Forecasting of New Cases of COVID-19 in Nigeria Using Autoregressive Fractionally Integrated Moving Average Models

Olumide Sunday Adesina[1*], Samson Adeniyi Onanaye[1], Dorcas Okewole[1] and Amanze C. Egere[1]

[1]*Department of Mathematical Sciences, Redeemer's University, Nigeria.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

Original Research Article	

Abstract

The emergence of global pandemic known as COVID-19 has impacted significantly on human lives and measures have been taken by government all over the world to minimize the rate of spread of the virus, one of which is by enforcing lockdown. In this study, Autoregressive fractionally integrated moving average (ARFIMA) Models was used to model and forecast what the daily new cases of COVID-19 would have been ten days after the lockdown was eased in Nigeria and compare to the actual new cases for the period when the lockdown was eased. The proposed model ARFIMA model was compared with ARIMA (1, 0, 0), and ARIMA (1, 0, 1) and found to outperform the classical ARIMA models based on AIC and BIC values. The results show that the rate of spread of COVID-19 would have been significantly less if the strict lockdown had continued. ARFIMA model was further used to model what new cases of COVID-19 would be ten days ahead starting from 31st of August 2020. Therefore, this study recommends that government should further enforce measures to reduce the spread of the virus if business must continue as usual.

_____

*Corresponding author: E-mail: adesinas@run.edu.ng, olumidestats@gmail.com;

1 Introduction

COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and was first recorded in China in 2019. The respiratory disease has spread globally from 2019 to the present time. COVID-19 was declared as a pandemic because it has proven to be a threat to human lives. The first confirmed case of COVID-19 in Nigeria was reported on 9 March 2020 as Nigeria Centre for Disease Control (NCDC).

COVID-19 pandemic is a disease affecting humans globally and is impacting on people's health which has resulted into millions of deaths as of mid-August 2020. At one period or the order, lockdown was enforced in Nigeria just as every other nation of the world with well-defined conditions. The first phase of lockdown in Nigeria which started on March 30th, 2020 was the period of total lockdown, and that was when significant increase in the number of people tested who positive to the COVID-19 was recorded, and there was little or no tangible clue about how to curtail the virus. So, government at all levels were obliged to provide palliative for citizenry in the period of lockdown though the palliative could only reach a few. Some Nigerians faced untold hardship in the period of lockdown because there was no income during the lockdown and palliative given by government limited, hence the need to partially ease the lockdown so that people can earn a living.

This study modelled what daily number of new positive cases of COVID-19 would have been if the strict lockdown continues and the actual value of new cases when the lockdown was relaxed with fewer strict conditions.

There are different mathematical models proposed by authors for modelling, detecting and forecasting cases of COVID-19 few of such are discussed below:

1.1 The SIR and SEIR model

The SIR and SEIR models are calculus based, and the popular among such models and the models are commonly used to model cases of the epidemic. One of these studies can be found in the work of [1]. Several authors have modified and extended the SIR and SEIR model. Following the acronym of the SIR model, the population is divided into three compartments; the susceptible (S), infected (I), recovered (R). Let $S, I,$ and R be the corresponding number of hosts in each population, then the changes that occur in the variables is represented by the system of differential equations as follows:

$$\frac{dS}{dt} = B - \lambda S - \mu S \tag{1}$$

$$\frac{dI}{dt} = \lambda S - \gamma I - \mu I \tag{2}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{3}$$

where B is the crude birth rate (births per unit time), μ is the death rate, and γ is the recovery rate. Assuming that the force of infection, λ, has the form,

$$\lambda = \beta \frac{I}{N} \tag{4}$$

where N is the total population size $(N = S + I + R)$. E added to the acronym SIR is the exposure compartment to make it SEIR. Other authors who adopted the SIR/SEIR model include [2-7]. Artificial intelligence and machine learning approach was equally used to model and forecast cases of COVID-19. The works of [8-15] shows the application of such.

1.2 The time series model

Time series model are basically used for forecasting, and various time series models can be found in the study by [16]. The works of [17-20] contains various time series model used to model and forecast the new cases of the novel COVID-19. In order to achieve the objective of this study, ARFIMA model was adopted for modelling and forecasting. The ARFIMA model was first introduced by Granger and [21, 22]. The order of ARFIMA is $(p, d, q), p$ is autoregressive order $AR(p), q$ is the moving average order $MA(q)$, and (d) is order of fractional differentiation. Among authors who applied time series to forecast cases of COVID-19 was [23]. The author used ARFIMA to model cases of COVID-19 in Algeria.

Time series data are usually first differenced if it is discovered that it is not stationary. A lot of time series data have too much long-range dependence so, they are rather classified as $I(0)$ rather than $I(1)$. Therefore, ARFIMA model is formulated to represent the long-range dependence series. It is important to conduct a Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [24] with a null hypothesis of $I(0)$ before implementing ARFIMA. The ARFIMA model generalises the ARIMA model's integer order of integration such that $'d'$ will be in the range $-0.5 < d < 0.5$. In time series, fractional integration is often used to describe long memory phenomena or long-range dependence relative to an ARIMA model which can have a short memory series. The remaining section of this paper is sectionalized as follows; section 2 is the materials and methods, section 3 is the results and discussion, and finally, section 3 is the conclusion.

2 Materials and Methods

2.1 The ARFIMA model

The Gaussian ARFIMA (p, d, q) may be expressed as

$$\phi(ﬧ)(1 - ﬧ)^d(y_t - \mu) = \Psi(ﬧ)\epsilon_t, \quad \epsilon_t \sim i.i.d \ (0, \sigma_\epsilon^2) \tag{5}$$

Where $ﬧ$ is the backward-shift operator, the AR lag is represented by $\phi(ﬧ) = 1 - \phi_1 ﬧ - \cdots - \phi_p ﬧ^p$, $and \ \Psi(ﬧ) = 1 + \vartheta_1 ﬧ + \cdots \vartheta_p ﬧ^p$ is the MA polynomial, $(1 - ﬧ)^d$ is the fractional differencing operator defined by

$$(1 - ﬧ)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)ﬧ^k}{\Gamma(-d)\Gamma(k + 1)} \tag{6}$$

The autocovariance function of the ARMA process with mean (μ) is expressed as

$$\gamma_i = E[(y_t - \mu)(y_{t-i} - \mu)] \tag{7}$$

And the variance matrix of the joint distribution of $y = (y_1, \dots, y_T)'$

And $V(y)$ can be expressed as

$$V(y) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{T-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} = \Sigma \tag{8}$$

$V(\boldsymbol{y})$ is a symmetric Toeplitz matrix, denoted by $T[\gamma_0, \ldots., \gamma_{T-1}]$. Under normality and

$$y \sim N_T(\mu, \Sigma) \tag{9}$$

In order to compute the autocovariances in (8), the log-likelihood is

$$\log L\ (d,\ \phi, \vartheta, \sigma_\epsilon^2) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}log\Sigma - \frac{1}{2}logz'\Sigma^{-1}z \tag{10}$$

Where $z = y - \mu$

The fractional parameter 'd' is allowed to assume any real value, and $\Gamma(\cdot)$ denoting the gamma (generalized factorial) function [20]. ARFIMA model is expressed as:

$$y_t = (1 - \daleth)^{-d}\big(\phi(\daleth)\big)^{-1}\Psi(\daleth)\epsilon_t \tag{11}$$

The process reveals short memory if $d = 0$. According to [25], if a series exhibits long memory, it is said to be integrated of order d, that is, it is an $I(d)$ process, with d a real number. This implies that it is neither stationary I(0) nor a unit root I(1) process. There are various methods of estimating ARFIMA model. In this study, the exact maximum likelihood estimation proposed by [26] was used for the estimation.

2.2 Data description

The data set used in this study was daily new positive cases of COVID-19 taken from Nigeria Center for Disease control, and can equally be found on the following link: https://ourworldindata.org/coronavirus/country/nigeria?country=~NGA. The data set are in two phases, first data set covers the period from 29th March 2020 to 29th June 2020 which basically represent the period of total and partially eased lockdown in Nigeria. The second phase covers from the period of 30th June 2020 to 30th August 2020, and this represents the period when the lockdown has been reasonably eased.

2.3 Parameter estimation and forecast

Software by [26] was used for the analysis and function in "tseries" package by [27] was used to conduct the KPSS test for stationarity, while functions in "arfima" package by [28] were used to conduct the ARFIMA analysis and ARFIMA of order (1, 0, 1) and (1, 0, 0) was conducted respectively for the two phases.

3 Results and Discussion

Table 1 shows the descriptive statistics of the two phases considered in this study, with phase one and phase two being 93 and 63 days respectively. The mean new cases when there was strict lockdown was computed as 263.2903 while mean new cases when the lockdown was eased was computed as 466.5324, Table 2 represents KPSS Test for Level Stationarity. The p-value <0.05 in both phases revealed that the data is not stationary.

Table 1. Descriptive statistics of new cases of COVID-19

	Phase 1 newcases	Phase 2 newcases
Mean	263.2903	466.5324
Standard Error	22.2683	17.77058273
Median	238	462.5
Mode	16	575
S.D	214.7482	139.9257084
Sample Variance	46116.8169	19579.20386
Kurtosis	-0.526942102	-0.511721814
Skewness	0.654959472	-0.30371194
Range	779	652
Minimum	0	138
Maximum	779	790
Sum	24486	28925
Count	93	62

Table 2. KPSS Test for level stationarity

	Phase 1	Phase 2
KPSS Level	2.224	1.059
Truncation lag parameter	3.000	3.000
p-value	0.010	0.010

Null hypothesis of $I(0)$, that is stationary.

3.1 Phase one results

Table 3. Model paremeter estimation

	ARFIMA (1, 0, 1)		ARFIMA (1, 0, 0)	
	Coefficient	S.E	Coefficient	S.E
Phi	-0.46418	0.39652	-0.0530	0.10886
Thetha	-0.37483	0.40662	-	-
Sigma^2	10073.1		10038	-

From Table 3, Phi is the autoregressive parameters in vector form, Theta is the moving average parameters in vector form, and sigma^2 is the desired variance for the innovations of the series.

Table 4. Model selection criteria

Model selection	ARFIMA (1, 0, 1)	ARFIMA (1, 0, 0)	ARIMA (1,0,0)	ARIMA (1,0,1)
AIC	868.1275	866.7616	1167.476	1111.821
BIC	880.7905	876.892	1177.607	1127.016

From Table 4, ARFIMA (1, 0, 0) shows to be better fit for the data since it gives a lower AIC and BIC value than ARFIMA (1, 0, 1) and the forecast value being compared with ARFIMA (1,0,1), ARIMA (1,0,0) and ARIMA (1,0,1) respectively The forecast value is given in Table 5.

The results show that the rate of spread of COVID-19 would have significantly been fewer if the lockdown continues. The graph obtained using Microsoft excel shows that there is a significant difference between the actual and forecast meaning that new cases of COVID-19 would have been less than report if lockdown was not eased. Using statistical tests such as t-test to compare the mean, F-test to Compare Standard Deviations, Mann-Whitney (Wilcoxon), W-test to compare medians, and Kolmogorov-Smirnov test to compare the

distributions of the two samples. For t-test we have, t = 2.55413   p-value = 0.0199268, for Standard deviations comparison; lockdown eased: [66.2276, 175.777], Standard deviation of lockdown forecast: [19.8931, 52.7991], and ratio of variances: [2.75295, 44.6216]. F-test to compare standard deviations as follows; F = 11.0834, P-value = 0.0013995. Mann-Whitney (Wilcoxon) W-test to compare medians; W = 20.0 $P - value = 0.0257479$. Two-sided large sample K-S statistic = 1.56525, Approximate P-value = 0.0148932.

Table 5. Forecast value and actual value of new cases of COVID-19

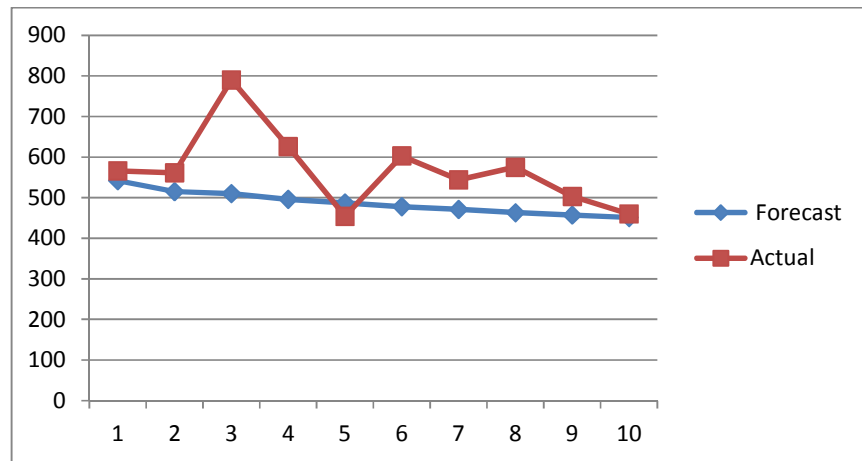| Date	Forecast	Actual eased lockdown
30/06/2020	542	566
01/07/2020	515	561
02/07/2020	510	790
03/07/2020	496	626
04/07/2020	487	454
05/07/2020	478	603
06/07/2020	471	544
07/07/2020	463	575
08/07/2020	457	503
09/07/2020	451	460

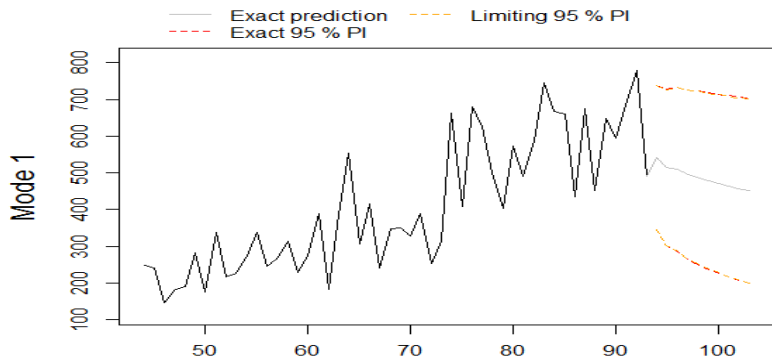**Fig. 1. Actual and forecast value supposing lockdown was not eased**

Fig. 2. Time series and prediction for AFIRMA (1, 0, 1)

Fig. 2 and Fig. 3 represents the time series and the prediction plot for ARFIMA (1, 0, 1) and ARFIMA (1, 0, 0) respectively. While Fig. 4 represents the difference in the fits. In Figs. 3 and 4, the thick broken red lines show the exact prediction at 95% Prediction interval, the lighter broken red lines show the limiting prediction at 95% Prediction interval, while the grey line shows the exact prediction.
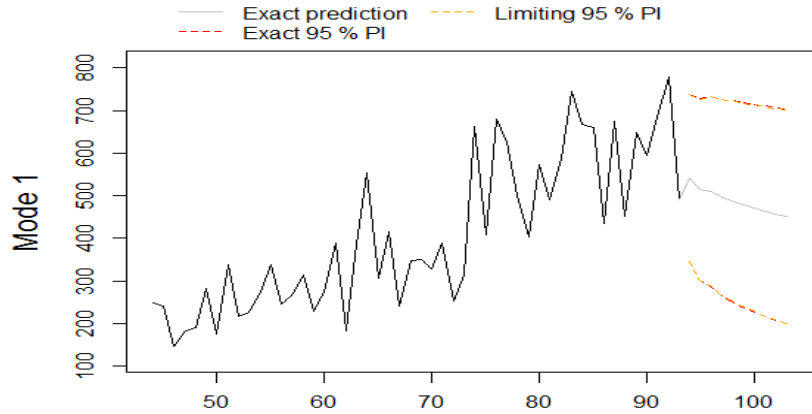
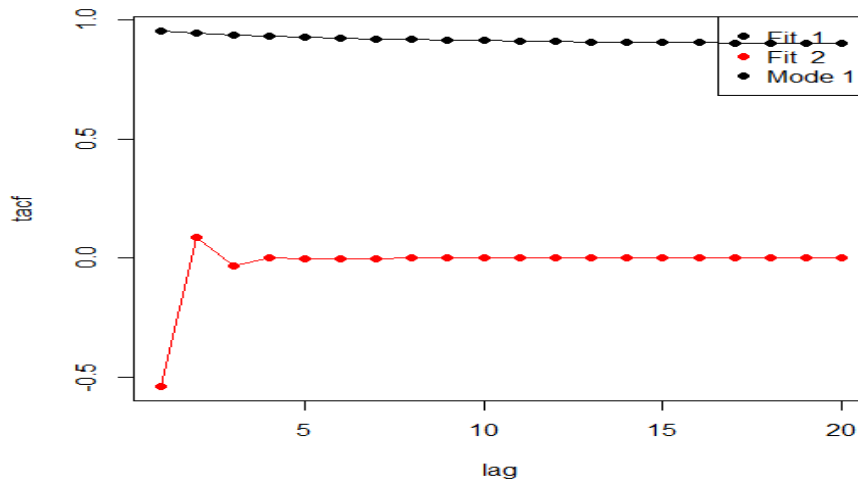

Fig. 3. Time series and prediction for AFIRMA (1, 0, 0)

Fig. 4. Comparing tacfs of fit 1 (1, 0, 1) and fit 2 (1, 0, 0).

Tacf plot is the theoretical autocorrelation functions (tacfs) of different models on the same data. Fig. 4 shows that ARFIMA (1, 0, 1) exhibits higher theoretical autocorrelation function (tacf) than ARFIMA (1, 0, 0).

3.2 Phase two results

Fig. 5 and Fig. 6 represents the time series and the prediction plot for ARFIMA (1, 0, 0) and ARFIMA (1, 0, 1) respectively. While Fig. 7 represents the difference in the tacfs of the fits. In Figs. 5 and 6, the thick broken red lines show the exact prediction at 95% Prediction interval, the lighter broken red lines show the limiting prediction at 95% Prediction interval, while the grey line shows the exact prediction.
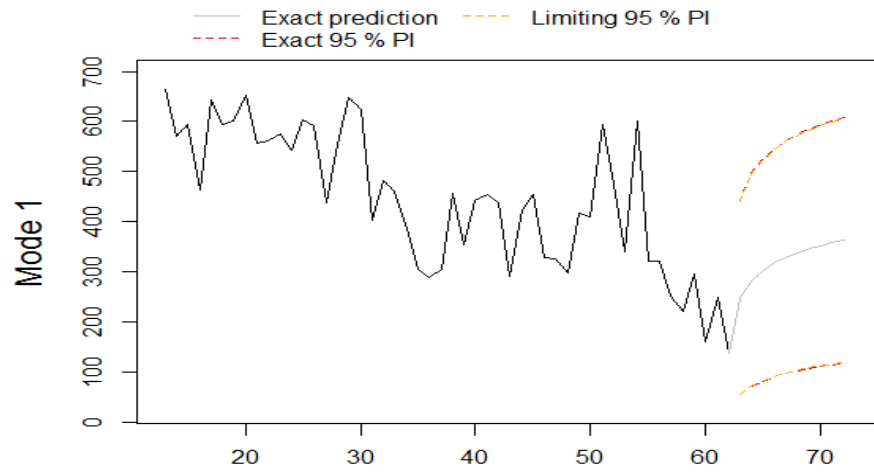
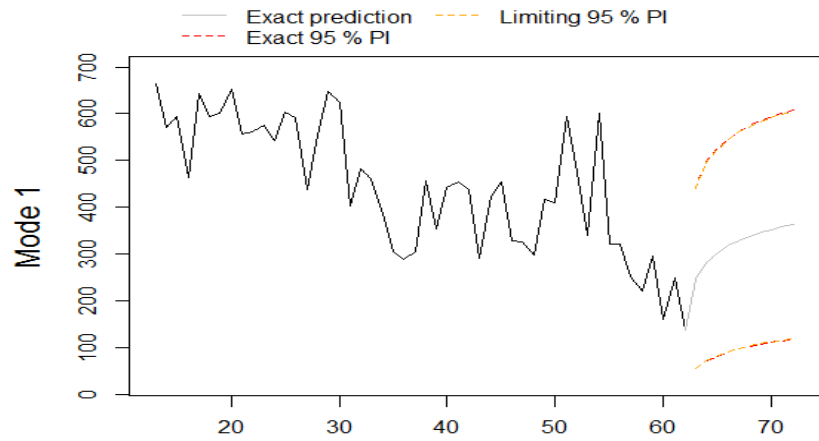**Fig. 5. Time series and prediction for AFIRMA (1, 0, 0)**

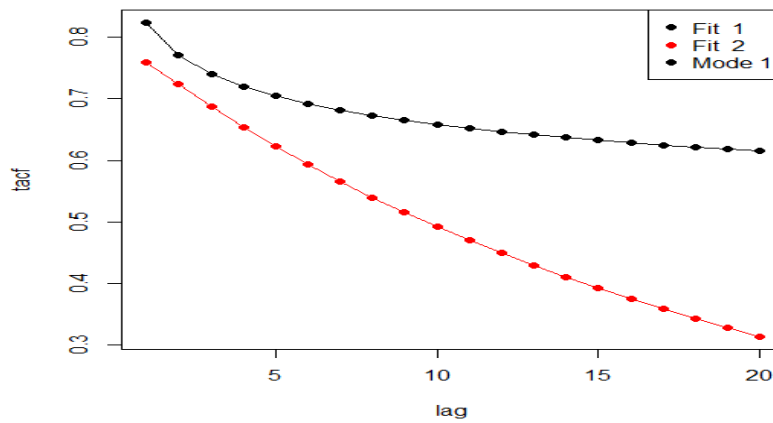**Fig. 6. Time series and prediction for AFIRMA (1, 0, 1)**

Fig. 7. Comparing the tacfs of fit 1 (1, 0, 0) and fit 2 (1, 0, 1)

Tacf plot is the theoretical autocorrelation functions (tacfs) of different models on the same data. Fig. 7 shows that ARFIMA (1, 0, 1) exhibits higher theoretical autocorrelation function (tacf) than ARFIMA (1, 0, 0).

Table 6. Model parameter estimation

	ARFIMA (1, 0, 1)		ARFIMA (1, 0, 0)	
	Coefficient	**S.E**	**Coefficient**	**S.E**
Phi	0.96961	0.05337	0.006723	0.15524
Thetha	0.42858	0.37972	-	0.06452
Sigma^2	9302.96		9653.47	-

From Table 6, Phi is the autoregressive parameters in vector form, Theta is the moving average parameters in vector form, and sigma^2 is the desired variance for the innovations of the series.

Table 7. Model selection criteria

Model selection	ARFIMA (1, 0, 1)	ARFIMA (1, 0, 0)	ARIMA (1,0,0)	ARIMA (1,0,1)
AIC	574.7652	577.064	763.333	752.9107
BIC	585.4009	585.573	771.8415	765.6735

AFIMA (1, 0, 1) proves to be better with the data in phase 2

From Table 7, ARFIMA (1, 0, 1) shows to be better fit for the data since it gives a lower AIC and BIC value than ARFIMA (1, 0, 1) and the forecast value being compared with ARFIMA (1,0,0), ARIMA (1,0,0) and ARIMA (1,0,1) respectively The forecast value is given in Table.

Table 8. Forecast of new cases of COVID-19 in Nigeria

Date	Forecast
31/08/2020	202
01/09/2020	205
02/09/2020	203
03/09/2020	201
04/09/2020	196
05/09/2020	191
06/09/2020	185
07/09/2020	180
08/09/2020	174
09/09/2020	167

Table 8 shows the forecast for the 10 days ahead for cases of COVID-19 in Nigeria starting from 31/08/2020 to 09/09/2020.

4 Conclusion

This study has been able to identify the advancement of mathematical sciences and its application in recent time. This study modelled and forecast new cases of COVID-19 using ARFIMA (p, d, q). The study revealed that if lockdown had continued in Nigeria beyond the day it was eased; fewer new positive cases of COVID-19 would have been recorded. Forecast for 10 days ahead from 31/08/2020 to 09/09/2020 was carried out using the ARFIMA $(1, 0, 0)$ and ARFIMA (1, 0, 1) model. The results reveal that ARFIMA (1, 0, 0) was a better fit for the data set used in phase one while ARFIMA (1, 0, 1) was a better fit for the data set used in phase two. The proposed ARFIMA model was compared with ARIMA (1, 0, 0), and ARIMA (1, 0,

1) and found to outperform the classical ARIMA models. The investigation on variation in performance of ARFIMA (p, d, q) resulting from different data set is suggested further studies.

Competing Interests

Authors have declared that no competing interests exist.

References

[1]   Usman BA, Kanumoori DSM, Marquez RAM. The impact of COVID-19 in Italy - A mathematical study; 2020.
        Available:https://www.researchgate.net/publication/342600721

[2]   Alemneh HT, Tilahun GT. Mathematical modeling and optimal control analysis of COVID-19 in Ethiopia. Preprint; 2020.
        Available:https://doi.org/10.1101/2020.07.23.20160473

[3]   Majia C. Modes of transmission of COVID-19 outbreak- a mathematical study; 2020.
        Available:https://doi.org/10.1101/2020.05.16.20104315

[4]   Shah K, Abdeljawad T, Mahariq I, Jarad F. Qualitative analysis of a mathematical model in the time of COVID-19. Hindawi BioMed Research International. 2020;Article ID 5098598:(1-11).
        Available:https://doi.org/10.1155/2020/509859

[5]   Kyrychko YN, Blyuss KB, Brovchenko I. Mathematical modelling of dynamics and containment of COVID-19 in Ukraine; 2020. medRxiv preprint.
        Available:https://doi.org/10.1101/2020.07.24.20161497

[6]   Madubueze CE, Akabuike, NM, Dachollom S. The role of mathematical model in curbing COVID-19 in Nigeria. medRxiv; 2020 preprint.
        Available:https://doi.org/10.1101/2020.07.22.20159210

[7]   Kumar S, Sharma S, Kumari N. Future of COVID-19 in Italy: A mathematical perspective, arXiv:2004.08588v1; 2020,

[8]   Panwara H, Guptaa PK, Siddiqui BMK, Morales-Menendez R, Singha V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos, Solitons and Fractals. 2020;138:109944.

[9]   Gozes O, Frid-Adar M, Sagie N, Zhang H, Ji W, Greenspan H. Coronavirus Detection and Analysis on Chest CT with Deep Learning; 2020 arXiv:submit/3118613 [cs.CV] 6 Apr 2020

[10]   Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine. 2020;121.
        Available:https://doi.org/10.1016/j.compbiomed.2020.103792.

[11]   Arumugam R, Rajathi M. A markov model for prediction of corona Virus COVID-19 in India- A Statistical Study. Journal of Xidian University; 2020;14(2):1422-1426.
        Available:https://doi.org/10.37896/jxu14.4/164

[12]   Alazab M, Awajan A, Mesleh A, Abraham A, Jatana V, Alhyari S. COVID-19 prediction and detection using deep learning. International Journal of Computer, Information Systems and Industrial Management Applications. 2020;12:168-181.
Available:www.mirlabs.net/ijcisim/index.html

[13]   Tamhane R, Mulge S. Prediction of COVID-19 outbreak using machine learning. International Research Journal of Engineering and Technology (IRJET). 2020:7(5).

[14]   Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. COVID-19 Outbreak prediction with machine learning, medrxiv; 2020.
Available:https://doi.org/10.1101/2020.04.17.20070094,
Available:https://www.medrxiv.org/content/10.1101/2020.04.17.20070094v1

[15]   Fayyoumi S, Idwan S, AboShindi H. Machine learning and statistical modelling for prediction of Novel COVID-19 Patients Case Study: Jordan. International Journal of Advanced Computer Science and Applications. 2020;11(5):2020.

[16]   Adesina OS, Agunbiade DA. Time series forecast models for foreign exchange rate in a developing economy. Assumption University-eJournal of Interdisciplinary Research (AU-eJIR). 2017;2(2):11-21.

[17]   Golinski P, Spencer, Modelling the COVID-19 Epidemic Using Time Series Econometrics; 2020.
Available:https://doi.org/10.1101/2020.06.01.20118612

[18]   Shrestha PM, Tha R, Neupane D, Adhikari K, Bhuju DR. Tracking and time series scenario of coronavirus: Nepal case. Applied Science and Technology Annals. 2020;1(1):42–47.

[19]   Nesteruk I. Statistics-based predictions of coronavirus epidemic spreading in mainland China. Innov Biosyst Bioeng. 2020;4(1):13–18.
DOI: 10.20535/ibb.2020.4.1.195074

[20]   Firmino PRA, de Sales JP, Gonçalves Júnior J, da Silva TA. A non-central beta model to forecast and evaluate pandemics time series, Chaos, Solutions and Fractals; 2020.
Available:https://doi.org/10.1016/j.chaos.2020.1102110960-0779/

[21]   Granger GWJ, Joyeux R. An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis, Wiley Blackwell. 1980;1(1):15-29.

[22]   Hosking JRM. Fractional differencing. Biometrika. 1981;68:165–176.

[23]   Balah B, Djeddou M. Forecasting COVID-19 new cases in Algeria using Autoregressive fractionally integrated moving average Models (ARFIMA); 2020.
Available:https://doi.org/10.1101/2020.05.03.20089615

[24]   Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics. 1992;54:159–178.

[25]   Baum CF. EC 823: Applied econometrics Boston College, Spring 2013 Christopher F Baum (Bc / Diw) Arima And Arfima Models.

[26]   Sowell F. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. Journal of Econometrics. 1992;53:165–188.

[27]   Core Team R. R A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2020.

[28]   Trapletti A, Hornik K. T series: Time series analysis and computational finance; 2019. R package version 0.10-47.

_____