International Journal of Medical Science and Applied Biosciences
ISSN: 2545-5893(Print) 2545-5877 (Online)
Volume 5, Number 2, June 2020
http://www.casirmediapublishing.com

SARIMA MODELLING OF DAILY LABORATORY CONFIRMED CASES OF CORONAVIRUS IN NIGERIA

Ette Harrison Etuk
Department of Mathematics
Rivers State University, Port Harcourt, Nigeria
Email: ettetuk@yahoo.com, ettehetuk@gmail.com, etuk.ette@ust.edu.ng

ABSTRACT: This study is an attempt to model daily confirmed cases of coronavirus in Nigeria. A time plot of the series shows an upward trend with some seasonality. It is tested for unit test and is shown to be non-stationary. Its difference shows evidence of stationarity. The correlogram of the difference shows significant spikes at the partial autocorrelation function at lags 1 and 12 and at its autocorrelation function at lags 1 and 13, with the lag 13 spike surrounded by spikes of comparable lengths in the same direction. This suggests an autoregressive fit of lags 1 and 12 and a moving average fit of lags 1, 13 and 14. A fit of the model shows that only the moving average lags are significant. A more specific SARIMA(0, 1, 1)x(0, 0, 1)$_{13}$ model is fitted to the series. This shows that the series may be regarded as a SARIMA(0, 1, 1)x(0, 0, 1)$_{13}$ case.
Key Words: Covid-19 pandemic, SARIMA modelling

INTRODUCTION

Coronavirus disease has attained a worldwide status as a pandemic. The prime case recorded in Nigeria appeared on February 27, 2020 and was confirmed at the Virology Laboratory of the Lagos State University Teachnig Hospital. He is from Italy and arrived Nigeria from Milan, Italy on a business trip. He was being managed at the Infectious Disease Hospital, Yaba, Lagos. The Federal Ministry of Health announced the first confirmed case in Nigeria on the next day, 28th February 2020. (Ehanire, 2020). The occurrence of a phenomenon like this is often an opportunity for researchers to model its incidence and in the case of a medical condition like this to proffer a curative solution to it. As published in Thisday newspaper, Gumel (2020) underscored this point saying that a researcher after studying the components of this phenomenon can model it using mathematical tools. Voice of Nigeria announces that Professor Maurice Iwu has claimed to have discovered a cure for it Ukoh(2020).

The approach of seasonal autoregressive integrated moving average modelling is to be adopted in this work. Proposed by Box and Jenkins (1976) it has been widely successfully applied to model seasonal time series. To mention a few, look at Etuk (2013), Mwanga *et al.* (2017) and Adams and Bamanga (2020). Here it is our intention to see the daily occurrence of this disease in Nigeria as a time series and model it accordingly. Section 2 dwells on Materials and Methods, section 3 on Results and Discussion and section 4 on the Conclusion.

MATERIALS AND METHODS

Data: The data used for this study are 64 values of daily cumulative laboratory confirmed cases of coronavirus-19 recorded by Nigerian Centre of Disease Control with website http://covid19.ncdc.gov.ng/ . They are displayed in the appendix.

Seasonal Autoregressive Integrated Moving Average Modellng

It is our intention to study this covid-19 phenomenon as a daly time series. It is hoped that it shall be modelled as a seasonal autoregressive integrated moving average model.

A time series $X_1, X_2, \ldots, X_n$ is said to follow an autoregressive moving average model if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q} \quad (1)$$

where $\{\varepsilon_t\}$ is a white noise process, the α's and β's are constants chosen such that the model (1) is both stationary and invertible. Model (1) is denoted as an ARMA(p, q). it may be written as an ARMA(p, q). It may be written as

$$A(L)X_t = B(L)\varepsilon_t \quad (2)$$

where $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p$ and $B(L) = 1 + \beta_1 L + \beta_2 L^2 + \ldots + \beta_q L^q$ and $L^k X_t = X_{t-k}$.

Hardly is a time series $\{X_t\}$ stationary. In that case Box and Jenkins (1976) proposed that differencing to an order d could render it stationary. Then if the difference of $\{X_t\}$, $\nabla^i(X_t)$, I = 1, 2, ..., d -1, is non-stationary and the d^{th} difference is stationary then a replacement of X_t by $\nabla^d(X_t)$, in (1) or (2) yields an autoregressive integrated moving average model of order (p, d, q) in X_t denoted by an ARIMA(p, d, q) model. Then $\nabla = 1 - L$.

If seasonal periodicity s is observed to happen in the series, assuming that there is a seasonal trend of period D, then the model becomes

$$\Phi(L^s)A(L)X_t = \Theta(L^s)B(L)\varepsilon_t \quad (3)$$

where $\Phi(L^s) = 1 - \varphi_1 L^s - \varphi_2 L^{2s} - \ldots - \varphi_P L^{Ps}$ and $\Theta(L^s) = 1 + \theta_1 L^s + \theta_2 L^{2s} + \ldots + \theta_Q L^{Qs}$. Model (3) is expressed as a seasonal autoregressive integrated moving average model of order (p, d, q)x(P, D, Q)$_s$ denoted by SARIMA(p, d, q)x(P, D, Q)$_s$. An indication of autoregressive order p is a significant spike at lag p on the partial autocorrelation function and an indication of a moving average order q is a significant spike on the partial autocorrelation function at lag q.

Computer Software: The software used for this work is eviews 10. It uses the least square approach too estimation of model parameters.

RESULTS AND DISCUSSION

The difference of the data yields daily laboratory confirmed cases of coronavirus in Nigeria. Its time plot is in figure 1. This shows a positive trend and some seasonality. The Augmented Dickey Fuller test of stationarity of Table 1 adjudges it as non-stationary, in which case differencing has to be done on it.

International Journal of Medical Science and Applied Biosciences
ISSN: 2545-5893(Print) 2545-5877 (Online)
Volume 5, Number 2, June 2020
http://www.casirmediapublishing.com

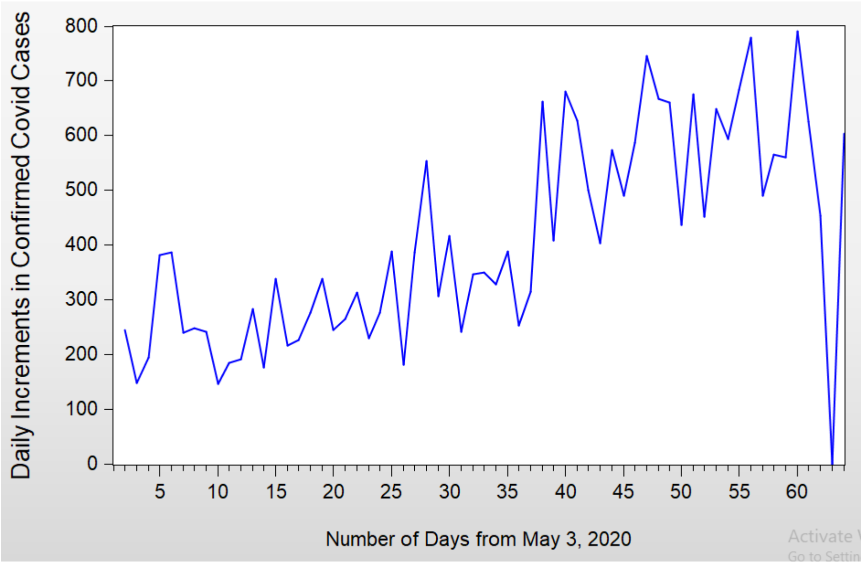


Figure 1: Time Plot of number of daily confirmed cases of cornavirus in Nigeria.

Table 1: Unit Root Test for confirmed cases

Null Hypothesis: DCOVID has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.519660	0.1159
Test critical values:	1% level	-3.542097	
	5% level	-2.910019	
	10% level	-2.592645	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DCOVID)
Method: Least Squares
Date: 07/05/20 Time: 06:15
Sample (adjusted): 4 64
Included observations: 61 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DCOVID(-1)	-0.280798	0.111443	-2.519660	0.0145
D(DCOVID(-1))	-0.400361	0.135526	-2.954133	0.0045
C	119.8430	49.07743	2.441916	0.0177

R-squared	0.320722	Mean dependent var	7.459016
Adjusted R-squared	0.297299	S.D. dependent var	171.7062
S.E. of regression	143.9366	Akaike info criterion	12.82455
Sum squared resid	1201629.	Schwarz criterion	12.92837
Log likelihood	-388.1489	Hannan-Quinn criter.	12.86524
F-statistic	13.69242	Durbin-Watson stat	2.035145
Prob(F-statistic)	0.000013		

The time plot of differences of confirmed cases in Figure2 shows a horizontal trend and some seasonality. The Augmented Dickey Fuller test on the series in Table 2 shows that they are stationary.

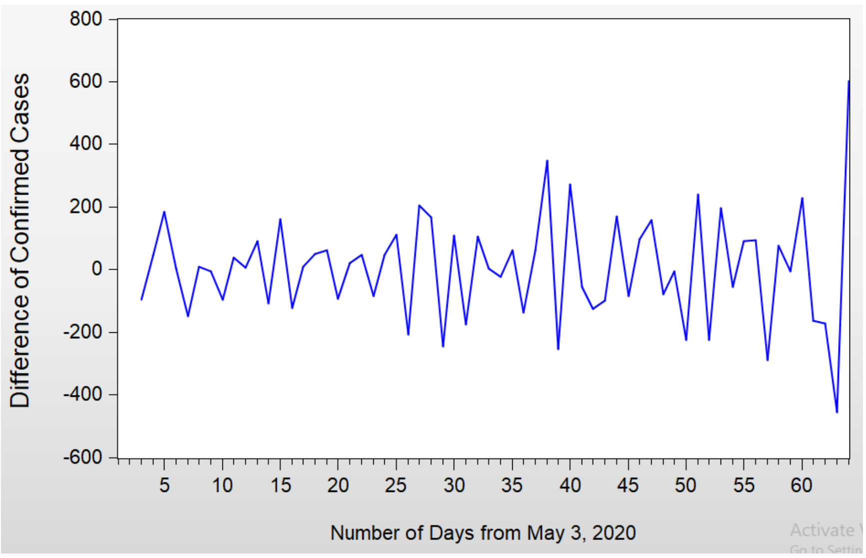


Figure 2: Difference of confirmed cases of coronavirus

Table 2: Unit Root Test on Difference of confirmed cases

International Journal of Medical Science and Applied Biosciences
ISSN: 2545-5893(Print) 2545-5877 (Online)
Volume 5, Number 2, June 2020
http://www.casirmediapublishing.com

Null Hypothesis: DDCOVID has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=10)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.31078	0.0000
Test critical values:	1% level	-3.542097	
	5% level	-2.910019	
	10% level	-2.592645	

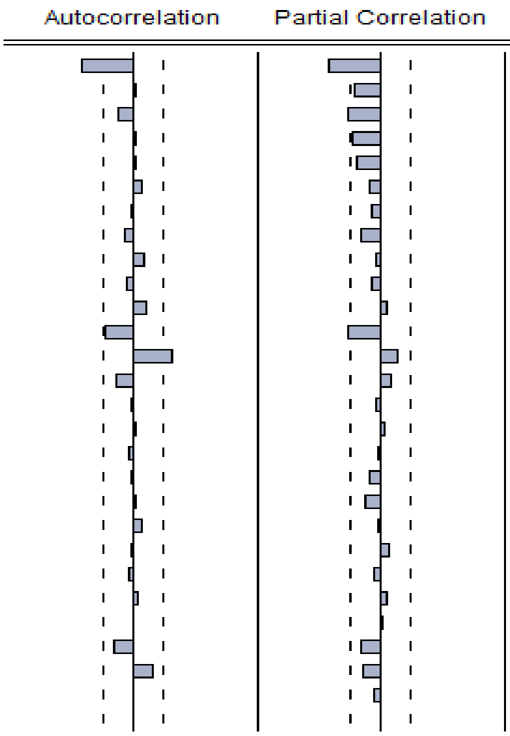
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DDCOVID)
Method: Least Squares
Date: 07/05/20 Time: 06:18
Sample (adjusted): 4 64
Included observations: 61 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DDCOVID(-1)	-1.554586	0.126278	-12.31078	0.0000
C	5.240674	19.25306	0.272200	0.7864

R-squared	0.719788	Mean dependent var	11.45902
Adjusted R-squared	0.715039	S.D. dependent var	281.5934
S.E. of regression	150.3194	Akaike info criterion	12.89564
Sum squared resid	1333160.	Schwarz criterion	12.96485
Log likelihood	-391.3170	Hannan-Quinn criter.	12.92276
F-statistic	151.5552	Durbin-Watson stat	2.118366
Prob(F-statistic)	0.000000		

The correlogram of the difference of the confirmed cases in Figure 3 below shows a positive spike at lags 1 and 12 for the partial autocorrelation function (PACF) and significant spikes at lags 1 and 13 in the autocorrelation function (ACF), the spike at lag 13 surrounded with comparative spikes, in direction and length, at lags 12 and 14, an indication of seasonality of lag 13. This has led to the hypothesis of an ARIMA(12, 1, 14) with the autoregressive lags 1 and 12 and the moving average lags 1, 13 and 14. A summary of the model fit in Table 3 shows that the only significant parameters are at the moving average lags of 1, 13 and 14, which is an indication of a seasonality of period 13.



	AC	PAC	Q-Stat	Prob
1	-0.442	-0.442	12.718	0.000
2	0.020	-0.218	12.746	0.002
3	-0.120	-0.267	13.707	0.003
4	0.013	-0.237	13.719	0.008
5	0.010	-0.196	13.725	0.017
6	0.067	-0.090	14.039	0.029
7	-0.024	-0.074	14.081	0.050
8	-0.069	-0.161	14.433	0.071
9	0.086	-0.039	14.983	0.091
10	-0.060	-0.078	15.256	0.123
11	0.101	0.043	16.051	0.139
12	-0.242	-0.263	20.690	0.055
13	0.332	0.139	29.613	0.005
14	-0.155	0.083	31.609	0.005
15	-0.014	-0.033	31.626	0.007
16	0.009	0.037	31.633	0.011
17	-0.043	-0.022	31.797	0.016
18	-0.011	-0.082	31.808	0.023
19	0.022	-0.137	31.853	0.032
20	0.081	-0.022	32.477	0.038
21	-0.024	0.076	32.532	0.052
22	-0.037	-0.065	32.667	0.067
23	0.025	0.062	32.730	0.086
24	-0.004	0.017	32.732	0.110
25	-0.165	-0.167	35.661	0.077
26	0.163	-0.152	38.589	0.053
27	0.006	-0.058	38.594	0.069
28	-0.007	0.007	38.599	0.088

Figure 3: Correlogram of difference of confirmed cases

Table 3: Estimation of ARIMA model for the difference of confirmed cases

Dependent Variable: DDCOVID
Method: Least Squares
Date: 07/05/20 Time: 05:51
Sample: 3 64
Included observations: 62
Convergence achieved after 46 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.039049	0.200023	0.195224	0.8459
AR(12)	-0.050128	0.173767	-0.288480	0.7740
MA(1)	-0.744949	0.154957	-4.807473	0.0000
MA(13)	0.348918	0.185932	1.876595	0.0658
MA(14)	-0.368165	0.150161	-2.451806	0.0174
SIGMASQ	15833.89	2767.697	5.720963	0.0000

R-squared	0.448331	Mean dependent var	5.790323
Adjusted R-squared	0.399075	S.D. dependent var	170.7991
S.E. of regression	132.4023	Akaike info criterion	12.74843
Sum squared resid	981701.3	Schwarz criterion	12.95428
Log likelihood	-389.2014	Hannan-Quinn criter.	12.82925
Durbin-Watson stat	1.972650		

This is a SARIMA(0, 1,1)X(0, 0,1)$_{13}$ model. A more specific estimation of the above-mentioned model is done on Figure 4 to have the model

International Journal of Medical Science and Applied Biosciences
ISSN: 2545-5893(Print) 2545-5877 (Online)
Volume 5, Number 2, June 2020
http://www.casirmediapublishing.com

$\nabla X_t = -0.7331\varepsilon_{t-1} + 0.3503\varepsilon_{t-13} - 0.3779\varepsilon_{t-14} + \varepsilon_t$

which is a SARIMA(0, 1, 1)x(0,0, 1)$_{13}$ model for the confirmed cases of coronavirus in Nigeria.

CONCLUSION

The daily confirmed cases of coronavirus in Nigeria have been shown to show a SARIMA(0, 1, 1)x(0,0, 1)$_{13}$ model. This model may be used to approximate its daily variation. Any study of the series may be based on this model.

Table 4: More specific Estimation of a SARIMA(0,1,1)x(0, 0, 1)$_{13}$ model for the confirmed cases

Dependent Variable: DDCOVID
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 07/05/20 Time: 06:24
Sample: 3 64
Included observations: 62
Convergence achieved after 35 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.733143	0.088343	-8.298805	0.0000
MA(13)	0.350301	0.170952	2.049119	0.0450
MA(14)	-0.377918	0.138582	-2.727031	0.0084
SIGMASQ	15839.22	2763.845	5.730863	0.0000

R-squared	0.448146	Mean dependent var	5.790323
Adjusted R-squared	0.419602	S.D. dependent var	170.7991
S.E. of regression	130.1214	Akaike info criterion	12.68608
Sum squared resid	982031.5	Schwarz criterion	12.82332
Log likelihood	-389.2685	Hannan-Quinn criter.	12.73996
Durbin-Watson stat	1.941012		

Inverted MA Roots	.92	.88+.29i	.88-.29i	.68-.64i
	.68+.64i	.31-.88i	.31+.88i	-.12-.93i
	-.12+.93i	-.54-.77i	-.54+.77i	-.83+.43i
	-.83-.43i	-.94		

REFERENCES

Adams, S. O. and Bamanga, M. A. (2020). Modeling and Forecasting Seasonal Behavior of Rainfall in Abuja, Nigeria; A SARIMA Approach. American Journal of Mathematics and Statistics, 10(1): 10 – 19.

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control, Holden Day, San Francisco.

Ehanire, O. (2020). First Case of Coronavirus Disease Confirmed in Nigeria. Nigerian Centre for Disease Control. https://ncdn.gov.ng accessed 01/07/2020.

Etuk, E. H. (2013). Seasonal Arima Modelling of Nigerian Monthly Crude Oil Prices. Asian Economic and Financial Review, 3(3): 333 – 340.

Gamel, A. B. (2020). Using Mathematics to Understand and Control the 2019 Novel Coronavirus Pandemic. https://www.thisdaylive.com/index.php/2020/05/03/using-mathematics-to-understand-and-control-the-2019-novel-cornavirus-pandemic/ accessed 08/06/2020.

Mwanga D., Ong'ala, J. and Orwa, G. (2017). Modelling Sugarcane Yields in the Kenya Sugar Industry: A SARIMA MODEL forecasting Approach. International Journal of Statistics and Applications, 7(6): 280 – 288.

Ukoh, E. (2020). Nigeria makes breakthrough on Covid-19 cure. https://www.von.gov.ng/nigeria-makes-breakthrough-on-covid-19-cure/ accessed 07/07/2020.

APPENDIX

Cumulative daily laboratory confirmed cases of Coronavirus in Nigeria (starting from 3 May 2020) read row wise

2558 2802 2950 3145 3526 3912 4151 4399 4641 4787 4971 5162 5445

5621 5959 6175 6401 6677 7016 7261 7526 7839 8068 8344 8733 8915

9302 9855 10162 10578 10819 11166 11516 11844 12233 12486 12801 1346 13873

14554 15181 15682 16085 16658 17148 17735 18480 19147 19808 20244 20919 21371

22020 22614 23298 24077 24567 25133 25694 26484 27110 27564 27564 28167

Source: NCDC Cornavirus COVID-19 Microsite http://covid19.ncdc.gov.ng/