FORECASTING THE MONTHLY REPORTED CASES OF HUMAN IMMUNE DEFICIENCY VIRUS (HIV) FROM 2019 TO 2023

(A CASE STUDY OF MURITALA MUHAMMAD SPECIALIST HOSPITAL SOKOTO, SOKOTO STATE)

BY

OSENI SULIAT KOYINSOLA

HND/23/STA/FT/0013

SUBMITTED TO

DEPARTMENT OF STATISTICS,

INSTITUTE OF APPLIED SCIENCES,

KWARA STATE POLYTECHNIC, ILORIN

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF HIGHER NATIONAL DIPLOMA (HND) IN STATISTICS.

JULY, 2025

CERTIFICATION

This is to certify that this work was carried out by OSENI SULIAT KOYINSOLA with Matric Number HND/23/STA/FT/0013 in the Department of Statistics, Kwara State Polytechnic, Ilorin.

_____ _____

MRS ELEPO T.A ***DATE***

(Project Supervisor)

_____ _____

MISS. AJIBOYE R.A ***DATE***

(Project Coordinator)

_____ _____

MRS ELEPO T.A ***DATE***

(Head of Department.)

_____ _____

(EXTERNAL EXAMINER) ***DATE***

DEDICATION

It’s a privilege to dedicate this project to Almighty God, the one that give me the opportunity to start and finish the research work. Also to my beloved parents for their love, care, prayer and financial support through my programme.

\

ACKNOWLEDGEMENT

Glory be to Almighty Allah for giving me the inspiration and opportunity, strength and wisdom to write this project. He is worthy of thanks and adoration.

My special thanks goes to my supervisor in person of **MRS ELEPO T.A**, despite her engagement in the department, she found time to go through every manuscript I presented to her, I thereby remain grateful and God will continue to bless you and your family in general (Amin). Also special thanks to her again, **MRS ELEPO T.A,** as the Head of Department for her advices throughout my stay in the Department and special thanks also go to all our Lecturers in Department of Statistics for giving us all the required knowledge.

My sincere appreciation goes to my parent **(MR AND MRS OSENI),** My big daddy **(MR IBRAHIM)** and big sister **(MARIAM)** for their amazing support both financially and Morally, May God bless them and make them eat the fruits of their labour (Amin).

Finally, my appreciation goes to Muritala Muhammad Specialist Hospital, Sokoto for cooperation during the data gathering exercise, May God bless you abundantly (Amen)

ABSTRACT

This research work is a statistical analysis of the monthly reported cases of human immunodeficiency virus (HIV) in Sokoto State, Nigeria between the periods 2019 – 2023. The data used was extracted from the Record Department of Muritala Muhammad Specialist Hospital Sokoto, Sokoto State. Time series analysis was employed to analyze, measure, compute variation and fluctuation in the number of HIV cases in Sokoto State. Descriptive Statistics method and the Autoregressive Integrated Moving Average (ARIMA) model were employed. The criteria of selection of models used were Akaike information Criteria (AIC), Schwartz Information Criteria (SIC) and Hanna-Quinn information criteria (HQC) of which ARIMA (0,1,1) and ARIMA (2,1,1) models were chosen as the best model. The Evaluation of models performance was estimated for both models and it was discovered that ARIMA (2, 1, 1) model is the most suitable model. The model was used to forecast for a year ahead.

Keywords: Human Immunodeficiency Virus (HIV), Time Series, Seasonality, Autoregressive Model, Moving Average Model, Autoregressive Integrated Moving Average (ARIMA)

TABLE OF CONTENTS

CHAPTER FOUR

CHAPTER FIVE

Chapter One

1.1 Introduction

The menace of HIV/AIDS has an alarming rate worldwide giving concerns to governments, non-governmental organizations and researchers. There has been numerous efforts by individuals, groups and organizations in past years to stem the rapid rate of the spread of this infection.

As the name implies, Human Immunodeficiency Virus (HIV) is a retrovirus that infects immune cells and destroys their functions. A retrovirus is a virus that uses RNA instead of DNA as its genetic material. It incorporates itself into the host cells' DNA by the use of an enzyme known as reverse transcriptase, this allows several copies of the virus to be produced in the host cells. Unlike certain other diseases, the human body cannot dispose of HIV. That indicates that once you get HIV, you have it forever. HIV is widespread all across the world, with Sub-Saharan Africa accounting for 70% of new infections each year (Awoleye & Thron, 2015; Kapila *et al*., 2016).

Heterosexual transmission is the most common method of transmission, which account for about 85% of all HIV infections. Southern Africa remains the epicenter of the pandemic, with significant rates of new HIV infections (Simon *et al*., 2006). Furthermore, studies revealed that Sub-Saharan Africa tends to bear a significant amount of the worldwide HIV burden. In 2010, Sub-Saharan Africa was home to 68% of all HIV-positive individuals. In 2012, it was projected that 3.5 million people in Nigeria live with the virus, putting Nigeria the third among countries with the highest burden of HIV

infection after India and South Africa (Awoleye & Thron, 2015). In 2021, 1.9 million Nigerians were living with HIV. In addition, 170,000 children up to the age of 14 were HIV positive (UNAIDS, 2021).

HIV is most commonly transmitted through sexual contact. It can also be passed from mother to child, during pregnancy or childbirth (by blood or fluid exposure), or through nursing. Non-sexual Transmission can also occur by sharing injection equipments, most commonly needles (Umunna & Olanrewaju, 2020).

In Nigeria, nearly 80% of new HIV infections are transmitted through heterosexual intercourse, with mother-to-child transmission following closely behind, 38% of new adult infections are attributable to female sex workers (FSW), injecting drug users (IDU), and men having sex with men (MSM), accounting for 3.5% of the adult population (Awoleye & Thron, 2015; Ummuna & Olanrewaju, 2020).

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 (World Health Organization, 2014). 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 (UNAIDS, 2013). Most cases of HIV infection in Nigeria occur via heterosexual means with epidemics more pronounced among the females. 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. 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. (Nigeria National Agency for the Control of AIDS (2012)

This study reviewed a discussion on the prevalence of HIV in Sokoto State, and developed a best model that predicts the monthly HIV cases in the State by means of the Autoregressive Integrated Moving Average (ARIMA) 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. HIV is found throughout the world and is prevalent in sub-Saharan Africa, accounting for 70% of new infections yearly (UNAIDS, 2013). Worldwide, an estimated 36.9 million people

are living with HIV and about 2 million people became newly infected in 2014 (Nigeria National Agency for the Control of AIDS (2010)).

Therefore, an epidemiological prediction of this virus is crucial to formulate policies and develop effective strategies; it is clearly advantageous to have a clear picture of its future occurrence, spread, and impact. This contributes to public health decision-making, resource allocation, and other intervention plans.

1.2 Statement of the problem

Despite ongoing efforts to control the spread of Human Immunodeficiency Virus (HIV), the numbers of reported cases continue to fluctuate, posing challenges for public health planning and intervention. Without accurate forecasting, it is difficult to anticipate trends, allocate resources effectively and implement timely prevention strategies. This study seeks to address the problem by forecasting the monthly reported HIV cases from 2019 to 2023 providing data –driven insights to support more effective public health response.

1.3 Significance of the Study

Forecasting the monthly reported cases of Human Immunodeficiency Virus (HIV) from 2019 to 2023 is crucial for understanding trends in the spread of the virus and evaluating the effectiveness of public health interventions. By analyzing these patterns, health authorities can allocate resources more efficiently, develop targeted prevention strategies and improve early detection efforts. This study also aids policymakers and healthcare providers in making informed decisions to reduce transmission rates, ultimately

contributing to better public health outcomes and progress toward HIV control and elimination goals.

1.4 Aim and Objectives

The aim of this study is to develop a best model that can predict the monthly HIV cases in Sokoto State. This is to be achieved through the following objectives:

1) to formulate time series models on the data collected;

2) to conduct a diagnostic check on the models formulated to determine the most suitable model;

3) to estimate the parameters of the various models and forecast the HIV prevalence.

1.5 Scope of the Study

The scope for any research work is usually discovered in order to give account to the extent of its inference. For this particular research the scope is restricted to the monthly reported cases of human immunodeficiency virus (HIV) from 2019 to 2023 in Muritala Muhammad Specialist Hospital Sokoto, Sokoto State and modeled the time series that can be used for future forecast for the data.

1.6 Sources of the data

The data used for this research was sourced from the Muritala Muhammad Specialist Hospital Sokoto, Sokoto State and the data is secondary since the method of collecting the data is documentary. The data was recorded in their daily register covering 60 months (5 years) from **2019-2023**

1.7 Definition of Terms

1. Forecasting

Forecasting refers to the process of using historical data and statistical or machine learning models to predict future values.

2. Human Immunodeficiency Virus (HIV)

HIV is a virus that attacks the body's immune system, specifically the CD4 cells (T cells), which help the immune system fight off infections. If not treated, HIV can lead to Acquired Immunodeficiency Syndrome (AIDS).

3. Monthly Reported Cases

These refer to the number of new HIV cases officially recorded and documented by health authorities or surveillance systems each month. These figures are typically collected through diagnostic testing and reporting by health institutions.

4. Time Series Data

This is a sequence of data points collected or recorded at regular time intervals—in this case, monthly.

5. Trend

A trend represents the long-term movement or direction in the time series data. In the context of HIV reporting, a trend may show whether case numbers are increasing, decreasing, or remaining stable over time.

6. Seasonality

Seasonality refers to regular, predictable fluctuations in data that recur at specific intervals, such as every month or season.

Chapter Two

Literature Review

Several studies have been written on HIV in Nigeria, the prevalence, the knowledge, awareness, factors responsible for its prevalence, and numerous further in the literature. However, only very few have been predictive in nature or forecast the prevalence or trend of HIV in Nigeria in the coming years.

Bashorun *et al*. (2014) carried out a cross-sectional study that involved pregnant women who attended Antenatal Clinics in 160 sentinel facilities for 10 years. The national and state prevalence were ascertained using unlinked-anonymous HIV testing (UAT), consecutive sampling, and Epi-Info. The burden of infection was also prognosticated using The Estimation and Projection Package with Spectrum. The results between the years 2001 and 2010 showed pivotal differences in the epidemic in different states. There was elevated HIV prevalence in eight countries and a reduction of more than 2% in six countries. States in the middle and south of Nigeria also had an advanced prevalence of HIV.

Badru *et al*. (2020) used cross-sectional data to examine comprehensive HIV knowledge, stigma, and comprehension of the threat of HIV among youthful adolescents. The data were obtained from the 2017 Akwa Ibom AIDS Indicator Survey. Data were scrutinized using descriptive statistics, Chi-square test or Fisher's exact test, and multivariable logistic retrogression models were used to determine associations with the study issues.

The results of the study indicated that there was significantly low comprehensive HIV knowledge among youthful adolescents, the majority of youthful adolescents reported stigmatizing tendencies towards people living with HIV, and most adolescents do not perceive themselves to be at threat of HIV, and HIV Prevalence among youthful adolescents was at 0.6.

Onovo *et al*. (2023) conducted a study to decide the level of state HIV seropositivity rates by scrutinizing the Nigeria HIV testing services (HTS) data for Nigeria from October 1, 2020, to September 30, 2021. The study used a Bayesian direct model with normal prior distribution and a Markov Chain Monte Carlo approach to estimate the HIV state-level prevalence for all the states in Nigeria, including FCT. The HIV seropositivity rates as the outcome variable were well acclimated for demographic, economic, biological, and societal covariates. The result shows a national HIV prevalence of 2.1% among grown-ups aged 15 – 49 times in Nigeria, which corresponds to roughly 2 million people living with HIV. It also shows a state-varied HIV prevalence in Nigeria, with Benue having the highest prevalence, followed by Rivers, Akwa Ibom, Edo, and Taraba, placing fourth and fifth independently. Jigawa had the smallest HIV prevalence.

Isichei *et al*. (2015) also carried out a cross-sectional study to examine the prevalence of HIV infection and associated risk factors among pregnant women in the Plateau state. 248 pregnant village women in prenatal clinics in five villages of Mangu Local

Government area of Plateau State, North Central Nigeria were recruited for the study. The results show a high prevalence of HIV in the area, at 3.2%. It also showed that there was an advanced rate of HIV infection among women who did not hesitate from sexual intercourse during gestation and women in polygamous marriages. Parous women were also noticed to be significantly more likely to be infected than nulliparous women.

Adeoye *et al*. (2021) assessed the prevalence of HIV among the crucial populations in Nigeria using HIV Testing Services handed by the Society for Family Health intervention, and data from the intervention between the times 2019 and 2020 were analyzed with SPSS interpretation 20. The results show an advanced prevalence of HIV among those 40 years and above, the male gender, and men who have coitus with men (MSM).

Box- Jenkins ARIMA fashion in prognosticating the number of adults who will be infected with HIV annually in Nigeria over the given period of 2019 to 2030 was applied by Nyoni and Nyoni (2020). The result of their study shows that the prevalence of new HIV infections in Nigerian adults will increase between the time 2019 and 2030 having an estimated figure of 101754 new HIV infections in Nigeria in 2019 and 114883 by 2030. However, this study is the only study in the literature that attempts to prognosticate the estimated rise and trend of HIV in Nigeria.

There remains a significant gap in the literature when it comes to forecasting the rise in HIV prevalence in Nigeria, underscoring the need for this study. This research aims to inform and influence programs focused on the prevention and control of HIV in the country. By applying an epidemiological prediction model such as the Autoregressive Integrated Moving Average (ARIMA), this study seeks to provide a clearer understanding of future HIV trends in Nigeria. The findings could play a crucial role in shaping national strategies aimed at reducing new infections and improving health outcomes for those living with HIV. This research intends to fill the current gap and serve as a valuable tool for public health planning and intervention in the study area. This study aims to model the prevalence of HIV infection among outpatients visiting Jibril Mai-Gwari I Memorial Hospital, Birnin Gwari, Kaduna State, Nigeria.

Apa-Ap *et al.,* (2017) 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-Jekins 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) × (0, 0,1)12.

Yu, H.-K., *et al.,* (2013) 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%.

Demissew (2015) 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.

He *et al.,* (2018) 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, Abdoulaye *et al.,* (2016) 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.

Kibunja *et al.,* 2014 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) × (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.

Chapter Three

Research Methodology

3.1 Introduction

This chapter is designed to explain the basic methodology of the research work. Hence the various materials and methods applied for the purpose of achieving the objectives of the study are going to be explained thoroughly below. Monthly reported cases of human immunodeficiency virus (HIV) from 2019 to 2023 in Muritala Muhammad Specialist Hospital Sokoto, Sokoto State were used for the study, Model identification, estimation and forecast will be carried out using the appropriate model.

3.2 Source of Data

The data used in this research work is a secondary data, It was extracted from the Record and Information Department of Muritala Muhammad Specialist Hospital Sokoto, Sokoto State from 2019-2023 which translate into 5 years.

3.3 Statistical Tools

Here it is tacitly assumed that information about the past is available in the form of numerical data. Ideally, at least 50 observations are necessary for performing time series analysis/modeling, as propounded by Box and Jenkins who were pioneers in time series modeling.

3.3.1Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

Autocorrelation function (ACF) and partial Autocorrelation function (PACF) of a stationary series are collected over time. The procedure usually serve as a prelude to selecting series of models that are subjected to statistical test to confirmation as the best and optimal model for a particular case. The methodology emphasized the use of some vital statistical approach such as determination of the model order, parameter estimation as well as model testing and forecast .behavior of ACF and PACF provide a suggestion of where the model can build from and corresponding model order. While use of ACF and PACF to determine model order is because both give sight information about the behavior of the time series. It is paramount and necessary requirement if a give series is cyclical, irregular, seasonal secular or process any of the two attributes must be stationary (Box and Jenkins 1970). When the series is stationary the order of the model that is the coefficients number can be determined.

3.3.2 Autoregressive (AR) model

Autoregressive model are base on the idea that the current value of the series, x_t can be explained as a function of p pass value $x_{t-1}, x_{t-2}, \ldots. x_{t-p}$ where p determines the number of steps into the pass needed to forecast the current value.

An autoregressive model of order p, abbreviated as AR (p) can be written as;

$$X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + ...\theta_p X_{t-p} + e_t \quad \ldots$$

(1)

where x_t represent the time series data at point t, $\theta_1, \theta_2 ....\theta_p$ are the parameters of the AR model and θ_p = 0 unless otherwise stated.

The letter p denotes the order of autoregressive model, defining how many previous values the current value is related to. The model is called autoregressive because the series is regressed on to past values of itself.

The error term $\{e_t\}$ in Equation (1) refers to the noise in the time series. Above, the errors were said to be iid. Commonly, they are also assumed to have a normal distribution with mean zero and variance δ^2.written as (NID $0, \delta^2$). The highest order p is referred to as the order of the model. By introducing the lag operator (back shift) B in equation (1) we obtain;

$$x_t = x_t B^0, \qquad X_{t-1} = x_t B^1,\ X_{t-2} = x_t B^2, ...x_{t-p} = x_t B^P$$

From the characteristics polynomial of order p, the model in log operators takes the following form

$$(1 - \theta_1 B - \theta_2 B^2 - ....\theta_P B^P) X_t = e_t \quad \ldots$$

(2)

The autoregressive operator $\theta(B)$ is defined to be

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_P B^P \quad \ldots(3)$$

More concise the AR model can be expressed as;

$$\theta(B) X_t = e_t \quad \ldots(4)$$

The mean of AR (2) model can be compute as follows (bear in mind that the model is stationary)

$X_t = m + \theta_1 X_{t-1} + \theta_2 X_{t-2} + ...\theta_p X_{t-p} + e_t$ Due to stationary condition

Mean = $\frac{m}{1 - \phi_1 - \phi_2 - ...\phi_p}$

and variance is defined as

Variance = $\left[\frac{1-\phi_2}{1+\phi_2}\right] \frac{\sigma_\varepsilon^2}{(1-\phi_2)^2 - \phi_2^2}$

3.3.3 Moving Average (MA) Model

A moving average model an alternative to the autoregressive representation in which the X_t on the left-hand side of the equation are assumed to be combined linearly, the moving average model of order q, abbreviated as MA (q) assumes the white noise (e_t) on the right –hand side of the defining equation are combined linearly to form the observed data .

A moving average model of order q, abbreviated as MA (q) can be written as; MA (q) can be written as;

$$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \quad \ldots$$

(5)

Where $\theta_1, \theta_2 ....\theta q$ are the moving average (MA) parameters in the model and ε_t, ε_{t-1}, $....\varepsilon_{t-q}$ are the white noise error terms $\theta_q = 0$ and ε_t is a Gaussian white noise series, with mean zero and variance δ_e^2 unless otherwise stated. The Moving Average model in lag operation B is of form

$$x_t = \varepsilon_t (1 + \theta_1 B + \theta_2 B^2 + ...\theta_q B^q) \quad \ldots (6)$$

The Moving Average operator $\theta(B)$ is of form;

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q \quad \ldots (7)$$

3.3.4 Autoregressive Moving Average (ARMA) models

This is a process of maxing Autoregressive and Moving Average model to form an Autoregressive Moving Average (ARMA) models for stationary time series. A general ARMA (p,q) model can be written as follows;

$$x_t = \theta_1 x_{t-1} + \theta_{t-2} + ... + \theta_p x_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ...\theta_q x_{t-q} \quad (8)$$

3.3.5 Autoregressive Integrated Moving Average (ARIMA) models

Autoregressive Integrated Moving Average (ARIMA) models are an extension to the class of ARMA models by adding the possibility to integrate a non-stationary process to a stationary. ARIMA (p,d,q) models are univariate time series models that consist of an autoregressive parameter (p), an order of integration (d) and moving average (q).

A process (x_t) is said to be an autoregressive integrated moving average (ARIMA) process if $\theta(B)\nabla^d x_t = (1-B)^d$ is the non- seasonal differencing operator of order d. to produce non-seasonal stationary of the d^{th} difference. Usually d = *1 or 2*.

3.3.6 Signal and Noise

The observed and recorded time series, say $\{X_n\}$, consists of two components:

1. **The signal**. This is the component of the data that contains information, Say $\{S_n\}$ this is the component of the time series that can be forecast.

2**. The noise**. This is the randomness that is observed, which may be due to numerous other variables affecting the signal, measurement imperfections, etc. Because the noise is random, it cannot be forecast.

3.3.7 Model Selection Criteria

When fitting models, there is the tendency of two or more models competing and for that reason it is appropriate to use good model selection criteria to select the most adequate model. In this study, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were the measures of goodness of fit that were employed to select the most adequate model. For a given data set, several competing models may be ranked according to their AIC, or BIC values with the one having the lowest information criterion value being the best.

In general the AIC, and BIC are given by;

AIC = 2k+nlog (RSS/n)

BIC = log (δ_e^2) +k/n xlog (n)

where

k is the number of parameters in the statistical model,

RSS is the residual sum of squares of the estimated model,

n is the number of observations in the data, δ_e^2 is the error variance.

3.3.8 Model Identification

In time series analysis, the most crucial steps are to identify and build a model based on the available data or information to suggest a suitable model to describe how the data has

been generated. If the data appears to be stationary no differencing is called for, and we identify d=0 where d is the order of difference of the series until its time plot appears to be stationary.

As mentioned earlier the input series of the ARIMA model need to be stationary, that is should have a constant mean, variance and autocorrelation through time. Therefore, usually the series' first need to be differenced until it is stationary. However, The first step in the identification of an ARIMA process is to examine the time plot if there is no seasonal variation. The three commonly used tools graphical methods in identification in time series are: - the time plot of the series, the time plot of autocorrelation at various lags (ACF) and the time plot of the partial autocorrelation function (PACF).

i. **Time plot of a series:-**This is a graphical representation of a time series of observed values $Y_1, Y_{2,} Y_3 ....., Y_K$ plotted against time observations. It is a useful tool for interpreting a set of autocorrelation co-efficient.

ii. **Time plot of Autocorrelation function:-**This is a graphical representation of the sample autocorrelation r_h versus h (the time lags). It is also said to be a plot that describes the correlation between values of the process at different point in time, as a function of the two times or of the time difference. we define the sample autocorrelation function to be the sequence of values

(1) $r_\tau = c_\tau / c_0, \qquad \tau = 0,1,...,T-1$, where

$$c_O = \frac{I}{N}\sum_{1}^{N}(y_t - \bar{y})^2 \text{ and } \quad c_t = \frac{1}{N}\sum_{t=1}^{N-K}(y_t - \bar{y})(y_{t-1} - \bar{y})$$

is the empirical auto covariance at lag τ and $\boldsymbol{c_0}$is the sample variance. One should note that, as the value of the lag increases, the number of observations comprised in the empirical auto covariance diminishes until the final element $C_t = N^{-1}(y_0 - \bar{y})(y_{t-1} - \bar{y})$.

3.3.9 Information Criterion

The Akaike information criterion is a measure of the relative goodness of fit of a statistical model. Akaike (1969, 73, 74) suggests measuring the goodness of fit for some particular model by balancing the error of the fit against the number of parameters in the model. It provides the measure of information lost when a given model is used to describe reality. It can be said to describe the tradeoff between bias and variance in model construction.

Hannan-Quinn information criterion (AIC) is a criterion for model selection. It is an alternative to Akaike information criterion (AIC).

Bayesian information criterion (BIC) or Schwarz criterion is another criterion for model selection among a finite set of models. BIC give a model with smaller order than AIC or AICc. The formula for the AIC, BIC, AIC and AICc are;-

$$AIC = 2 * K - 2 * \ln(L)$$

$$BIC = \ln(n) * k - 2 * In(L)$$

$$HQC = n \log(\frac{RSS}{n}) + 2 * k \log n$$

$$AIC_C = AIC + \frac{2(K+1)(K+2)}{n-k-2}$$

where k is the number of parameters in the statistical model.

L is the maximized value of the likelihood function for the estimated model. AICc is a modification of AIC by Hurvich and Tsai (1989) while Burnham and Anderson (1989) insist on using AICc regardless of sample size since it converge to AIC as n is large and n = the sample size.

3.4 Augmented Dickey-Fuller (ADF) test for Stationarity

In statistics and econometrics, an augmented Dickey–Fuller test (ADF) is a test for a unit root in a time series sample. It is an augmented version of the Dickey-Fuller test for a larger and more complicated set of time series models. If the null hypothesis is rejected it implies that the series is stationary. Otherwise, the unit root exists and is simply means that the series is non-stationary.

3.4.1 KPSS Test for Stationarity

The integration properties of a series may also be investigated by testing the null hypothesis that the series is stationary against a unit root. Kwiatkowski et al (1992) have derived a test for this pair of hypotheses. The null hypothesis is rejected when the test statistic is greater than appropriate significance level.

There are four important steps in analyzing time series data. These are broadly given as;

Description: When presented with a time series problem, the first step in the analysis is to packet the data and obtains simple descriptive measure of the main properties of the series. For some series such obvious feature and fairly simple model dominate the variation, for other series more sophisticated technique will be required to provide an adequate analysis.

Explanation: When observations are taken one, two or more variables, it may be possible to use the variation in one time series to explain in another series. This may lead to an understanding of the mechanism which generate a given time series, For example if it is of interest to see how sales are infected by price and economic conditions, analysis will allow hidden pattern (e.g. trends to be revealed)

Prediction: Given an observed time series, one may want to predict the future value of the services, this is an important takes in sales forecasting, and in the analysis of the economic and industrial time series. Prediction is closely related to control problem in many

situations. For example if one can predict that a manufacturing process is going to be more off target, then an appropriate correction action to be taken.

Control: When a time series is generated it measures the quality of manufacture process. The aim of the analysis may be controlling. The process control of several different kind change invariable over a period of time are considered to be a result of combine impact of force that constantly at work.

3.5 Forecasting

Forecasting is an important activity in economic, commerce, marketing and various branches of science, this project is concern with forecasting method based on the use of time series analysis. Suppose we have an observed time series X_1, X_2 ……., X_n and wish to forecast future values such as $X_N + hX_N + h$ the integer h is called the lead time or the forecast horizon (h for horizon) and the forecast of $X_h + h$+ h. made at time N for h step ahead will be denoted by XN(h). It is essential to specify both the time the forecast is made and the lead time.

3.6 Data Presentation

Table 1: The monthly data of record cases of HIV/AIDs from January 2019 to December 2023

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
2019	48	53	46	48	43	45	51	37	44	72	36	68
2020	55	41	59	65	34	60	34	49	69	47	50	64
2021	42	29	44	63	47	45	43	40	38	46	41	32
2022	42	33	32	25	53	127	78	135	72	81	95	79
2023	61	67	55	32	118	86	87	55	46	50	57	42

Chapter Four

Data Analysis and Discussion of Results

4.1 Introduction

This chapter deals with the analysis and discussion of the results obtained from the study. The chapter is sub-divided into three main headings namely; preliminary analysis, further analysis and discussion of results

4.2 Descriptive Statistics

This section explains the descriptive statistics of the data on Average of The monthly data of record cases of HIV/AIDs in Muritala Muhammad Specialist in Sokoto. The maximum (Max) and minimum (Min) values for the HIV/AIDS for the entire period were 135.00 and 25.00 respectively as shown in Table 4.2. Also, the monthly data of a record cases of HIV/AIDs in the entire period was Positive skewed and platy-kurtic in nature with the average and coefficient of variation (CV) being 55.600 and 40.723% respectively.

Table 4.1 Descriptive Statistics for the monthly data of a record case of HIV/AIDs in Muritala Muhammad Specialist in Sokoto, from Jan, 2019 to December, 2023

Variable	Mean	Min.	Max.	CV (%)	Skewness	Kurtosis
HIV/AIDs cases	55.600	25.000	135.000	40.723	1.641	2.910

The monthly record of HIV/AIDs cases for the various months indicates that, the highest average of monthly record of HIV/AIDs cases is in the month of June with 72.6 and the least average occurred in the month of February with 44.6 as shown in Table 4.3. In

terms of the maximum and minimum HIV/AIDs record cases, in the month of August and April had the highest and lowest values recorded cases respectively. The month of August has the largest variability followed by May as shown by their coefficient of variations (CV) in Table 4.3.Again, it was observed that HIV/AIDs cases have negative skewed in the month of March, April and December respectively while other months were negative skewed and also, observed that the number of recorded HIV/AIDS patient from January to December were leptokurtic in nature.

Table 4.2 Monthly Descriptive Statistics of the number of HIV/AIDs cases from Jan, 2019 to Dec. 2023

Month	Mean	Min.	Max.	CV (%)	Skewness	Kurtosis
January	49.6	42.0	61.0	16.78	0.367	-1.411
February	44.6	29.0	67.0	34.79	0.481	-1.199
March	47.2	32.0	59.0	22.29	-0.343	-1.047
April	46.6	**25.0**	65.0	38.53	-0.123	-1.646
May	59.0	34.0	118.0	**57.11**	1.342	0.053
June	**72.6**	45.0	127.0	47.82	0.797	-0.854
July	58.6	34.0	87.0	38.99	0.260	-1.601
August	63.2	37.0	**135.0**	**64.51**	1.387	0.097
September	53.8	38.0	72.0	28.93	0.310	-1.730

October	59.2	46.0	81.0	27.33	0.499	-1.575
November	55.8	36.0	95.0	41.86	1.075	-0.309
December	57.0	32.0	79.0	34.02	-0.256	-1.469

Having explored the general theory of ARIMA model in the preceding section, this section is dedicated to fitting the ARIMA of model to the monthly recorded of HIV/AIDS patients, with a view to achieve the pre-stated aim and objectives of the research work. The data employed in this study comprise of 60 monthly observations of the monthly recorded of HIV/AIDS in Muhammad Muritala Specialist Hospital in Sokoto, Sokoto State from January, 2019 to December, 2023.

4.3 Further Analysis

To identify the model of any time series data, it is necessary to identify the process of data generation and to achieve this, a study of pattern and behavior was done by plotting the time plot. ACF and PACF of the series.

4.3.1 Time plot of the number of HIV/AIDS cases .

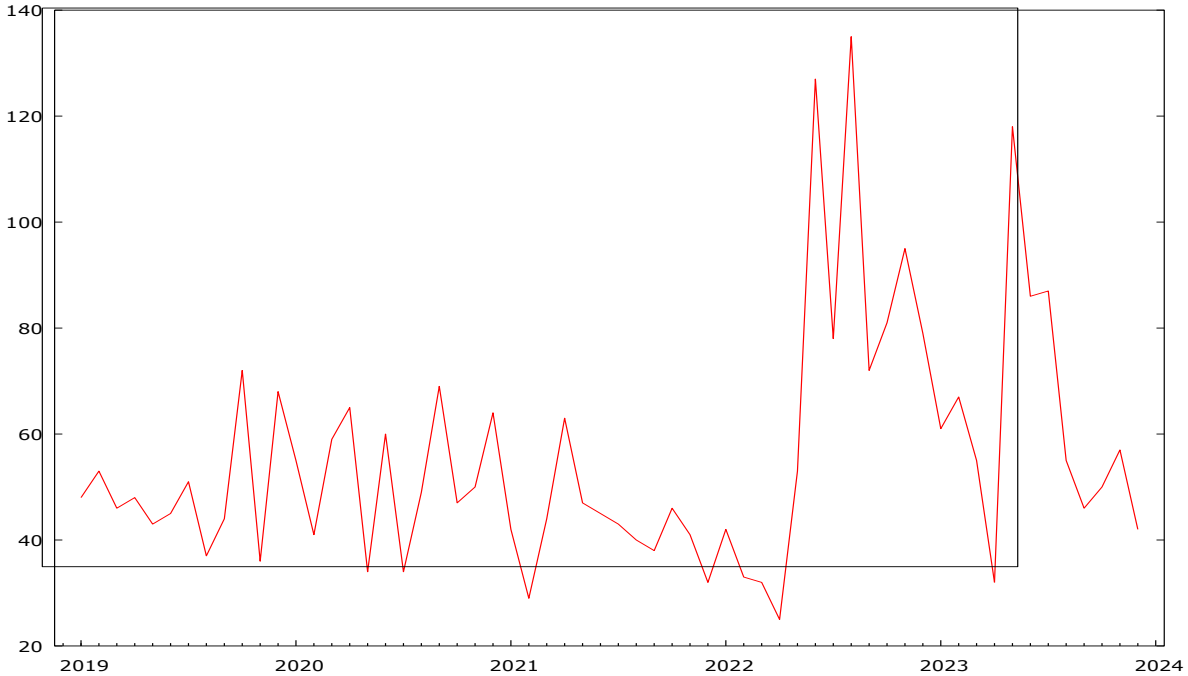

Figure 4.1 Time plot of the original series.

It could be observed that August, 2022 recorded the highest the number of cases of HIV/AIDs while April, 2022 recorded the lowest the number of cases of HIV/AIDs for the period of study. The monthly recorded of patients with HIV/AIDS showed a fluctuation pattern, suggesting that the mean and variance of the number of HIV/AIDs cases in Specialist Hospital in Sokoto have been changing over time. This indicates that, there is non-stationary, the mean and variance. It can be removed by taking differencing of the data as shown in Figure 4.1.

4.3.2 ACF and PACF of the number of HIV/AIDS cases

After the preliminary analysis was observed that the data can be modelled using Box-Jenkins procedure. Hence, in this section, the correlation structure of the number of HIV/AIDS cases using ACF and PACF are examined as shown in figure 4.2 below. The most striking feature of this correlogram is that the autocorrelation coefficients at various lags are very high, (at lag 1, 2 and 3); these are individually statistically significantly different from zero, out of the 95% confidence bounds. This is the typical correlogram of a non-stationary time series. The autocorrelation starts at a very high value and a decline (spikes down) very slowly toward zero as the lags lengthens, showing a purely MA series. After the first two lags, the PACF drops dramatically, and most PACFs after lag 2 are statistically insignificant, showing an AR of order 2.

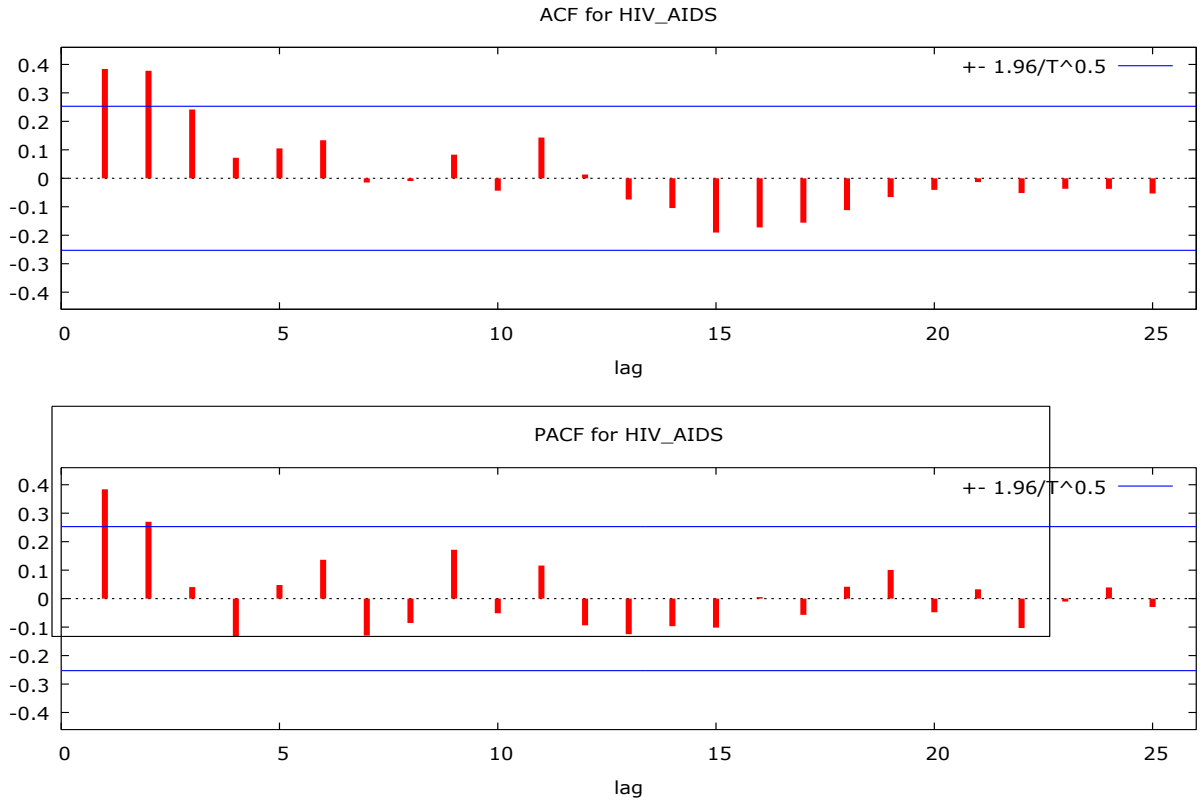

Figure 4.2: ACF and PACF plots of monthly HIV/AIDs.

4.3.3 Test for Unit Roots of number of HIV/AIDS cases

The KPSS test was carried out, the result showed that the data is not stationary before first differencing since the KPSS t-statistic is greater than the critical value at 5% and 10% levels of significance, implying that the null hypothesis (H_0) was rejected. Similarly, ADF test was carried out, the results are shown in the Table 4.4, we have strog evidence to fail to reject H_0,which says there is presence of unit root in the data before

first difference since the p-values are greater than the value of alpha at $1\%, 5\%\ and\ 10\%$ level of significance.

Table 4.3 ADF and KPSS tests of monthly recorded cases of HIV/AIDs

Test	**Without trend**					
KPSS	**Test Statistics**	**Critical value**		**Test Statistics**	**Critical value**	
	0.473	1%	5%	0.2258	1%	5%
		10%			10%	
		0.727	0.469 0.351		0.214	0.149
					0.121	
ADF	**Without constant**		**With constant**		**With constant and trend**	
	Test statistic	**p-value**	**Test statistic**	**p-value**	**Test statistic**	**p-value**
	-0.1957	0.6159	-3.3274	0.0263	-0.5262	0.0618

4.3.4 Time series plot of first differenced series.

Since the time series is not stationary, we have to make it stationary before we can apply the Box-Jenkins methodology. This can be done by differencing the series once. The figure 4.3 shows the time plot of the first difference of the original series. The diagram that the plots the series after one differencing shows that the variability of the series appears to be stable. The time plot of the series appears to be stationary for both mean and variance suggesting that the time series is stationary..

We can also see this visually from the ACF and PACF correlogram given below.

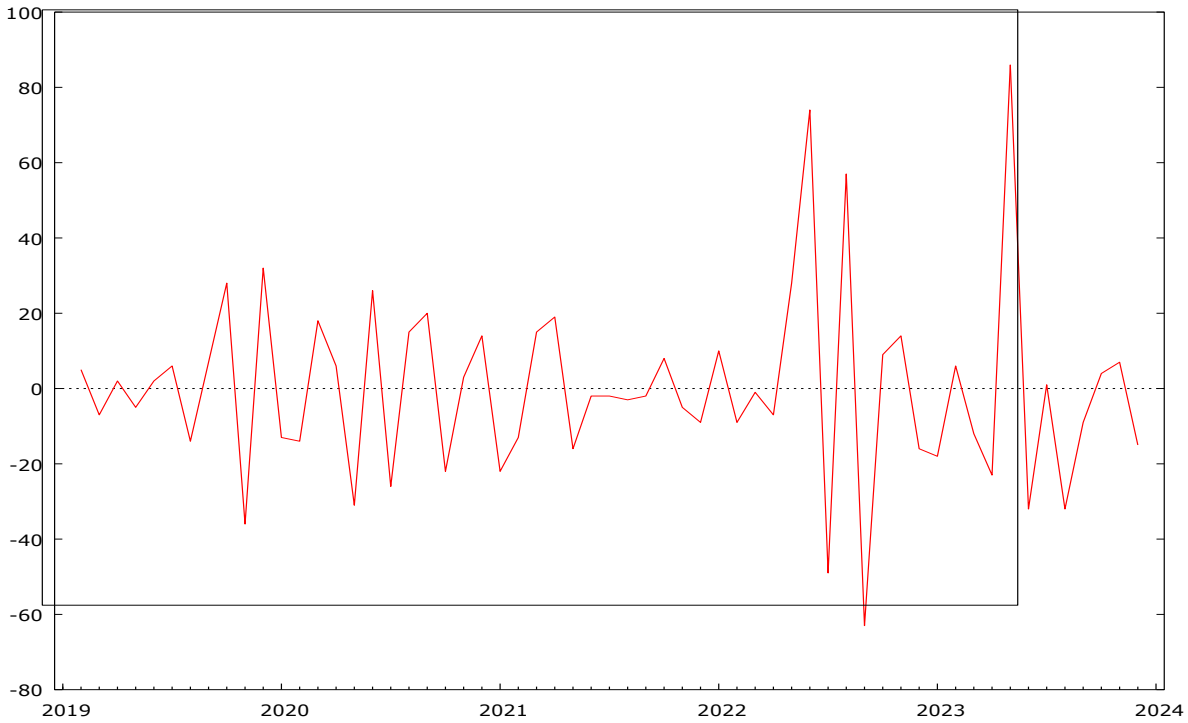

Figure 4.3: Time series plot of first differenced series

4.3.5 ACF and PACF first differenced series

From figure 4.4 shows the estimated ACF and PACF of series' first difference. Thus, the ACFs at lag 1. 10 and 11 seem statistically different from zero (at the 95% confidence limit, those lags are asymptotic and so can be considered approximate), but at all other lags, they are not statistically different from zero i.e. the ACF decayed after lag 1, 10 and 11. Also, the PACF plot is statistically different from zero only at lag 1, and 8; all other lags are not different from zero (at the 95% confidence limits). It could be seen that the

data is stationary and that there is no existence of seasonality. This suggest that AR and MA exists in the time series data. Hence, the series is non-seasonal but integrated and therefore the best model that will fit the data is ARIMA model.

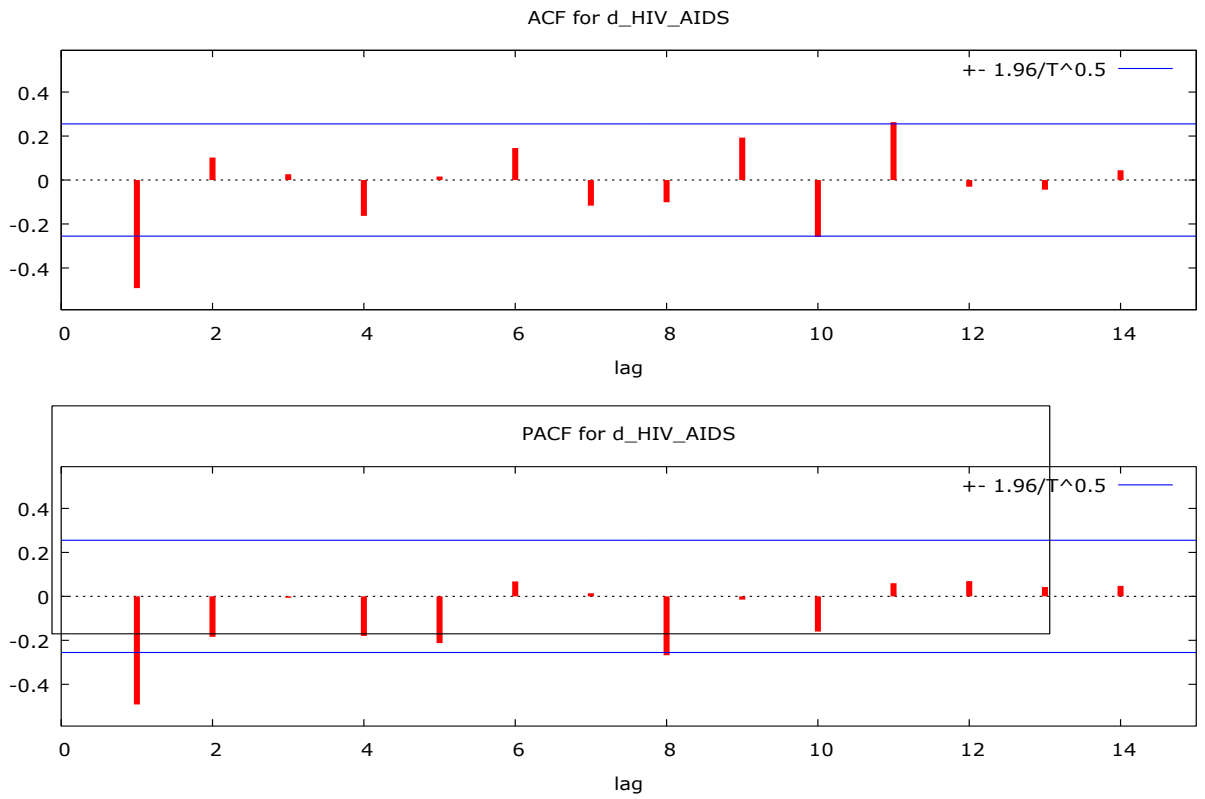

Figure 4.4: ACF and PACF first differenced series

4.3.6 Test for Unit Roots of number of HIV/AIDS cases

From Table 4.5, the data is stationary at first differencing, since the KPSS t-statistic is less than the critical value at 1%, 5% and 10% level of significance, implying that we fail to reject H_0, which claims stationary on the data at first difference. Similarly, from

Table 4.5, we have strong evidence to rejectH_0, which state that there is no presence of unit root in the data at first difference since the p-values are less than the value of alpha (α) at 1%, 5% and 10% level of significance.

Table 4.4 KPSS and ADF tests first difference of monthly recorded cases of HIV/AIDs

Test	**Without trend**					
KPSS	**Test Statistics**	**Critical value**		**Test Statistics**	**Critical value**	
	0.0469	1%	5%	0.0420	1%	5%
		10%			10%	
		0.727	0.469 0.351		0.214	0.149
					0.121	
ADF	**Without constant**		**With constant**		**With constant and trend**	
	Test statistic	**p-value**	**Test statistic**	**p-value**	**Test statistic**	**p-value**
	-3.6775	0.0001	-3.6334	0.0051	-3.5754	0.0319

4.4 MODEL IDENTIFICATION AND SELECTION

Since after first difference, the data is said to be stable. Thus, $d = 1, p = 0,1,2$ and $q = 0, 1,2,3$ the various tentative models identified for the number recorded cases of HIV/AIDs. Table 4.6 shows the results of model selection, 11 models were tested on the Alkaike information criteria (AIC); Bayesian information criteria (BIC); Hanna-Quinn

information criteria (HQC), and two models were selected for further examination namely: ARIMA (0, 1, 1) and ARIMA (2, 1, 1) models since they have the minimum values of AIC, BIC, and HQC compared to other models.

Table 4.5: Result of ARIMA model identification and selection

Model	AIC	BIC	HIC
ARIMA (0, 1, 1)	**534.239***	**540.472***	**536.672***
ARIMA (0, 1, 2)	536.205	544.516	539.449
ARIMA (0, 1, 3)	533.451	543.839	537.506
ARIMA (1, 1, 0)	537.322	543.555	539.755
ARIMA (1, 1, 1)	534.744	543.054	537.987
ARIMA (1, 1, 2)	538.198	548.586	542.253
ARIMA (1, 1, 3)	534.041	546.506	538.907
ARIMA (2, 1, 0)	537.339	545.649	540.583
ARIMA (2, 1, 1)	**533.147***	**543.535***	**537.202***
ARIMA (2, 1, 2)	535.129	547.595	539.996
ARIMA (2, 1, 3)	535.599	550.142	541.276

4.4.1 Model Estimation

Table 4.6 displays the parameter estimates of the ARIMA (0, 1, 1) model. Observing the p-values of the parameters of the model, it can be seen that the Moving Average

components were highly significant at the 1% level. The model appears to be the best model among the proposed models. Similarly, from Table 4.8, the p-value for the constant term is greater than 5% level of significance, therefore we cannot reject the null hypothesis and conclude that its statistically insignificant to the model. While the parameters phi_1 (ϕ_1), phi_2 (ϕ_2) and theta_1 (θ_1) are statistically significant to the model because the p-values are less than 5% and 10% level of significance ($\alpha = .05$ & 0.1), therefore, we reject the null hypothesis.

Table 4.6: The result of ARIMA (0, 1, 1) model estimation for the number of HIV/AID cases

Variable	Coefficient	Standard Error	Z-Statistic	P-value
Constant	0.064596	1.02443	0.06306	0.9497
Theta_1 (θ_1)	-0.636676	0.131881	-4.828	1.38e-06***

The estimated ARIMA (0, 1, 1) model for the number of HIV/AIDs cases is given by;

$$Y_t = 0.064596 + \varepsilon_t - 0.636676\varepsilon_{t-1} \quad (4.1)$$

Table 4.7: The result of ARIMA (2, 1, 1) model estimation for the number of HIV/AID cases

Variable	Coefficient	Standard Error	Z-Statistic	P-value
Constant	0.348273	0.281779	1.236	0.2165
Phi_1(ϕ_1)	0.262463	0.128214	2.047	0.0407**
Phi_2(ϕ_2)	0.245564	0.128077	1.917	0.0552*
Theta_1 (θ_1)	-1.0000000	0.0486555	-20.55	1.38e-06***

The estimated ARIMA (2, 1, 1) model for the number of HIV/AIDs cases is given by;

$$Y_t = 0.348273 + 0.262463y_{t-1} + 0.245564y_{t-2} + \varepsilon_t - 1.000000\varepsilon_{t-1} \quad (4.2)$$

4.4.2 Model Checking

Model checking is done through analyzing the residuals from the fitted model. In time series modeling, the selection of the best model fit to the data is directly related to whether residual analysis is performed well. One of the assumptions for this is that, for a good model, the residuals must follow a white noise process, that is, if the model fits the data well, the residuals are expected to be random; independent and identically distributed following the normal distribution (Bollerslev, et. al (1992), Muhammad, et. al (2002) and Ngailo E., 2011).

To ensure that, the fitted ARIMA (0, 1, 1) and ARIMA (2, 1, 1) model is adequate, both the Ljung-Box and ARCH-LM tests were performed. The Ljung-Box test as shown in

Table 4.8 and 4.9 revealed that, the residuals of the model were free from serial correlation at lags 12, 24, 36, and 48 since the p-values of test statistic exceeds the 5% significance level at all these lags. This indicates that the mean of the residuals of the model were finite. Further, the ARCH-LM test also shown in Table 4.8 and 4.9 revealed that, the residuals of the models were free from conditional heteroscedasticity, since the ARCH-LM test fails to reject the null hypothesis of no ARCH effect in the residuals of the equation at the 5% significance level. This shows that the residuals of the models were uncorrelated, thus have zero mean and have a constant variance over time; hence are white noise series.

Table 4.8:Ljung-Box Test and ARCH-LM Test of ARIMA (0, 1, 1) Model for number of HIV/AID cases

Model	Lag	Ljung-Box Test		ARCH-LM Test	
		Test statistic	P-value	Test Statistic	P-value
ARIMA (0,1,1) ARIMA (0,1,1)	12	6.1295	0.909	7.68085	0.809547
	24	20.0956	0.691	18.0994	0.79816

Table 4.9: Ljung-Box Test and ARCH-LM Test of ARIMA (2, 1, 1) Model for number of HIV/AID cases

		Ljung-Box Test	ARCH-LM Test

Model	Lag	Test statistic	P-value	Test Statistic	P-value
ARIMA (2,1,1) ARIMA (2,1,1)	12 24	7.1295 25.0956	0.520 0.893	9.60297 18.0672	0.650747 0.799736

4.4.3 Residual ACFs and PACFs of the selected models

The estimated residual ACF and PACF are shown in figure 4.5 and 4.6 below. The correlogram of both autocorrelation and partial autocorrelation gives the impression that the residuals estimated are purely random [i.e. the residual ACF and PACF seem statistically different from zero at some lags, but all other lags are not statistically different from zero]. Hence, there may not be any need to look for any other ARIMA models

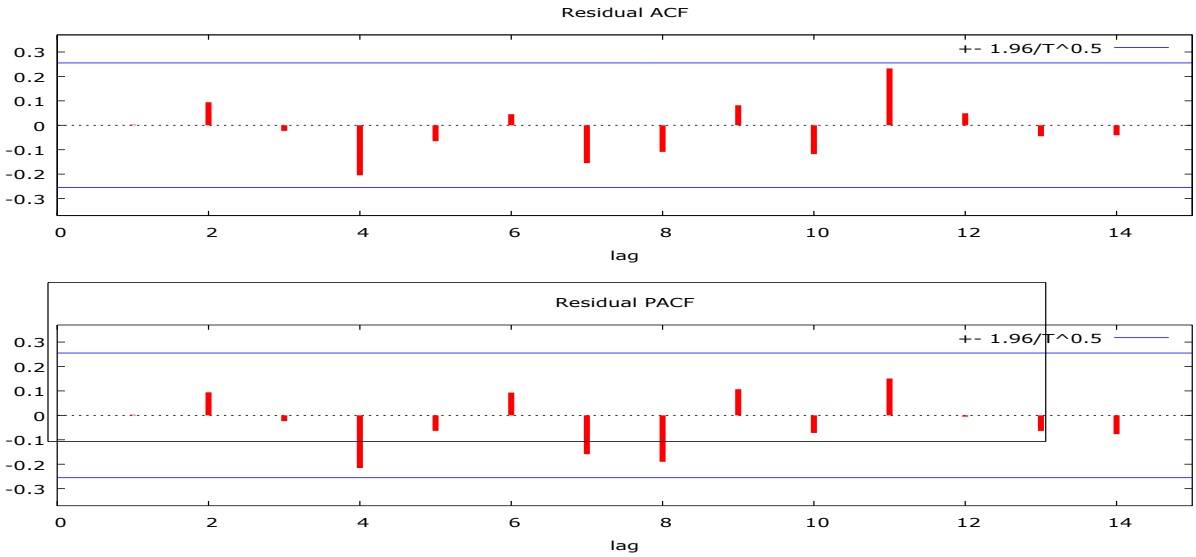

Figure 4.5: The Residual ACF and PACF of ARIMA (0,1,1) model for HIV/AIDs cases

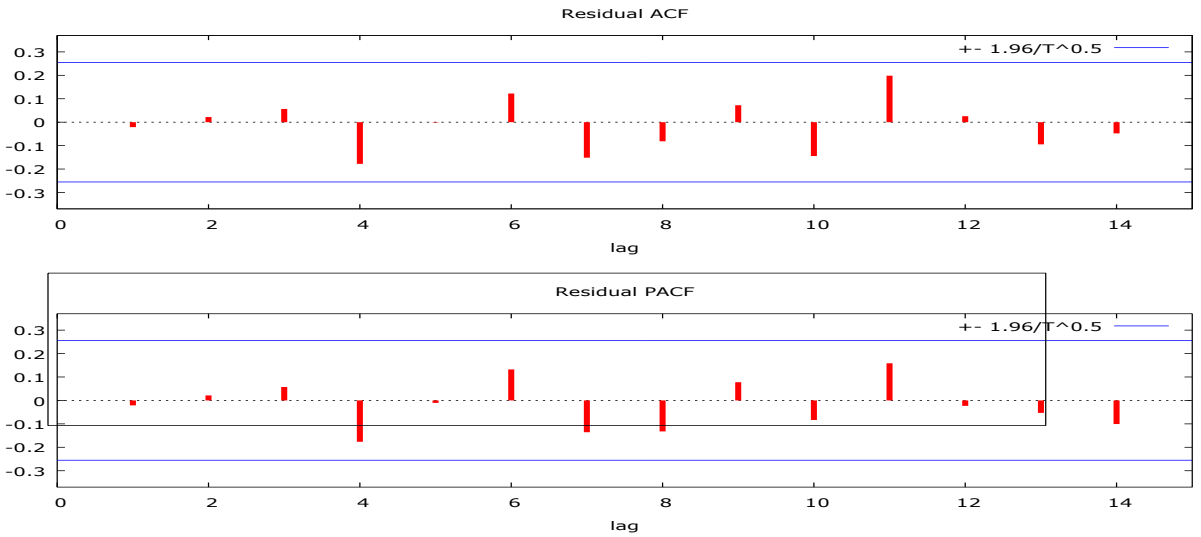

Figure 4.6: The Residual ACF and PACF of ARIMA (2,1,1) model for HIV/AIDs cases

4.5 Forecasting

In time series, forecasting is a mathematical way of estimating future values using present and historical values of the series (Aidoo, 2010). Forecasting as described by Box and Jenkins (1976) provide basis for economic and business planning, inventory and production control and optimization of industrial process.

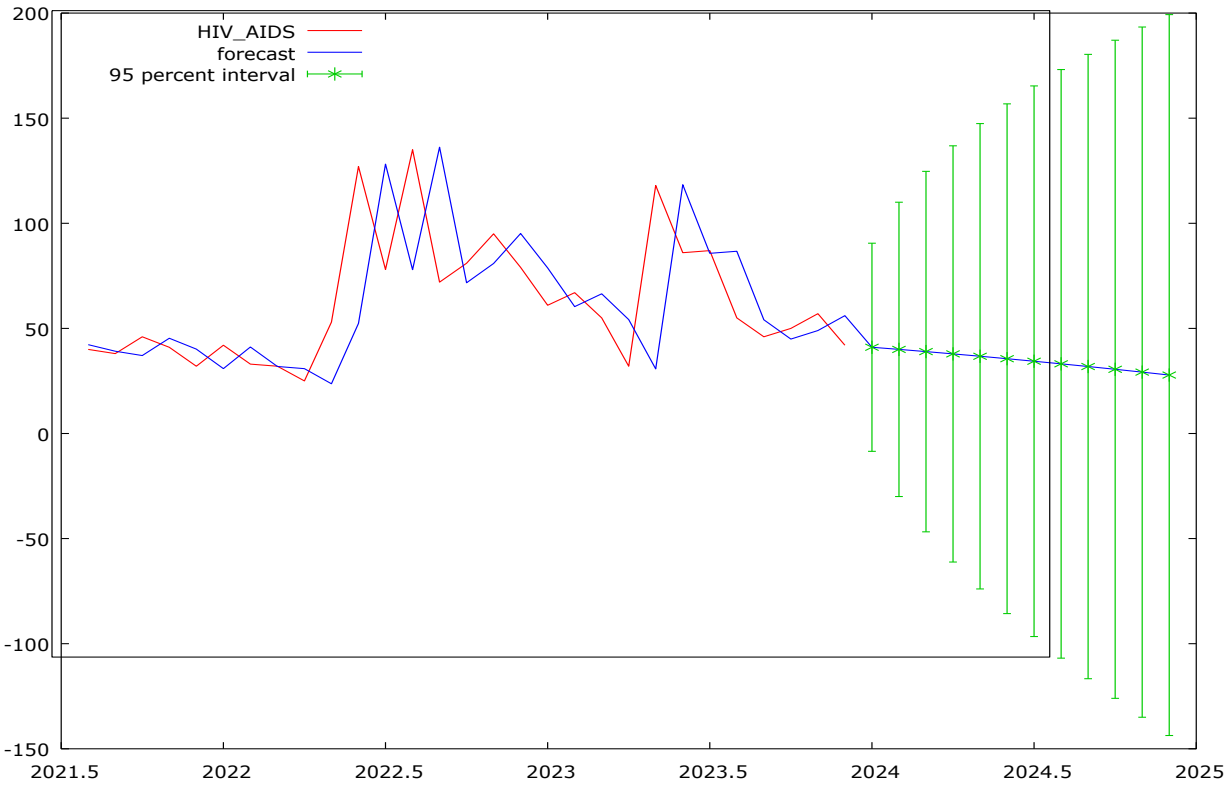

Figure 4.7: ARIMA (0, 1, 1) plot of 12 Months Forecasted result for number of HIV/AIDs cases

Figure 4.8: ARIMA (2, 1, 1) plot of 12 Months Forecasted result for number of HIV/AIDs cases

The figures 4.7 and 4.8 displays the transformed and the forecasted values of the monthly recorded of number of HIV/AIDs cases in Specialist Hospital in Sokoto produced by the ARIMA (0,1,1) and ARIMA (2, 1, 1) model for the next 12 months, that is, from January, 2024 to December, 2024. The figures also displayed how the forecasted values behave. It can be confirmed that the models somehow fit the data well.

Table 4.10: Result of ARIMA (0, 1, 1) 12 Months forecasted values (for 95% CI)

TIME	Prediction	std. error	95% confidence interval
2024:01	41.02	25.257	-8.48 - 90.52
2024:02	40.01	35.718	-30.00 - 110.01

2024:03	38.95	43.746	-46.79 - 124.69
2024:04	37.86	50.513	-61.14 - 136.87
2024:05	36.74	56.475	-73.95 - 147.43
2024:06	35.57	61.866	-85.68 - 156.83
2024:07	34.37	66.823	-96.60 - 165.34
2024:08	33.13	71.436	-106.88 - 173.14
2024:09	31.86	75.770	-116.65 - 180.36
2024:10	30.54	79.868	-126.00 - 187.08
2024:11	29.19	83.766	-134.99 - 193.37
2024:12	27.80	87.491	-143.67 - 199.28

Table 4.11: Result of ARIMA (2, 1, 1) 12 Months forecasted values (for 95% CI)

TIME	Prediction	std. error	95% confidence interval
2024:01	48.04	21.592	5.73 - 90.36
2024:02	45.86	23.567	-0.33 - 92.05
2024:03	44.76	26.764	-7.69 - 97.22
2024:04	44.38	29.703	-13.84 - 102.60
2024:05	43.37	32.123	-19.59 - 106.33
2024:06	42.54	34.493	-25.06 - 110.15
2024:07	41.67	36.683	-30.22 - 113.57
2024:08	40.75	38.747	-35.19 - 116.70
2024:09	39.82	40.711	-39.97 - 119.62
2024:10	38.86	42.583	-44.60 - 122.32
2024:11	37.87	44.376	-49.10 - 124.85
2024:12	36.86	46.100	-53.49 - 127.22

4.5.1 Model Validation Based on Forecasting Power

Table 4.12 Forecasting Evaluation of models performance

Performance Indicators	ARIMA (0, 1, 1)	ARIMA (2, 1, 1)
Mean Squared Error	448.95	406.23
Root Mean Squared Error	21.188	20.155
Mean Absolute Error	14.217	13.906
Mean Percentage Error	-8.7772	-10.831
Mean Absolute Percentage Error	25.361	26.108

The smaller the value of the error, the better is the forecasting performance of the model. From the Table 4.12, it could be seen that both the two selected models have shown good result (minimum performance indicators). But forecasting of ARIMA (2, 1, 1) model is more closer to the actual series. Therefore, the predictive power of ARIMA (2, 1, 1) model is better and suitable for monthly number of HIV/AIDS cases, as such the model best fit the data.

Chapter Five

Summary of Findings, Conclusion and Recommendations

5.1 Summary of Findings

The result of the descriptive statistics of the data on monthly data of recoded of HIV/AIDs cases in Specialist Hospital in Sokoto is shown in Table 4.2. The maximum (Max) and minimum (Min) values for the number of HIV/AID cases for the entire period were 135 and 25 respectively as shown in Table 4.2. Also, the Monthly recorded cases of HIV/AIDs cases for the entire period was positively skewed and platy-tokurtic in nature with the average and coefficient of variation (CV) being 55.600 and 40.737% respectively.

The average number of recorded HIV/AIDs cases for the various months indicates that, the highest average of monthly record of HIV/AIDs cases is in the month of June with 72.6 and the least average occurred in the month of February with 44.6 as shown in Table 4.3. In terms of the maximum and minimum HIV/AIDs record cases, in the month of August and April had the highest and lowest values recorded cases respectively. The month of August has the largest variability followed by May as shown by their coefficient of variations (CV) in Table 4.3.Again, it was observed that HIV/AIDs cases have negative skewed in the month of March, April and December respectively while other months were negative skewed and also, observed that the number of recorded HIV/AIDS patient from January to December were leptokurtic in nature.

To forecast the values of a time series, the basic Box-Jenkins strategy is as follows: We first observed the time plot and observed that there is strongly trend in the series. After we examined the series for stationarity, this step is done by computing the ACF and PACF and or by formal unit root analysis and found out that the series was not stationary, given the ADF **-0.1957** with p-value **0.6159**. We need to make it stationary, by take differencing the series once and re-plot the time plot and examined the series again, this is done by computing the ACF and PACF and or by formal unit root analysis and found out that the series was stationary, given the ADF **-3.6775** with p-value **0.0001**. To forecast, I selected three the models that have minimum of AIC, BIC and HQC and test for residuals by checking for white noise in each model, by Jarque-Bera test and Ljung-Box Q Test, which show that all the models are good for forecasting but we need one with minimum of Mean Squared Error, Root Mean Square Error, Mean Absolute error, Mean Percentage Error and Mean Absolute Percentage Error by done that, ARIMA (0, 1, 1) and ARIMA (2, 1, 1) were selected, with the ARCH-LM test, test statistic, **9.60297** with p-value **0.650747** at lag 12 and the Ljung-Box Q=**7.1295** with p-value **0.520,** we do not reject the null hypothesis and therefore concluded that the residuals are normally distributed.

5.2 Conclusion

The study has presented model, Autoregressive Integrated Moving Average ARIMA approach to forecast monthly number of recorded HIV/AIDs cases. The stages in the model building (that is the identification, estimation and checking) strategy has been explored and utilized. Based on minimum AIC, BIC and HQC values, the best fit ARIMA are (0, 1, 1) and ARIMA (2, 1, 1). After estimation of the parameters of selected models, a series of diagnostic and forecast accuracy tests were performed. Having satisfied all the model assumptions, ARIMA (2, 1, 1) models is judged to be the best model for forecasting in both the Box-Jenkins and Volatility modeling, and therefore these model was used for a twelve (12) month's forecast.

5.3 Recommendations

Based on the on the subject matter rising from the research, the following recommendations are made

- The government should implement community outreach programs during months with historically higher case reports
- Awareness, Prevention and Voluntary testing campaigns should be organized particularly targeting high-risk groups
- The Government should encourage routine testing as part of general health check-ups
- The Government should establish monitoring systems to ensure patients remain in care and achieve viral suppression

- The Hospital should foster collaboration between the State, Ministry of Health, Local and International NGOs to support data-driven HIV control efforts
- The Hospital should regularly update the forecasting model with new data to enhance prediction accuracy

References

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

Adeoye, O., Alau, K., Salome, Chika-Igbokwe, S., Nwaogu, P., Adu, R., Odeh, R., & Omoregie, G. (2021). Correlates of HIV prevalence among key population in Nigeria. *Journal of AIDS and HIV Research, 13*(2), 22-27.

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

Awoleye, O. J., & Thron, C. (2015). Determinants of human immunodeficiency virus (HIV) infection in Nigeria: A synthesis of the literature. *Journal of AIDS and HIV Research, 7*(9), 117-129. https://doi. org/10.5897/JAHR2015.0338

Badru, T., Mwaisaka, J., Khamofu, H., Agbakwuru, C., Adedokun, O., Pandey, S. R., & Torpey, K. (2020). HIV comprehensive knowledge and prevalence among young adolescents in Nigeria: evidence from Akwa Ibom AIDS indicator survey, 2017. *BMC Public Health, 20*(1), 45. https://doi.org/10.1186/s12889- 019-7890-y

Bashorun, A., Nguku, P., Kawu, I., Ngige, E., Ogundiran, A., Sabitu, K., & Nsubuga, P. (2014). A description of HIV prevalence trends in Nigeria from 2001 to 2010: what is the progress, where is the problem? *The Pan African Medical Journal, 18*(Suppl 1).

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.

Nigeria National Agency for the Control of AIDS (2012) Global AIDS Response: Country Progress Report. GARPR, Abuja.

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.

Isichei, C., Brown, P., Isichei, M., Njab, J., Oyebode, T., & Okonkwo, P. (2015). HIV Prevalence and Associated Risk Factors Among Rural Pregnant Women in North Central Nigeria. *American Journal of Health Research, 3*(1), 18-23. https://doi.org/10.11648/j. ajhr.20150301.14

Joint United Nations Programme on HIV/AIDS (UNAIDS). (2016). *HIV and AIDS estimates of children aged 0 to 14 in Nigeria*. https://www.unaids.org/en/ regionscountries/countries/nigeria.

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

Kapila A., Chaudhary S., Sharma, R. B., Vashist, H., Sisodia, S. S., Gupta, A. (2016). A REVIEW ON: HIV AIDS. *Indian Journal of Pharmaceutical and Biological Research, 4*(3), 69-73. https://doi.org/10.30750/ ijpbr.4.3.9

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.

Nyoni, S. P., & Nyoni, M. T. (2020). Adults Newly Infected with HIV in Nigeria a Box-Jenkins ARIMA Approach. *JournalNX*, 297-307.

Onovo, A. A., Adeyemi, A., Onime, D., Kalnoky, M., Kagniniwa, B., Dessie, M., & Meri, H. (2023). Estimation of HIV prevalence and burden in Nigeria: a Bayesian predictive modelling study. *EClinicalMedicine, 62*.

Simon, V., Ho, D. D., & Karim, Q. A. (2006). HIV/ AIDS epidemiology, pathogenesis, prevention, and treatment. *The Lancet, 368*(9534), 489-504. https:// doi.org/10.1016/s0140-6736(06)69157-5

Umunna, N. C., & Olanrewaju, S. O. (2020). Forecasting the Monthly Reported Cases of Human Immunodeficiency Virus (HIV) at Minna Niger State, Nigeria. *Open Journal of Statistics, 10*(3), 494-515. https://doi.org/10.4236/ojs.2020.103030

UNAIDS (2013) UNAIDS Report on the Global AIDS Epidemic 2013. Geneva.

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.

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