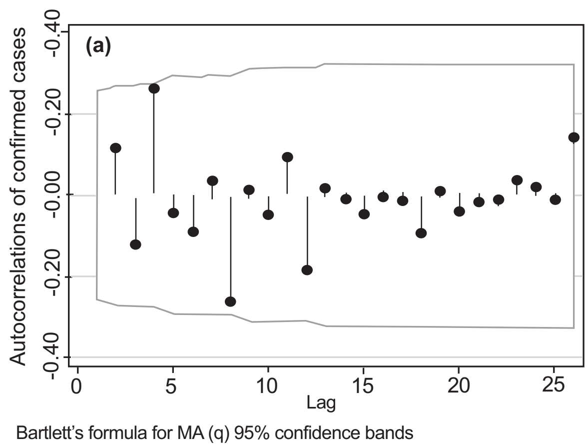

Fig. 2(a). ACF of confirmed COVID-19 cases after first difference.

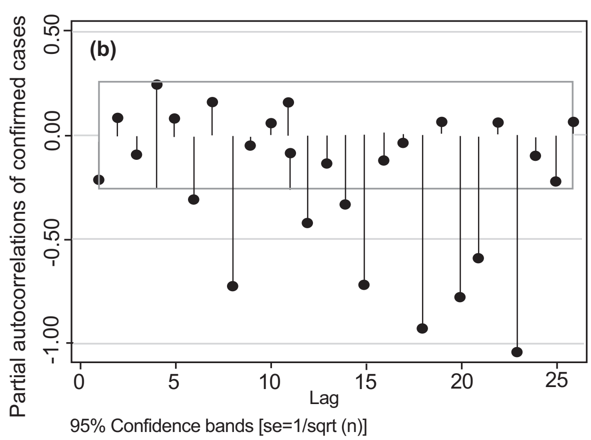

Fig. 2(b). PACF of confirmed COVID-19 cases after first difference.

Table 2. Augmented Dickey-fuller test for stationarity

Series	Df	ADF	5% CV	Mackinnon P-value	Order of integration	Decision
Obs. confirmed cases	58	-2.337	-2.924	0.160	I(0)	Accept
First difference	57	-8.967	-1.674	0.000*	I(1)	Reject
Obs. discharged	58	-2.270	-2.924	0.182	I(0)	Accept
First difference	57	-9.459	-2.925	0.000*	I(1)	Reject
Obs. death	58	-2.855	-2.924	0.051	I(0)	Accept
Third difference	55	-11.694	-2.925	0.000*	I(3)	Reject

Note: * imply significant and thus stationarity at alpha = 5%.

spikes as orders of AR and MA terms in the ACF and PACF plots of the three series and displayed in Fig. 2(a-c).

Figures 2(a-f) are demonstrate gradually decreasing Autocorrelation function (ACF) and Partial Auto-Correlation Function (PACF) which sharply dropped simultaneously after p significant lags, while in Fig. 2(c), ACF and PACF show gradually decaying patterns with p and q significant lags. From these, an array of candidate models are obtained and presented in Table 3.

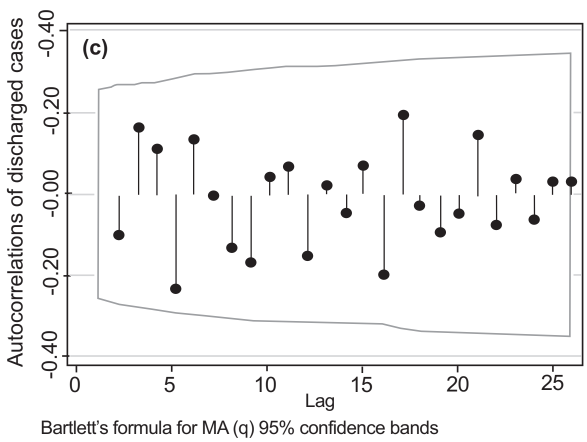

Fig. 2(c). ACF of discharged COVID-19 cases after first difference.

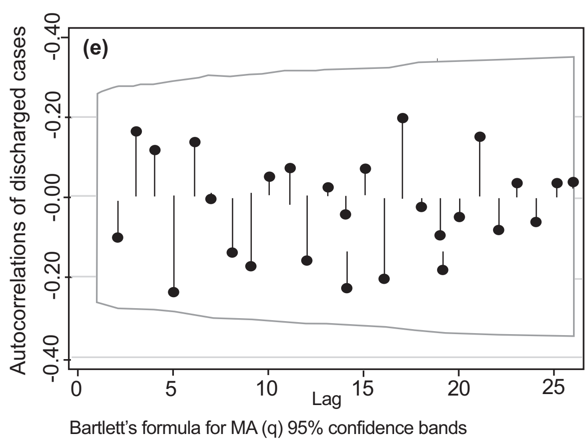

Fig. 2(e). ACF of COVID-19 deaths after third difference.

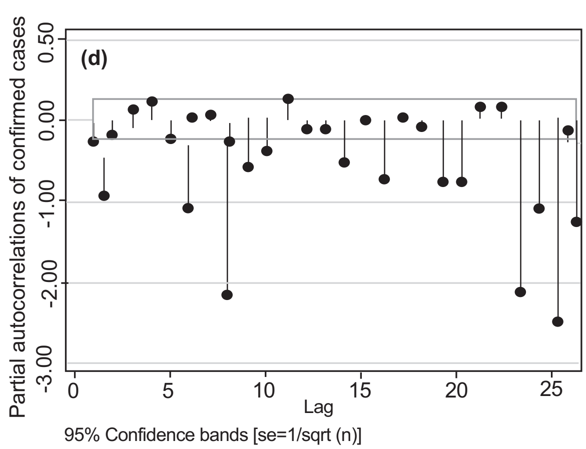

Fig. 2(d). PACF of discharged COVID-19 cases after first difference.

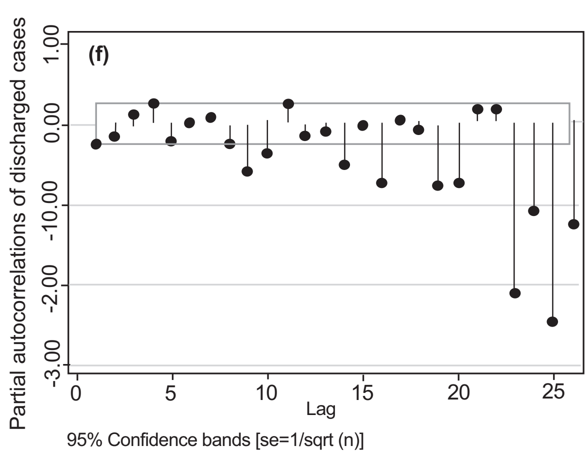

Fig. 2(f). PACF of COVID-19 deaths after third difference.

Table 3. Possible fitted models for confirmed, discharged and death cases

Series	Models	AIC	SIC	Best model
Confirmed cases	(1,1,0)	1010.064	1016.140	ARIMA (4,1,0)
	(2,1,0)	1007.893	1015.995	
	(3,1,0)	997.408	1007.535	
	(4,1,0)	**996.255**	**1006.407**	
	(5,1,0)	998.020	1012.197	
	(6,1,0)	997.500	1013.703	
	(7,1,0)	996.290	1014.518	
	(8,1,0)	996.280	1016.533	
	(9,1,0)	993.313	1015.592	
Discharged	(1,1,0)	1006.502	1012.578	ARIMA (3,1,0)
	(2,1,0)	990.607	998.708	
	(3,1,0)	**980.644**	**990.771**	
	(4,1,0)	982.503	994.655	
	(5,1,0)	982.575	996.753	
	(6,1,0)	982.962	999.165	
	(7,1,0)	984.950	1003.177	
	(8,1,0)	985.735	1005.989	
	(9,1,0)	987.492	1009.771	
	(10,3,1)	981.102	1005.406	
Deaths	(1,3,1)	578.086	584.053	ARIMA (7,3,1)
	(2,3,1)	550.230	560.175	
	(3,3,1)	534.881	544.826	
	(5,3,1)	527.119	541.0421	
	(6,3,1)	528.123	546.024	
	(7,3,1)	**516.086**	**533.987**	
	(8,3,1)	517.164	537.054	

Note: Bold implies best-fitted model with minimum AIC and SIC values.

From the array of candidate models presented in Table 3, the best-fitted models for forecasting each of the three series were found to be ARIMA (4,1,0), (3,1,0) and (7,3,1) for confirmed, discharged and death cases respectively since they have the least AIC and SIC scores. The best models obtained were fitted to each of the three data sets and the result is presented in Table 4.

From Table 4, estimates of the three best (ARIMA) models were found to be statistically significant since their corresponding P-values were less than the chosen critical value at alpha=5% level. The best-fitted models were eventually employed to forecast future COVID-19 confirmed (infection), discharged (recovery) and death rates for 29 weeks respectively at 95% confidence level. The forecast is available in Table 5.

Table 5 indicates that future confirmed and discharged cases will decline significantly while there will be zero COVID-19 death by the 67th week, 14 weeks after the introduction of vaccine. Apart from this, forecast model values of confirmed, discharged and death cases were superimposed on actual values' plot with the results displayed in Fig. 3(a-c).

Figures 3(a-c) show that the actual (observed) values (of confirmed, discharged and deaths) intersect the predicted values throughout the 58 week period of observation. Also, little variation exists between forecasted and actual values and the forecasted series are within the 95% confidence bounds. Figure 3(c) essentially showed that the death pattern flattens (out)

Table 4. Parameter estimates of the best ARIMA models

Cases	Models	Variables	Coefficients	s.e	Z	P>|Z|	95% C.I. Lower limit	upper limit
Confirmed	ARIMA (3,1,0)	Constant	-15.395	62.313	-0.205	0.805	-137.527	106.736
	AR1	-1.028	.093	11.03	0.000*	-1.211	-.845	
	AR2	-.805	.159	-5.07	0.000*	-1.116	-.494	
	AR3	-.708	.188	-3.77	0.000*	-1.077	-.340	
	AR4	-.683	.167	-3.99	0.001*	-1.427	.061	
Discharged	ARIMA (4,1,0)	Constant	-13.635	56.169	-0.24	0.808	-123.724	96.453
	AR1	-1.075	.111	-9.69	0.000*	-1.292	-.857	
	AR2	-.929	.140	-6.62	0.000*	-1.204	-.655	
	AR3	-.475	.138	-3.43	0 .001*	-.746	-.204	
Death	ARIMA (7,3,1)	Constant	0.004	0.018	0.2	0.404	-0.033	0.040
	AR1	-2.163	0.137	-15.76	0.000*	-2.432	-1.894	
	AR2	-2.836	0.294	-9.66	0.000*	-3.411	-2.261	
	AR3	-2.837	0.432	-6.57	0.000*	-3.683	-1.991	
	AR4	-2.506	0.554	-4.52	0.000*	-3.593	-1.420	
	AR5	-1.998	0.707	-2.83	0.005*	-3.384	-0.610	
	AR6	-1.292	0.554	-2.33	0.020*	-2.379	-0.206	
	AR7	-0.494	0.239	-2.06	0.039*	-0.963	-0.025	
	MA1	-1.000	0.197	-5.07	0.000*	-1.387	-0.613	

Note: *depicts statistical significance at alpha = 5% level.

Table 5. Forecast of COVID-19 infection for the next 29 weeks according to ARIMA models with 95% confidence interval

	Confirmed ARIMA (3,1,0)			Discharged ARIMA (4,1,0)			Death ARIMA (7,3,1)		
Week	Prediction	LL	UL	Prediction	LL	UL	Prediction	LL	UL
59	558	435.527	679.794	256	145.807	365.989	86	85.309	85.765
60	1023	900.904	1145.171	2550	2439.857	2660.039	42	41.409	41.865
61	890	768.365	1012.632	1073	-1182.890	-962.710	20	19.825	20.281
62	1049	-1171.55	-927.283	1387	-1496.720	-1276.53	10	9.644	10.099
63	569	-691.348	-447.080	1228	1118.294	1338.475	5	4.416	4.872
64	469	346.745	591.012	431	321.169	541.351	2	2.169	2.625
65	424	301.941	546.208	994	-1104.080	-883.896	1	0.791	1.247
66	190	-312.196	-67.929	36	-73.915	146.267	1	0.410	0.866
67	385	-507.507	-263.239	633	522.709	742.891	0	-0.063	0.393
68	66	-56.457	187.811	289	-398.910	-178.729	0	-0.004	0.453
69	206	83.857	328.125	343	-452.610	-232.428	0	-0.193	0.263
70	0	-121.971	122.297	288	178.377	398.558	0	-0.101	0.355
71	169	-290.834	-46.567	98	-11.890	208.292	0	-0.145	0.311
72	49	-170.763	73.505	258	-368.484	-148.303	0	-0.124	0.332
73	72	-49.800	194.467	2	-108.206	111.976	0	-0.134	0.321
74	26	-96.536	147.731	144	33.990	254.171	0	-0.129	0.327
75	64	-186.019	58.249	81	-191.347	28.834	0	-0.132	0.325
76	52	-173.970	70.297	95	-205.047	15.134	0	-0.131	0.326
77	9	-113.372	130.895	62	-48.410	171.772	0	-0.131	0.325
78	13	-109.583	134.684	13	-96.936	123.246	0	-0.131	0.325
79	25	-147.013	97.255	74	-183.892	36.290	0	-0.131	0.325
80	36	-157.697	86.570	10	-119.759	100.422	0	-0.131	0.325
81	13	-135.369	108.899	25	-84.786	135.395	0	-0.131	0.325
82	2	-124.224	120.043	31	-140.672	79.509	0	-0.131	0.325
83	14	-136.143	108.124	34	-143.599	76.583	0	-0.131	0.325
84	24	-145.834	98.434	5	-105.118	115.063	0	-0.131	0.325
85	18	-140.102	104.165	7	-117.197	102.985	0	-0.131	0.325
86	11	-132.715	111.553	29	-138.598	81.583	0	-0.131	0.325
87	13	-134.894	109.374	13	-122.654	97.527	0	-0.131	0.325

Note: LL represents lower limit and UL represents upper limit.

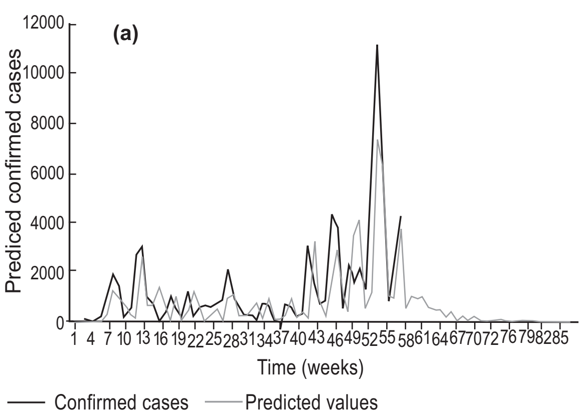

Fig. 3(a). Forecast model values of confirmed cases superimposed on the actual values.

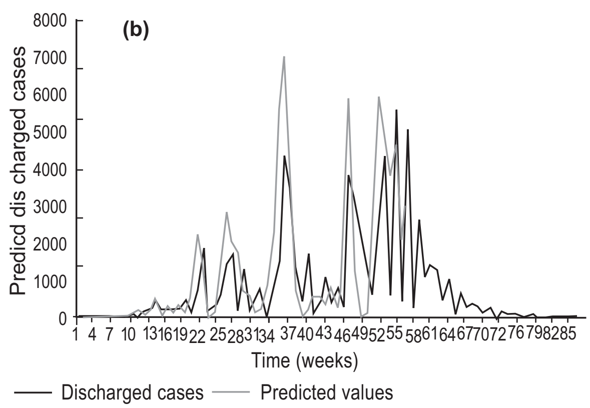

Fig. 3(b). Forecast model values (of discharged COVID-19 cases) superimposed on the actual values.

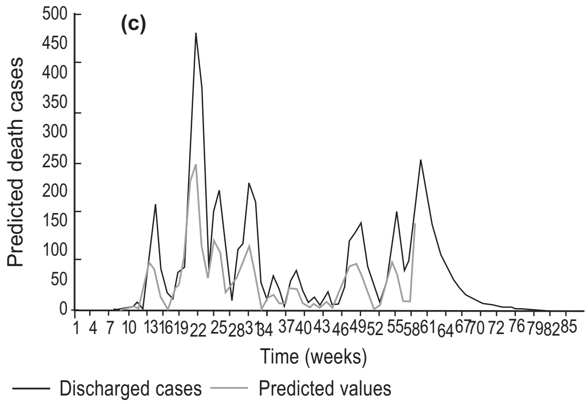

Fig. 3(c). Forecast model values of COVID-19 deaths superimposed on the actual values.

after a few weeks. Finally, the Portmanteau test was employed to test whether the time series model provides a good fit for the COVID-19 or otherwise. Table 6 shown the result of the test.

The results obtained in Fig. 3(a-c) and Table 6 suggest that the predicted values fitted well with actual values. The ARIMA models also provide a perfect fit for each of the three cases (since P>0.05).

Table 6. Portmanteau test of the models

Variable	Degree of freedom	Portmanteau (Q) statistic	P-value
Confirmed cases	28	21.737	0.703
Discharged cases	28	29.684	0.281
Death cases	28	60.140	0.100

Conclusion

The study examined Nigeria's COVID-19 trend during the introduction of the COVID-19 vaccine using the Box-Jenkins methodology. ARIMA models were applied to confirmed cases, discharged cases and COVID-19 deaths for 58 consecutive weeks (from February 28th, 2020 to April 11th, 2021 after they had met stationarity conditions. The best ARIMA models were obtained for the cases and a 29-week forecast was made based on these models. The forecasts show a sharp decline in the number of confirmed and discharged cases, while zero COVID-19 death was recorded by the 67th week (8 weeks into the forecast) which apparently was 14 weeks after the introduction of vaccine in Nigeria. The findings in this study also indicate that the respective mean and median weekly survival rates of 0.777 and 0.804 are relatively high since they are greater than 0.5, which may be attributed to the impact of vaccine introduction. This study reveals that government efforts in combating the spread of the virus are yielding the intended results. It is therefore, recommended that strategies aimed at expanding the existing vaccination levels through public awareness and enlightenment programmes will eliminate the perceived lack of trust, myth and misconception about the COVID-19 vaccine and encourage more vaccination.

Authors' contributions. O.Y. conceived and initiated the research idea. O.Y. and B.T. developed the theory and performed the computations. O.Y., B.T., J. S.O., T.O. K.A and O.F. verified the analytical methods. O.Y., B.T, J. and S.O, wrote the manuscript with input from all authors. O.Y. supervised the findings of this work. All authors discussed the results and contributed to the editing of the final manuscript.

Conflict of Interest. The authors declare that they have no conflict of interest.

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19:** 716-723. http://dx.doi.org/10.1109/TAC.1974.1100705

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F. 2020. The proximate origin of SAR-CoV-2. *Natural Medicine*, **26:** 450-452. DOI: 10.1038/s41591-020-0820-9.

Aronu, C.O., Ekwueme, G.O., Sol-Akubude, V.I., Okafor, P.N. 2021. Coronavirus (COVID-19) in Nigeria: survival rate, *Scientific African*, **11:** e00689 https://doi.org/10.1016/j.sciaf.2020.e00689

Arora, P., Kumar, H., Panigrahi, B.K. 2020. Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India. *Chaos Solitons Fractals*, **139:** 110017. doi:10.1016/j.chaos.2020.110017

Ayinde, K., Lukman, A.F., Rauf, R.I., Alabi, O.O., Okon, C.E. 2020. Modeling Nigerian COVID-19 cases: a comparative analysis of models and estimators. *Chaos Solitons Fractals*, **138:** 109911. https://doi.org/10.1016/j.chaos.2020.109911

Baleanu, D., Abadi, M.H., Jajarmi, A., Vahid, K.Z., Nieto, J.J. 2022. A new comparative study on the general fractional model of COVID-19 with isolation and quarantine effects. *Alexandria Engineering Journal*, **61:** 4779-4791. https://doi.org/10.1016/j.aej.2021.10.030

Bastos, S.B., Cajueiro, D.O. 2020. Modelling and forecasting the early evolution of the COVID-19 pandemic in Brazil. *Scientific Reports*, **10:** 19457. https://doi.org/10.1038/s41598-020-762571

Betthäuser, B.A., Bach-Mortensen, A.M., Engzell, P. 2023. A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, **7:** 375-385. doi: 10.1038/s41562-022-01506-4

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. 2015. *Time Series Analysis. Forecasting and Control*, pp. 712, John Wiley & Sons Inc., Hoboken, New Jersey, USA.

Box, G.E.P., Jenkins, G.M. 1976. *Time Series Analysis: Forecasting and Control*, pp. 575, revised edition Holden Day, San Francisco. CA, USA.

Chu, J. 2021. A statistical analysis of the novel corona-virus COVID-19 in Italy and Spain, *PLoS ONE*, **16:** 1-36. https://doi:10.1371/journal.pone. 0249037

Fanelli, D., Piazza, F. 2020. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals*, **134:** 109761. doi:10.1016/j.chaos.2020.109761

Garba, S.M., Lubuma, J.M., Tsanou, B. 2020. The transmission dynamics of COVID-19 pandemic in south Africa. *Mathematical Biostatistics*, **328:** 108441. doi:10.1016/j.mbs.2020.108441

Holmes, E.C., Goldstein, S.A., Rasmussen, A.L., Robertson, D.L., Crits-Christoph, A., Wertheim, J.O., Anthony, S.J., Barclay, W.S., Boni, M.F., Doherty, P.C., Farrar, J., Geoghegan, J.L., Jiang, X., Leibowitz, J.L., Neil, S.J.D., Skern, T., Weiss, S.R., Worobey, M., Andersen, K.G., Garry, R.F., Rambaut, A. 2021. The origins of SAR-CoV-2: a critical review. *Cell*, **84:** 4848-4856. https://doi.org/10.1016/j.cell.2021.08.017

Ibekwe, T.S., Fasunla, A.J. 2020. Telemedicine in otorhinolaryngological practice during COVID-19 pandemic. *Nigerian Medical Journal*, **61:** 111-113. doi:10.4103/nmj.NMJ_201_20

James, A., Dalal, J., Kousi, T., Vivacqua, D., Câmara, D.C.P., Dos Reis, I.C., Mesa, S.B., Ng'ambi, W., Ansobi, P., Bianchi, L.M., Lee, T.M. 2022.. An in-depth statistical analysis of the COVID-19 pandemic's intial spread in the WHO African region. *BMJ Global Health*, **7:** e007295. doi:10.1136/bmjgh-2021-007295

Ljung, G.M., Box, G.E.P. 1978. On a measure of lack of fit in time series models, *Biometrika*, **65:** 297-303. Doi:10.1093/biomet/65.2.297

Lukman, A.F., Rauf, R.I., Abiodun, O., Oludoun, O., Ayinde, O., Ogundokun, R.O. 2020. COVID-19 prevalence estimation: four most affected African countries, *Infectious Disease Modelling*, **5:** 827-838. https://doi.org/10.1016/j.idm.2020.10.002

Mmaduakor, C.O., Adejuwon, S.O., Adeyefa, E.A., Ajewole, O.R., Ngwu, B.A., Kuboye, J.O. 2022. Transmission rate of Coronavirus (COVID-19) in Nigeria. *Savanna Journal of Basic and Applied Sciences*, **4:** 1-10. http://www.sjbas.com.ng

Mohammadi, H., Rezapour, S., Jajarmi, A. 2022. On the fractional SIRD mathematical model and control for the transmission of COVID-19: the first and the second waves of the disease in Iran and Japan. *ISA Transactions*, **124:** 103-114. doi:10.1016/j.isatra.2021.04.012

Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S., Wakefield, J. 2023. The WHO estimates excess mortality associated with the Covid-19 pandemic. *Nature*, **613:** 130-137. https://doi.org/10.1038/s41586-022-05522-2

Nigeria Centre for Disease Control and Prevention. 2020. https://ncdc.gov.ng

Nishiura, H., Linton, N.M., Akhmetzhanov, A.R. 2020. Serial interval of coronavirus (COVID-19) Infections. *International Journal of Infectious Diseases*, **93:** 284-286. doi:10.1016/j.ijid.2020.02.060.epub 2020mar4

NPHCDA, 2021. Nigeria Primary Health Care Development Agency Cited 2022 October, 28. https://nphcda.vaccination.gov.ng/vaccination-sites

Odekina, G., Adedayo, A., Imaga, O.F. 2022. Modeling and forecasting the third wave of COVID-19 incidence rate in Nigeria using vector autorgressive model approach. *Journal of the Nigerian Physical Sciences*, **4:** 117-122. doi:10.46481/jnsps.2022.431

Ogundokun, R.O., Lukman, A.F., Kibria, G.B., Awotunde, J.B., Aladeitan, B.B. 2020. Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, **5:** 543-548. https://doi.org/10.1016/j.idm.2020.08.003

Rajeshbhai, S.P., Dhar, S.S., Shalabh. 2022. Fourth wave of COVID-19 in India: statistical forecasting.

medRxiv, doi:10.13140/RG.2.2.32048.02566

Perone, B. 2020. An ARIMA model to forecast the spread and final size of COVID 2019 epidemic in Italy. *SSRN Electronic Journal*, doi.org/10.1101/2020.04.27.20081539

Sarkar, K., Khajanchi, S., Nieton, J.J. 2020. Modeling and forecasting of COVID-19 pademic in India. *Chaos Solitons Fractals*, **139:** 110049. DOI:10.1016/j.chaos.2020.110049

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, **6:** 461-464. http://dx.doi.org/10.1214/aos/1176344136

Stubinger, J., Schneider, L. 2020. Epidemiology of Corona-virus COVID-19: forecasting the future incidence in different countries. *Health Care*, **8:** 99. doi:10.3390/healthcare8020099

Takele, R. 2020. Stochastic modelling for predicting COVID-19 prevalence in east African countries. *Infectious Disease Modelling*, **5:** 598-607. DOI:10.1016/j.idm.2020.08.005

Wuhan Municipal Health and Health Commission's Briefing on the Current Pneumonia Epidemic Situation in our City. 2020. (WMHC) http://wjw.wuhan.gov.cv/front/web/showDetail/2019123108989

Yaffee, R.A., McGee, M. 2000. *Introduction to Time Series Analysis and Forecasting: With Applications in SAS and SPSS*. Academic Press, USA. https://www.researchgate.net/publication/234789625

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W. 2020. A novel coronavirus from patients with pneumonia in China. *New England Journal of Medicine*, **382:** 727-733. doi:10.1056/NEJMoa2001017