American Journal of Mathematics and Statistics 2022, 12(2): 15-21
DOI: 10.5923/j.ajms.20221202.01

Markov Switching Autoregressive Modelling of the COVID-19 Daily Cases and Deaths in Nigeria

Benson T. I.[1], Biu E. O.[2], Nwakuya M. T.[2,*]

[1]Department of Mathematics Statistics, Ignatius Ajuru University of Education, Rivers State Nigeria
[2]Department of Mathematics and Statistics, University of Port Harcourt, Nigeria

Abstract This article modeled the daily COVID-19 infected cases and deaths of Nigeria using a Markov switching model. Structural breaks and Stationarity of the daily COVID-19 cases and deaths series were investigated. Unit root and unit root structural break tests were applied where evidence of breaks exists. The results show that each of the series was found to be a nonlinear and nonstationary series with evidence of a structural break. The results of the unit root tests in the presence of structural breaks indicated that each of the series was two significant wave changes. Consequently, a Markov switching AR (MSM(2)-AR(1)) model with two regimes was fitted to the data having established its suitability in modelling the series. Finally, a transition probability matrix between the expected number of COVID-19 infected cases and death cases was obtained.

Keywords Markov Switching AR Model, Structural Breaks, COVID-19, Nonstationary Series, Volatility, Nonlinear Series

1. Introduction

There is a need to carry out statistical analysis on the new coronavirus in the world and as well as in Nigeria. Since the epidemic breakouts, a large number of enterprises have faced closures, employment has become more difficult, and people's lives have been greatly affected. Focusing only on the statistical analysis, however, does not solve the problem. But the government needs to adopt the statistical results and interpretation in putting the measures to control the spread of the coronavirus. Since the outbreak of the epidemic, there have been several studies to understand the trend of the virus and predict or forecast the next possible behavior of the virus in order to help in policy making to combat the virus. Some of the studies on COVID-19 include the work of Ogundokun *et al.,* (2020) who in their study showed that the government of Nigeria made the right decision in enforcing travel restrictions because it was discovered that people's travel history and contacts increased their chances of contracting COVID-19. Maleki *et al.,* (2020) proposed an effective technique in predicting confirmed and recovered COVID-19 infections all over the world. Roy *et al.,* (2021) in their study of the epidemiologic pattern in the prevalence and incidence of COVID-2019 in India, showed that west and south of the Indian district are highly vulnerable for COVID-2019.

These studies proposed different models for the virus which varies from time series models to growth models. Adedeji *et al* (2021) in their paper examined the dynamic effect of COVID-19 on four major oil prices concerning China and Nigeria using the structural vector autoregressive method. Zhao *et al* (2020) predicted Covid-19 spread in African countries using the Maximum-Hasting (MH) parameter estimation method and the modified Susceptible Exposed Infectious Recovered (SEIR) model. Kartikasari *et al* (2020) evaluate the temporal COVID-19 infection behavior among HCW-D, HCW-A, and non-HCW using ARFIMA Model while Balah and Djeddou (2020) used Autoregressive fractionally integrated moving average Models (ARFIMA) to forecast Covid-19 daily new cases in Algeria. Lahmiri and Bekiros (2021) employed ARFIMA and FIGARCH models to examine long memory (self-similarity) in digital currencies and international stock exchanges before and during the COVID-19 pandemic. Tatar *et al* (2021) Analyzed excess deaths during the COVID-19 pandemic in the state of Florida using seasonal autoregressive integrated moving average time-series modeling and historical mortality trends. Although these models provide results, there is a risk that these models may not actually reflect the actual properties of the data which might lead to either underestimation or overestimation of the data. Hence, it is important to properly examine the properties of the data to determine its nature and based on that determine the appropriate models that can be associated with such dataset and then proceed to obtain the best model that fits the data for proper and adequate modelling and

* Corresponding author:
maureen.nwakuya@uniport.edu.ng (Nwakuya M. T.)
Received: Mar. 7, 2022; Accepted: Mar. 25, 2022; Published: Aug. 15, 2022
Published online at http://journal.sapub.org/ajms

forecasting for optimal policy making.

From the foregoing, it is obvious that the total number of COVID-19 infected cases and deaths in Nigeria have not been studied using a regime-based model. Testing for nonlinearity and structural break is an important step in time series analysis since it helps researchers to determine whether to use each of the linear and nonlinear time series models. In this paper, we demonstrate the usefulness of the regime-switching Markov switching autoregressive (MS-AR) approach in examining the COVID-19 total infected cases and total deaths in Nigeria. We seek to establish evidence of nonlinearity in the total number of COVID-19 infected cases and deaths due to common regime switching behavior occasioned by structural breaks. The remainder of the paper is organized as follows: Section 2 introduces other methodologies and the Markov switching vector autoregressive specifications. Section 3 presents the data analysis and results while Section 4 concludes the paper.

2. Methodology

2.1. Unit Roots and Stationarity Tests: (Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test)

The KPSS test is a Unit root test that tests for the stationarity of a given series around a deterministic trend. It is used to test whether we have a deterministic trend or stochastic trend: KPSS test is conducted under the following hypothesis:

H_0: $Y_t \sim I(0)$ trend stationary

H_1: $Y_t \sim I(1)$ stochastic trend

If the null hypothesis is not rejected, it means that there is evidence that the series is trend stationary. The process is as follows:

- Regress y_t on a constant and time trend. Get Ordinary Least Square residuals $\hat{u}$
- Calculate the partial sum of the residuals:

$$S_t = \sum_{i=1}^{t} \hat{u}_i \quad (1)$$

- Compute the KPSS test statistic;

$$KPSS = T^{-2} \sum_{t=1}^{T} \frac{S_t}{S_u^2}; \quad (2)$$

where S_u^2 is the estimate of the long-run variance of the residuals.

- Reject H_0 when KPSS is large

2.2. Nonlinearity and Structural Break Tests

The BDS and Tsay tests were adopted for testing nonlinearity in the series. The test of Brock et al. (1996) known as the BDS test, takes the concept of the correlation integral and transforms it into a formal test statistic which is asymptotically distributed as a standard normal variable under the null hypothesis of independent and identical distribution against an unspecified alternative Brooks C. (1996). Considering a time series x_t with h histories, $x_t^h = (x_t, x_{t+1}, \dots, x_{t-m+1})$, The correlation integral at embedding dimension h can be estimated by;

$$C_h(\epsilon, T) = \frac{2}{T_m\,(T_m - 1)_{t<s}} \sum I_\epsilon\left(x_t^h, x_s^h\right) \quad (3)$$

where the parameter h is the embedding dimension, T is the sample size and $T_m = T - m + 1$, is the maximum number of overlapping vectors that we can form with a sample of size T, I_ϵ is an indicator function that is equal to one if $\|x_t^h - x_s^h\| < \epsilon$ and equal to zero otherwise. A pair of vectors x_t^h, x_s^h is said to be ϵ apart, if the maximum-norm $\|.\|$ is greater or equal to ϵ. Using this relation the BDS test statistic is defined as;

$$BDS(h, \epsilon) = \sqrt{T} \frac{C_h(\epsilon,T) - [C_1(\epsilon)]^h}{\sigma_h(\epsilon,T)} \quad (4)$$

Where $\sigma_h(\epsilon, T)$ is the standard deviation of $\sqrt{T}(C_h(\epsilon, T) - [C_1(\epsilon)]^h)$.

The Tsay test of Tsay (1989), tests for evidence of threshold nonlinearity under the null hypothesis of linearity in the dataset. This test is an improvement of Keenan's test for nonlinearity. Keenan (1985) proposes a nonlinearity test for time series that uses $\hat{x}_t^2$ only and modifies the second step of the reset test to avoid multicollinearity between $\hat{x}_t^2$ and x_{t-1}. Tsay test improved on the Keenan's test by allowing disaggregated nonlinear variables, that is all cross products of $X_{t-i}X_{t-j}, i, j = 1, \dots, p$. Hence generalizing Keenan test by explicitly looking for quadratic serial dependence in the data.

The procedure is as follows;

Firstly, using any selection criterion, select the value p of the number of lags involved in the regression, then fit X_t on $(1, X_{t-1}, \dots, X_{t-p}, i, j = 1, \dots, p$ to obtain the fitted values $\hat{X}_t$, the residuals set $\hat{e}_t$ and the residual sum of squares SSR. Then regress the product $X_{t-i}X_{t-j}$ on $(1, X_{t-1}, \dots, X_{t-p}, i, j = 1, \dots, p$, to obtain the residual set $\hat{\xi}_t$. The calculate;

$$\hat{\vartheta}_t = \frac{\sum_{t=p+1}^{n} \hat{e}_t\, \hat{\xi}_t}{\sum_{t=p+1}^{n} \hat{\xi}_t} \quad (5)$$

The test statistics is the obtained as;

$$F^* = \frac{(n-2p-2)\hat{\xi}_t}{SSR - \hat{\xi}_t} \quad (6)$$

This test statistic F^* is asymptotically distributed as $F_{m,n-m-p-1}$ where $m = \frac{p(p-1)}{2}$. Once there is evidence of nonlinearity, we proceed to check for the presence of structural break.

2.3. Unit Roots and Stationarity Tests

Kim and Perron, (2006) stated that in testing for unit root, it is necessary to consider the effects of structural breaks during unit root testing. Structural breaks play significant roles in that it can lead to misspecification of deterministic function of the auxiliary regression that is used for either testing for unit root or variance stationarity.

In this study, we consider the structural break unit root test of Zivot and Andrews (1992) and Perron (1989). Perron uses a modified Dickey-Fuller (DF) unit root tests that includes dummy variables to account for one known, or exogenous structural break. The break point of the trend function is fixed (exogenous) and chosen independently of the data. Perron's (1989) unit root tests allows for a break under both the null and alternative hypothesis. These tests have less power than the standard DF type test when there is no break. However, Perron (1989) points out that they have a correct size asymptotically and is consistent whether there is a break or not. Moreover, they are invariant to the break parameters and thus their performance does not depend on the magnitude of the break. Zivot and Andrews (1992) endogenous structural break test is a sequential test which utilizes the full sample and uses a different dummy variable for each possible break date. The break date is selected where the t-statistic from the ADF test of unit root is at a minimum (most negative). Consequently, a break date will be chosen where the evidence is least favorable for the unit root null.

2.4. Markov Switching Vector Autoregressive Model Specification

The Markov Switching Model of Hamilton (1989), also known as the Regime Switching Model, is one of the most popular nonlinear time series models in the literature. This model involves multiple structures that can characterize the time series behaviors in different regimes or states or episodes. A novel feature of the Markov switching model is its Markovian property, which states that the probability of transitioning to any particular state is dependent solely on the current state, and not on the sequence of state that preceded it.

In this work we considered modeling the Markov Switching Autoregressive (MS-AR) process. We specifically consider a two regime-switching. We intend to model the regime S_t as the outcome of an unobserved 2-state Markov chain with S_t independent of v_t for all t. The MS-AR model for two regimes with autoregressive (AR) process of order k is as follows:

$$x_t = \begin{cases} c_1 + \sum_{i=1}^{k} \eta_{1,i} x_{t-i} + v_t \;\; if \;\; S_t = 1 \\ c_2 + \sum_{i=1}^{k} \eta_{2,i} x_{t-i} + v_t \;\; if \;\; S_t = 2 \end{cases} \quad (7)$$

Where the regimes are represented by S_t. The intercept and parameters of AR part are dependent of S_t at time t. Given that we considered two regimes, we label regime 1 as S_1 representing the period of first wave of COVID-19 and we label regime 2 as S_2 representing the 2nd wave of COVID-19. The transitions of the S_t are presumed to be ergodic and intricate 1st order Markov-process with transition probabilities

$$p(s_t = 2|s_{t-1} = 1) = \upsilon_1, \; p(s_t = 1|s_{t-1} = 2) = \upsilon_2 \quad (8)$$

The transition matrix P captures the probability of switching, where;

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

The nearer the probability P_{ij} is to one the longer it takes to shift to the next regime. The diagonal element of the matrix of the transition probabilities contains important information on the expected duration on the state of the regime.

Let D be defined as the duration of state j; we have:

$$D = 1, if \; S_t = j \;\&\; S_{t+1} \neq j; \Pr[D = 1] = 1 - P_{jj} \quad (9)$$

$$D = 2, if \; S_t = S_{t+1} = j \;\&\; S_{t+2} \neq j; \Pr[D = 2] = P_{jj}\,(1 - P_{jj}) \quad (10)$$

Then the expected duration of regime j can be derived as;

$$E(D) = \sum_{j=1}^{\infty} j \Pr(D = j) = \frac{1}{1-P_{jj}} \quad (11)$$

As stated by Oluwasegun, A. A. et al (2020).

3. Results and Discussion

In this section, we present the results and discussion on the daily total infected cases and total deaths of COVID-19 data for Nigeria from 28th February 2020 to 30th September 2021.

3.1. Time Plot of the Series

A look at the time series plot of the total COVID-19 cases and deaths in Figure 1 shows that the total COVID-19 Cases and Deaths of Nigeria respectively has an upwards trend and is non-stationary.

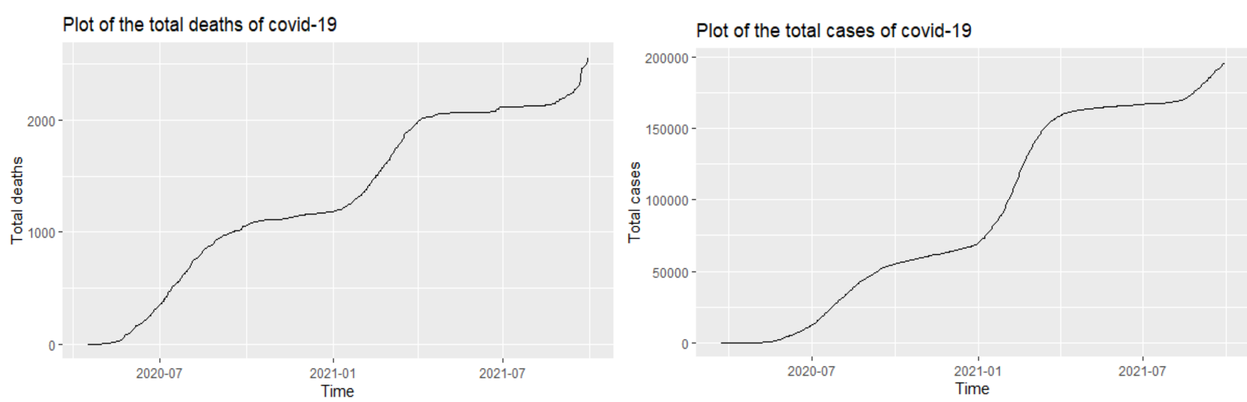


Figure 1. Time series plot of the total COVID-19 cases and deaths of Nigeria

Table 1. KPSS unit root test results based on total COVID-19 cases series

Series	Level	Trend Stationary
Case	7.9305	0.6279
1st Dcase	0.5687	0.5050
2ndDdcase	0.0516*	0.0413*

From Table 1, at 0.05 significance level, the 2nd difference of the total cases is stationary while the original series and the 1st difference of the original series are nonstationary.

Table 2. KPSS unit root test results based on total COVID-19 deaths series

Series	Level	Trend Stationary
Death	7.7308	0.7322
1stDdeath	0.2751	0.2502
2ndDdeath	0.0223*	0.0176*

From Table 2, at 0.05 significance level, the 2nd difference of the total deaths is stationary while the original series and the 1st difference of the original series are nonstationary.

Figure 2 shows the time plots of the differenced series which strongly indicates evidence of non-stationarity, volatility clustering and change in structure hence we proceed to test for nonlinearity and structural break in the next sub section.

3.2. Nonlinearity and Structural Break Tests for the Returns

In this subsection, we present the results of the two nonlinearity tests. From Table 3, the results corresponding to the BDS and Tsay tests show that there is evidence of nonlinearity in both series. Furthermore, the Zivot-Andrew and the Perron breakpoint tests in Figure 3 and Figure 4 showed evidence of one structural break. The identified break dates are 30th December, 2020 for Zivot-Andrew test, 6th December, 2020 for Perron test each for the total COVID cases and for the total COVID deaths, 31st March, 2021 and 5th August, 2020 were the identified break dates from the Zivot-Andrew and Perron tests respectively.

Table 3. Nonlinearity test results

Test series	**BDS**	**Tsay**
Case	89.4878(0.0000)	63.96(0.0000)
Death	77.3279(0.0000)	1.331(0.2492)

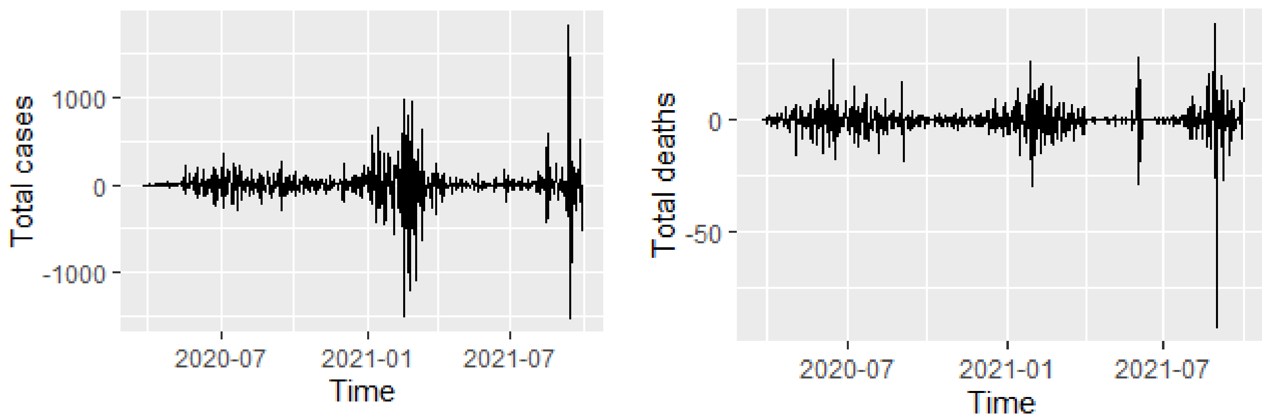


Figure 2. Time series plot of the 2nd difference of the two series

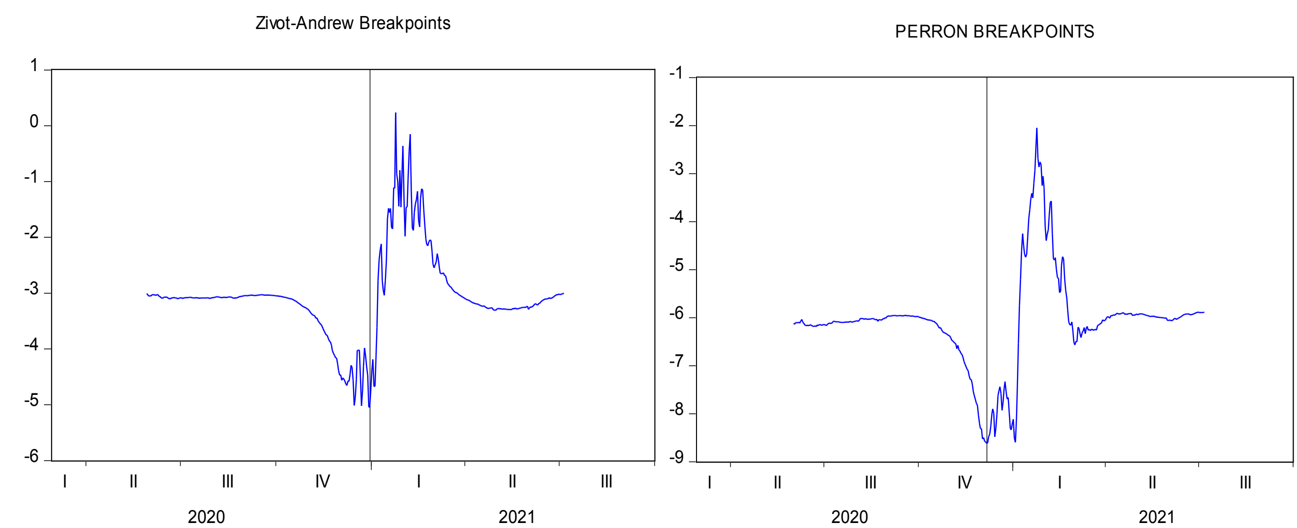


Figure 3. Plot based on Zivot-Andrew and Perron test for the Total COVID-19 Cases

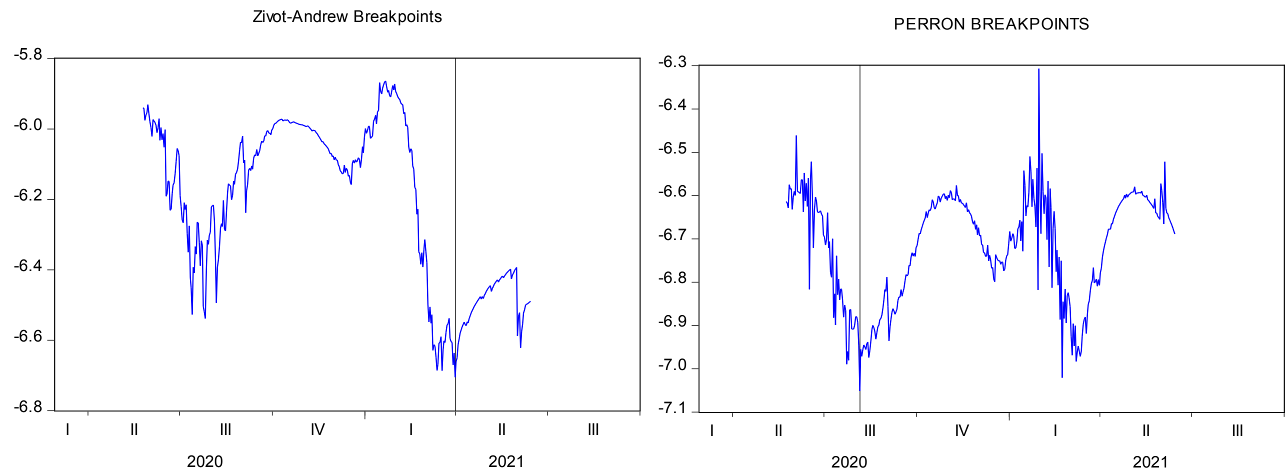


Figure 4. Plot based on Zivot-Andrew and Perron test for the Total COVID-19 Deaths

3.3. Model Estimation

The possible number of Covid-19 waves to be considered in the MSM model is determined on the basis of recursive local fitting as suggested in Gharleghi *et al.* (2014). The plot of the recursive residuals about the zero point, plus and minus two standard errors is shown in Figure 5.

The two figures below show that there is a significant wave change on the two occasions. Hence we proceed to fit a MSM(2)-AR(1) model for both the total infected cases and total death cases. The results of the fitted MSM(2)-AR(1) model is shown in the table below.

Table 4. The MSM(2)-AR(1) results for the Total COVID-19 Infected Cases

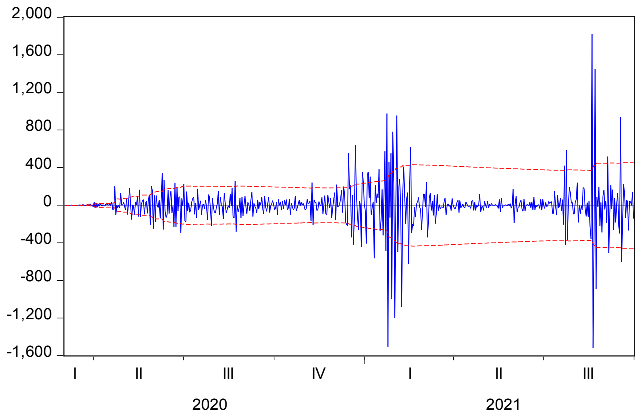
Total Cases	Estimates (Std. Error)		t-statistics (p-value)
$\mu(s_t = 1)$	297.5222 (38.93363)		7.6418 (0.0000)***
$\mu(s_t = 2)$	118.9386 (24.86355)		4.7837 (0.0000)***
$\mu_1(s_t = 1)$	1.008035 (0.000468)		2153.294 (0.0000)***
$\mu_1(s_t = 2)$	1.000817 (0.000176)		5685.382 (0.0000)***
$\sigma^2(s_t = 1)$	5.367854		
$\sigma^2(s_t = 2)$	0.031533		
AIC	13.78179		
LogLik	-3989.719		
p_{ij}	$s_t = 1$	$s_t = 2$	$E(D_{st})$
$s_t = 1$	2.643845	0.376484	15.06719
$s_t = 2$	-3.933099	0.386669	52.06500

The table 4 above shows the estimated parameters of the infected cases. It shows that the expected number of cases for the 1st wave rises by 297% daily, while that of the 2nd wave rises by 118% daily. Also, the variance in 1st wave is higher than that of the 2nd wave which implies more volatile during the 1st wave than during the 2nd wave. This result high volatility of 1st wave can be attributed to the period of the first total lockdown while the low volatility of 2nd wave can be attributed to the period just before the end of the partial lockdown. Also, the expected duration for the 1st wave was about 15 days while in the 2nd wave it was about 52 days.

Table 5. The MSM(2)-AR(1) results for the Total COVID-19 Deaths

Total Deaths	Estimates (Std. Error)		t-statistics (p-value)
$\mu(s_t = 1)$	7.101443 (0.848436)		8.3700 (0.0000)***
$\mu(s_t = 2)$	1.737148 (0.805501)		2.1566 (0.0310)***
$\mu_1(s_t = 1)$	1.002741 (0.000772)		1298.982 (0.0000)***
$\mu_1(s_t = 2)$	1.000179 (0.000496)		2017.380 (0.0000)***
$\sigma^2(s_t = 1)$	1.7545		
$\sigma^2(s_t = 2)$	0.0321		
AIC	6.461789		
LogLik	-1789.377		
p_{ij}	$s_t = 1$	$s_t = 2$	$E(D_{st})$
$s_t = 1$	3.817865	0.563658	46.50694
$s_t = 2$	-4.475754	0.579185	88.86086

From table 5 the estimated parameters of the 1st wave shows that the expected number of death rises by 7.1% daily while it rises by 1.7% during the 2nd wave. Also, the variance in 1st wave is higher than that of the 2nd wave which implies that the total COVID-19 deaths is more volatile during the 1st wave than during the 2nd wave. This high volatility of the 1st wave can be attributed to the period of the first total lockdown while the low volatility of the 2nd wave can be attributed to the period just before the end of the partial lockdown. Also, the expected duration of 1st wave was lower than that of the 2nd wave, with about 46 days expected duration while for 2nd wave its about 88 days.



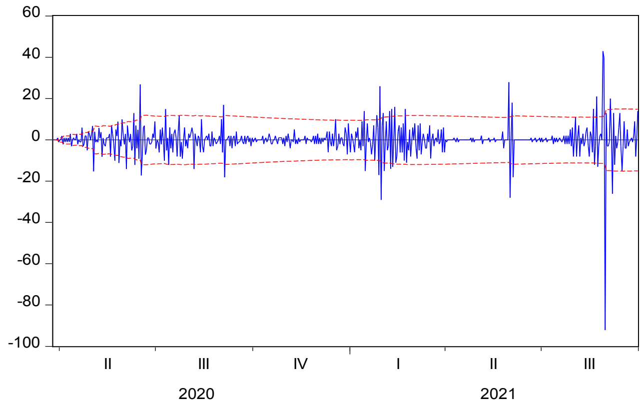


Figure 5. Recursive Residual Plot for total infected cases and total death cases

The expected number of COVID-19 infected cases and Death cases matrix is given below;

$$p= \begin{matrix} & infected & Deaths \\ infected & 15.067 & 52.065 \\ Deaths & 46.507 & 88.861 \end{matrix}$$

And the transition probability matrix between the Expected numbers of COVID-19 infected

Cases and Death cases, is;

$$p= \begin{matrix} & infected & Deaths \\ infected & 0.224 & 0.776 \\ Deaths & 0.343 & 0.654 \end{matrix}$$

This transition probability matrix can be used to predict the first, second, third periods etc. between the two waves for the number of COVID-19 infected and death cases.

4. Conclusions

This paper examined the nonlinear modelling of COVID-19 daily infected cases and deaths in Nigeria using the MS-AR model. The model estimates that the expected number of infected cases rises by 297% daily during the first wave of COVID-19 while it rises by 118% daily during the second wave of the pandemic. Also, the volatility of the two waves show that the COVID-19 infected cases in the first wave was more volatile than that of the second wave of the pandemic. The transition probabilities showed that in the first wave, the COVID-19 infected cases expected duration is around 15 days and 52 days in the second wave. For the COVID-19 deaths, the model estimates that the expected number of infected cases rises by 7.1% daily during the first wave of COVID-19 while it rises by 1.7% daily during the second wave of the pandemic. Also, the volatility of the two waves show that the COVID-19 infected cases in the first wave was more volatile than that in the second wave of the pandemic.

REFERENCES

[1] Adedeji, A. N., Ahmed, F. F., & Adam, S. U. (2021). Examining the dynamic effect of COVID-19 pandemic on dwindling oil prices using structural vector autoregressive model. Energy, 230, 120813.

[2] Brocks, C. (1996). Testing for non-linearity in daily sterling exchange rates. Applied Financial Economics, Vol 6, pp 307-317.

[3] Brock, W., Dechert, D., Sheinkman, J., and LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), 197-235.

[4] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357-384.

[5] Kartikasari, P., Yasin, H., & Di Asih, I. M. (2020). ARFIMA Model for Short Term Forecasting of New Death Cases COVID-19. In E3S Web of Conferences (Vol. 202, p. 13007). EDP Sciences.

[6] Keenan, D. (1985). A tukey nonadditivity-type test for time series. Biometrika 72, 39–44.

[7] Lahmiri, S., & Bekiros, S. (2021). The effect of COVID-19 on long memory in returns and volatility of cryptocurrency and stock markets. Chaos, Solitons & Fractals, 151, 111221.

[8] Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K. H. (2020). Time series modeling to forecast the confirmed and recovered cases of COVID-19. Travel medicine and infectious disease, 37, 101742.

[9] Oluwasegun A. A., Seno, A. and Omorogbe, J. A. (2020). Markov Regime-Switching Autoregressive Model of Stock Market Returns in Nigeria. CBN Journal of Applied Statistics Vol. 11, No. 2, pp 65-83.

[10] Ogundokun, R. O., Lukman, A. F., Kibria, G. B., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modeling of COVID-19 confirmed cases in Nigeria. Infectious Disease Modelling, 5, 543-548.

[11] Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361-401.

[12] Roy, S., Bhunia, G. S., & Shit, P. K. (2021). Spatial prediction of COVID-19 epidemic using ARIMA techniques in India. Modeling earth systems and environment, 7(2), 1385-1391.

[13] Tatar, M., Habibdoust, A., & Wilson, F. A. (2021). Analysis of excess deaths during the COVID-19 pandemic in the state of Florida. American journal of public health, 111(4), 704-707.

[14] Tsay, R. S. (1989). Testing and modelling threshold autoregressive processes. *Journal of the Royal Statistical Society B*, 84, 231-240.

[15] Zhao, Z., Li, X., Liu, F., Zhu, G., Ma, C., & Wang, L. (2020). Prediction of the COVID-19 spread in African countries and implications for prevention and control: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal, and Kenya. Science of the Total Environment, 729, 138959.

[16] Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics*, 20(1), 25-44.