



Modeling and Forecasting the Third wave of Covid-19 Incidence Rate in Nigeria Using Vector Autoregressive Model Approach

Gabriel O. Odekina, Adedayo F. Adedotun*, Ogbu F. Imaga

Department of Mathematics, Covenant University Ota, Ogun, Nigeria

Abstract

Modeling the onset of a pandemic is important for forming inferences and putting measures in place. In this study, we used the Vector autoregressive model to model and forecast the number of confirmed covid-19 cases and deaths in Nigeria, taking into account the relationship that exists between both multivariate variables. Before using the Vector Autoregressive model, a co-integration test was performed. An autocorrelation test and a heteroscedasticity test were also performed, and it was discovered that there is no autocorrelation at lags 3 and 4, as well as no heteroscedasticity. According to the findings of the study, the number of covid-19 cases and deaths is on the rise. To forecast the number of cases and deaths, a Vector Autoregressive model with lag 4 was used. The projection likewise shows a steady increase in the number of deaths over time, but a minor drop in the number of confirmed Covid-19 cases.

DOI:10.46481/jnsps.2021.431

Keywords: Covid-19, Vector Autoregressive model, Akaike Information Criterion, Final Prediction Error, Hannan Quinn Information Criterion

Article History :

Received: 07 October 2021

Received in revised form: 10 February 2022

Accepted for publication: 10 February 2022

Published: 28 February 2022

©2022 Journal of the Nigerian Society of Physical Sciences. All rights reserved.
Communicated by: T. Latunde

1. Introduction

Coronavirus disease has now been declared a global pandemic. The first case in Nigeria was discovered on February 27, 2020, and was confirmed at the Lagos State University Teaching Hospital's Virology Laboratory. In late December 2019, the coronavirus disease known as a form of severe acute respiratory syndrome was initially discovered in the city of Wuhan, China and it was acknowledged as a pandemic by [1] on the 11th of March 2020 after infecting over 118,000 people globally. This virus has been the main worry of doctors of medicine, community health experts, and researchers of all fields. Much international public/ community health inventiveness is being executed

and swift research of the biology of the virus and pathogenesis of the virus are being conducted in research institutes all around the globe.

The virus which has spread at an exponential rate all over the world has negatively affected the healthcare system in many countries. The covid-19 pandemic is one of the worst pandemics mankind has ever been confronted after the Spanish flu pandemic in 1918, which caused the deaths of about 50 million people at a time the world's population was around 2 billion. Economic and social interruption triggered by the pandemic is overwhelming. Disease prevention and control are eager for disease prediction guidance. Effective models for short-term forecasting have a pivotal role to develop strategic planning methods in the public health system. Under the guidance of the prediction model, we know the severity and the trends of

*Corresponding author tel. no: +2348055711272

Email address: adedayo.adedotun@covenantuniversity.edu.ng
(Adedayo F. Adedotun)

the pandemic under different strategies [2]. According to [3], those who would suffer in the medium-term as a result of actions made to prevent the spread of COVID-19 are the influence of the virus on the socioeconomic determinants of health, as well as its consequences on the next generation.

In Nigeria, the first known and confirmed case of Covid-19 was documented on the 27th of February 2020 in Lagos state according to [2]. After the index case on the 27th of February, the number of confirmed cases has been on the rise with the earliest reported death case on 22nd March 2020. Due to the speedy escalation in the number of cases in Nigeria, the Federal government had to enforce total lockdown in Lagos, Abuja, and Ogun state. Some states which were not included in the total lockdown by the federal government also had lockdown enforced by the State Government to curtail the rise of the deadly virus. With the outbreak of Covid-19, a lot of studies have been carried out in various science disciplines to either reduce the spread or control the increasing trend of the disease. Therefore, to manage and comprehend the epidemic, various approaches of estimation, modeling, and forecasting have been introduced.

Based on five deep learning approaches, [4] conducted a relative analysis to forecast the new number of covid-19 reported cases and retrieved instances. Long Short-Term Memory (LSTM), Gated recurrent units (GRUs), basic Recurrent Neural Network (RNN), Variational Auto-Encoder (VAE) algorithms, and Bidirectional LSTM (BiLSTM) algorithms were utilized for the global prediction of COVID-19 cases with a small amount of data. Their research is based on daily verified instances and the number of cases retrieved from six countries: China, Spain, Italy, Australia, the United States, and France. When the performance of each model was tested, it was discovered that VAE had a higher predicting precision than the other models.

Forecasting the Coronavirus (Covid-19) cases and deaths, [5] proposed the approach of statistical time series to model and forecast the short period behavior of Covid-19. They assumed a trend that is multiplicative which aims to capture the persistence of the two variables predicted (number of cases and mortality rate) as well as their uncertainty. The anticipated time series model showed an excellent level of precision and ambiguity as additional data were collected. In a study by [6], the effect of total lockdown on Covid-19 prevalence rate and death rate in China was investigated, and it was concluded that lockdown is effective in lowering the incidence rate and mortality rate.

The widespread increase in covid-19 and death as a result of corona-virus infection has been predicted by [7]. They also looked at a time series model that was used to forecast the number of confirmed and recovered coronavirus cases. The error distributions were carefully designed as a two-member scale combination of classical (TP-SMN) models, with the best match carefully selected. The chosen model was used to forecast the global number of diseases and fatalities caused by covid-19. The study [8] looked at epidemic data and statistics with a focus on Covid-19 and found that lockdown measures in Italy, Spain, and China, as well as the closure of firms in Hubei that provided non-essential services, were beneficial approaches.

In the study carried out on the coronavirus (covid-19) in

Spain and Italy by [9], two simple mathematical epidemiological models were applied where it was observed that the log-linear regression yielded an improved result and basic estimate of the everyday incidence for both countries. [10] studied the gender-based covid-19 prevalence rate and death rate in Nigeria. In the study, a Wilcoxon signed-rank test was adopted to examine disparity in the sex distributions of the daily prevalence. In the work of [11], the autoregressive integrated moving average was adopted to forecast the covid-19 incidence rate in India where an increasing tendency in the number of coronavirus cases was observed.

The Vector autoregressive model and Co-integrated Vector Autoregressive models are time series models used for multivariate time series data set. A lot of research has been carried out through the adoption of this model which takes into account the linear dependence that exists among the variables. For example, [12] adopted the Co-integrated Vector Autoregressive model in modeling Wind speed along with some selected meteorological variables.

Using the (VAR) Model, a time-series analysis was proposed to investigate the impact of environmental pollution on mortality in Nigeria. The data set passes the stationarity test, indicating that the data is steady and that the VAR model would fit well. Furthermore, environmental pollution has a considerable impact on mortality in Nigeria, according to the study [13].

A study conducted in the United State by [14] used a VAR model to predict the covid-19 prevalence rate in the U.S. The result of the research stated that the situation of the pandemic will get shoddier if there is no active control.

Vector Autoregressive Integrated Moving Average analysis (VARIMA) to establish the relationship between the number of deaths due to covid-19 and the number of new cases of covid-19 in the country. The AICC was used to select the best model after it fulfilled all the assumptions [15].

Modeling the outbreak of a pandemic is pertinent for inference-making and implementation of policies. In this paper, we adopted the Vector autoregressive model in modeling and forecasting the number of covid-19 cases and deaths in Nigeria.

In the next section (Section 2), we describe the data we used in the analyses and the VAR model and analysis plan for the research. The Results section (Section 3) provides the prediction results by VAR modeling and an internal validation/evaluation of the model. Section 4 discusses the model performance, further improvement, and comparison with other models.

2. Material and Methods

2.1. Data source

The data used for this study is daily data on the number of Covid-19 cases and death obtained from <https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.xlsx>. The methodology used for this paper is given below;

2.2. Vector Autoregressive Model

Consider a k-dimensional vector autoregressive model of order 2 given below, where $a_t \sim N(0, \Sigma)$, Φ_1 and Φ_2 are $k \times k$

matrices of unknown coefficients. Let $S_t(t)$ denote the price of the risk free asset at time t and the model is given as follows

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + a_t \tag{1}$$

Subtracting y_{t-1} from both sides of equation (1), and adding $\Phi_2 y_{t-1}$ to the right-hand side of equation (1), we have

$$\Delta y_t = \Phi y_{t-1} + \Gamma \Delta y_{t-1} + a_t, \quad t = 1, 2, \dots, T \tag{2}$$

where $\Phi = (\Phi_1 + \Phi_2 - I_{k \times k})$. Equation (2) is referred to as the vector error correction model, otherwise called the co-integrated vector autoregressive model. Suppose we have a multivariate variable consisting of variables P and Q, according to [16], equation (2) can be written in matrix form as;

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} \Delta P_{t-1} \\ \Delta Q_{t-1} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

In the formulation of the vector error correction model in equation (2), there are three cases of interest to be considered which are;

1. Rank(Φ) = 0 implies y_t is not cointegrated and the vector error correction model in equation (2) reduces to a vector autoregressive model in (1)
2. Rank(Φ) = k , then y_t contains no unit root. That is, y_t is stationary and $I(0)$, where k is the total number of variables.
3. $0 < \text{rank}(\Phi) = m < k$, then there is at least one stationary linear combination of the variables which is the co-integrating relation. In this case, $(\Phi) = \alpha\beta^1$, where α = vector of adjustment coefficients and β = co-integrating vector.

For equation (2) to hold, Φ should be of a reduced rank [17].

2.3. Augmented Dickey Fuller test (ADF)

The Augmented Dickey fuller test statistics were used to test the stationarity of the data. Considering the hypothesis; $H_0: \theta_1 = 1$ Vs $H_0: \theta_1 < 1$, The ADF test statistic is given as;

$$ADF = \frac{\widehat{\beta} - 1}{Std(\widehat{\beta})} \tag{3}$$

2.4. Akaike Information Criterion

The AIC for selecting the underlying vector autoregressive (p) is given as

$$AIC = \ln \left(\left| \widehat{\Omega}_i \right| \right) + \frac{2p^2 i}{n} \tag{4}$$

Where n is the size of the sample, p is the parameter numbers,

$$\widehat{\Omega} = \frac{1}{n} \sum_{n=i+1}^n \widehat{a}_i' (\widehat{a}_i) \tag{5}$$

a_t is the error term

2.5. Final Prediction Error (FPE)

The FPE for selecting the underlying vector autoregressive (p) is given as

$$FPE = det \left(\frac{1}{N} \sum_1^N e(t, \widehat{\theta}_N) (e(t, \widehat{\theta}_N))^T \right) \left(\frac{1 + d/N}{1 - d/N} \right) \tag{6}$$

Where N is the number of values in the estimation data set, $e(t)$ is the n by 1 vector of prediction errors, θ_N represents the estimated parameters, d is the number of estimated parameters.

2.6. Hannan-Quinn Information Criterion (HQIC)

The HQIC for selecting the underlying vector autoregressive (p) is given as

$$HQIC = -2L_{max} + 2k \ln(\ln(n)) \tag{7}$$

where L_{max} the log-likelihood, k is the number of parameters, and n is the number of observations

2.7. Lagrange Multiplier Statistic

The Autoregressive Conditional Heteroscedastic (ARCH) Lagrange Multiplier (LM) is used to test the hypotheses of homoscedasticity. This involves regressing the squared residuals on the conditional mean equation which may be an autoregressive, or moving average model. For example, considering an ARMA (1,1) process,

$$\begin{aligned} r_t &= \phi r_{t-1} + \phi e_{t-1} + e_t, \\ e_t^2 &= \beta_0 + \beta_1 e_{t-1}^2 + \beta_2 e_{t-2}^2 + \dots + \beta_q e_{t-q}^2 \end{aligned} \tag{8}$$

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots \beta_q = 0 \text{ vs } H_1 : \beta_i > 0$$

$$LM_E = TR^2 \tag{9}$$

Where T = Sample size, R^2 = R squared

2.8. Residual Autocorrelation

The Box-Pierce statistic which was proposed by [18] will be used to test the autocorrelation in the residuals. H_0 : No autocorrelation up to order k vs H_1 : There is autocorrelation up to order k . The statistic for the test is given as

$$Q = n \sum_{j=1}^k r_j^2 \tag{10}$$

Where n = Sample size, r = Autocorrelation at lag j

2.9. Normality Test

The Jarque-Berra test was used to decide if the error correction model is Gaussian distributed. The test is used to measure the discrepancy in Skewness and Kurtosis of a variable compared to those of the Gaussian distributions.

H_0 : The variable is distributed normally vs H_1 : The variable is not distributed normally.

$$JB = \frac{M - p}{6} \left[S^2 + \frac{(L - 3)^2}{4} \right], \tag{11}$$

where

M= number of observations,

p=Number of estimated parameters,

S= Skewness,

L=Kurtosis.

The condition is to reject the null hypotheses if the p-value \leq level of significance.

3. Data Analysis and Interpretation

3.1. Time plot

The foremost step during the study of the data is to generate the time plot of the variables. Figure 1 is a time plot of the amount of Covid-19 confirmed cases in Nigeria and the number of Covid-19 related deaths in Nigeria. The graph shows an increasing inclination (trend) in the number of confirmed covid-19 cases and deaths in the country with the first wave occurring early in the year 2020, the second wave in the late year 2020, and the third wave which is the greatest occurring in the early year 2021. There seems to be no decrease in the pandemic

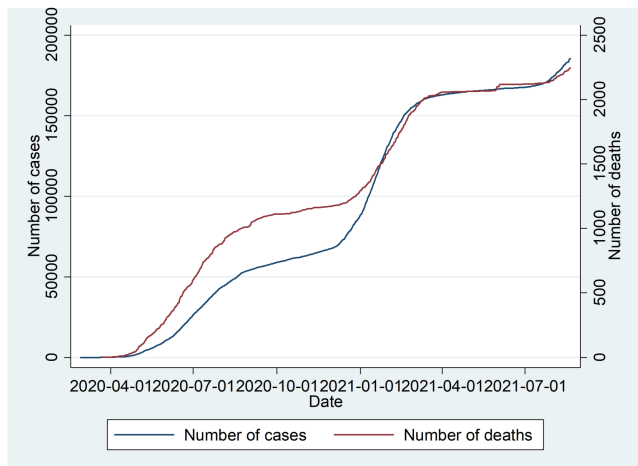


Figure 1: Time plot of Covid-19 reported cases and deaths

3.2. Descriptive Statistics

Table 1 gives a descriptive summary of statistics of the number of confirmed covid-19 cases and covid-19 deaths. It is observed that there is a high level of variation in the data obtained. The difference between the three measures of central tendencies

(mean, median, and mode) in the number of confirmed covid-19 cases shows a departure from normality, the same with the number of covid-19 deaths. This departure from normality can be ignored considering the reasonably large sample size.

Table 1: ADF Descriptive statistics

	Number of cases	Number of deaths
Mean	88164.28	1294.141
Median	66523.00	1180.00
Standard Deviation	64682.12	706.9598
Variance	4183776688	499792.141
Range	185570	2246

3.3. Stationarity test

From table 2, it is seen that the null hypothesis of non-stationarity could not be rejected for the number of confirmed Covid-19 cases as our p-value is greater than our level of significance 0.05, hence the need for differencing. H_0 : The series is non-stationary vs H_1 : The series is stationary

Table 2: ADF Stationarity test

	Test Statistic	1% critical value	5% critical value
Z(t)	1.371	-3.430	-2.860

MacKinnon approximate p-value for Z(t) = 0.9970 From table 3, it is seen that the null hypothesis of non-stationarity was rejected after taking the third difference as our p-value is less than our level of significance 0.05. The series is stationary after the first differencing. H_0 : The series is non-stationary vs H_1 : The series is stationary

Table 3: ADF Stationarity test

	Test Statistic	1% critical value	5% critical value
Z(t)	-6.469	-3.430	-2.860

MacKinnon approximate p-value for Z(t) = 0.0000 From table 4, it is seen that the null hypothesis of non-stationarity was rejected for the number of Covid-19 deaths as our p-value of 0.0002 is less than our level of significance 0.05 which means the series is stationary. H_0 : The series is non-stationary vs H_1 : The series is stationary

MacKinnon approximate p-value for Z (t) = 0.0002

3.4. Lag selection and model estimation

A lag of order 4 was selected based on the following information criterion used where it is observed that the minimum occurs on the fourth lag.

Using equation 2 on the Covid-19, the model for the number of deaths and number of cases are estimated and presented in equations 12 and 13 respectively with the model summary presented in tables 6 and 7. From table 6, it is evident that for every increase in the number of deaths, there is a 0.0012504, 0.0024392 and 0.0011357 upsurge in the number of reported

Table 4: ADF Stationarity test

	Test Statistic	1% critical value	5% critical value
Z(t)	-4.559	-3.430	-2.860

Table 5: Akaike Information Criterion

Lag order	AIC	FPE	HQIC
1	19.3886	902429	30.7735
2	19.1106	683387	19.4081
3	18.9981	610694	19.143
4	18.9129*	560827*	18.9713*

covid-19 cases at lag 2, 3, and 4 respectively with a -0.0008647 decrease in the number of covid-19 cases at lag 1 with their respective confidence intervals. Table 7 also gives the vector autoregressive model for the number of covid-19 cases and it is also evident that for every increase in the number of cases, there is a 2.924165 and 1.410195 upsurge in the number of deaths at lag 1 and lag 3 respectively with a -4.285465 and -0.280125 decrease at lag 3 and lag 4 respectively.

Table 8 gives the normality test of the disturbances as not being normally distributed. However, with practically huge sample sizes, the contravention of the Gaussian hypothesis ought not to cause any setback [19]

$$\begin{aligned} \text{Number of deaths} = & -0.0008647y_{t-1} + 0.0024392y_{t-2} \\ & + 0.0011357y_{t-3} + 0.0012504y_{t-4} \end{aligned} \quad (12)$$

$$\begin{aligned} \text{Number of cases} = & 2.924165y_{t-1} - 4.285465y_{t-2} \\ & + 1.410195y_{t-3} - 0.280125y_{t-4} \end{aligned} \quad (13)$$

3.5. Lagrange Multiplier Test for Autoregressive Conditional Heteroscedasticity (ARCH) and Autocorrelation

Table 9 reports the heteroscedasticity test of the number of confirmed Covid-19 cases and deaths in Nigeria. The test result shows that the null hypothesis of no ARCH effect was not rejected. That is, there is no ARCH effect. Table 10 gives the Box-Pierce statistic for the autocorrelation test of the 4 lags which shows no autocorrelation is depicted across the four lags.

$$H_0 : \text{no ARCH effects} \vee s H_1 : \text{ARCH}(p) \text{ disturbance}$$

$$H_0 : \text{no autocorrelation at lag order}$$

3.6. Forecast Precision

Figure 2 shows a forecast of the number of covid-19 cases and deaths for the next one year (365 days). From the graph, it can be seen that there is a sharp rise in the Covid-19 mortality rate alongside a slight decrease in the number of confirmed cases with 95% confidence interval. This shows the future expectation of the current pandemic

Table 6: Vector Autoregressive Model

Death	Coefficient	P-Value	Standard error	95% confidence interval
Number of cases				
L1	-.0008647	0.393	.0010122	-.0028487 .0011192
L2	0.0024392	0.033	.0011437	.0001977 .0046808
L3	0.0011357	0.334	.0011767	-.0011705 .003442
L4	0.0012504	0.258	.001105	-.0009153 .0034161
Constant	1.888731	0.000		

Table 7: Vector Autoregressive Model

Number of cases	Coefficient	P-Value	Standard error	95% confidence interval
Number of deaths				
L1	2.924165	0.131	1.935567	-0.8694764 6.717806
L2	-4.285465	0.141	2.913769	-9.996347 1.425417
L3	1.410195	0.630	2.930209	-4.332909 7.153298
L4	-0.280125	0.988	1.91525	-3.781833 3.3725808
Constant	20.04022	0.457		

Table 8: Jarque-Berra test

Variables	Chi-square	df	Prob>Chi2
Number of cases	5102.057	2	0.00000
Number of deaths	1381.003	2	0.00000
All	6483.060	4	0.00000

4. Conclusion

The surge of covid-19 has crippled the health care system in Nigeria and other parts of the world. The need to model and study the incidence rate of covid-9 cannot be overemphasized as it is pertinent for concrete decision-making. There has been a sharp rise in the figures of covid-19 cases and death as depicted by the time plot. The time plot of the number of cases and deaths shows an “S” shape which indicates the increase in the pandemic. A Vector Autoregressive model of lag 4 was adopted which was used to make a forecast on the number of cases and death. Moreover, an autocorrelation test and a test of heteroscedasticity were carried out where it was observed that there exists no autocorrelation at lag 3 and lag 4 and there exists no heteroscedasticity. A Jarque-Berra test of normality of the disturbances was done on the Vector Autoregressive model

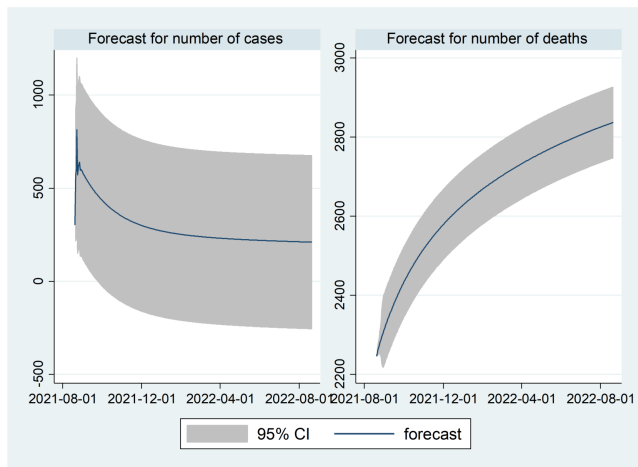


Figure 2: Forecast of the number of Covid-19 reported cases

Table 9: Testing for ARCH effect

Lags (p)	Chi2	Prob> Chi2
4	1.035	0.3089

Table 10: Box-Pierce statistic test of autocorrelation

Lag	Chi2	Prob>Chi2
1	12.2810	0.31327
2	42.5130	0.08523
3	86.0806	0.42145
4	53.4515	0.21436

which indicates a departure from normality. However, this result can be ignored for a reasonably large sample size of at least 30 according to [19]. The forecast also reveals an upward trend in the number of deaths with a slight decrease in the number of infections. This is an indication that in the future, the spread may be reduced however, there will be a high mortality rate resulting from this pandemic. Though the cause of death can also be attributed to some other factors such as age or other underlying ailments the patient may have which could have been triggered by Covid-19. This can be an area of further research.

Concerning the above findings, the following recommendations have been made; (1) The government needs to put further measures in place to curtail the spread of the virus and aim towards flattening the curve, (2) an awareness should be created for individuals to be properly enlightened and get vaccinated, (3) the government should also render proper health care to infected individuals to reduce the mortality rate.

References

- [1] "World Health Organization."
- [2] "Nigeria Center for Disease Control." (2020).
- [3] K. Temitope, et al., "Gut microbial Diversity Is Reduced by the Probiotic VSL#3 and Correlates with Decreased TNBS-Induced Colitis. Inflammatory Bowel Diseases", **17** (2011) 289. <http://dx.doi.org/10.1002/ibd.21366>.
- [4] Z. Abdelhafid, H. Fouzi, D. Abdelkader, & S. Ying, "Deep learning methods for forecasting covid-19 time series data", (2020):www.elsevier.com/locate/chaos.
- [5] F. Petropoulos, S. Makridakis & N. Stylianou, "COVID-19: Forecasting confirmed cases and deaths with a simple time series model", International Journal of Forecasting (2020), <https://doi.org/10.1016/j.ijforecast.2020.11.010>.
- [6] M. F. Alexandre, D. C. Antonio, Daniela Cristina Moreira Marculino Figueiredo, Marc Saez, & Andrés Cabrera León, "Impact of lockdown on covid-19 incidence and mortality in China: An interrupted time series study", (2020). <http://dx.doi.org/10.2471/BLT.20.256701>.
- [7] M. A. Mohsen, R. M. B Mohammad, H. H. D Mohammad, & P. E Kim-Hung, "Time series modelling to forecast the confirmed and recovered cases of COVID-19", (2020):www.elsevier.com/locate/chaos.
- [8] D. A. Hoseinpour, M. Alizadeh, P. Derakhshan, P. Babazadeh, & A. Jahandideh, "Understanding epidemic data and statistics: A case study of COVID-19," J Med Virol. 2020;1–15. <https://doi.org/10.1002/jmv.25885>.
- [9] J. Chu, "A statistical analysis of the novel coronavirus (COVID-19) in Italy and Spain", PLoS ONE 16: e0249037 (2021). <https://doi.org/10.1371/journal.pone.0249037>.
- [10] O. O. Olusola-Makinde, & O. S Makinde, "COVID-19 incidence and mortality in Nigeria: gender based analysis", PeerJ 9:e10613 DOI 10.7717/peerj. (2021) 10613.
- [11] T. Hiteshi, R. Prabhat, C. Tanmoy, & S. Vandana, "Coronavirus (covid-19): ARIMA based time series analysis to forecast near future", (2020). <https://arxiv.org/abs/2004.07859>.
- [12] H. G Dikko , G.O Odekina & T. Dahiru, "Wind speed Modeling with some selected meteorological variables", *Journal of Nigerian Association of Mathematical Physics.* **43**(2017) 189.
- [13] A. F. Adedotun, O. G.Obadina, O. S, Adesina, K.O. Omosanya,& R.J. Dare, " Statistical analysis of the effect of environmental degradation on mortality rate: A Vector Autoregressive Model Approach", *International journal of Science and Engineering Invention*, (2018) 2455.
- [14] W. Qinan, Z. Yaomu & C. Xiaofei , "A Vector Autoregression Prediction Model for COVID-19 Outbreak", [Stat. AP](2021) arXiv: 2102.04843v1.
- [15] A. Meimela, S. S Lestaris, F. Mahdy, T. Toharudin & B. N. Ruchjana, "Modeling of covid-19 in Indonesia using vector autoregressive integrated moving average", Journal of Physics: Conference Series, **1722**.
- [16] R. S. Tsay, "Analysis of financial time series", John Wiley and Sons (2005).
- [17] S. Johansen, "Statistical analysis of cointegration vectors", Journal of Economic Dynamics and Control, **12**(1988) 231.
- [18] G. E. P. Box & D. A. Pierce, "Distribution of residual correlations in autoregressive-integrated moving average time series models", Journal of the American Statistical Association, **65**(1970) 1509.
- [19] A. Ghasemi & S. Zahediasl, "Normality tests for statistical analysis: a guide for non-statisticians", International journal of endocrinology and metabolism **10** (2012) 486.
- [20] H. Zhao & X. Rong, "Portfolio selection problem with multiple risky assets under the constant elasticity of variance model", Mathematics and Economics **50** (2012) 179.
- [21] X. Xiao, K. Yonggui & Kao, "The optimal investment strategy of a DC pension plan under deposit loan spread and the O-U process", (2020). Preprint submitted to Elsevier.
- [22] M. Jonsson & R. Sircar, "Optimal investment problems and volatility homogenization approximations", Modern Methods in Scientific Computing and Applications **75** (2002) 255.