International Journal of Data Science and Analysis
2022; 8(6): 182-186
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20220806.12
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

Self-Exciting Threshold Autoregressive Modelling of COVID-19 Confirmed Daily Cases in Nigeria

Nwakuya Maureen Tobechukwu[1, *], Biu Oyinebifun Emmanuel[1], Benson Tina Ibienebaka[2]

[1]Department of Mathematics and Statistics, University of Port Harcourt, Choba, Nigeria

[2]Department of Mathematics Statistics, Ignatius Ajuru University of Education, Obukuru, Nigeria

Email address:

maureennwakuya@uniport.edu.ng (Nwakuya Maureen Tobechukwu)

[*]Corresponding author

To cite this article:

Nwakuya Maureen Tobechukwu, Biu Oyinebifun Emmanuel, Benson Tina Ibienebaka. Self-Exciting Threshold Autoregressive Modelling of COVID-19 Confirmed Daily Cases in Nigeria. *International Journal of Data Science and Analysis*. Vol. 8, No. 6, 2022, pp. 182-186. doi: 10.11648/j.ijdsa.20220806.12

Received: November 1, 2022; **Accepted**: November 15, 2022; **Published**: November 23, 2022

Abstract: This article proposed the modelling of the daily COVID-19 confirmed cases in Nigeria using a Self-Exciting Threshold Autoregressive (SETAR) model. Coronavirus also known as Covid-19 first appeared in Wuhan in December 2019 and quickly spread across the world and became a major phenomenon confronting humanity today. Since the outbreak of COVID-19, several models have been introduced to study the virus and recommend appropriate policy direction to tackle the pandemic. Due to the nonlinear behavior of the series, the Self-Exciting Threshold Autoregressive (SETAR) model was adopted. The series was found to be nonstationary series which was differenced twice to achieve stationarity. The series exhibited nonlinearity with evidence of a structural break. A SETAR (2, 4, 1) model was identified as the most fitted model to the data. Furthermore, the identified SETAR nonlinear model was used to obtain a one month period forecasts for the daily confirmed COVID-19 cases. The forecast accuracy measure were used to verify that SETAR (2, 4, 1) was the best fitted model and forecast for the month of January 2023 was presented. The result also evidenced that the number of daily confirmed cases is expected to increase from 281,526 cases in year 2022 to 312,776 cases in year 2023.

Keywords: Self-Exciting Threshold Autoregressive (SETAR) Model, Structural Breaks, COVID-19, Nonstationary Series, Nonlinear Series

1. Introduction

The Coronavirus also known as Covid-19 first appeared in Wuhan in December 2019. It has quickly spread across the world and became a major phenomenon confronting humanity today. While the Coronavirus endangers peoples' lives and safety, it has also severely affected the economies of several countries. Because of the epidemic, many businesses have had to close, employment has become more difficult, and peoples' lives have been greatly negatively impacted. During the early stages of the outbreak, the number of cases doubled approximately every seven and a half days. The African continent confirmed its first case of COVID-19 in Egypt on 14th of February 2020 and that of Sub-Shara African, Nigeria, was first recorded on 27th February 2020 diagnosed from a 44-year-old Italian citizen in Lagos state [2]. According to the World Health Organization [16], COVID-19 situation report of 11th October 2021, there were 237,383,711 confirmed cases. One major use of modelling is for policy direction. Hence, lack of adequate models often lead to wrong policy formulations or directions by institutions or governments. Since the outbreak of COVID-19, several models have been introduced to study the virus and recommend appropriate policy direction to tackle the pandemic. The importance of the implementation of proposed statistical models as policy directions cannot be overemphasized. 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 actual nature and based on that determine the different models that can be associated with such datasets.

Then proceed to obtain the best model that fits the data for proper and adequate modelling and forecasting for optimal policy making, planning and control in the Nigeria.

In other to model nonlinear behavior of time series data, allowing the flow of the existence of different regimes with different dynamics for the different regimes is most natural. Some of the models that assume that the dynamic behavior of the time series in each regimen is obtained by an autoregressive (AR) model include; threshold AR (TAR), self-exciting threshold AR (SETAR) and smooth transition AR models (STAR). This paper implements the Self-Exciting Threshold Autoregressive (SETAR) in modeling COVID-19 daily confirmed cases.

1.1. Self-Exciting Threshold Autoregressive (SETAR) Model

Among the traditional types of nonlinear time series models with threshold experience is a Self-Exciting Threshold Autoregressive (SETAR) models. The SETAR model is an extension of autoregressive models that are applied in time series modelling to give a form of flexibility in the model parameters through regime-changing attitude. This model is inspired by several nonlinear traits that are frequently seen in practice such as imbalance in the downward and upward design of a time series process. It uses piecewise linear models to obtain a better approximation of the conditional mean equation. Unlike the traditional piecewise linear model, this model allows the change in models to occur in time and space by using threshold space to enhance the linear approximation. Relative to other nonlinear models SETAR model's evaluation and postulation, are simple and easy to understand.

1.2. The Structure of the SETAR Model

In situations where the threshold variable in the TAR modelling is obtained as the delayed point of the relevant series, the SETAR models are said to be valid. A SETAR $(d, p_i, i = 1, \dots, k)$ with k regimes is presented by Tong H. and Lim K. [12] as;

$$Y_t = \begin{cases} \beta_0^{(1)} + \sum_{j=1}^{p^{(1)}} \beta_j^{(1)} Y_{t-j} + e_t^{(1)} \; if \; Y_{t-d} < \vartheta_1 \\ \beta_0^{(2)} + \sum_{j=1}^{(2)} \beta_j^{(2)} Y_{t-j} + e_t^{(2)} \; if \; \vartheta_1 < Y_{t-d} \leq \vartheta_2 \\ \vdots \\ \beta_0^{(k)} + + \sum_{j=1}^{p^{(k)}} \beta_j^{(k)} Y_{t-j} + e_t^{(k)} \; if \; \vartheta_{k-1} \leq Y_{t-d} \end{cases} \quad (1)$$

where, k stands for number of regime of the SETAR model, $p^{(j)}$ is order of the autoregressive within jth regime, ∀ = 1, 2, …,; ϑ is the threshold value ($-\infty = \vartheta_0 < \vartheta_1 < \vartheta_2 \dots < \vartheta_{k-1} < \vartheta_k = \infty$); Y_{t-d} is the threshold variable that propels the regime changes. The d parameter is the delay parameter where $d < p$ given that d is a non-negative integer, $e_t^{(j)}$ is a white noise and $e_t^{(j)} \sim iid(0, \sigma_j^2$ given $\sigma_j^2 < \infty$. The superscripts in the model specify positions of the regime for each regime.

2. Literature Review

Tong H, and Yeung I. [13] implemented SETAR modelling to examine the data on the assets market price, Feng H. and Liu J. [6] in a study conducted on the GDP of Canada covering the period between 1965 and 2000 period used the SETAR model. Ayden D., Güneri and Öznur İşçi [1] did a comparative analysis on Export Volume Index and Domestic Producer Price Index Series in Turkey, and evaluated the performances of Hybrid AR, SETAR and ARM models. The result presented AAR-SETAR as the best model for the series in question. Other authors that have engaged SETAR models to describe several observed characteristics of time series data include; Tong H. [9] and Tong H. [10] and this can also be found in literature by Waiter L. and Richardson S. [15] on applications of epidemiology while Clements M. and Smith J. [4] applied it in modelling of water pollution. Tong H. [11] mentions numerous areas of application of the SETAR model, Tong H. [11] and de George J. [5] validated the model and the statistical properties. Nwakuya M. and Biu E. [7] modelled COVID-19 deaths using the autoregressive fractional integrated moving average (ARFIMA) with d parameter estimated using 4 different methods; Exact Maximum Likelihood (EML) Estimator and Non-Linear Least Square Estimator (NLS) and two semiparametric method; Geweke/Porter-Hudak estimator (GPH) and Smoothed Periodogram (SPERIO) and the EML was seen to produce the best d parameter value 0.499812 and ARFIMA (4, 0.499812, 2) was selected for forecast. The results showed that COVID-19 daily deaths are expected to decrease from 2,087 deaths in 2022 to 1,346 deaths in 2023. Benson et al [3] fitted Markov switching AR (MSM(2)-AR(1)) to COVID-19 daily cases and deaths. The result showed evidence of structural breaks with two significant wave changes for both series. 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.

3. Methodology

This research was conducted using Covid-19 daily confirmed cases in Nigeria from Jan 2020 to Sept. 2022. In applying the SETAR models it is very crucial to test for linearity and stationarity. In literature there are various linearity tests for non-linear time series, for this research we applied the TSAY test and the ADF test for stationarity. The structural break was identified by plots based on Zivot-Andrew and Perron test. A SETAR (p, d) model was obtained based on the Mean Square Error (MSE) and Akaike Information Criteria (AIC). The Ljung test was carried out and the correctly specified SETAR model was used for forecasting.

3.1. Augmented Dickey-Fuller (ADF) Test for Stationarity

The ADF test is among the group of tests known as Unit Root Test. The ADF test is appropriate for testing the stationarity of a time series. Unit root is an attribute of a time series that makes it non-stationary. Said S. and Dickey D. [8] augment the basic autoregressive unit root test to accommodate general ARMA (p, q) models with unknown orders and their test is referred to as the augmented DickeyFuller (ADF) test. The ADF test tests the null hypothesis that a time series y_t is I (1) (that is $\Theta = 1$) against the alternative that it is I (0), assuming that the dynamics in the data have an ARMA structure. The ADF is usually applied to the regression;

$$\Delta y_t = \alpha + \beta_t + \Theta y_{t-1} + \vartheta_1 \Delta y_{t-1} + \cdots + \vartheta_{p-1} \Delta y_{t-p+1} + e_t \quad (2)$$

Where alpha is a constant and beta is the coefficient of the time trend and p is the lag order of the autoregressive progress. The test statistics of the ADF has the following form;

$$ADF_t = \frac{\widehat{\Theta}}{SE(\Theta)} \quad (3)$$

Where $\widehat{\Theta}$ is the least square coefficient estimate of the lagged variable and $SE(\Theta)$is its standard error.

3.2. Tsay Test

Tsay test is used for testing nonlinearity. Tsay H. [14] test makes use of arranged autoregression and recursive estimation to derive an alternative test for threshold nonlinearity. Tsay's test for nonlinearities is performed by the following way using only the first and second-order regression terms.

(i) Regress Y_t on vector $[1, Y_{t-1}, \ldots, Y_{t-k}]$ and obtain the residual estimate $\hat{e}_t$. The regression model is then $Y_t = \mathrm{B}_t \rho + e_t$

Where $\mathrm{B}_t = [1, Y_{t-1}, \ldots, Y_{t-k}]$ is the vector consisting of the past values of Y and $\rho = \{\rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(k)}\}^T$ is the first-order autoregressive parameter vector, where k presents the order of the model and $n = [k+1, \ldots, sample\ size]$.

(ii) Regress the vector H_t on B_t and obtain the residual estimate vector $\hat{X}_t$. The regression model is; $\mathrm{H}_t = \mathrm{B}_t H + X_t$

Where H_t is a vector of length $(1/2)K(K+1)$. The transpose of $\mathrm{H}_t\, H_t'$ are obtained from the matrix $[Y_{t-1}, \ldots, Y_{t-k}]'[Y_{t-1}, \ldots, Y_{t-k}]$ by stacking the column elements on and below the main diagonal. The second-order regression parameter matrix is denoted by H, and n = [K + 1,..., sample size].

(iii) Regress e_t on $\hat{X}_t$ and obtain the error $\hat{e}_t$: given $\hat{e}_t = \hat{X}_t \beta + e_t$, where β is the regression parameter matrix of two residuals.

(iv) Let $\hat{F}$ be the F ratio of the mean square of regression to the mean square of error:

$$\hat{F} = \frac{(\sum \hat{X}_t \hat{e}_t)(\sum(\hat{X}_t)'(\hat{X}_t))^{-1}}{(1/2)K(K+1)\sum(\hat{e}_t)^2} \times \left(\sum(\hat{X}_t)'\,\hat{e}_t\right)\left(n - K - \frac{1}{2}K(K+1) - 1\right) \sim F_{\upsilon_1, \upsilon_2}$$

This is used to describe the value of rejection of the null hypothesis of linearity. This follows an F distribution with $\upsilon_1 = (1/2)K(K+1)$ and $\upsilon_2 = sample\ size - (1/2)K(K+3) - 1$.

3.3. Ljung and Box Test

The Ljung-Box test is a test used for checking for the presence of autocorrelation in a time series. It is a function of a piled up of autocorrelations of a series ρ_j, to any identified time lag k.

The test statistic is presented as;

$$LB = n(n+2)\sum_{j=1}^{k} \frac{\rho_j^2}{n-j} \sim \chi^2_{(k)} \quad (4)$$

Where: n is size of sample, k is length of lags and $\rho$is sample autocorrelation coefficient.

The hypothesis is as follows:

H_o: The residuals are independently distributed

H_1: The residuals are not independently distributed

4. Results

4.1. Time Plots of the Data

Here, we present the plots of time series data including the auto correlation and partial auto correlation function plots of the total COVID-19 Cases and Deaths in Nigeria.

Time Plots of the Differenced Series

Here, we present the plots of the 1st and 2nd differenced time series data including the auto correlation and partial auto correlation function plots for the 2nd differenced times series of the total COVID-19 Cases in Nigeria.

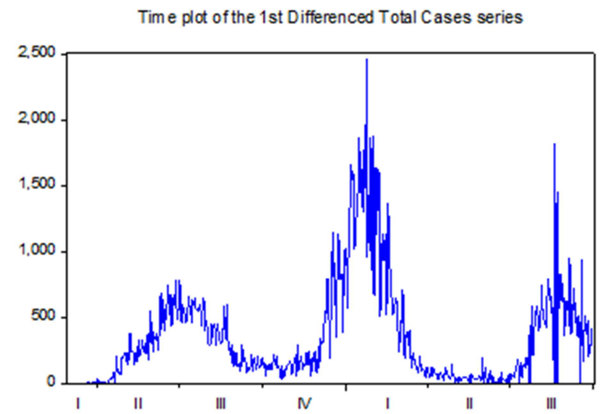

Figure 1. *Represents time plot of the 1st differenced total daily confirmed COVID-19 cases of Nigeria.*

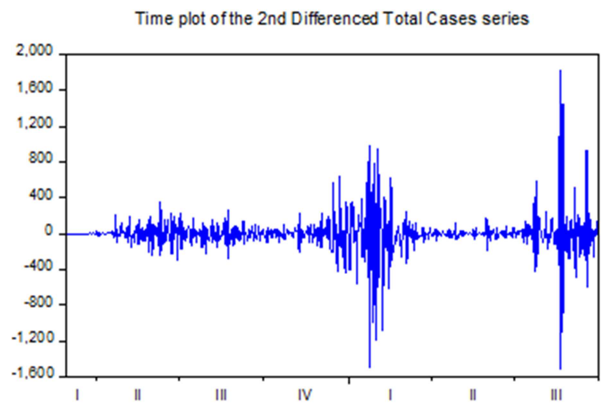

Figure 2. *Represents time plot of the 2nd differenced total daily confirmed COVID-19 cases of Nigeria.*

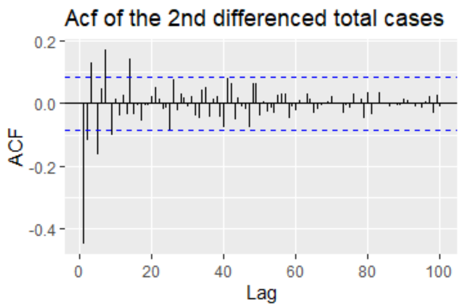

Figure 3. *Represents ACF of the 2nd differenced total daily confirmed COVID-19 cases of Nigeria.*

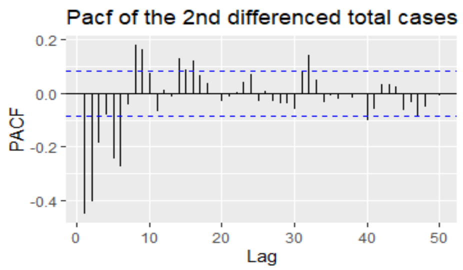

Figure 4. *Represent PACF of the 2nd differenced total daily confirmed COVID-19 cases of Nigeria.*

Figures 1 and 2 present the 1st and 2nd differenced series. The 1st differencing didn't achieve stationarity so a second differencing was done to achieve stationarity as seen in figure 2. While figure 2 and figure 3 presents the ACF and PACF plots with significant lags.

4.2. Unit Root Test

The ADF unit root tests were employed to check the total daily confirmed COVID-19 cases for stationarity.

Table 1. *The ADF Test of the Total daily confirmed COVID-19 Cases Series.*

Augmented Dickey-Fuller Test			
Series	**Level Stationary**		
	Level	**1st Diff**	**2nd Diff**
Total daily confirmed cases	-61102 (0.01)	-1.5381 (0.7696)	-8.1193 (0.01)**

Footnote:** means significant at 5%, p-values are in parentheses

The table above shows the ADF test for the series at levels; first difference, and second difference. The result shows significant of the 2nd differencing, that stationarity, agreeing with the plot in figure 2.

Table 2. *Nonlinearity Test Results*

Test Series	**Tsay statistic**
Total confirmed cases	63.99 (0.0000)**

Footnote: ** 5% significance, *** 10% significance

We proceed to fit a SETAR model since the Tsay test above shows evidence of threshold nonlinearity.

4.3. Model Fitting

Equations (5) shows the fitted SETAR models for the total COVID-19 Cases and deaths in Nigeria respectively.

$$\hat{x}_t = \begin{cases} \left.\begin{matrix} 7.6032 + 1.4027(Y_{t-1}) + 0.0693(Y_{t-2}) - 0.4102(Y_{t-3}) - 0.0613(Y_{t-4}) \\ std.E.(15.934892)(0.061933)(0.106979)(0.112301)(0.065674) \end{matrix}\right\} if\ Y_{t-1} < 127024 \\ \left.\begin{matrix} -18.4738 + 1.2899(Y_{t-1}) - 0.2458(Y_{t-2}) + 0.4894(Y_{t-3}) - 0.5332(Y_{t-4}) \\ std.E.(126.302750)(0.046576)(0.080726)(0.078968)(0.044759) \end{matrix}\right\} if\ Y_t \geq 127024 \end{cases} \tag{5}$$

The results show a threshold of 127024.

Table 3. *The Ljung-Box Diagnostic Test Statistic.*

Model	**Test Statistic**	**P-value**
SETAR (2,4,1)	2.5709	0.1988

This result in table 3 above shows that the model is correct, since we have an insignificant P-value.

4.4. Forecasting

Table 4. *Forecast Accuracy Measures for the SETAR Models.*

variables	**Models**	**RMSE**	**MAE**	**MAPE**
Total Cases	SETAR (2; 4, 1)	213.0368	111.2289	0.7386

Table 5. *Forecast for January 2023 using the SETAR Models.*

Dates	**Forecast values SETAR (2, 4, 1)**	**Low 95% C. I.**	**High 95% C. I.**
1/1/2023	300901	254588	347214
1/2/2023	301318	254289	348347
1/3/2023	301734	253985	349484
1/4/2023	302151	253678	350624
1/5/2023	302568	253367	351768
1/6/2023	302984	253053	352916
1/7/2023	303401	252735	354067
1/8/2023	303818	252413	355222
1/9/2023	304234	252088	356380
1/10/2023	304651	251760	357542
1/11/2023	305068	251428	358707
1/12/2023	305484	251092	359876
1/13/2023	305901	250753	361049
1/14/2023	306318	250411	362225
1/15/2023	306734	250065	363404
1/16/2023	307151	249715	364587
1/17/2023	307568	249362	365773
1/18/2023	307984	249006	366962
1/19/2023	308401	248646	368155
1/20/2023	308818	248283	369352
1/21/2023	309234	247917	370551
1/22/2023	309651	247547	371754
1/23/2023	310068	247174	372961
1/24/2023	310484	246798	374171
1/25/2023	310901	246418	375384
1/26/2023	311318	246035	376600
1/27/2023	311734	245649	377819
1/28/2023	312151	245259	379042
1/29/2023	312568	244867	380268
1/30/2023	312984	244471	381498
1/31/2023	313401	244071	382730

Table 6. *Expected Forecast Values for the Year 2022 and 2023, Using SETAR Models.*

Expected COVID-19	**Forecast values SETAR (2,4,1)**	**Low 95% C. I.**	**High 95% C. I.**
Expected COVID-19 for 2022	281526	262225	300827
Expected COVID-19 for 2023	312776	244215	381336

5. Conclusion

This research work modelled the COVID-19 confirmed cases in Nigeria from Feb. 2021 to Sept. 2022 employing the Self-Exciting Threshold Autoregressive model (SETAR), a non-linear time series. According to the findings, the time series plots for the COVID-19 confirmed cases showed evidence of non-stationarity, nonlinearity and structural changes. Stationarity was achieved after 2^{nd} differencing. Thereafter, the SETAR (2; 4, 1) for the total confirmed cases was identified and fitted to the series. Results showed that the threshold value was 127024 numbers of cases. This implies that the total COVID-19 cases reached 127,024 active cases before a structural break. The total expected forecast values for the year 2022 and 2023 was presented. In conclusion the study showed evidence of nonlinearity and threshold nonlinearity for the series which was then modelled using SETAR model. The series showed evidence of structural breaks. Furthermore, the SETAR models obtained showed that the total daily confirmed COVID-19 cases reached 127,024 active cases before entering the second regime or wave of the pandemic. The SETAR (2; 4, 1) model has obtained as the a good model for the forecast, a forecast using the model showed that the number of confirmed cases is expected to increase from 281,526 cases in year 2022 to 312,776 cases in year 2023.

References

[1] Aydın D., Güneri and Öznur İşçi (2015). Time Series Prediction Using Hybridization of AR, SETAR and ARM Models. International Journal of Applied Science and Technology, 5 (6).

[2] Ayinde, K., Lukman, A. F., Rauf, R. I., Alabi, O. O., Okon, C. E. & Ayinde, O. E. (2020). Modeling Nigerian Covid-19 cases: A comparative analysis of models and estimators. *Chaos, Solitons & Fractals*, *138 (2020)*, 109-119.

[3] Benson T. I., Biu E. O. and Nwakuya M. T. (2022). Markov Switching Autoregressive Modelling of the COVID-19 Daily Cases and Deaths in Nigeria. American Journal of Mathematics and Statistics, 12 (2); 15-21. DOI: 10.5923/j.ajms.20221202.01

[4] Clements, M. P. and Smith, J. (2001). The performance of alternative forecasting methods for SETAR models, International Journal of Forecasting 13; 463-75.

[5] de Goojier, J. G. (2001). On threshold moving-average models. Journal of Time Series Analysis, 19; 1–18. MR1624163.

[6] Feng, H., Liu, J. (2002). A SETAR Model for Canadian GDP: Non-linearities and Forecast Comparisons. University of Victoria, Working Paper EWP 0206, ISSN 1485-6441.

[7] Nwakuya M. T. and Biu E. O (2022). ARFIMA Modelling And Long Memory: The Case Of COVID19 Daily Deaths In Nigeria. International Journal of Innovative Mathematics, Statistics & Energy Policies 10 (4); 47-59.

[8] Said, S. E. and Dickey, D. A. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. Biometrika, 71; 599-607.

[9] Tong, H. (1978). On a threshold model, Pattern Recognition and Signal Processing, Amsterdam: Sijthoff & Noordhoff, 101-41.

[10] Tong, H. (1983). Threshold models in non-linear time series analysis, Lecture Notes in Statistics, No. 21, Heidelberg: Springer, 62-64.

[11] Tong, H. (1990). Non-linear Time Series: A Dynamical System Approach. Oxford: Oxford University Press.

[12] Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 42 (3), 245-292.

[13] Tong, H. & Yeung, I. (1991). On tests for Self-exciting Threshold Autoregressive-Type Non-linearity in Partially Observed Time Series. Applied Statistics, 40; 43-62.

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

[15] Watier, L. and Richardson, S. (1995). Modelling of an epidemiological time series by a threshold autoregressive model. The Statistician 44 (3); 353-364.

[16] WHO (2020). WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020. WHO. https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020.