Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2019, 6(9):120-129**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

A Time Series Model of HIV Prevalence in Nigeria

O. Abu, M.A. Emeje

Department of Mathematics and Statistics, Federal Polytechnic, Idah, Nigeria
Corresponding author's email address abuonoja2008@yahoo.com

Abstract HIV/AIDS has been a global dreadful infection/disease. Understanding the patterns of HIV spread over the years and projecting into the future is important. In this paper, we studied autoregressive integrated moving average (ARIMA) models with a view to fitting them to HIV prevalence data in Nigeria and forecasting into the near future. The preliminary analysis using time series, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots suggested ARIMA models: ARIMA (3,1,0), ARIMA (5,1,0) and ARIMA (7,1,0). The stochastic simulation of these models was carried out. Residual analysis using LBQ test showed that these models were adequate for forecasting. Further, model performance tests using mean absolute error (MAE) and mean square error (MSE) showed that ARIMA (7,1,0) was the best for forecasting. The forecast results show that there would be 1.5%, 0.97% and 0.71% prevalence in 2016, 2017 and 2018 respectively. The findings of this study show that HIV prevalence in Nigeria is expected to drop in the near future, and suggest that the Federal Government of Nigeria should be consistently implementing their current control programs.

Keywords HIV/AIDS, ARIMA models, stochastic simulation, residual analysis

1. Introduction

The human immune-deficiency virus (HIV) together with the associated acquired immune deficiency syndrome (AIDS) is still a global threat [1-2].

The basic routes of HIV transmission between persons are well understood. The major routes are sexual (heterosexual and homosexual) and mother-to-child [2-3].

Several intervention methods are available. These range from sex abstinence, use of condoms, education and use of antiretroviral drugs and counseling.

As pointed out in Williams *et al* [2], the development of antiretroviral drugs to treat HIV has been a singular scientific achievement. Between 1995 and 2009 an estimated 14.4 million life-years has been gained globally among adults on ART but the rate of new infections is unacceptably high and still exceeds the number of people starting ART each year.

As presented in casels *et al* [3], ART reduces viral load and the probability of transmission. It also reduces HIV/AIDS-related mortality and, therefore, increases the life expectancy of infected individuals.

Stochastic models of HIV have been proposed and studied by researchers. For example, Peterson *et al* [4] applied Monte-Carlo simulation technique in a population of intravenous drug users.

Greenhalgh and Hay [5] studied a mathematical model of the spread of HIV/AIDS among injecting drug users. Dalal *et al* [6] examined a stochastic model of AIDS and condom use. Dalal, *et al* [7] also studied a stochastic model for internal HIV dynamics. Ding *et al* [8] carried out risk analysis for AIDS control based on a stochastic model with treatment rate. Tuckwell and Le Corfec [9] studied a stochastic model for early HIV-1 population dynamics. Waema and Olowofeso [10] studied a mathematical model for HIV transmission using generating function approach.

In another development, ARIMA models were recently used to fit and forecast disease data. We do not try to be encyclopedic. Zhou et al [11] constructed a hybrid model to forecast the prevalence of schistosomiasis in humans in Yangxin County. Promprou *et al* [12] fitted ARIMA models to forecast dengue haemorrhagic fever cases in Southern Thailand. Trottier and Philippe [13] carried out a univariate time series analysis on pertusis, mumps, measles and rubella based on Box-Jenkins or Auto-Regressive Integrated Moving Average (ARIMA) model. Takyi *et al* [14] fitted an ARIMA model to malaria cases in Ejisu- Juaben Municipality. Jere and Moyo [15] fitted an ARIMA model to monthly malaria cases in Zambia from 2009 to 2013. ARIMA models for disease data can also be found in Malinga [16], Garett [17], Imran *et al* [18] and Jiang [19].

The plan of this paper is as follows. Introductory part is presented in section 1. The methodology is presented in section 2. Section 3 is devoted to presentation of results. Discussion and conclusive remarks are passed in sections 4 and 5 respectively.

2. Materials and Methods

Data Sources

The data used for this study was obtained from ANC report in 2015.

The ARIMA Model

In this study, we use autoregressive integrated moving average (ARIMA) model which is a generalization of an ARMA model for our analysis. These models can be fitted to time series data either to better investigate the data or to forecast future points in the series. The model is generally referred to as an ARIMA (p, d, q) model, where p , d and q are non-negative integers that refer to the order of the autoregressive, integrated and moving average components of the model respectively. ARIMA models constitute an important component of Box-Jenkins approach to time series modeling [20].

Given a time series of data $\{y_t\}$ where t is an integer index and the y_t are real numbers, then an $ARIMA\ (p, d, q)$ model is given by:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Or

$$(1 - \sum_{i=1}^{p} \alpha_i L^i) y_t = (1 + \sum_{i=1}^{q} \theta_i L^i) \varepsilon_t \qquad (1)$$

Here, L is the lag operator, the α_i are the parameters of the autoregressive component of the model, the θ_i are the parameters of the moving average component and ε_t are error terms which are assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean. Assuming that the polynomial

$(1 - \sum_{i=1}^{p} \alpha_i L^i)$ has a unitary root of multiplicity d, then it can be written as:

$$(1 - \sum_{i=1}^{p} \alpha_i L^i) = (1 + \sum_{i=1}^{p-d} \theta_i L^i)(1 - L)^d \qquad (2)$$

An $ARIMA\ (p, d, q)$ process expresses this polynomial factorization property and is given by:

$$(1 - \sum_{i=1}^{p} \theta_i L^i)(1 - L)^d y_t = (1 + \sum_{i=1}^{q} \theta_i L^i) \varepsilon_t \qquad (3)$$

ARIMA models are used for observable non-stationary processes y_t that have some clearly identifiable trends [20].

3. Results and Analysis

Descriptive Statistics

Mean	median	max	min	range	Std. Dev.	Skewness	kurtosis
4.2163	4.4500	5.8000	1.8000	4.0000	1.0345	-0.6264	2.5766

Graphical properties of the series

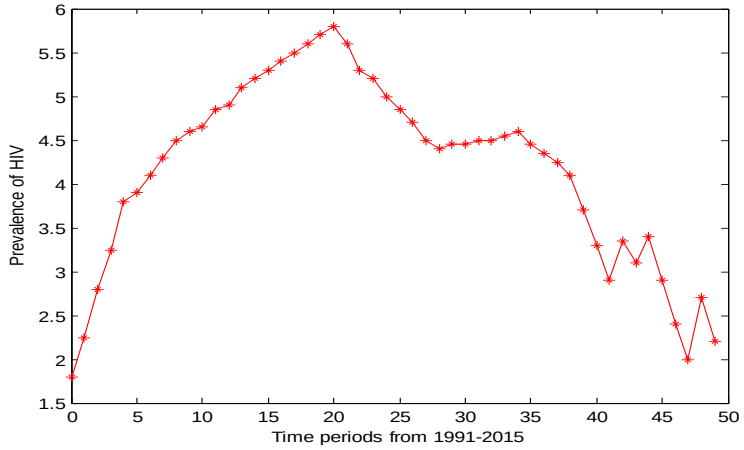

Figure 1: Time series plot of HIV prevalence in Nigeria

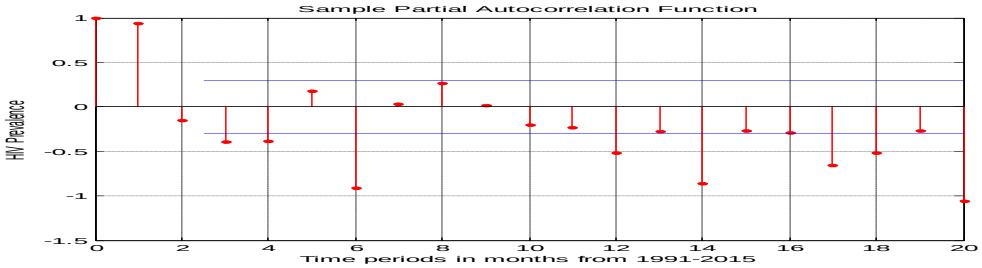

Figure 2: Autocorrelation function plot for HIV prevalence data

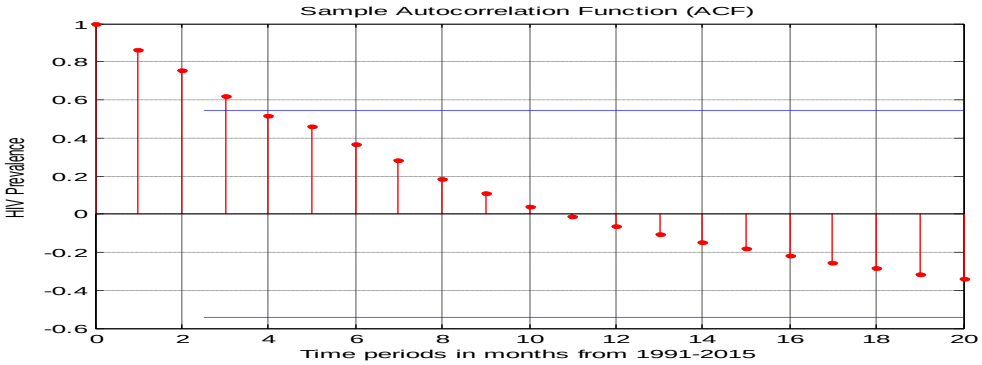

Figure 3: Partial autocorrelation function plot for HIV prevalence data

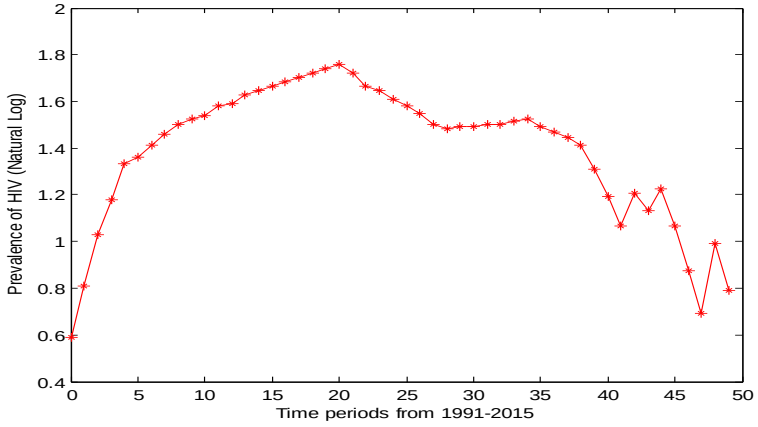

Figure 4: Time series plot of HIV prevalence in Nigeria (Natural log)

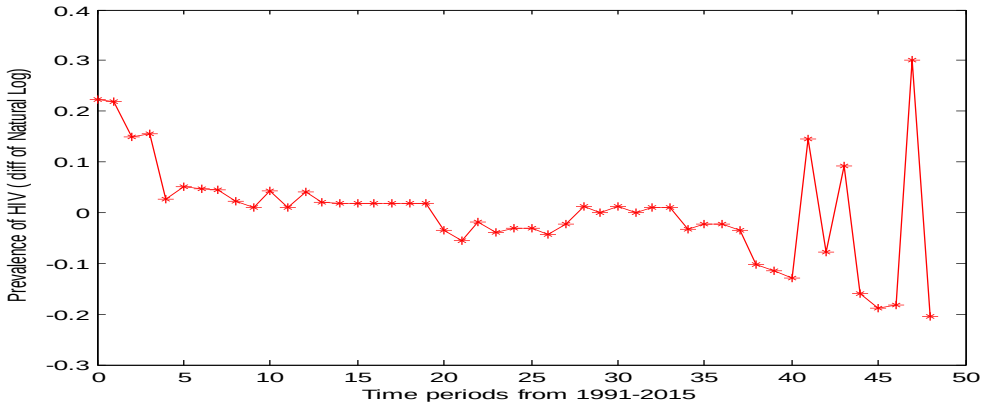

Figure 5: Time series plot of HIV prevalence in Nigeria (Diff. of Natural log)

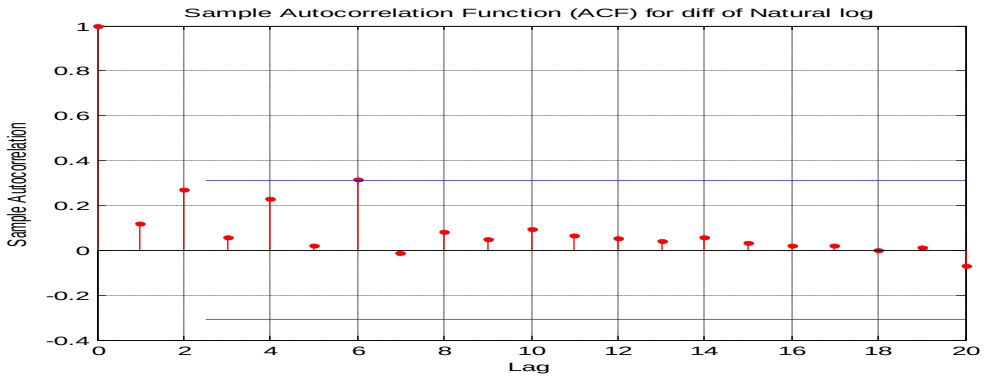

Figure 6: Autocorrelation function plot for HIV prevalence data (Diff. of Natural log)

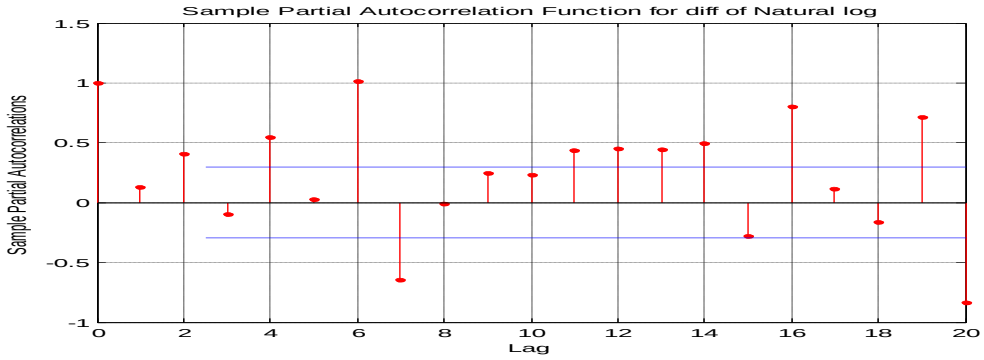

Figure 7: Partial autocorrelation function plot for HIV prevalence data (Diff. of Natural log)

Table 1: ADF Test for stationarity of HIV prevalence data (Diff. of Natural log)

H	p-Value	Test Stat	Critical Value
1	0.0010	-53.9823	-19.2264

Model Identification

Considering Figures 6 and 7, we suggest ARIMA(3,1,0), ARIMA(5,1,0) and ARIMA(7,1,0). Therefore, the ARIMA models are given as folows.

$(1-\sum_{i=1}^{p}\alpha_i L^i)(1-L)^d y_t = (1+\sum_{j=1}^{q}\theta_j L^j)\varepsilon_t$ (4)

For $p=3, d=1$ and $q=0$, (1) becomes

$y_t = (1+\alpha_1)y_{t-1} + (\alpha_2-\alpha_1)y_{t-2} + (\alpha_3-\alpha_2)y_{t-3} - \alpha_3 y_{t-4} + \varepsilon_t$ (5)

Thus, substituting the estimated parameters gives

$y_t = 1.0120y_{t-1} + 0.3466y_{t-2} + 0.2229y_{t-3} - 0.2778y_{t-4} + \varepsilon_t$ (6)

For $p=4, d=1$ and $q=0$, (1) becomes

$$y_t = (1+\alpha_1)y_{t-1} + (\alpha_2-\alpha_1)y_{t-2} + (\alpha_3-\alpha_2)y_{t-3} + (\alpha_4-\alpha_3)y_{t-4} + (\alpha_5-\alpha_4)y_{t-5} - \alpha_3 y_{t-6} + \varepsilon_t \quad (4)$$

Substituting the estimated parameters gives

$$y_t = 0.9880y_{t-1} + 0.3466y_{t-2} + 0.2029y_{t-3} - 0.1680y_{t-4} + 0.5101y_{t-5} - 0.8796y_{t-6} + \varepsilon_t \quad (7)$$

For $p = 7, d = 1$ and $q = 0$, (1) becomes

$$y_t = 0.9880y_{t-1} + 0.3466y_{t-2} + 0.2029y_{t-3} - 0.1680y_{t-4} + 0.5101y_{t-5} - 0.8883y_{t-6} - 0.6051y_{t-7} + 0.5964y_{t-8} + \varepsilon_t \quad (8)$$

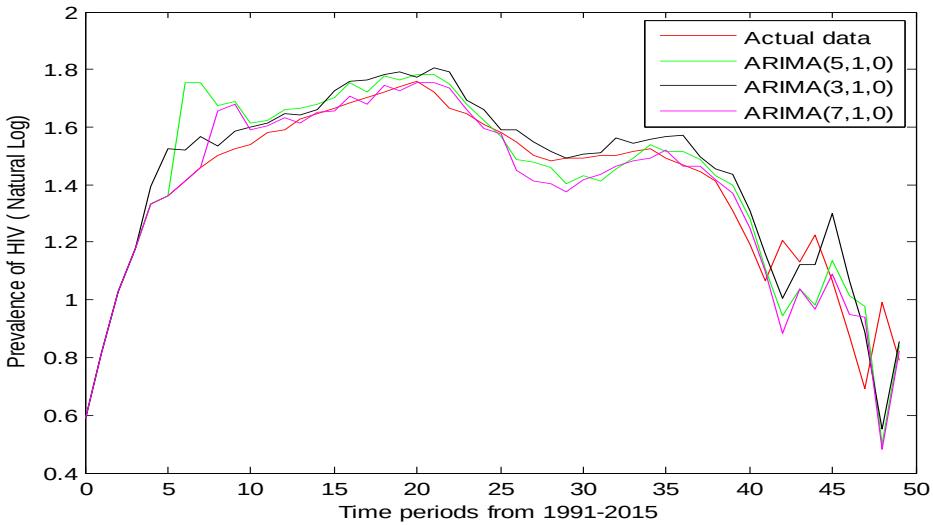

Figure 8: Plot of Actual data and ARIMA Models

Residual Analysis

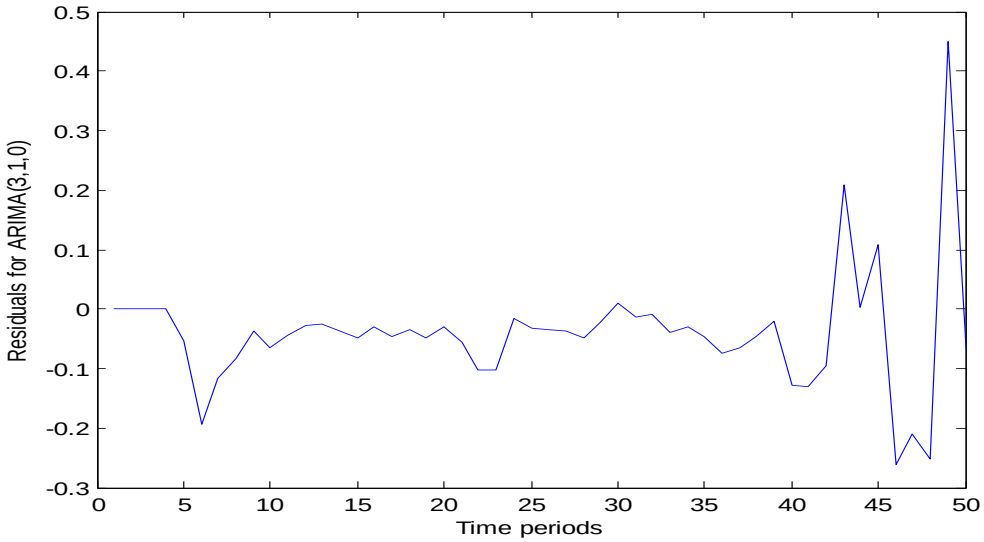

Figure 9: Plot of residuals for ARIMA (3,1,0) Model

Table 2: LBQ Test for ARIMA (3,1,0) residuals

H	p-Value	Q stat	Critical Value
0	0.5059	19.2454	28.4120

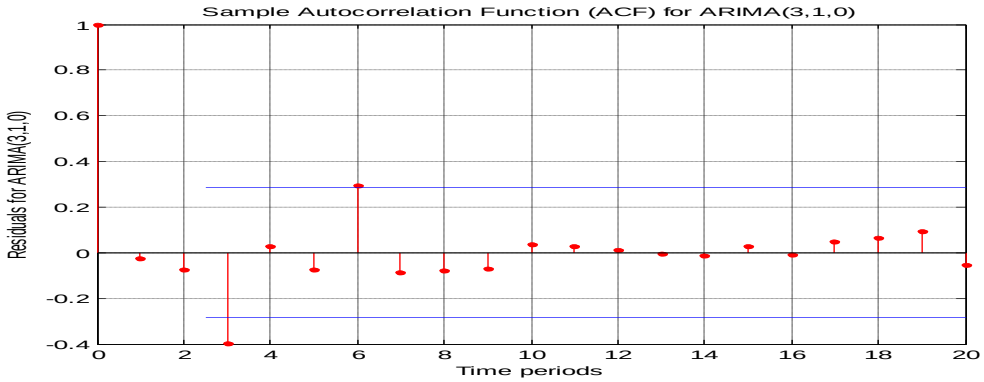

Figure 10: Autocorrelation function plot for ARIMA (3,1,0) residuals

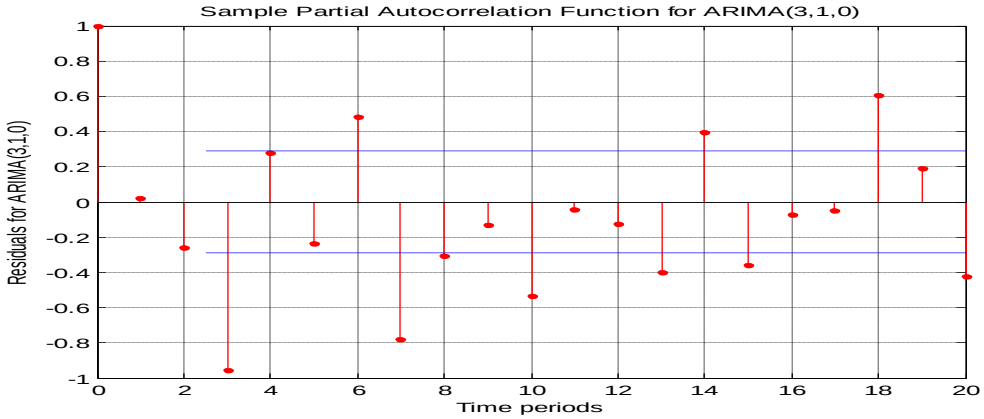

Figure 11: Partial autocorrelation function plot for ARIMA (3,1,0) residuals

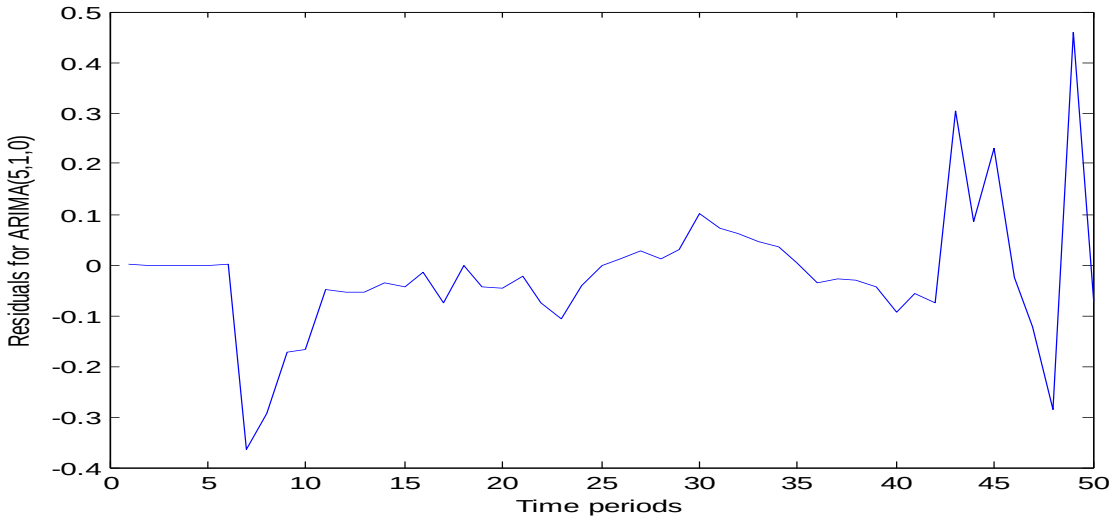

Figure 12: Plot of residuals for ARIMA (5,1,0) Model

Table 3: LBQ Test for ARIMA (5,1,0) residuals

H	p-Value	Q stat	Critical Value
0	0.9840	8.9067	28.4120

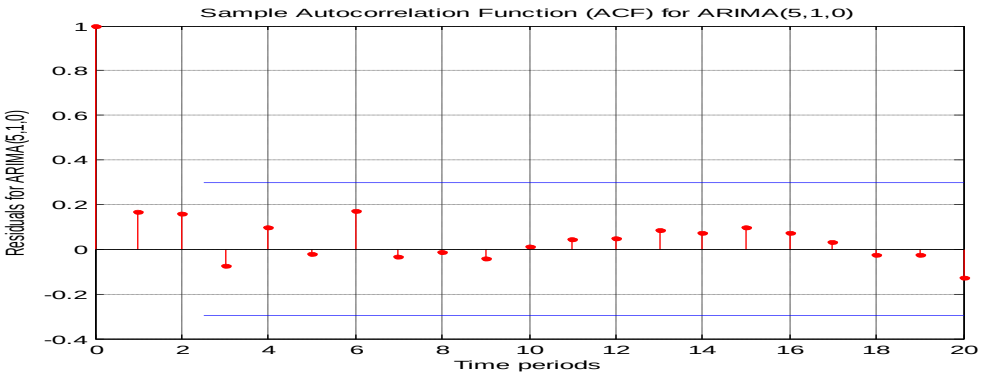

Figure 13: Autocorrelation function plot for ARIMA (5,1,0) residuals

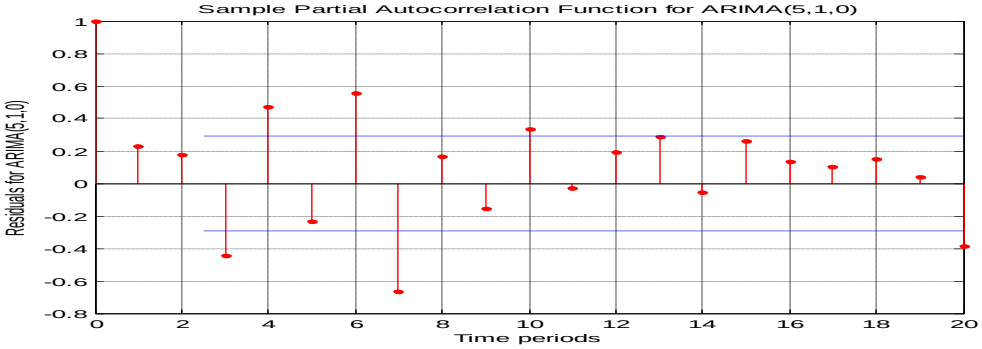

Figure 14: Partial autocorrelation function plot for ARIMA (5,1,0) residuals

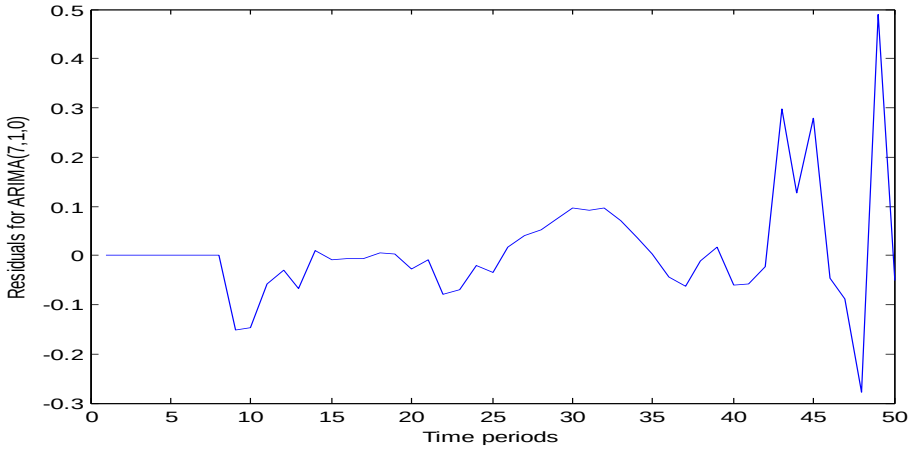

Figure 15: Plot of residuals for ARIMA (7,1,0) Model

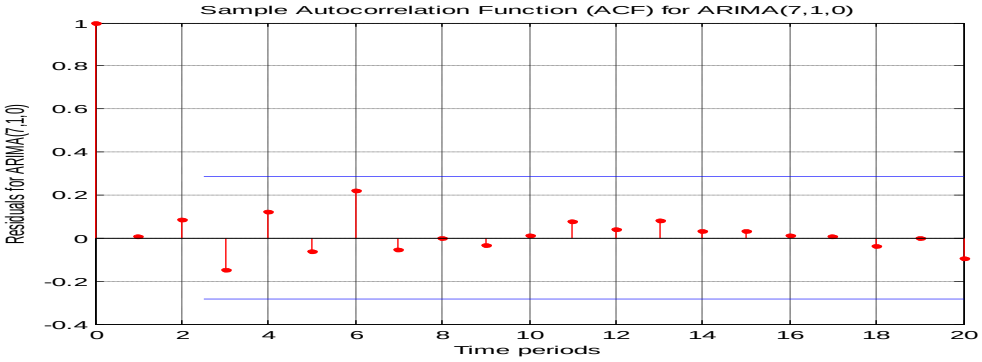

Figure 16: Autocorrelation function plot for ARIMA (7,1,0) residuals

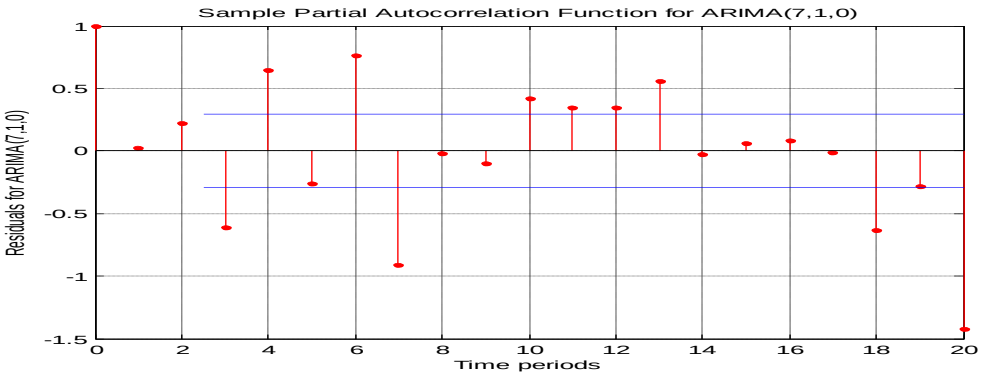

Figure 16: Partial autocorrelation function plot for ARIMA (7,1,0) residuals

Table 4: LBQ Test for ARIMA (7,1,0) residuals

H	p-Value	Q stat	Critical Value
0	0.9491	10.8877	28.4120

Performance Metrics

Table 5: Performance Metrics

Model	MAE	MSE
ARIMA(3,1,0)	0.0714	0.0140
ARIMA(5,1,0)	0.0791	0.0154
ARIMA(7,1,0)	0.0620	0.0130

4. Discussion

This paper studies autoregressive integrated moving average (ARIMA) models with a view to applying them to fit HIV prevalence data in Nigeria and forecasting into the future.

This study begins with the preliminary analysis of HIV data. First, we did the time series plot of the data, followed by the corresponding ACF and PACF plots. The results are shown in Figures 1, 2 and 4 respectively. Figure 1 shows clearly that there are trends in the times series data and that the data set exhibits non-stationarity. This view is buttressed by the ACF and PACF plots of the data. See Figures 2 and 3.

We further transformed the data by taking the natural logarithm of the original data and doing the time series plot. The result is shown in Figure 4. This result shows that there is non-stationarity in the natural log form of the data. We did another transformation by taking the first order difference of the log of the data. The plots of this transformed series, ACF and PACF are shown in Figures 5, 6 and 7. At this stage, stationarity is now achieved. This assertion is ascertained by augmented Dickey-Fuller test conducted whose result is presented in Table 1.

Based on Figures 6 and 7, ARIMA(3,1,0), ARIMA(5,1,0) and ARIMA(7,1,0) were selected. The plot of the stochastic solutions of these models and the actual data is shown in Figure 8.

The goodness-of-fit and model adequacy check was carried out via residual analysis. The plots for ARIMA (3,1,0), ARIMA(5,1,0) and ARIMA(7,1,0) residuals are shown in Figures 9, 12 and 15 respectively. The corresponding ACF and PACF plots for the above models residuals are shown in Figures 10, 11, 13, 14, 16 and 17 respectively. The LBQ tests for these models were conducted and the results are shown in Tables 2, 3 and 4 respectively. The results of this residual analysis show that the models fit the data well and are adequate for forecasting.

We further did performance metric test using mean absolute error (MAE) and mean square error (MSE). The results show that ARIMA (7,1,0) model is the best for forecasting and as such used. The forecast result shows that HIV prevalence is expected to drop in the near future and would be subsequently eliminated if the current intervention programs are consistently implemented.

5. Conclusion

In this study, we studied autoregressive integrated moving average (ARIMA) models to fit HIV prevalence data in Nigeria from 1991 to 2015 and thereafter use to forecast into the future. The suitable models identified are ARIMA(3,1,0), ARIMA(5,1,0) and ARIMA(7,1,0) which are respectively

$y_t = 1.0120y_{t-1} + 0.3466y_{t-2} + 0.2229y_{t-3} - 0.2778y_{t-4} + \varepsilon_t,$

$y_t = 0.9880y_{t-1} + 0.3466y_{t-2} + 0.2029y_{t-3} - 0.1680y_{t-4} + 0.5101y_{t-5} - 0.8796y_{t-6} + \varepsilon_t$ and

$y_t = 0.9880y_{t-1} + 0.3466y_{t-2} + 0.2029y_{t-3} - 0.1680y_{t-4} + 0.5101y_{t-5} - 0.8883y_{t-6} - 0.6051y_{t-7} + 0.5964y_{t-8} + \varepsilon_t$

These three models are adequate for forecasting. However, ARIMA (7,1,0) is the best forecasting based on performance metric test conducted. The forecast result shows that HIV prevalence in the future is expected to drop if the current control programs are consistently implemented.

Reference

[1]. Mukandavire, Z., Das, p., Chiyaka, C. and Nyabadza, Z.(2010). Global analysis of an HIV/AIDS epidemic model. World Journal of Modeling and Simulation, vol. 6 no. 3, pp. 231-240. ISSN1746-7233, England, UK.

[2]. Williams, B., Lima, V. and Gouws, E. (2011). Modeling the impact of antiretroviral therapy on the epidemics of HIV. Current HIV Research, Bentham Science Publishers Ltd.

[3]. Casels, S., clark, S. J. and Morris, M. (2008). Mathematical models for HIV transmission dynamics: Tools for social and behavioural science research. J. Acquir Immune Defic synd. Volume 47, Supplement 1. Lippincott Williams and Wilkins.

[4]. Peterson, D. Willard, K. Altmann, M. Gatewood, L. and Davidson, G.(1990). Monte Carlo simulation of HIV infection in an intravenous drug user community. Journal of Acquired Immune Deficiency Syndromes, 3 (11), pp. 1086-1095.

[5]. Greenhalgh, D. and Hay. G. (1997). Mathematical modelling of the spread of HIV/AIDS amongst injecting drug users. IMA Journal of Mathematics Applied in Medicine and Biology, 14, pp.1138.

[6]. Dalal, N. Greenhalgh, D. and Mao, X. (2007). A Stochastic model of AIDS and condom use. Journal of Mathematical Analysis and Applications, 325, pp. 36-53.

[7]. Dalal, N. Greenhalgh, D. and Mao, X. (2008). A Stochastic model for internal HIV dynamics. Journal of Mathematical Analysis and Applications, 341pp.1084-1101.

[8]. Ding, Y., Xu, M. and Hu, L. (2009). Risk analysis for AIDS control based on a stochastic model with treatment rate. Human and Ecological Risk Assessment, 15 (4), pp. 765-777.

[9]. Tuckwell, H. and Le Corfec, E. (1998). A Stochastic model for early HIV-1 population dynamics. Journal of Theoretical Biology, 195 (4), pp.451-463.

[10]. Waema, R. and Olowofeso, O. E. (2005). Mathematical modeling for human immunodeficiency virus (HIV) transmission using generating function approach. Kragujevac J. Sci. 27: 115-130.

[11]. Zhou, L., Yu, L., Wang, Y., Lu, Z., Tian, L., Tan, L., ... & Liu, L. (2014). A hybrid model for predicting the prevalence of schistosomiasis in humans of Qianjiang City, China. PLoS One, 9(8), e104875.

[12]. Promprou, S.,Jaroensutasinee,M. and Jaroensutasinee, K. (2006). Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models. Dengue Bulletin, Vol 30, 99-106.

[13]. Trottier H. and Philippe P. (2006), "Univariate time series analysis of pertussis, mumps measles and rubella: Theory and Methods", The Internet Journal of Infectious Diseases, ISSN: 2528-7836, Volume 3 Number 4.

[14]. Takyi Appiah, S., Otoo, H., & Nabubie, I. B. (2015). Times series analysis of malaria cases in Ejisu-Juaben Municipality. Int. J. Sci. Technol. Res, 4, 220-226.

[15]. Jere, S. and Moyo, E. (2016). Modelling Epidemiological Data Using Box-Jenkins Procedure. Journal of Statistics, 6:295-302.

[16]. Malinga, J.K. (2015). Forecasting Malaria Case Admissions In Three Kenyan Health Facilities. M.Sc. Thesis, University of Nairobi.

[17]. Garrett, L.C. (2012). Using Box-Jenkins Modeling Techniques to Forecast Future Disease Burden and Identify Disease Aberrations in Public Health Surveillance Report. Ph.D. Thesis, Western Michigan University.

[18]. Imran, M, Nasir, J.A. and Zaidi, S.A.A. (2014). Forecasting of New Cases of TB, using Box-Jenkins Approach. JUMDC 5(2):37-42

[19]. Jiang,, H., Michael Livingston, Robin Room, Paul Dietze, Thor Norström6 and William C. Kerr (2014). Alcohol Consumption and Liver Disease in Australia: A Time Series Analysis of the Period 1935–2006. Alcohol and Alcoholism Vol. 49, No. 3, pp. 363–368.

[20]. Kuhe, D.A., Chiawa, M.A. and Aboiyar, T. (2016). A Time Series Model of Poverty Incidence in Nigeria. Journal of Scientific and Engineering Research, 3(2):261-283.