

Forecasting the Monthly Reported Cases of Human Immunodeficiency Virus (HIV) at Minna Niger State, Nigeria

Nwanne Christiana Umunna, Samuel Olayemi Olanrewaju

Department of Statistics, University of Abuja, Abuja, Nigeria

Email: olanrewaju.samuel@uniabuja.edu.ng

How to cite this paper: Umunna, N.C. and Olanrewaju, S.O. (2020) Forecasting the Monthly Reported Cases of Human Immunodeficiency Virus (HIV) at Minna Niger State, Nigeria. *Open Journal of Statistics*, 10, 494-515.

<https://doi.org/10.4236/ojs.2020.103030>

Received: April 8, 2020

Accepted: June 7, 2020

Published: June 10, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

There has been a moderate increase in newly diagnosed HIV-infected Minna populace, which calls for serious attention. This study used time series data based on monthly HIV cases from January 2007 to December 2018 taken from the statistical data document on HIV prevalence recorded in General Hospital Minna, Niger State. The methodology employed to analyze the data is based on mathematical models of ARMA, ARIMA and SARIMA which were computed and diagnosed. From the results of parameter estimation of the models, ARMA(2, 1) model was the best model among the other ARMA models using information criteria (AIC). Diagnostic test was run on the ARMA(2, 1) model where the results show that the model was adequate and normally distributed using Box-Lung test and Q-Q plot respectively. Furthermore, ARIMA of first and second differences was estimated and ARIMA(1, 0, 1) was the best model from the result of the AIC and diagnostic test carried out which revealed that the model was adequate and normally distributed using Box-Lung and Q-Q plot respectively. Furthermore, the results obtained in the ARMA and ARIMA models were used to arrive at a combined model given as $ARIMA(1, 0, 1) \times SARIMA(1, 0, 1)_{12}$ which was subsequently estimated and found to be adequate from the result of the Box-Lung and Q-Q plot respectively. Post forecasting estimation and performance evolution were evaluated using the RMSE and MAE. The results showed that, $ARIMA(1, 0, 1) \times SARIMA(1, 0, 1)_{12}$ is the best forecasting model followed by ARIMA(1, 0, 2) on monthly HIV prevalence in Minna, Niger state.

Keywords

Human Immunodeficiency Virus, Autoregressive Moving Average, Autoregressive Integrated Moving Average, Seasonal Autoregressive

1. Introduction

HIV infection has spread over the last 30 years and has a great impact on health, welfare, employment and criminal justice sectors; affecting all social and ethnic groups throughout the world. Recent epidemiological data indicate that HIV remains a public health issue that persistently drains our economic sector having claimed more than 25 million lives over the last three decades [1]. The estimated overall number of People Living with HIV (PLWHIV) by the end of 2014 was approximately 36.9 (34.3 - 41.4) million and Sub-Saharan Africa was the most affected region, having 25.8 (24.0 - 28.7) million PLWHIV and 66% of all people with HIV infection living in the region (Yi, 2007). Of all people living with HIV globally, 9% of them live in Nigeria [2]. Most cases of HIV infection in Nigeria occur via heterosexual means with epidemics more pronounced among the females [3]. The country already burdened by political instability and endemic political corruption as a result of almost 33 years of military rule now seems prepared to “wipe out” the virus within a few decades [3]. Notwithstanding the progress in institutional reforms and political commitment to tackle the disease, the country has seen more citizens placed on life-saving medication of active antiretroviral therapy (AART) to increase the survival of such HIV seropositive individuals [3].

This study reviewed a discussion on the prevalence of HIV in Minna, Niger State and developed a best model that predicts the monthly HIV cases in Minna by means of the Seasonal Autoregressive Integrated Moving Average (SARIMA) with Box-Jenkins Method. HIV which stands for “Human Immunodeficiency Virus” is a serious disease that is caused by a virus that spread through the body fluids which attacks the body immune system just like cancer and can lead to death. Dissimilar to some different infections, the human body can’t dispose of HIV. That implies that once you have HIV, you have it forever [2]. HIV is found throughout the world and is prevalent in sub-Saharan Africa, accounting for 70% of new infections yearly [2]. Worldwide, an estimated 36.9 million people are living with HIV and about 2 million people became newly infected in 2014 [4].

The earliest report of HIV dates back to 1981 with five cases of *Pneumocystis carinii* pneumonia in healthy young homosexual men in Los Angeles, CA. At the time, it was described as “cellular-immune dysfunction” related to “sexual contact” [5]. Since then, tremendous efforts have been made worldwide for the diagnosis, control and prevention of HIV. Thirty-five million people are currently living with human immunodeficiency virus (HIV) globally. While 9.7 million infected people are receiving antiretroviral therapy, 2.3 million people are newly infected every year. Transmission via semen is one of the most prevalent me-

thods of HIV-1 transmission, accounting for up to 80% of new infections every year.

In the majority of cases, HIV is a sexually-transmitted infection. However, HIV can also be transmitted from a mother to her child, during pregnancy or childbirth (through blood or fluid exposure), or through breastfeeding. Non-sexual transmission can also occur through the sharing of injection equipment such as needles.

Today, scientists are still working to find a treatment for HIV and the recent studies show that a new vaccine will be developed by 2025 [6]. These are quite promising studies for the whole world. However, it is important to understand people who are living with that virus are also struggling with social, economic and psychological problems. UNAID and the National Agency for the Control of AIDS estimate that there are 1.9 million people living with HIV in Nigeria (Punch News Paper).

Results from the Nigeria HIV/AIDS Indicator and Impact Survey (NAISS) indicate a national HIV prevalence in Nigeria of 1.5% among adults aged 15 - 49 years. The survey revealed an improvement in the national prevalence rate from 3.4% in 2012 to 1.9% in 2018.

The President of Nigeria, Muhammadu Buhari early last year (2019) launched the Revised National HIV and AIDS Strategic Framework 2019-2021, which will guide the country's future response to the epidemic.

Aim and Objectives

The general objective of this study is to develop a best model that can predict the monthly HIV cases in Minna. This is to be achieved through the following Specific objectives:

- 1) Formulate time series models on the data collected.
- 2) Conduct a diagnostic check on the models formulated to determine the most suitable model.
- 3) Estimate the parameters of the various models and forecast the HIV prevalence.

2. Empirical Framework and Theoretical Issues

A few related works of the use of SARIMA methodology to model epidemic incidence include the following; [7] worked on forecasting monthly cases of Human immunodeficiency syndrome (HIV) of the Philippines. The researchers utilized advanced statistical tool in developing the model using univariate Box-Jenkins method in forecasting the HIV cases per month. The result showed that monthly cases of HIV in the Philippines had an upward trend. The researchers came up with the best model based on AIC which is $(2, 1, 0) \times (0, 0, 1)_{12}$.

[8] used HIV infection data from 1985 to 2012 to fit ARIMA models. Akaike Information Criterion and Schwartz Bayesian Criterion statistics were used to

evaluate the constructed models. Estimation was via the maximum likelihood method. To assess the validity of the proposed models, the mean absolute percentage error (MAPE) between the number of observed and fitted HIV infections from 1985 to 2012 was calculated. The fitted ARIMA models were used to forecast the number of HIV infections from 2013 to 2017 and the result showed that the fitted number of HIV infections was calculated by optimum ARIMA(2, 2, 1) model from 1985-2012 and the number was similar to the observed number of HIV infections, with a MAPE of 13.7%.

[9] conducted a study with the aim of formulating a model to determine the trend, prevalence and projecting HIV/AIDS epidemics in Ethiopia. Data were obtained from UNAIDS and Ministry of Health bulletin in Ethiopia. The data was analyzed using Autoregressive Integrated Moving Average (ARIMA) time series analysis model and the ARIMA(2, 3, 2) appeared to be providing the best fit for the observed data.

[10] worked on Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China. The study aims to describe the epidemiology of influenza viruses among children in Wuhan, China during the past nine influenza seasons (2007-2015) and to predict the positive rate of different types of influenza virus in the future. Their study suggests that the ARIMA model can be used to forecast the positive rate of different types of influenza virus.

The estimated results of model showed that Peads incoming is influenced by seasonal variation of data, [11] works on Energy Consumption Forecasting Using Seasonal ARIMA with Artificial Neural Networks Models. The quarterly energy consumption of the United States from January 1973 to June 2015 is used. It aimed to forecast the residential energy consumption in U.S. using the Box-Jenkins methodology and Artificial Neural Network approach and compared their results in order to know the best model for predicting energy consumption in U.S. From their results they concluded that the forecasting accuracy is not quite significant. But, the performance of ANN model is better than SARIMA model in terms of forecasting accuracy from the test data using MAE and MAPE, the opposite result happens for MSE. While the SARIMA model fits better the historical data (training data) than ANN models using all performance parameters.

[12] also worked on Forecasting Precipitation Using SARIMA Model: A Case Study of Mt. Kenya Region. Two objectives were formulated from their research which is to determine the forecasted values of precipitation in Mt. Kenya region and also to determine the accuracy of the SARIMA model in forecasting precipitation in the same region. Monthly data collected from Kenya meteorological department covering a period of 1995 to 2010 for wind data and 1970 to 2011 for precipitation data but will be limited to the available wind data. SARIMA models were fitted and the least AIC and BIC value was picked which is SARIMA(1, 0, 1) \times (1, 0, 0)₁₂ that turns out to be the best model since it has the least values of the information criteria and forecasting evaluation was conducted using the RMSE.

3. Research Methodology

3.1. Research Design

The research design adopted for this study is a descriptive and Box-Jenkins research design. Descriptive survey design is a research design in which data is collected consistently to explain and predict the given situation. For this purpose, non-seasonal Box Jenkins approach is used to find the best fitted, the best forecasting model and the accuracy of the forecasting values are checked by comparing residuals. The steps of the suggested model and its forecasting can be explained in the following steps. Determining whether the time series is stationary or not is a very important concept before making any inferences in time series analysis. Therefore, Augmented Dickey Fuller (ADF) and Phillips-Person (PP) tests will be used to check the stationarity of the data series. There are several methods that can be used to fit a time series model, among them, ARMA, ARIMA, and SARIMA model which will be used on the stationary data of this study.

3.2. Population of the Study and Research Sample

The study was carried out based on monthly data on HIV prevalence as secondary data, which was collected from document based on January 2007 to December 2018 retrievable document from the Statistical data record on HIV prevalence from the record of Communicable diseases in Minna general hospital for both male and female.

3.3. Method of Data Collection

Documentary evidence constitutes the instrument of data collection. The major sources of data are from Minna general hospital Statistical record on communicable diseases. The data for this study are secondary monthly HIV data sourced from the General hospital Minna in Niger state from January 2007 to December 2018.

3.4. Technique of Data Analysis and Model Specification

The advances in Time Series enable researchers to use those techniques in their analysis to re-analyze the traditional rotation analysis applied in earlier studies [13]. The central idea behind model identification is a time series derived from ARIMA process which has some sort of theoretical autocorrelation properties. Fitting the empirical autocorrelation patterns with the theoretical ones helps to identify the potential tentative model for the given time series data. In this step, transformation of observed time series to stationary is inevitable.

The software that was used for the test is Eviews 4.0 version.

3.5. Autoregressive Moving Average (ARMA) Models

We can have combinations of the two processes to give a new series of models called ARMA(p, q) models. The Autoregressive model (AR) and moving average

(MA).

Where

AR of order p is:

$$X_n = m + e_n + \phi_1 X_{n-1} + \phi_2 X_{n-2} + \cdots + \phi_p X_{n-p} \quad (3.4)$$

for $n \geq 0$, where $\{e_n\}$ $n \geq 0$ is a series of independent, identically distributed (iid) random variables, and m is a constant.

MA of order q is:

$$X_n = m + e_n + \theta_1 e_{n-1} + \theta_2 e_{n-2} + \cdots + \theta_q e_{n-q}, \quad (3.5)$$

for $n \geq 1$ where $\theta_1, \dots, \theta_q$ are real numbers and m is a real number.

The general form of the ARMA(p, q) models where p is used for the number of autoregressive components, and q for the number of moving average components is written as:

$$X_n = m^1 + \sum_{k=1}^p \phi_k X_{n-k} + \sum_{j=1}^q \theta_j e_{n-j} + e_n, \quad n \geq 0, \quad (3.6)$$

where $\{X_n\}$ $n \geq 1$, is some constant, and the ϕ_k and θ_j are defined as for AR and MA models respectively.

3.6. Autoregressive Integrated Moving Average (ARIMA) Models

Autoregressive (AR), Moving Average (MA) or Autoregressive Moving Average (ARMA) models in which differences have been taken are collectively called Autoregressive Integrated Moving Average or ARIMA models. A time series $\{Y_t\}$ is said to follow an integrated autoregressive moving average model if the d^{th} difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. If $\{W_t\}$ follows an ARMA(p, q) model, we say that $\{Y_t\}$ is an ARIMA(p, d, q) process. For example, for practical purposes, we can usually take $d = 1$ or at most 2.

Consider then an ARIMA($p, 1, q$) process. With $W_t = Y_t - Y_{t-1}$, we have

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (3.7)$$

Or, in terms of the observed series,

$$Y_t - Y_{t-1} = \phi_1 (Y_{t-1} - Y_{t-2}) + \phi_2 (Y_{t-2} - Y_{t-3}) + \cdots + \phi_p (Y_{t-p} - Y_{t-p-1}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (3.8)$$

3.7. Seasonal Autoregressive Integrated Moving Average (SARIMA) Models

The ARIMA model (3.7) is for non-seasonal non-stationary data. A purely seasonal time series is the one that has only seasonal AR or MA parameters. Seasonal autoregressive models are built with parameter called seasonal autoregressive (SAR) parameters. The SAR parameters represent the autoregressive relationships that exist between time series data separated by multiples of the number of periods per season. Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is known as the Seasonal ARIMA (SARIMA) model. In this model seasonal differencing of appropriate order is

used to remove non-stationarity from the series. A first order seasonal difference is the difference between an observation and the corresponding observation from the previous year and is calculated as $X_t = Y_t - Y_{t-s}$. For monthly time series $S = 12$ and for quarterly time series $S = 4$. This model is generally termed as the SARIMA(p, d, q) \times (P, D, Q)_s.

For a seasonal time series of order s , [14] proposed that $\{X_t\}$ be modelled by:

$$A(L)\Phi(L^s)\nabla_s^d X_t = B(L)\Theta(L^s)\varepsilon_t \quad (3.9)$$

where the series must have been subjected to seasonal differencing D times and non-seasonal differencing d times, $\nabla_s = 1 - L^s$, being the seasonal differencing operator. Moreover, $\Phi(L)$ and $\Theta(L)$ are the seasonal autoregressive and moving average operators respectively. These seasonal operators are polynomials in L .

Suppose that $\Phi(L) = 1 + \phi_1 L + \phi_2 L^2 + \dots + \phi_P L^P$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_Q L^Q$, then the time series $\{X_t\}$ is said to follow a multiplicative seasonal autoregressive integrated moving average model of orders p, d, q, P, D, Q and s , designated $(p, d, q) \times (P, D, Q)_s$ SARIMA model.

4. Presentation of Result and Finding

To really come out with a good forecasting model of the HIV Prevalence Recorded in General Hospital Minna (2007-2018) data, ARMA, ARIMA and SARIMA models were fitted to the series. Furthermore, this section also explains the behavior of the rate of contracting HIV in Minna general hospital of Nigeria, test for unit root, specification of the models, estimation of the parameters of the forecasting model using the above model, selection of the best competing forecasting models using AIC while forecast evaluation of these models using Root Mean Square Error, Mean Absolute Error and Mean Absolute Percentage Error and forecast plot for seasonal models were critically looked into.

4.1. Descriptive Statistics of the HIV Data

In this section, we discuss empirical results beginning with preliminary analysis conducted with the aim to determine the normality of the data. Skewness, kurtosis and Jarque-Bera show the normality of the distribution. A distribution is said to be normal when skewness is approximately zero and kurtosis is three. Also, the probability of the Jarque-Bera statistics tells whether the series is normal or not. The null hypothesis of the Jarque-Bera test says that the distribution is a normal one. Therefore, if the probability is less than 0.05, we reject the null hypothesis and conclude that the distribution is not normal (Table 1).

Furthermore, from the Jarque-Bera test for normality of each of the variables, it was observed in the above table that the variables "HIV prevalence" p -value is less than 0.1 (10%) level of significance and not at 5% level. Thus, the enter variable is normally distributed at 10% level of significance. This is a strong factor of the fundamental assumptions of the application of ARMA, ARIMA and SARIMA models. Hence, data differencing transformation is considered in order to correct for the normality assumption violation (Table 2).

Table 1. Descriptive statistics of the HIV prevalence recorded in General Hospital Minna (2007-2018).

Statistics	HIV Prevalence
Mean	85.51389
Median	80.00000
Maximum	228.0000
Minimum	0.000000
Std. Dev.	46.75049
Skewness	0.487059
Kurtosis	2.559941
Jarque-Bera	6.855339
Probability	0.032463
Sum	12314.00
Sum Sq. Dev.	312542.0
Observations	144

Table 2. Augmented dickey-fuller test of stationarity (ADF) of the HIV Prevalence Recorded in General Hospital Minna (2007-2018) data.

Null Hypothesis: HIV has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 0 (Automatic—based on SIC, maxlag = 13)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.411370	0.0029
Test critical values:		
1% level	-4.023506	
5% level	-3.441552	
10% level	-3.145341	

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

4.2. Parameter Estimation of ARMA Models and Models Selection

Table 3 shows the results of parameter estimation and model selection for the ARMA models, where results of the different estimation parameter of ARMA were estimated with most of the parameter significant at 1% and 5%. AIC was used to select the best model that will be used for ARIMA and SARIMA model because it is the combination of AR and MA model. From the AIC, ARMA(2, 1) was selected to be the best model since it has the smallest AIC. With this selection, our ARIMA model will be AR(2) and MA(1) while the integrated difference will be of one (1) and two (2).

4.3. Diagnostic Tests for ARMA Models

Using the best model in **Table 2**, the result of **Table 3** shows the *P*-value for ARMA(2, 1) indicates there is no evidence that the residuals are dependent. This further confirms that the ARMA(2, 1) model is adequate.

Table 3. Parameter estimation of ARMA models and models selection.

	ARMA(1, 1)	ARMA(2, 1)	ARMA(1, 2)	ARMA(2, 2)
Intercept	-527.6081	-489.1910	-618.0471	-489.2602
AR1	-0.2819*	-0.7670*	-0.9625*	-0.7684*
AR2	-	-0.6447*	-	-0.6443*
MA1	-0.7131**	-0.2148**	0.1985*	-0.2130
MA2	-	-	-0.7837**	-0.0041
Log Likelihood	-2149.62	-2111.91	-2153.72	-2111.91
AIC	24.8858	24.6035	24.9447	24.6152
BIC	24.9406	24.6768	25.0176	24.7067

* at 1%, ** at 5%.

Figure 1 presents the trends analysis of the monthly data on HIV prevalence during the period of 2007 to 2018. The HIV prevalence started in January 2007 at a very slow prevalence rate. Until about September, 2008 when there was a sharp increase on the prevalence from 50 units to about 170 units. This clearly suggests an outbreak in the HIV virus. Although a relative decline in this trend was similarly observed as from July 2009 through to mid-year 2012. Another sharp increase in the trend is also observed in November 2012 but declined to almost zero in May 2015. With a steady gradual steady increase observed from march 2016 till date. This shows that if something is not done immediately the trend will go out of control.

4.4. Parameter Estimation of AR, MA, ARMA AND SARMA Models and Models Selection

Table 3 shows the results of parameter estimation and model selection for the **AR, MA, ARMA & SARIMA** models, where results of the different estimation parameter of the models were estimated with most of the parameter significant at 1% and 5%. AIC was used to select the best model. The models **AR, MA, ARMA AND SARIMA were considered** because the data set is in stationary at its original state and thus requires no differencing and transformation. Hence, the order and combination of the AR and MA component of the model is determined from the Correlogram plot below (**Table 4**).

These plots are used to choose the order parameters for candidates ARMA model. The simple moving average (MA) model is a parsimonious time series model used to account for very short-run autocorrelation. It does have a regression like form, but here each observation is regressed on the previous innovation, which is not actually observed. A weighted sum of previous and current noise is called Moving Average (MA) model.

Model identification started with autocorrelation analysis. Plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) (**Figure 2**) showed only the first lag of the ACF was significant (*i.e.* laying outside the grey



Figure 1. HIV Prevalence recorded in General Hospital Minna (2007-2018) Trend.

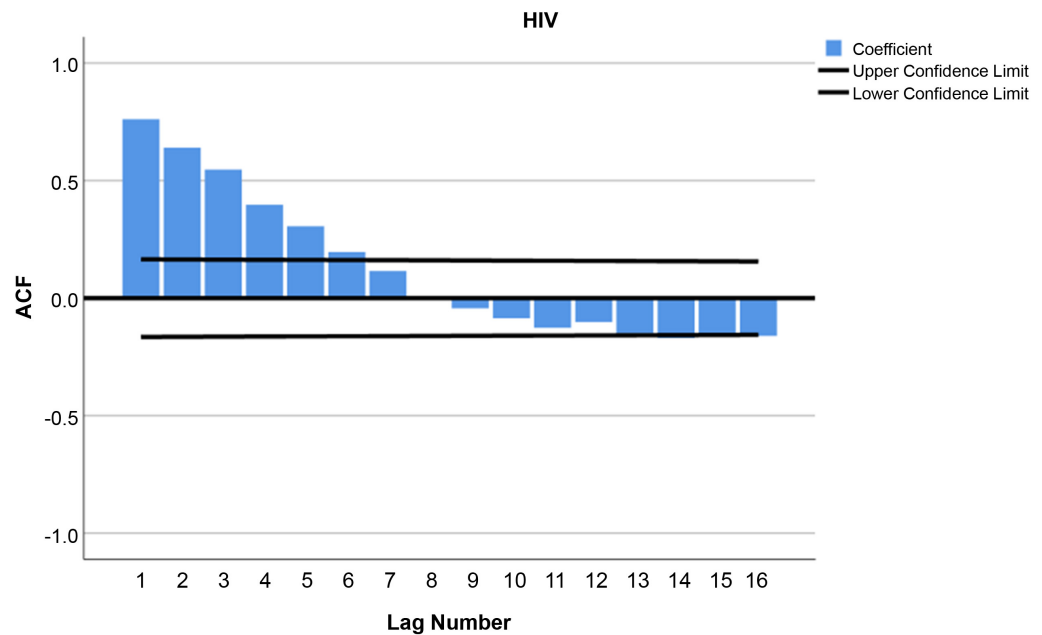


Figure 2. Plot of ACF and PACF of ARMA model.

Table 4. Correlogram plot.

ACF and PACF Model Description		
Model Name		MOD_5
Series Name	1	HIV
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		12
Maximum Number of Lags		16
Process Assumed for Calculating the Standard Errors of the Autocorrelations		Independence (white noise) ^a
Display and Plot		All lags

95% CI band). It was also observed that the first few lags of ACF did not decay with time. Based on the autocorrelation structure, several potential models were

Table 5. Candidate models proposed.

Candidate models	
MA(1)	ARMA(1, 4)
MA(2)	ARMA(1, 5)
MA(3)	ARMA(1, 6)
MA(4)	SARIMA(1, 0, 1) (1, 0, 1)
MA(5)	SARIMA(1, 0, 2) (1, 0, 2)
MA(6)	SARIMA(1, 0, 3) (1, 0, 3)
AR(1)	SARIMA(1, 0, 4) (1, 0, 4)
ARMA(1, 1)	SARIMA(1, 0, 5) (1, 0, 5)
ARMA(1, 2)	SARIMA(1, 0, 6) (1, 0, 6)

identified.

ACF plots display correlation between a series and its lags. In addition to suggesting the order of differencing, ACF plots can help in determining the order of the MA(q) model. Thus, as observed from the ACF plots we have MA(1, 2, 3, 4, 5, 6).

Based on the ACF/PACF plots the following candidate models was proposed (Table 5).

The candidate model with the smallest value of the residual sums of squares is the model that best fit the data at hand. Also, using order selection strategy proposed in Hannan and Rissanen (1982) and used by [15] and [16], the model with the least Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) is the best among other models under consideration.

4.5. Parameter Estimation for Candidate Models Codes and Summary Using R-Console

```
> library(forecast)
> library("ggplot2")
> library("forecast")
> library("tseries")
> data = ts(read.csv("data.hiv.csv", header = TRUE, stringsAsFactors = FALSE))
> ma1 <- arima(data, order = c(0, 0, 1))
> ma2 <- arima(data, order = c(0, 0, 2))
> ma3 <- arima(data, order = c(0, 0, 3))
> ma4 <- arima(data, order = c(0, 0, 4))
> ma5 <- arima(data, order = c(0, 0, 5))
> ma6 <- arima(data, order = c(0, 0, 6))
> summary(ma1)
Call:
arima(x = data, order = c(0, 0, 1))
Coefficients:
```

```

      ma1      intercept
      0.6401    85.2213
s.e.    0.0594    4.8635
sigma^2 estimated as 1273: log likelihood = -719.33, aic = 1444.66

```

```
> summary(ma2)
```

```
Call:
```

```
arima(x = data, order = c(0, 0, 2))
```

```
Coefficients:
```

```

      ma1      ma2      intercept
      0.6542    0.3323    85.0295
s.e.    0.0869    0.0709    5.5283

```

```
sigma^2 estimated as 1125: log likelihood = -710.43, aic = 1428.87
```

```
> summary(ma3)
```

```
Call:
```

```
arima(x = data, order = c(0, 0, 3))
```

```
Coefficients:
```

```

      ma1      ma2      ma3      intercept
      0.6543    0.4557    0.4208    85.0665
s.e.    0.0847    0.0761    0.0764    6.4169

```

```
sigma^2 estimated as 939.9: log likelihood = -697.68, aic = 1405.37
```

```
> summary(ma4)
```

```
Call:
```

```
arima(x = data, order = c(0, 0, 4))
```

```
Coefficients:
```

```

      ma1      ma2      ma3      ma4      intercept
      0.6724    0.5009    0.5015    0.1576    85.0292
s.e.    0.0812    0.0890    0.0862    0.0740    7.0642

```

```
sigma^2 estimated as 912.2: log likelihood = -695.54, aic = 1403.08
```

```
> summary(ma5)
```

```
Call:
```

```
arima(x = data, order = c(0, 0, 5))
```

```
Coefficients:
```

```

      ma1      ma2      ma3      ma4      ma5      intercept
      0.6746    0.5279    0.5277    0.2259    0.1656    84.9291
s.e.    0.0869    0.1014    0.0925    0.0848    0.0793    7.6515

```

```
sigma^2 estimated as 884.3: log likelihood = -693.31, aic = 1400.63
```

```
> summary(ma6)
```

```
Call:
```

```
arima(x = data, order = c(0, 0, 6))
```

```
Coefficients:
```

```

      ma1      ma2      ma3      ma4      ma5      ma6      intercept
      0.6370    0.493    0.5412    0.2747    0.2829    0.1507    84.9344
s.e.    0.0871    0.100    0.0939    0.0884    0.1044    0.0952    8.1966

```

sigma^2 estimated as 869.9: **log likelihood = -692.15, aic = 1400.31**

```
> ar1 <- arima(data, order = c(1,0,0))
```

```
> summary(ar1)
```

Call:

```
arima(x = data, order = c(1, 0, 0))
```

Coefficients:

	ar1	intercept
	0.7637	84.4499
s.e.	0.0532	10.3551

sigma^2 estimated as 900.1: **log likelihood = -694.55, aic = 1395.1**

```
> arma1<-arima(data, order = c(1, 0, 1))
```

```
> arma2<-arima(data, order = c(1, 0, 2))
```

```
> arma3<-arima(data, order = c(1, 0, 3))
```

```
> arma4<-arima(data, order = c(1, 0, 4))
```

```
> arma5<-arima(data, order = c(1, 0, 5))
```

```
> arma6<-arima(data, order = c(1, 0, 6))
```

```
> summary(arma1)
```

Call:

```
arima(x = data, order = c(1, 0, 1))
```

Coefficients:

	ar1	ma1	intercept
	0.8448	-0.1980	84.4697
s.e.	0.0555	0.0981	12.3268

sigma^2 estimated as 878.1: **log likelihood = -692.79, aic = 1393.58**

```
> summary(arma2)
```

Call:

```
arima(x = data, order = c(1, 0, 2))
```

Coefficients:

	ar1	ma1	ma2	intercept
	0.8311	-0.2073	0.0587	84.5462
s.e.	0.0640	0.1039	0.1051	12.0457

sigma^2 estimated as 876.1: **log likelihood = -692.63, aic = 1395.27**

```
> summary(arma3)
```

Call:

```
arima(x = data, order = c(1, 0, 3))
```

Coefficients:

	ar1	ma1	ma2	ma3	intercept
	0.768	-0.1151	0.026	0.1704	84.6665
s.e.	0.096	0.1311	0.109	0.0967	11.0960

sigma^2 estimated as 858: **log likelihood = -691.16, aic = 1394.33**

```
> summary(arma4)
```

Call:

```
arima(x = data, order = c(1, 0, 4))
```

Coefficients:

	ar1	ma1	ma2	ma3	ma4	intercept
	0.8049	-0.1446	0.0054	0.1729	-0.0938	84.6406
s.e.	0.0928	0.1231	0.1021	0.0932	0.1069	11.4072

sigma^2 estimated as 853.3: **log likelihood = -690.78, aic = 1395.57**

> summary(arma5)

Call:

arma(x = data, order = c(1, 0, 5))

Coefficients:

	ar1	ma1	ma2	ma3	ma4	ma5	intercept
	0.7282	-0.0884	0.0604	0.2236	-0.0561	0.1425	84.8513
s.e.	0.1266	0.1438	0.1100	0.0963	0.1093	0.1021	11.1302

sigma^2 estimated as 841.4: **log likelihood = -689.83, aic = 1395.66**

> summary(arma6)

Call:

arma(x = data, order = c(1, 0, 6))

Coefficients:

	ar1	ma1	ma2	ma3	ma4	ma5	ma6
intercept							
	0.6865	-0.0478	0.0853	0.2518	-0.0288	0.1512	0.0439
	84.9150						
s.e.	0.1731	0.1839	0.1286	0.1198	0.1288	0.1054	0.0970
	10.9617						

sigma^2 estimated as 840.2: **log likelihood = -689.73, aic = 1397.46**

> sarma1<-arma(data, order = c(1, 0, 1), seasonal = list(order = c(1, 0, 1), period = 12))

> sarma2<-arma(data, order = c(1, 0, 2), seasonal = list(order = c(1, 0, 2), period = 12))

> sarma3<-arma(data, order = c(1, 0, 3), seasonal = list(order = c(1, 0, 3), period = 12))

> sarma4<-arma(data, order = c(1, 0, 4), seasonal = list(order = c(1, 0, 4), period = 12))

> sarma5<-arma(data, order = c(1, 0, 5), seasonal = list(order = c(1, 0, 5), period = 12))

> sarma6<-arma(data, order = c(1, 0, 6), seasonal = list(order = c(1, 0, 6), period = 12))

> summary(sarma1)

Call:

arma(x = data, order = c(1, 0, 1), seasonal = list(order = c(1, 0, 1), period = 12))

Coefficients:

	ar1	ma1	sar1	sma1	intercept
	0.8399	-0.1750	-0.6316	0.7791	84.4652

```

s.e.    0.0552    0.0977    0.3538    0.3155    13.1139
sigma^2 estimated as 845.2: log likelihood = -690.59, aic = 1393.17
> summary(sarma2)
Call:
arima(x = data, order = c(1, 0, 2), seasonal = list(order = c(1, 0, 2), period =
12))
Coefficients:
      ar1      ma1      ma2      sar1      sma1      sma2      intercept
      0.8245     -0.1879    0.0717     -0.6498    0.8021    0.0059    84.5483
s.e.    0.0630    0.1054    0.1039    0.6473    0.6575    0.1714    12.8811
sigma^2 estimated as 841.7: log likelihood = -690.35, aic = 1396.7
> summary(sarma3)
Call:
arima(x = data, order = c(1, 0, 3), seasonal = list(order = c(1, 0, 3), period =
12))
Coefficients:
      ar1      ma1      ma2      ma3      sar1      sma1      sma2
sma3
      0.7548     -0.0856    0.0523    0.1840     -0.1493    0.3056     -0.0128
      0.1610
s.e.    0.0966    0.1308    0.1067    0.0948    0.5515    0.5475    0.1286
      0.1154
      intercept
      83.514
s.e.    13.396
sigma^2 estimated as 812.1: log likelihood = -688.04, aic = 1396.08
> summary(sarma4)
Call:
arima(x = data, order = c(1, 0, 4), seasonal = list(order = c(1, 0, 4), period =
12))
Coefficients:
      ar1      ma1      ma2      ma3      ma4      sar1      sma1
sma2      sma3
      0.7795     -0.1115    0.034     0.1849     -0.0602    0.5665     -0.4278
     -0.1129    0.1898
s.e.    0.0991    0.1308    0.106     0.0937    0.1185    1.4295    1.4192
     0.2226    0.1184
      sma4      intercept
      -0.1457    83.3802
s.e.    0.2379    12.8059
sigma^2 estimated as 808.5: log likelihood = -687.79, aic = 1399.58
> summary(sarma5)
Call:

```

```
arima(x = data, order = c(1, 0, 5), seasonal = list(order = c(1, 0, 5), period = 12))
```

Coefficients:

	ar1	ma1	ma2	ma3	ma4	ma5	sar1
sma1							
	0.7210	-0.0708	0.0679	0.2325	-0.0364	0.1106	0.2541
	-0.1284						
s.e.	0.1306	0.1528	0.1115	0.1026	0.1226	0.1020	1.3827
	1.3780						

	sma2	sma3	sma4	sma5	intercept
	-0.0660	0.1698	-0.0931	-0.0366	83.5674
s.e.	0.1874	0.1119	0.2567	0.1631	12.2876

sigma^2 estimated as 802.2: **log likelihood = -687.19, aic = 1402.38**

```
> summary(sarma6)
```

Call:

```
arima(x = data, order = c(1, 0, 6), seasonal = list(order = c(1, 0, 6), period = 12))
```

Coefficients:

	ar1	ma1	ma2	ma3	ma4	ma5	ma6
sar1							
	0.6824	-0.0383	0.0986	0.2631	-0.0172	0.1171	0.0532
	-0.3471						
s.e.	0.1673	0.1804	0.1330	0.1244	0.1307	0.1025	0.1065
	NaN	NaN					

	sma2	sma3	sma4	sma5	sma6	intercept
	-0.0877	0.1773	-0.1227	-0.0133	0.0171	83.6397
s.e.	NaN	0.0892	NaN	0.1252	NaN	12.6138

sigma^2 estimated as 801.4: **log likelihood = -687.08, aic = 1406.16**

Estimated value of the parameter of the best model

```
> summary(sarma1)
```

Call:

```
arima(x = data, order = c(1, 0, 1), seasonal = list(order = c(1, 0, 1), period = 12))
```

Coefficients:

	ar1	ma1	sar1	sma1	intercept
	0.8399	-0.1750	-0.6316	0.7791	84.4652
s.e.	0.0552	0.0977	0.3538	0.3155	13.1139

sigma^2 estimated as 845.2: **log likelihood = -690.59, aic = 1393.17.**

The result shows the estimation of the best model and also identifies the significance of its parameter. Based on the computed value of the coefficient for each parameter and its standard error, the absolute quotient value of the AR1, MA1, SAR1, SMA1 respectively, is greater than 0.05, it means that there is statistical sufficient evidence to say that the parameters are significant (**Table 6**).

Table 6. Candidate models performance summary based on the Akaike information criterion (AIC).

sn	MODEL	log likelihood	Akaike info criterion (AIC)	Model Rank
1	MA(1)	-719.33	1444.66	3
2	MA(2)	-710.43	1428.87	6
3	MA(3)	-697.68	1405.37	12
4	MA(4)	-695.54	1403.08	11
5	MA(5)	-693.31	1400.63	10
6	MA(6)	-692.15	1400.31	13
7	AR(1)	-694.55	1395.1	5
8	ARMA(1, 1)	-692.79	1393.58	2
9	ARMA(1, 2)	-692.63	1395.27	7
10	ARMA(1, 3)	-691.16	1394.33	4
11	ARMA(1, 4)	-690.78	1395.57	8
12	ARMA(1, 5)	-689.83	1395.66	16
13	ARMA(1, 6)	-689.73	1397.46	17
14	SARIMA(1, 0, 1) (1, 0, 1) ₁₂	-690.59	1393.17	1*
15	SARIMA(1, 0, 2) (1, 0, 2) ₁₂	-690.35	1396.7	15
16	SARIMA(1, 0, 3) (1, 0, 3) ₁₂	-688.04	1396.08	18
17	SARIMA(1, 0, 4) (1, 0, 4) ₁₂	-687.79	1399.58	14
18	SARIMA(1, 0, 5) (1, 0, 5) ₁₂	-687.19	1402.38	19
19	SARIMA(1, 0, 6) (1, 0, 6) ₁₂	-687.08	1406.16	9

*The best performing model.

Figure 3 shows the residual plot of the best model created as part of residual diagnostics of the model. This shows that the variance of the error term are seems to be constant. It also shows that the average of the residual is approximately equal to zero.

Figure 3 further shows the residual analysis to identify the normality of error terms. Since the computed p -value of Jarque-Bera test with p -value is greater than 0.05 level of significance, there is statistical evidence not to reject or fail to reject the null hypothesis of the normality of error term. This means that the error term is normally distributed.

Table 7 shows the residual analysis in identifying the independency of error term for Autoregressive Conditional Heteroskedasticity (ARCH). Since the computed p -value Box-Ljung test is equal to 0.1846 which is greater than the assigned alpha 5%, there is a statistical sufficient evidence to say that the error term is independent.

Figure 4 shows the Independency of error term generalized autoregressive conditional heteroskedasticity (GARCH) (informal way). It is however noticed

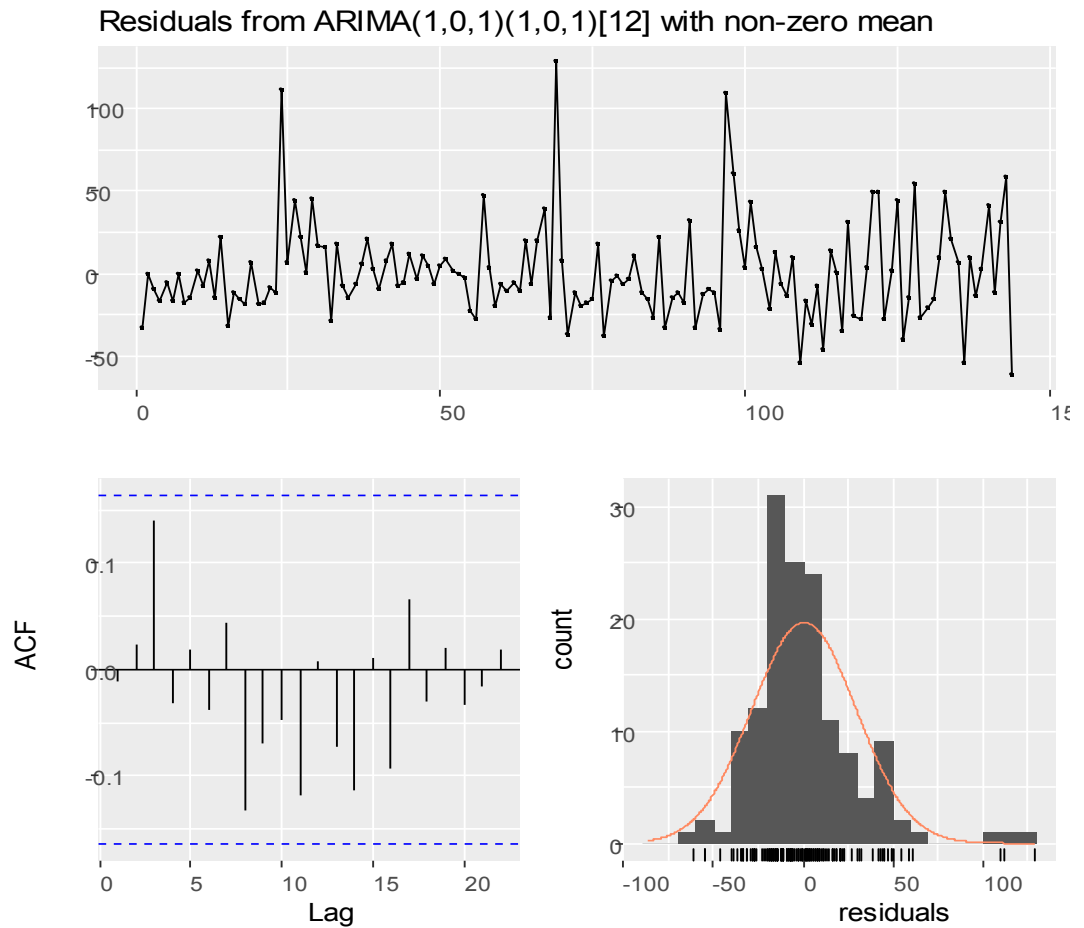


Figure 3. Independence of error term autoregressive conditional heteroskedasticity.

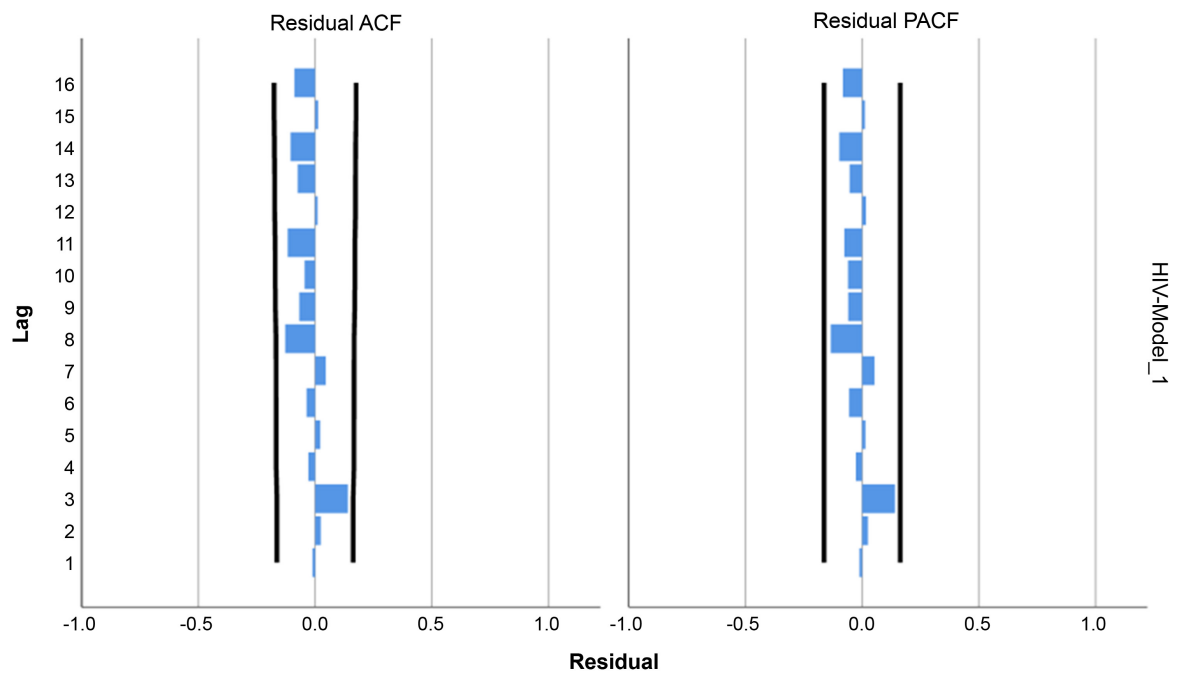


Figure 4. ACF and PACF of independency of residual.

that no spike hits the line at any lag, this strongly suggests that the model is free of white noise (Figure 5).

4.6. Forecast with the Fitted Model

One of the objectives of fitting and selecting the best model from AR/MA/ARMA/SARIMA model to data is to be able to forecast its future values. The model that best fits the data going by the various statistics given in Table 8 below is SARIMA(1, 0, 1) \times (1, 0, 1)₁₂.

Figure 6 shows the point forecast (blue), it indicates that the forecasted value from the created model has an increasing and decreasing trend from 2019 January-2019 October with a semi-continuous increase in January, 2019 till October, 2019.

Table 7. Ljung-box test. Independency of error term for Autoregressive Conditional Heteroskedasticity (ARCH).

data: Residuals from ARIMA(1, 0, 1)(1, 0, 1) [12] with non-zero mean						
Model	Number of Predictors	Model Fit statistics		Ljung-Box Q (18)		Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
HIV-Model_1	0	0.603	7.5221	5	0.1846	0

Total lags used: 10.

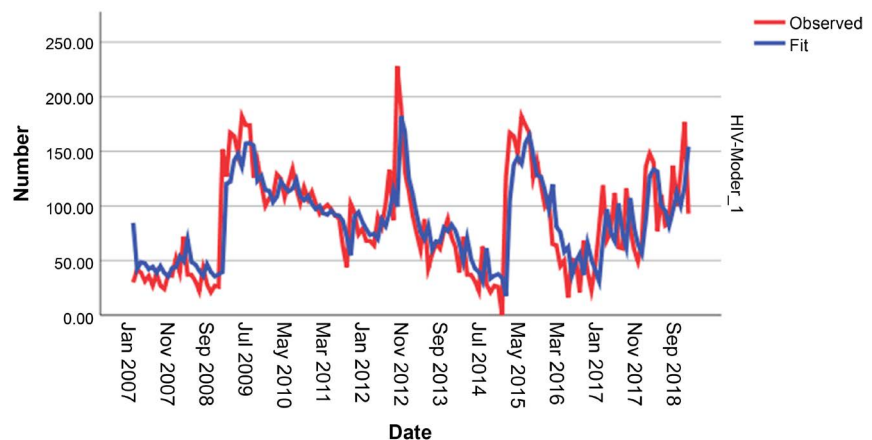


Figure 5. Plots of the observed and the forecast trend of HIV prevalence data.

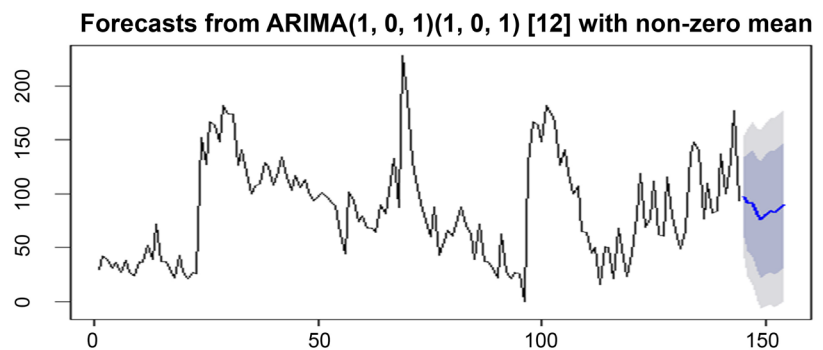


Figure 6. Forecasted value from 2007 to 2019.

Table 8. Forecast of data using SARIMA(1, 0, 1) \times (1, 0, 1)₁₂.

		Forecast data				
	Date	Point.Forecast	Lo.80	Hi.80	Lo.95	Hi.95
2019	JANUARY	96.94	59.68	134.20	39.96	153.93
	FEBRUARY	91.71	46.97	136.45	23.28	160.14
	MARCH	91.21	41.87	140.56	15.75	166.68
	APRIL	80.85	28.50	133.20	0.79	160.91
	MAY	76.16	21.79	130.53	-6.99	159.31
	JUNE	80.46	24.71	136.20	-4.80	165.72
	JULY	83.64	26.94	140.34	-3.08	170.36
	AUGUST	82.33	24.97	139.70	-5.40	170.06
	SEPTEMBER	85.62	27.79	143.45	-2.82	174.06
	OCTOBER	89.81	31.66	147.96	0.87	178.75

The fitted number of HIV infections was calculated by optimum SARIMA(1, 0, 1) model from 2019 January-2019 October. The fitted number or the inbound forecast was similar to the observed number of HIV cases.

5. Summary

This study revealed that SARIMA(1, 0, 1) (1, 0, 1)₁₂ without drift is the best fit mathematical model forecasting monthly cases of Human Immunodeficiency Virus (HIV) of Minna population. Time series data which is monthly HIV new cases in Minna General Hospital (year 2007-2018) was used. Models such as ARMA, ARIMA and SARIMA were used with a monthly dataset from “January 2007”, to “December, 2018”. The preliminary analysis of the data obtained shows that the distribution of the monthly HIV cases in Minna is stationary at first difference and result of Jarque-Bera statistic revealed that Minna HIV data is not normally distributed as the probability-values is less than 1% and 5%. The Parameter of the ARMA models and Models selection were estimated with most of the parameter significant at 1% and 5%. AIC was used to select the best model that was used for ARIMA and SARIMA models because it is the combination of AR and MA model. From the AIC, ARMA(1, 1) was selected to be the best model since it has the smallest AIC. The diagnostic test shows that ARMA(1, 1) shows no evidence that the residual is dependent, also the Q-Q plot result confirmed that the model is normally distributed.

More so, ARIMA of first and second difference were estimated and ARIMA(1, 0, 1) was the best model from the result of the AIC and diagnostic test carried out which revealed that the model was adequate and normally distributed using Box-Lung and Q-Q plot respectively. From the results of the parameter estimated, most of the parameters were significant and SARIMA(1, 0, 1) was selected to be the best model since it has the smallest AIC. A diagnostic test also

was evaluated which confirms that SARIMA(1, 0, 1) is an adequate model because the residual is not dependent and the Q-Q plot is normally distributed.

Furthermore, estimating the SARIMA model, shows that the parameter are significant at 1% and 5% and the diagnostic test indicate that SARIMA(1, 0, 1) \times (1, 0, 1)₁₂ without drift is an adequate model since there is no evidence of dependent in the residual of the model and the Q-Q plot is normally distributed. The monthly HIV cases in Minna time series were normal on its level but stationary at first difference. The range of monthly cases that occurred from year 2007 to 2018 is from 147 to 845 cases and the highest peak happened in May 2009 and May 2015 with 182 cases.

6. Conclusion

The following conclusions are derived from the findings presented:

- 1) The monthly HIV cases from 2017 to 2018 show an increasing trend, somewhat have a cycle and seasonality as well.
- 2) It found out that the highest increase of the HIV cases is on November 2012 to September 2013 and the highest decrease of the HIV cases is on January 2007 to September 2008.
- 3) The best model that can predict the HIV monthly cases is SARIMA(1, 0, 1) \times (1, 0, 1)₁₂ without drift.
- 4) The forecasted value of the created model has moderate increasing trend.
- 5) The average forecasted value is half of the actual value from January 2007.

Therefore, in this study based on the seasonal pattern of HIV prevalence in Minna, the SARIMA model is proposed as a useful tool for monitoring prevalence. The results of the study will be beneficial specifically to Niger State Government for prevention and control of HIV and Nigeria Government.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] World Health Organization Fact Sheet (2014) Global Update on the Health Sector Response to HIV. Geneva.
- [2] UNAIDS (2013) UNAIDS Report on the Global AIDS Epidemic 2013. Geneva.
- [3] Nigeria National Agency for the Control of AIDS (2012) Global AIDS Response: Country Progress Report. GARPR, Abuja.
- [4] Nigeria National Agency for the Control of AIDS (2010) United Nations General Assembly Special Session (UNGASS) Country Progress Report. Nigeria: January 2008 to December 2009.
- [5] Kee, M.K., Lee, J.H., Chu, C., *et al.* (2009) Characteristics of HIV Seroprevalence of Visitors to Public Health Centers under the National HIV Surveillance System in Korea: Cross Sectional Study. *BMC Public Health*, **9**, Article No. 123.
<https://doi.org/10.1186/1471-2458-9-123>

- [6] Fritzer, F., Gabriel, M. and Johann, S. (2002) Forecasting Austrian HICP and Its Components Using VAR and ARIMA Models. Working Papers 73, Oesterreichische National Bank (Austrian Central Bank).
- [7] Apa-Ap, R. and Tolosa, H.L. (2017) Forecasting the Monthly Cases of Human Immunodeficiency Virus (HIV) of the Philippines. *Indian Journal of Science and Technology*, **11**, 1-10. <https://doi.org/10.17485/ijst/2018/v11i47/121923>
- [8] Yu, H.-K., *et al.* (2013) Forecasting the Number of Human Immunodeficiency Virus Infections in the Korean Population Using the Autoregressive Integrated Moving Average Model. *Osong Public Health and Research Perspectives*, **4**, 358-362. <https://doi.org/10.1016/j.phrp.2013.10.009>
- [9] Demissew, T.G. (2015) Modeling and Projection of HIV/AIDS Epidemics in Ethiopia Using ARIMA. Master's Thesis, University of Nairobi College of Physical and Biological Sciences, School of Mathematics, Nairobi.
- [10] He, Z.R. and Tao, H.B. (2018) Epidemiology and ARIMA Model of Positive-Rate of Influenza Viruses among Children in Wuhan, China: A Nine-Year Retrospective Study. *International Journal of Infectious Diseases*, **74**, 61-70. <https://doi.org/10.1016/j.ijid.2018.07.003>
- [11] Abdoulaye, C., Wang, F. and Liu, X. (2016) Energy Consumption Forecasting Using Seasonal ARIMA with Artificial Neural Networks Models. *International Journal of Business and Management*, **11**, 231-243. <https://doi.org/10.5539/ijbm.v11n5p231>
- [12] Kibunja, H.W., Kihoro, J.M., Orwa, G.O. and Yodah, W.O. (2014) Forecasting Precipitation Using SARIMA Model: A Case Study of Mt. Kenya Region. *Mathematical Theory and Modeling*, **4**, 50-58.
- [13] Dickey, D. and Fuller, W. (1997) Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, **74**, 427-431. <https://doi.org/10.1080/01621459.1979.10482531>
- [14] Box, G.P. and Jenkins, G.M. (1976) Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.
- [15] Eni, D. and Adesola, A.W. (2013) Sarima Modelling of Passenger Flow at Cross Line Limited, Nigeria. *Journal of Emerging Trends in Economics and Management Sciences*, **4**, 427-432.
- [16] Yi, J., Du, C.T., Wang, R.H., *et al.* (2007) Applications of Multiple Seasonal Autoregressive Integrated Moving Average (ARIMA) Model on Predictive Incidence of Tuberculosis. *Chinese Journal of Preventive Medicine*, **41**, 118-121.