FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 6 No. 4, August, 2022, pp 1 - 5
DOI: https://doi.org/10.33003/fjs-2022-0604-1008

STATISTICAL MODELLING OF COVID-19 CASES IN NIGERIA WITH A NEGATIVE BINOMIAL AUTOREGRESSIVE MODEL

Atigbi, A. O., Oguntade, E. S., *Oladimeji, D. M.

Department of Statistics, University of Abuja, Nigeria

*Corresponding authors' email: oladimejidaamilare@yahoo.com

ABSTRACT

Coronavirus disease (COVID-19) is a deadly global pandemic caused by a virus of the family *coronaviridae*. It is an infectious disease which affects respiratory systems and causes people to experience mild to moderate symptoms and sometimes severe cases of the disease which usually resulted into death especially among those patients with other comorbidity conditions and elderly with immunosenescence effects. Nigeria registered its index case of COVID-19 on 27th February 2020. Subsequently, the number of reported cases were on increasing trend. Numerous studies on modelling the sporadic increase cases or spread of SARS-COV-2 using different methodologies were documented in literature. However, issues relating to over-dispersed problem and the presence of autocorrelation were not well handled in such methods. In this present study, the modelling of the spread of SARS-CoV-2 in Nigeria was done using a Negative Binomial Autoregressive model. Study data were collected on a daily basis from the update released by the Nigeria Centre for Disease Control from 1st April, 2020 to 29th May, 2021. The results showed that the number of confirmed SARS-CoV-2 cases increased comparatively between April-2020 to June-2020. However, the number of reported cases dropped steadily between July-2020 to Nov-2020. The data were over-dispersed and the presence of autocorrelation was observed. The results revealed that among the four NBAR estimated candidate models, NBAR (1) returned the lowest Akaike Information Criterion. Thus, NBAR (1) is the most parsimonious NBAR model for the data. Therefore, NBAR (1) can be used in predicting daily cases of SARS-CoV-2 in Nigeria.

Keywords: Autoregressive model, Negative Binomial, Nigeria, Overdispersion, SARS-CoV-2

INTRODUCTION

The Corona virus disease (COVID-19) is a global disease that has ravaged the world since its interception in 2019 at Wuhan China. This pandemic which started as local outbreak in year 2019 later graduated to regional epidemics and global pandemic in year 2020. COVID-19 is a severe acute respiratory syndrome corona virus 2 (SARS-COV-2). This disease caught the global community unprepared. Thereby, it has affected the global and state's economics, lifestyle, tourism, education and other sectors negatively. SARS-COV-2 is a highly transmittable and pathogenic viral infection caused by SARS-CoV-2 (Muhammad *et al*., 2020). A corona virus belongs to the genus 'corona virus' of the '*corona viridae'* family.

The disease transmitted rapidly across states, regions and the globe with a considerable impact on global morbidity, mortality and healthcare utilization. The number of cases continues to rise throughout the globe (especially with the present omicron variant) contributing a serious menace to the public health. World Health Organization (WHO) reported more than 315.3 million confirmed cases of SARS-COV-2, including 5,510,174 deaths in more than 213 countries, areas and territories (WHO, 2022). The first case of Africa's COVID–19 was reported in Egypt on 14 February 2020 while Nigeria registered its index case on 27 February 2020 (Adeniran, Oguntade, Anjorin, & Ajagbonna, 2021). After the index case was discovered in Ogun state Nigeria, the number of new cases of COVID–19 reported by the National Centre for Disease Control (NCDC) are steadily on increase per day. There were about 7,731,116 confirmed cases in Africa as at 13th January, 2022 with Nigeria accounting for about 249,586 of these cases. The total number of deaths reported in Nigeria as at 13th January, 2022 was 3,092 deaths (WHO, 2022). The pandemic caused the largest global recession in the history of the world at large, with more than a third of the global population at the time on lockdown. Nigeria was not left out as the country remains one the first five most hit countries in Africa. Social distancing, no large gathering and use of nose mask among others were mandated and no business could operate. The economy of Nigerian was affected negatively as no activities could hold, all schools were ordered to shut down and many other predicting factors to country's gross domestic product (GDP) were on hold. Hence, needs to model the SARS-COV-2 in order to foresee and prevent any spike that might occur in the nearest future.

Numerous studies on modelling the sporadic increase or spread of SARS-COV-2 using different methodologies were documented in literature (Katoch & Sidhu, 2021; Sharma & Nigam, 2020). While a considerable number of studies explored and utilized autoregressive integrated moving average (ARIMA) for forecasting (Katoch & Sidhu 2021; Tandon, 2020), others considered the use of regression model for relationships among influencing variables and predictions (Chu, 2021; Sharma & Nigam, 2020). Sharma and Nigam (2020) appraised covid-19 cases with a combination statistical procedures to model COVID 19 cases. The authors adopted a linear regression model, ARIMA and Holt–Winters models to provide the forecast of COVID-19 cases in India. The study found ARIMA (5, 2, 5) model as the best-fitting model for COVID-19 data in the study domain. Katoch and Sidhu (2021) explored Box-Jenkins procedures to analyse the temporal dynamics of the COVID-19 outbreak. Katoch and Sidhu suggested an unending increase in the number of confirmed COVID-19 cases in India in the near future.

In this present study, the modelling of the spread of SARS-CoV-2 in Nigeria was done using a Negative Binomial Autoregressive model. A negative Binomial distribution is a discrete probability distribution that models the number of successes in a sequence of independent and identically, distributed Bernoulli trials. This distribution is commonly used to describe the distribution of count of data. It's flexibility and presence of an additional parameter in this discrete distribution make it a robust distribution in handling an over-dispersion problem commonly encounter in count data (Oguntade, Shohaimi, Nallapan, Lamidi-sarumoh, & Salari, 2020). A statistical scenario in which an observed distributional variance is higher than the variance of a theoretical model, then over-spreading is said to occur. This phenomenon in real life data is not exempted for the reported COVID cases. Recent literature documented the use of a negative binomial autoregressive model for time series count observations such as stock returns and in handling the number of earthquakes (Liu, Li & Zhu, 2019).

It is on this backdrop that this study focuses on modeling the spread of SARS-CoV-2 in Nigeria using a Negative Binomial Autoregressive model.

STUDY DESIGN

This study was a retrospective study centered on modeling the spread of SARS-CoV-2 in Nigeria using a Negative Binomial Autoregressive model. Data on SARS-COV-2 cases were gathered on a daily basis from the update released by the NCDC.

Data Collection

The data was collected on a daily basis from the update released by the NCDC between 1st April, 2020 to 29th May, 2021 and the total after every month was calculated. The data was further defined to be a monthly time series data based on the stated adjustments. (https://covid19.ncdc.gov.ng)

A Negative Binomial Distribution

The negative binomial distribution, like the normal distribution, arises from a mathematical formula in equation 1 below. It is commonly used to describe the distribution of count data, such as the numbers of parasites in blood specimens, number of daily occurrences of a disease like cases of COVID-19. Also, like the normal distribution, it can be completely defined by just two parameters, its mean (m) and shape parameter (k). However, unlike the normal distribution, the negative binomial distribution does not naturally result from the use of large samples, nor does it arise from a single casual model.

The probability mass function of the negative binomial distribution is

$$f(\boldsymbol{k},\boldsymbol{r},\boldsymbol{p}) = \boldsymbol{Pr}(\boldsymbol{X}=\boldsymbol{k}) = \binom{\boldsymbol{k+r-1}}{\boldsymbol{r-1}}(\boldsymbol{1-p})^{\boldsymbol{k}}\boldsymbol{P}^{\boldsymbol{r}} \quad (1)$$

where r is the number of success, k is the number of failures, and P is the probability of success.

A Negative Binomial Autoregressive Model (NBAR)

Let P(λ) represents the Poisson distribution with parameter λ, γ(δ, β, c) the gamma distribution with degree of freedom δ, non-centrality parameter β and scale parameter c, and NB(δ, β) the negative binomial (NB) distribution with positive parameters δ, β. This parameterization relies on the interpretation of the NB distribution as a Poisson distribution with gamma stochastic intensity, with δ the degree of freedom of the gamma distribution, and β the scale of this intensity. The process (X_t) is defined jointly with a real positive intensity process (Y_t) in the following way

i) The conditional distribution of X_{t+1} given Y_{t+1}, X_t is poisson $P(\beta X_{t+1})$.

ii) the conditional distribution of Y_{t+1}, X_t, Y_t is centered gamma with shape parameter $\delta + \mathrm{X}_t$, and scale parameter c;γ($\delta + \mathrm{X}_t$,0,c),

where $\mathrm{X}_t = (\mathrm{X}_t, X_{t-1}, \ldots, \mathrm{X}_1, \ldots)$, $\mathrm{Y}_t = (\mathrm{Y}_t, Y_{t-1}, \ldots, \mathrm{Y}_1)$, and β, δ, c are positive scalars. In other words (Y_{t+1}) *is* the stochastic intensity of the count X_{t+1} at time t +1,which depends on past count X_t, thus we have a causal chain.

$\ldots \mathrm{X}_t \longrightarrow \mathrm{Y}_{t+1} \longrightarrow \mathrm{X}_{t+1} \longrightarrow \mathrm{Y}_{t+2} \ldots$

A Conditional Over-dispersion

Because of its stochastic intensity, the NBAR process (X_t) features conditional over-dispersion.

The conditional variance is

$$V[X_{t+1}|X_t] = \mathrm{E}\big[V[X_{t+1}|X_{t,}Y_{t+1}]|X_t\big] + V\big[E[X_{t+1}|X_{t,}Y_{t+1}]|X_t\big] = \beta c(\delta + X_t) + \beta^2 c^2(\delta + X_t) \quad (2)$$

Thus, the coefficient of (conditional) over-dispersion is:

$$\frac{V[X_{t+1}|X_t]}{\mathrm{E}[X_{t+1}|X_t]} = 1 + \beta c > 1.$$

It is constant independent of the conditioning variable X_t. The larger the serial correlation $\rho = \beta c$ the more important the conditional over-dispersion.

As a comparison, it is easily checked that a Poisson process is conditionally under-dispersed, since for such a process:

$$V[X_{t+1}|X_t] = p(1-p)X_t + \lambda \leq \mathrm{E}\big[X_{t+1}|X_{t,}\big] = pX_t + \lambda,$$

Whereas a Poisson autoregressive features neither conditional under-dispersion, nor conditional over-dispersion, since:

$$\mathrm{E}[X_{t+1}|X_t] = \lambda_0 X_t + \lambda_1 = \mathrm{V}\big[X_{t+1}|X_{t,}\big]$$

Stationarity condition

The stationarity of univariate count process is essential before a valid model can be ascertained. Firstly, a necessary condition for mean-variance stationarity is that the means $\mu_1 = \mathrm{E}[X_{1,t}]$ and $\mu_2 = \mathrm{E}[X_{2,t}]$ are positive and finite. In this case they satisfy the following system, obtained by taking expectations in equation

$(1 - \beta_1 c_3 \alpha_1 - k_1 c_1)\mu_1 = \beta_1 c_3 \delta_3 + k_1 c_1 \delta_1 + \beta_1 \alpha_2 c_3 \mu_2,$

$(1 - \beta_2 c_3 \alpha_2 - k_2 c_2)\mu_2 = \beta_2 c_3 \delta_3 + k_2 c_2 \delta_2 + \alpha_1 \beta_2 c_3 \mu_1$

Thus a necessary condition for mean-variance stationarity is:

$1 - k_1 c_1 - \alpha_1 \beta_1 c_3 > 0$, and $1 - k_2 c_2 - \alpha_2 \beta_2 c_3 > 0$.

The process (X_t) is strictly stationary if and only if all the eigenvalues of matrix:

$$M := \begin{pmatrix} \frac{\partial \alpha_1(0,0)}{\partial u_1} & \frac{\partial \alpha_1(0,0)}{\partial u_2} \\ \frac{\partial \alpha_2(0,0)}{\partial u_1} & \frac{\partial \alpha_2(0,0)}{\partial u_2} \end{pmatrix}$$ are in modulus smaller than 1.

RESULTS

Table 1 presents the descriptive statistics of Nigeria confirm SARS-CoV-2 cases. As observed from Table 1, the mean, median, maximum and minimum of the SARS-CoV-2 cases were 444, 304, 2464 and 5 respectively for the time period examined. The variance value for the data is 178291. As observed the SARS-CoV-2 variance is larger than its means thus, there could be over-dispersion in the series. Figure 1 presents the time series plot of the data.

Table 1. Summary Statistics of Confirmed SARS-COV-2 Cases in Nigeria

	Confirmed SARS-CoV-2 cases
Mean	444
Median	304
Maximum	2464
Minimum	5
Variance	178291
Observations	424

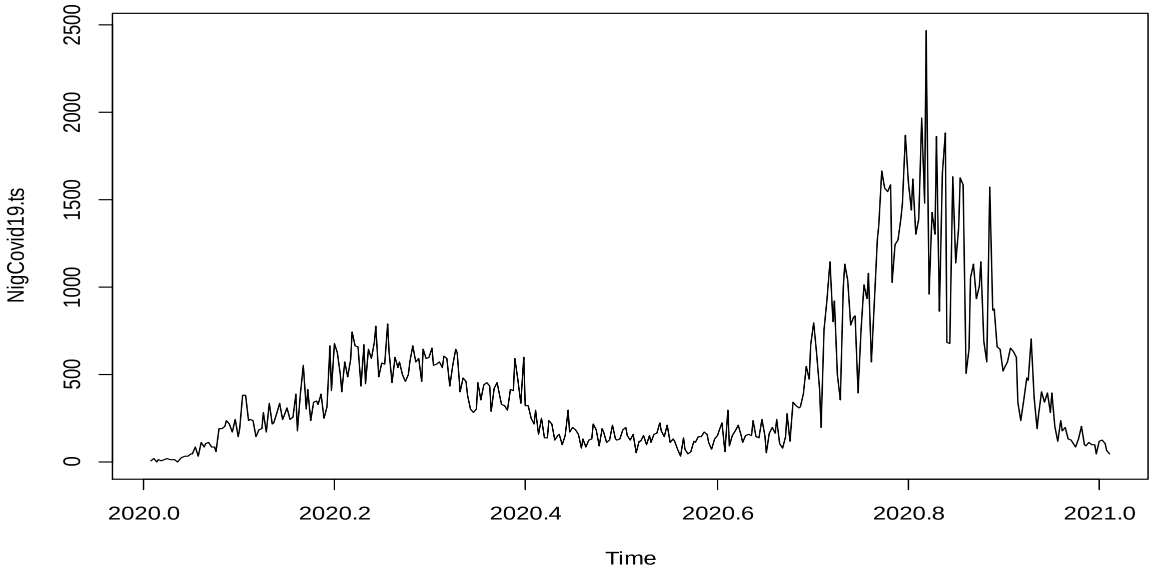

Figure 1. Time Series Plot of the Confirmed SARS-CoV-2 Cases in Nigeria

As observed from Figure 1, the confirmed SARS-CoV-2 cases increased comparatively between April-2020 to June-2020. The cases dropped steadily between July-2020 to Nov-2020. This could be attributed to the effectiveness of the Federal Government SARS-CoV-2 Guidelines such as Wearing of Masks, Social Distance, regular Hand-washing, etc. Remarkably, the cases increased exponentially between December 2020 to January 2021. Afterwards, the confirmed SARS-CoV-2 cases dropped relatively.

Result of Over-Dispersion Test

The Over-Dispersion test results of the confirmed SARS-CoV-2 cases revealed that there is presence of Over-dispersion in the confirmed SARS-CoV-2. The test returns a p-value of <0.0001 that is less than 0.05 level of significance.

Table 2. Box- Ljung Test for SARS-COV-2 in Nigeria

Model	X-square	DF	P-value
NigSARS-CoV-2	1538.6	6	<0.0001

Table 2 displays the autocorrelation test results of the confirmed SARS-CoV-2 data. Among other results, the test returns a p-value of <0.0001 that is less than 0.05 level of significance. Thus. there is presence of autocorrelation in the residuals of the confirmed SARS-CoV-2 data. Consequently, the results of both tests; Over-Dispersion and Autocorrelation suggest suitability of Negative Binomial type-models such as Negative Binomial Autoregressive Model (NBAR).

Table 3. Unit Root Test for SARS-COV-2 in Nigeria

Difference	Dickey-fuller	P-value
I(0)	-0.5918	0.9774
I(1)	-7.3468	0.01

Table 3 reports the Augmented Dickey-Fuller (ADF) test results of the SARS-CoV-2 series. The ADF-test returns a p-value (0.9774) that is greater than 0.05 level of significance. The results deduced that the null hypothesis (i.e. non-stationary) cannot be rejected at level which indicates the presence of unit root. However, the ADF test results returned a p-value (0.01) that is lesser than 0.05 significance level deducing stationarity at 1st difference.

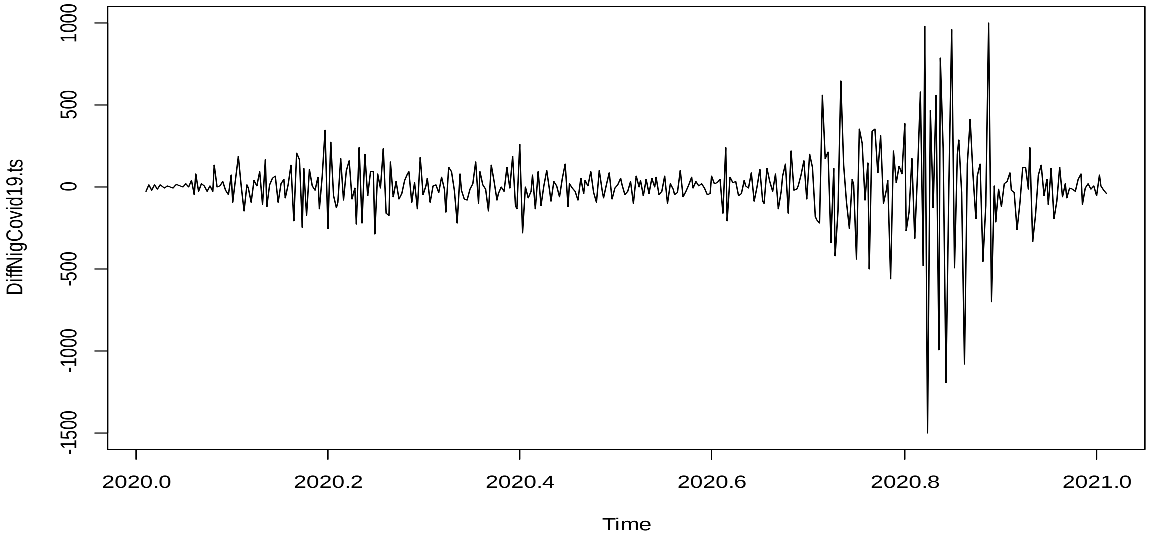

Figure 3. First Difference Time Series Plot of The SARS-CoV-2 Cases in Nigeria

Table 4 presents the summary of the estimated NBAR models with their respective Akaike information criterion (AIC) values. Using this specification measure, four NBAR candidate models were estimated. NBAR (1) returned with the lowest AIC (5203) while NBAR (4) gave the highest value of AIC. Thus, NBAR (1) returned as the most parsimonious NBAR models for the confirmed SARS-CoV-2 series.

Table 4. NBAR Model Selection for SARS-COV-2 in Nigeria

NBAR	AIC
NBAR(1)	5203
NBAR(2)	5431
NBAR(3)	5232
NBAR(4)	5521

Following the aforementioned, Table 5 presents the summary statistics of the estimated NBAR (1) model and the coefficients for the AR (1) and Intercept using the generalized method of moment (GMM) approach. The AR (1) and Intercept (1.21 and 6.10 respectively) were found to be significant at 0.01 significant level. Table 5 also shows a theta estimate of 1.24. Thus, the NBAR (1) model for the confirmed SARS-CoV-2 cases is given in 3.

$$NigCovid19_t = 6.096 + 1.24NigCovid19_{t-1} \quad (3)$$

Table 5. NBAR (1) Model Estimation for SARS-COV-2 in Nigeria

COEFFICIENT	ESTIMATE	STD ERROR	Z VALUE	P-VALUE
INTERCEPT	6.096	0.1277	13.234	<0.0001
AR1	1.2123	0.1873	7.2434	<0.0001
THETA	1.2367			
AIC	5203			

Additionally, the NBAR (1) model was diagnosed for goodness of fit. The autocorrelation function (ACF) plots of the model residuals showed that none of the ACF points is greater than 0.5 significant level. Similarly, Ljung-Box test results confirmed the absence of autocorrelation in the series ($p > 0.05$).

DISCUSSION

This study aimed at modelling the country confirmed SARS-CoV-2 cases using a Negative Binomial Autoregressive (NBAR) approach. The study examined daily data span from 1st April 2020 to 29th May 2021. The number of confirmed SARS-CoV-2 cases increased comparatively between April-2020 to June-2020. However, number of cases dropped steadily between July-2020 to Nov-2020. Also, the preliminary analysis revealed that there was presence of over dispersion in the data (the variance is not equal to the mean). This is in agreement with a related work on malaria cases in Nigeria (Oguntade *et al*, 2020). The authors observed over-dispersion problem, thus adopted a negative Binomial generalized linear model instead of Poisson model which was not tenable in such situation.

The distribution of the confirmed SARS-CoV-2 cases revealed a non-stationarity of the series at level form as there were random spikes. This persistent rise in number of confirmed cases in COVID-19 observed in the study area within the study period may be due to lack of proper usage of protective measures and adherent to COVID 19 protocols

(Katoch & Sidhu, 2021). Similarly, the time series plots and augmented dickey-fuller tests of the original series indicated a non-stationarity of the series, hence series was differenced to attain stationarity. The choice of parsimony model agrees with a related study elsewhere in India (Sharma & Nigam, 2020). The authors selected the best model based on a similar criterion.

CONCLUSION

Empirical results revealed that among the four NBAR estimated candidate models, NBAR (1) returned as the lowest AIC. Thus, NBAR (1) is the most parsimonious NBAR model for the series. The diagnostic goodness of fit results for selected models revealed no any trace of serial correlation in the NBAR (1) residuals. Therefore, NBAR (1) can be used in predicting daily cases of SARS-CoV-2 in Nigeria.

REFERENCE

Adeniran, L. A., Oguntade, E. S., Anjorin, T. S., & Ajagbonna, P. (2021). Epidemiological status of coronavirus diseases and the remedy potentials of medicinal plants in Africa, *13*(March), 18–32. https://doi.org/10.5897/JPP2020.0581

Chu J (2021) A statistical analysis of the novel coronavirus (SARS-COV-2) in Italy and Spain PLoS One 16(3).

Katoch R and Arpit Sidhu A (2021). An Application of ARIMA Model to Forecast the Dynamics of COVID-19 Epidemic in India. *Global Business Review* 1-14.

Liu, M. Li Q. and Zhu, F. (2019): Threshold negative binomial autoregressive model. *J. of Theo. and Appl. Stat.*, 53 (1-25).

Muhammad A, suliman K, Abeer K, Nadia B, Rabeer S. (2020), SARS-COV-2 Infection: Emergence, Transmission, and Characteristics of Human Corona Viruse, vol.24 (91-98).

Oguntade, E. S., Shohaimi, S., Nallapan, M., Lamidi-sarumoh, A. A., & Salari, N. (2020). Statistical Modelling of the Effects of Weather Factors on Malaria Occurrence in Abuja , Nigeria. *Int. J. Environ. Res. Public Health*, *17*(3474;), 1–12.

Sharma and Nigam (2020), Modelling and Forecasting of COVID-19 Growth Curve in India. *Trans India Natl Acad Eng.; 5(4):697-71.*

Tandon. H, Prabhat. R , Tanmoy. C, Vandana. S, (2020), Coronavirus (SARS-COV-2): ARIMA based time-series analysis to forecast near future. *Quantitative Biology-Population and Evolution* Arxis.2004.07859(q.bio)

WHO, (2022): WHO Coronavirus (COVID-19) Dashboard with Vaccination Data, World Health Organization. https://covid19.who.int/