African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)
www.abjournals.org

VECTOR AUTOREGRESSIVE MODELS FOR MULTIVARIATE TIME SERIES ANALYSIS ON COVID-19 PANDEMIC IN NIGERIA[1]

Ajao I.O[1], Awogbemi C.A[2] and Ilugbusi A.O[1]

[1]Department of Mathematics and Statistics, The Federal Polytechnic, Ado-Ekiti, Nigeria
[2]Department of Statistics, National Mathematical Centre, Abuja, Nigeria

ABSTRACT: *In this paper, we have been able to use vector autoregressive (VAR) models for modeling and forecasting covid-19 variables with special focus on Nigeria cases from 1st march to 10th June 2020. At lag of order 2, the hypothesis of non-stationary is rejected at 5% level for all the multivariate variables using the augmented Dickey Fuller and Phillips-Perron unit root tests. The Granger causality test results indicate that there is a bivariate causal relationship among the variables by rejecting the null hypothesis of no Granger causality. The determinants of confirmed cases, new cases, and total deaths from covid-19 are generally significant at 5% level with p-value 0.0001 in each of the three derived models. The criteria AIC and log-likelihood implemented on the models confirmed that the VAR model of order 2 gives a better model for predictions and forecasts of covid-19 cases in Nigeria. This paper recommends a suitable model for handling multivariate time series data and suggests a reliable approach for forecasting future cases of covid-19 variables in the country and help health policy makers in finding solution to the unceasing upward trend in the cases of the pandemic.*

KEYWORDS: VAR Model, Covid-19 Variables, Stationarity, Forecasts, Granger Causality

INTRODUCTION

On 31 December 2019, the World Health Organization (WHO) was formally notified about a cluster of cases of pneumonia in Wuhan City, home to 11 million people and the cultural and economic hub of central China. By 5 January, 59 cases were known and none had been fatal (WHO, 2020). Ten days later, WHO was aware of 282 confirmed cases, of which four were in Japan, South Korea and Thailand (WHO, 2020). There had been six deaths in Wuhan, 51 people were severely ill and 12 were in a critical condition. The virus responsible was isolated on 7 January and its genome shared on 12 January (WHO, 2020). The cause of the severe acute respiratory syndrome that became known as COVID-19 was a novel coronavirus, SARS-CoV-2. The rest is history, albeit history that is constantly being rewritten: as of 12 May, 82,591 new cases of COVID-19 worldwide were being confirmed daily and the death rate was over 4200 per day (WHO, 2020).

The Federal Ministry of Health has confirmed a coronavirus disease (COVID-19) case in Lagos State, Nigeria. The case, which was confirmed on the 27th of February 2020, is the first case to be reported in Nigeria since the beginning of the outbreak in China in January 2020. (NCDC, 2020). The spread of novel Corona Virus Disease (COVID-19) in Nigeria continue to record significant increase as the latest statistics provided by the Nigeria Centre for Disease Control reveal (NCDC, 2020).

[1] Paper presented at the International E-Conference on COVID-19 Global Impacts, 20-21 July, 2020.

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)
www.abjournals.org

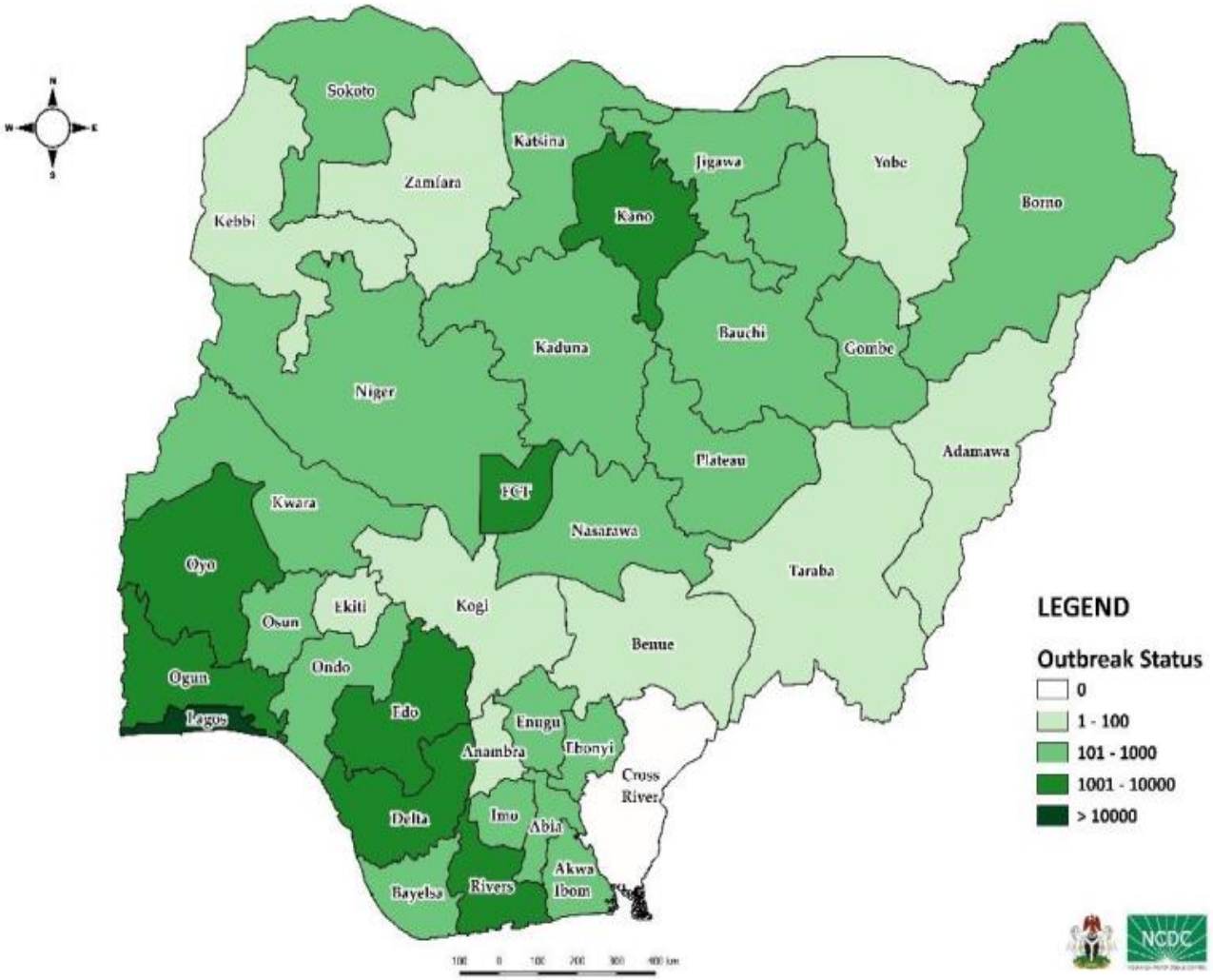

Fig.1: Map of Nigeria Showing 35 States and FCT Affected by COVID-19 as at June 10 2020

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)
www.abjournals.org

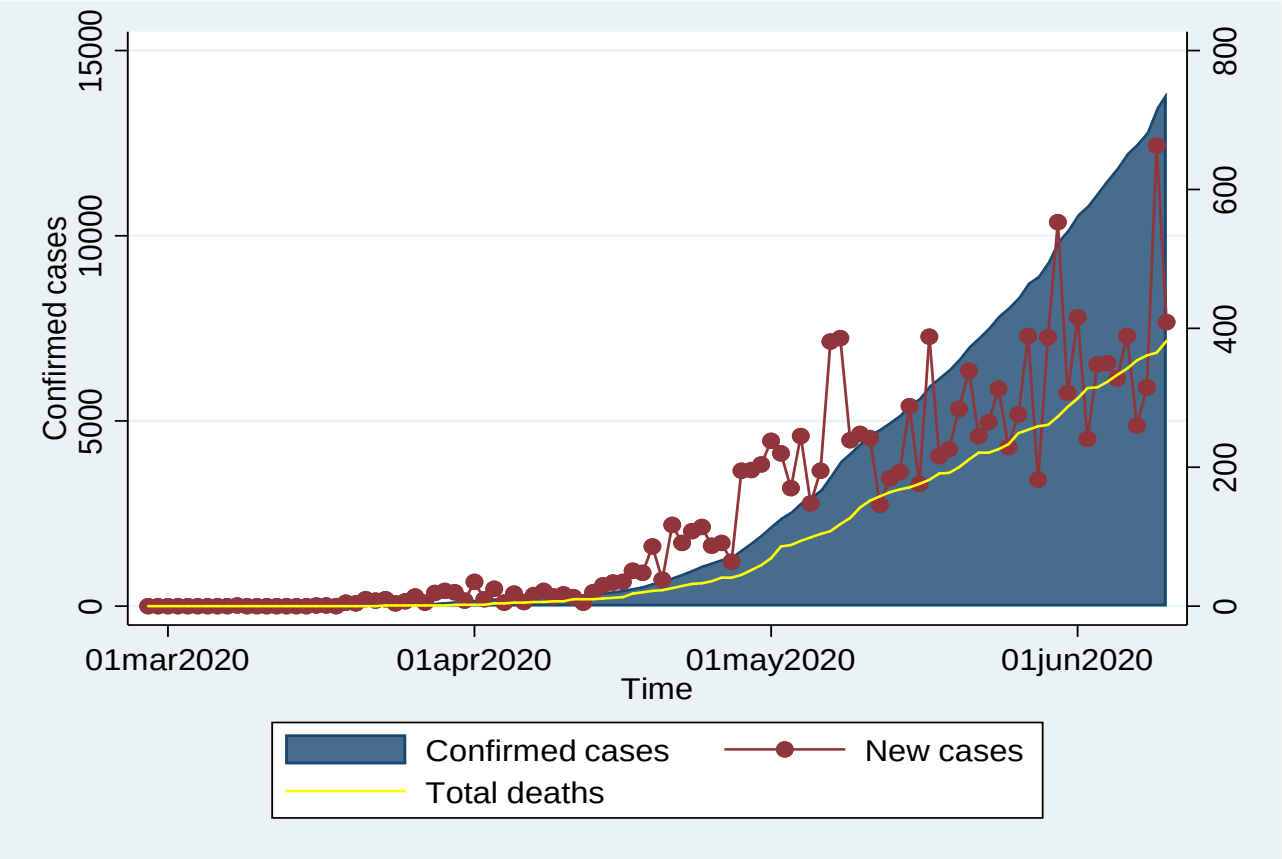

Fig.2: Chart Showing Confirmed, New, and Death Cases of Covid-19 in Nigeria from Feb. 28 to June 10

Vector Autoregression (VAR) Model

The vector autoregression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models. Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model. VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables. More rigorous treatments can be found in Hamilton (1994), Lutkepohl (2005), and Amisano and Giannini (1997). Stock and Watson (2001) provide an excellent nonmathematical treatment of vector autoregressions and their role in macroeconomics. Becketti (2013) provides an excellent introduction to VAR analysis with an emphasis on how it is done in practice.

When there are no constraints placed on the coefficients, the VAR(*p*) is a seemingly unrelated regression model with the same explanatory variables in each equation. As discussed in Lutkepohl (2005) and Greene (2008,), performing linear regression on each equation produces the maximum likelihood estimates of the coefficients. The definitive technical reference for VAR models is Lutkepohl (1991) and updated surveys of VAR techniques are given in works of Watson (1994); Lutkepohl (1999); and Waggoner and Zha (1999). Applications of VAR models to financial data are given in works of Hamilton (1994a; 1994b); Campbell, Lo, and MacKinlay (1997); Mills (1999); and Tsay (2001).

When building a VAR model, the following steps can be used. The Akaike Information Criterion (AIC) have been used to identify the order, then estimate the specified model by using the least squares method (if there are statistically insignificant parameters, the model should be re-estimated by removing these parameters), and finally use the *Qk* (m) statistic of the residuals to check the adequacy of a fitted model. The time series Y_t follows a VAR(*p*) model, if it satisfies

$$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \ldots + \emptyset_p Y_{t-p} + \alpha_{t,} \quad p > 0 \quad \ldots\ldots(1)$$

where, Y_t is a vector of the dependent variable $\emptyset_0$ is a *k*-dimensional vector; and α_t is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix Σ. Covariance matrix Σ must be positive definite; otherwise, the dimension of Y_t can be reduced. The error term, α_t is a multivariate normal and $\emptyset_j$ are $k \times k$ matrices. Using the back-shift operator *B*, the VAR(*p*) model can be written as:

$$\left(1 - \emptyset_1 \mathrm{B} - \ldots - \emptyset_p B^p\right) Y_t = \emptyset_0 + \alpha_t \quad \ldots\ldots(2)$$

where, *I* will be the $k \times k$ identity matrix. In a compact form, it is as follows:

$$\emptyset(\mathrm{B}) Y_t = \emptyset_0 + \alpha_t \quad \ldots\ldots(3)$$

where, $\emptyset(\mathrm{B}) = 1 - \emptyset_1 \mathrm{B} - \ldots - \emptyset_p B^p$ is a matrix polynomial, if Y_t is weakly stationary, then it reduces to:

$$\mu = E(Y_t) = \left(1 - \emptyset_1 B - \cdots - \emptyset_p\right)^{-1} \emptyset_0 = [\emptyset(1)]^{-1} \emptyset_0 \quad \ldots\ldots(4)$$

Provided that the inverse exists, since determinant of [Φ(1)] is different from zero.

then the VAR(*p*) model becomes:

$$\tilde{Y}_t = \emptyset_1 \tilde{Y}_{t-1} + \ldots + \emptyset_p \tilde{Y}_{t-p} + \alpha_{t,} \quad \ldots\ldots(5)$$

This results can be obatined as:

$$Cov(Y_t, \alpha_t) = \Sigma, the\ covariance\ matrix\ of\ a_i$$

$$Cov(Y_{t-1}, \alpha_t) = 0, \text{for } 1 > 0$$

$$\Gamma_l = \emptyset_1 \Gamma_{l-1} + \ldots + \emptyset_p \Gamma_{l-p} + \alpha_{l,} \quad \text{for } 1 > 0 \ldots\ldots(6)$$

The equation (6) is a multivariate version of Yule–Walker equation and it is called the moment equation of a VAR(*p*) model. The concept of partial autocorrelation function of a univariate

series can be generalized to specify the order *p* of a vector series. Consider the following consecutive VAR models: (Hossain, Kamruzzaman, and Ali, 2015)

$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \alpha_t$ .....(7)

$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \alpha_t$ .....(8)

$Y_t = \emptyset_0 + \emptyset_1 Y_{t-1} + \ldots + \emptyset_p Y_{t-p} + \alpha_t$ .....(9)

The ordinary least squares (OLS) method is used for estimating parameters of these models. This is called the multivariate linear regression estimation in multivariate statistical analysis (Tsay, 2001)

Data Analysis

The data used for this research was obtained from the website of the Nigeria Centre for Diseases Control (NCDC). All analyses were done using *R* version 4.0.0 and STATA version 15

Tests of Stationarity

Stationarity was achieved at d = 2 for the three variables. The Augmented Dickey-Fuller Test were carried out to test for the significance of stationarity of the data sets. (Dickey & Fuller, 1979), and Phillips-Perron test (PP) (Phillips & Perron, 1998).

Table 1: Augmented Dickey-Fuller and Phillips-Perron Tests

Variable	Lag order	Dickey-Fuller	P-value	Phillips-Perron	P-value
Confirmed cases	4	-6.7765	0.012	-127.89	0.01
New cases	4	-8.4926	0.001	-141.20	0.01
Total deaths	4	-8.2840	0.002	-110.63	0.01

The null hypothesis of non-stationarity is rejected in all the cases, this shows that the lagged series are stationary at 5% level of significance.

Granger Causality Test

This test is necessary in order to ascertain the cause of one or the other variable in the data set

Table 2: Granger test of causality

F-Test	df1	df2	p-value
3.5548	4	285	0.0075

Therefore, the null hypothesis stating that "lagged values of *confirmed cases* do not cause *new cases* and *total deaths*" can be rejected at 5% level of significance since the p-value 0.03268 is less than 0.05. This implies that *confirmed cases* Granger-causes *new cases* and *total deaths* of covid-19 in Nigeria.

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)

www.abjournals.org

Estimation of Parameters for the Models

Tables 3-5 presents the estimates of the parameters for the formulation of the models from the datasets. All are significance are measured at 5% level.

Table 3: Estimation Results for Equation: Confirmed Cases

	Estimate	Std. Error	t value	P	R-sq	Model-P
confirmedcase.L1	**2.6593**	**1.2953**	**2.053**	**0.0428**	**0.9998**	**2.2e-16**
newcase.L1	**-1.5041**	**1.2806**	**-1.175**	**0.2431**		
totaldeaths.L1	**-0.6610**	**2.6048**	**-0.254**	**0.8002**		
confirmedcase.L2	**-1.7110**	**1.3018**	**-1.314**	**0.1919**		
newcase.L2	**0.0900**	**0.1362**	**0.661**	**0.5102**		
totaldeaths.L2	**3.0310**	**2.5817**	**1.174**	**0.2433**		
const	**-25.6444**	**18.5806**	**-1.380**	**0.1708**		
trend	**1.4953**	**0.5786**	**2.584**	**0.0113**		

Table 4: Estimation results for equation: new cases

	Estimate	Std. Error	t value	P	R-sq	Model-P
confirmedcase.L1	**1.7069**	**1.3083**	**1.305**	**0.195**	**0.8350**	**2.2e-16**
newcase.L1	**-1.5641**	**1.2935**	**-1.209**	**0.230**		
totaldeaths.L1	**-0.8260**	**2.6310**	**-0.314**	**0.754**		
confirmedcase.L2	**-1.7639**	**1.3149**	**-1.341**	**0.183**		
newcase.L2	**0.1008**	**0.1375**	**0.733**	**0.466**		
totaldeaths.L2	**3.3924**	**2.6077**	**1.301**	**0.196**		
const	**-25.5337**	**18.7679**	**-1.360**	**0.177**		
trend	**1.4968**	**0.5844**	**2.561**	**0.012**		

Table 5: Estimation results for equation: total deaths

	Estimate	Std. Error	t value	P	R-sq	Model-P
confirmedcase.L1	**-0.0001**	**0.0505**	**-0.002**	**0.9981**	**0.9996**	**2.2e-16**
newcase.L1	**0.0303**	**0.0499**	**0.606**	**0.5460**		
totaldeaths.L1	**0.8996**	**0.1016**	**8.857**	**4.95e-14**		
confirmedcase.L2	**0.0029**	**0.0508**	**0.059**	**0.9533**		
newcase.L2	**0.0109**	**0.0053**	**2.058**	**0.0424**		
totaldeaths.L2	**-0.0216**	**0.1007**	**-0.215**	**0.8304**		
const	**-0.6740**	**0.7246**	**-0.930**	**0.3546**		
trend	**0.0292**	**0.0226**	**1.292**	**0.1994**		

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)

www.abjournals.org

The coefficients for a variable are listed in the estimate column. The L1 and L2 attached to each variable name indicate that they are lag 1 and lag 2 variables.

Using the notations t = time (days), C = confirmed cases, N = new cases, and T = total deaths. The equation for confirmed cases is

$$\hat{C}_t = -0.6740 + 0.0216\text{t} - 0.0001C_{t-1} + 0.0303N_{t-1} + 0.8996T_{t-1} + 0.0029C_{t-2} + 0.0109N_{t-2} - 0.0216T_{t-2} \quad \ldots\ldots (10)$$

The equation for new cases is

$$\hat{N}_t = -25.5337 + 1.4968t + 1.7069C_{t-1} - 1.5641N_{t-1} - 0.8260T_{t-1} - 1.7639C_{t-2} + 0.1008N_{t-2} + 3.3924T_{t-2} \quad \ldots\ldots (11)$$

The equation for total deaths is

$$\hat{T}_t = -25.6444 + 1.4953t + 2.6593C_{t-1} - 1.5041N_{t-1} - 0.6610T_{t-1} - 1.7110C_{t-2} + 0.0900N_{t-2} + 3.0310T_{t-2} \quad \ldots\ldots (12)$$

Using the above derived models, the following forecasts (in table 6) can therefore be made easily, and be represented in fig. 2 as extension to the actual series

Table 6: Ten Days Forecasts made From the Models

Forecasts	Confirmed cases	New Cases	Total Deaths
Jun-11	14302	429	397
Jun-12	14722	419	408
Jun-13	15162	440	419
Jun-14	15610	447	431
Jun-15	16068	458	443
Jun-16	16535	467	455
Jun-17	17012	476	468
Jun-18	17499	486	481
Jun-19	17996	497	494
Jun-20	18504	507	507

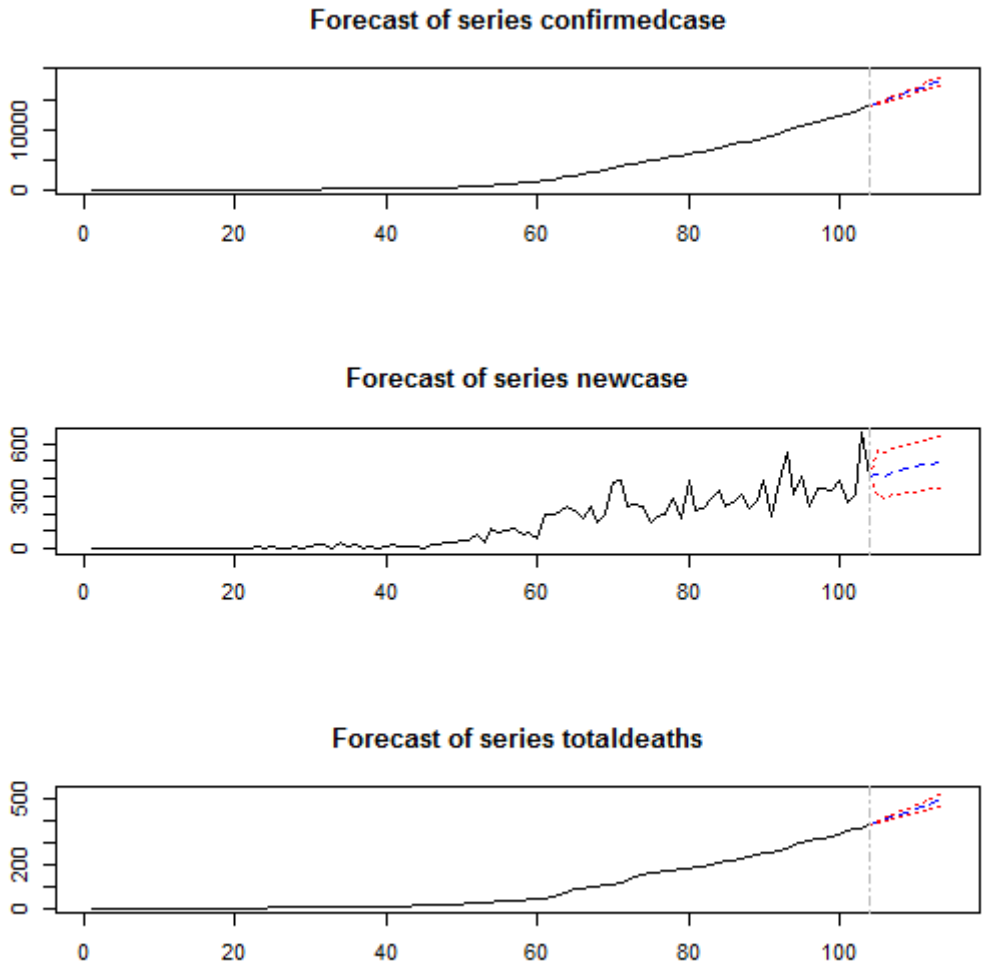

Fig. 3: Plot Showing Actual and Forecasts Values for 10 Days with CI Bounds

It is obvious from the forecasts made that the model is effective, because the forecast values follow the general pattern in the series

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)
www.abjournals.org

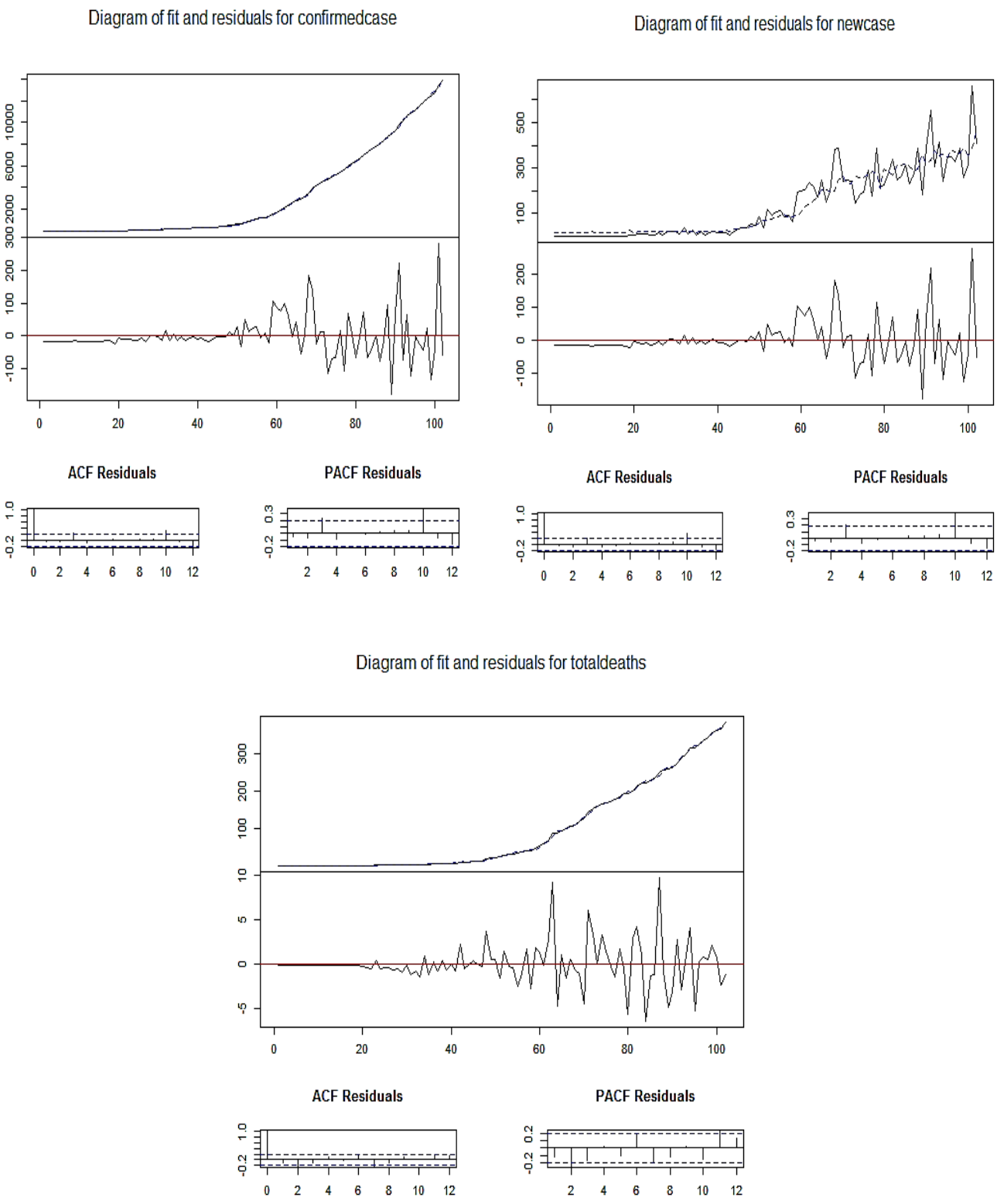

Fig. 4: Fits and Residual Plots

African Journal of Biology and Medical Research
ISSN: 2689-534X
Volume 3, Issue 2, 2020 (pp. 171-181)

www.abjournals.org

Model Diagnostics

The above charts (fig. 4) show the actual, the fitted lines and the residuals. For the plots on confirmed cases, the actual and the fitted are closely knitted together, signifying a good model. For new cases, the fitted passes through the actual values, this also shows that the model predicts well. Lastly, for total deaths, the difference between the fitted and the actual is not obvious, this informs us that the model predicts reliably. The residuals clustering around zero in all the three cases indicate normality of the residual values and adequacy of the models for future forecasts.

Table 7: VAR Models Selection Criteria

Order of p	AIC	Log-Likelihood
1	2256.99	-1113.5
2	2242.76	-1097.4

Using the AIC criterion, the VAR of order 2 model gives a better precision, therefore should be made use in estimating the parameters and forecasting.

DISCUSSION OF RESULTS

The general upward movements noticed in *confirmed*, *new*, and *death* cases of covid-19 in Nigeria as represented in fig. 2 above is alarming. With respect to the data coverage of this paper, the sudden rise started with *confirmed* cases on the 18th March, 2020 and maintains the upward trend until June 10, 2020. Using the Augmented Dickey Fuller and Phillips-Perron unit root tests, the null hypothesis of non-stationarity is rejected in all the cases of the variables, this shows that the lagged series are stationary at 5% level of significance. The Granger causality test reveal that *confirmed cases* Granger-causes *new cases* and *total deaths* of covid-19 in Nigeria. Using *confirmed cases* as the response variable it is discovered that there is a general significant relationship among the predictors and the response variable with p-value 2.2e-16. Having a critical examination on the estimates displayed in table 3, it will be seen that only *confirmedcase.L1* and trend are significant having p-values 0.0428 and 0.0113 respectively. The significance of relationship is also visible in the other models as revealed in tables 4 and 5. Trend is the only one significant in the second model while *newcase.L2* is the only one in the second one. The real, fitted and the residual plots displayed in fig. 4 show that the models are well fitted and that the forecasts in table 6 are reliable. The criteria AIC and log-likelihood implemented on the model confirmed that the VAR model of order 2 gives a better model for predictions and forecasts of covid-19 cases in Nigeria.

CONCLUSION AND RECOMMENDATION

Using the results from the vector autoregressive analysis for multivariate time series carried out on covid-19 cases in Nigeria, it can be concluded that the VAR model of order 2 gives a better model suitable for predicting and forecasting future occurrences of *confirmed cases*, *new cases*, and *total deaths* of pandemic in Nigeria. It is therefore recommended that researchers interested in modelling the pandemic employ the model for reliable predictions. Furthermore, the government should intervene in curbing the ever-increasing cases of the pandemic to save that population at risk.

REFERENCES

Amisano, G., and C. Giannini. (1997). *Topics in Structural VAR Econometrics*. 2nd ed. Heidelberg: Springer.

Becketti, S. (2013). *Introduction to Time Series Using Stata*. College Station, TX: Stata Press.

Campbell, J., Lo, A., & MacKinlay, C. (1997). *The econometrics of financial markets*. Princeton: Princeton University Press.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74,* 427-431.

Greene, W. H. (2008). *Econometric Analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J. D. (1994a). *Time series analysis.* Princeton: Princeton University Press.

Hamilton, J. D. (1994b). State space models. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of econometrics*. Amsterdam: Elsevier.

Hossain, A., Kamruzzaman, M., and Ali, M. (2015). Vector Autoregressive (VAR) Modeling and Projection of DSE. *Chinese Business Review*, June 2015, Vol. 14, No. 6, 273-289

Lutkepohl, H. (1999). *Vector autoregressions* (Unpublished manuscript, Institutfür Statistik und Ökonometrie, Humboldt-Universitat zu, Berlin).

Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. New York: Springer.

Mills, T. C. (1999). *The econometric modeling of financial time series* (2nd ed.). Cambridge: Cambridge University Press.

NCDC (2020): http://covid19.ncdc.gov.ng. Accessed July, 7 2020

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika, 75,* 335-346.

Stock, J. H., and M. W. Watson. (2001). Vector autoregressions. *Journal of Economic Perspectives* 15: 101–115.

Tsay, R. S. (2001). *Analysis of financial time series.* Berlin: Springer-Verlag

Waggoner, D. F., & Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics, 81*(4), 639-651.

World Health Organization. Coronavirus disease 2019 (COVID-19). Situation Report – 113. 12 May 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed July 7 2020

World Health Organization. GCM teleconference – Note for the Records. 10 January 2020. Subject: Pneumonia in Wuhan, China. Available from: https://www.who.int/blueprint/10-01-2020-nfr-gcm.pdf?ua=1. Accessed July 7 2020

World Health Organization. Novel coronavirus (2019-nCoV). Situation Report – 1. 21 January 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4 Accessed July 7 2020

World Health Organization. Teleconference of the R&D Blueprint GCM. 20 January 2020. Pneumonia of unknown etiology in Wuhan China. Available from: https://www.who.int/blueprint/priority-diseases/key-action/20-01-2020-nfr-gcm.pdf?ua=1. Accessed July 7 2020