FORCASTING MODEL ON HIV INFECTIONS

(Case study of Ado/Odo- Ota Local Govt Ogun State Nigeria.)

Ilo H.O[1], Obanla O.James[2] , Hassan F.A.[3]

Department of Statistics, Ogun State Institute of Technology, Igbesa Ogun State, Nigeria.
Email Address: alameda444@yahoo.com

Abstract

This paper work is on HIV infection forcasting since the increase in the rate of recorded cases of HIV/AIDS in Adolodo-ota Local Government, Ogun State. The Box –Jenkins methods are used to analyse the problem . the statistical software used to analyse it are MINTAB and R Statistical Package.

The findings shows that there is an increase in the number of HIV Patients for some period of time but later exhibit a decreasing pattern.

Lastly, the appropriate fitted model for the data was Autoregressive [AR(I)] and number of HIV patients was stationary.

Keywords: Autocorrelation, Autoregressive, Auto-covariance, Auto-regressive Integrated moving average, stationary and Box-Jenkins.

INTRODUCTION

HIV is a virus, and AIDS is the condition the virus may cause. HIV is a virus that can lead to immune system deterioration. The term "HIV" stands for human immunodeficiency virus. The name describes the virus: Only humans can contract it, and it attacks the immune system. As a result, the immune system is unable to work as effectively as it should.

Our immune systems can completely clear many viruses in our bodies, but that's not the case with HIV. Medications can control HIV successfully by interrupting its viral life cycle.

Acquired Immune Deficiency Syndrome (AIDS) is the most unwelcome visitor of the 20th century .It has been disturbing and posing a great threat to human race and world population. Since the discovery of AIDS in 1981

1.1 STATEMENT OF THE PROBLEM

Over the last decade, there has been alarming increase in the rate of recorded cases of HIV/AIDS in our various hospitals especially in Ogun state. This has led to serious questions on the minds of the citizens in particular and the professionals in our medical field in general. However make it a complicated disease with a very serious economic and social effect on the victims. This study would focus on knowing the rate at which HIV/AIDS spread and suggest possible solutions in preventing the spread of the disease.

1.2 AIM AND OBJECTIVES OF THE STUDY

The aim of this study is to know the rate at which HIV/AIDS spread each year and to provide preventive measures against the spread of the disease. To achieve this aim, the following are the objectives

1. To examine the pattern of the data collected.
2. To formulate an appropriate model for the data collected.
3. To predict the future occurrence of HIV/AIDS in the period of 10years.

1.3 REVIEW OF STATISTICAL TOOL

In order to model the stochastic mechanism that generates the observed series, a vital assumption which is called stationary is needed. The basic idea of stationarity can be described as the probability

laws governing the process which do not change with time (Akgun, 2003). A time series is said to be strictly stationary if the statistical properties of the time series are unaffected by a change of time origin. Moreover, a time series is said to be covariance stationary if the first and second order moments of the time series are unaffected by a change of time origin. A strict stationary process is always a covariance process while a covariance process is a strict stationary process if and only if the covariance process is normally distributed. In practice, it is enough to take covariance stationary processes and in this thesis, the term stationary corresponds to covariance stationary (Sariaslan, 2010).

Using the fact that covariance stationary process does not depend on time but rather depends on time intervals, the auto-covariance function can be calculated easily. As a result of this, the autocorrelation function and the partial autocorrelation function, which have important roles to detect a time series model, can be found in a simple way. The Autocorrelation function is a useful tool for determining the moving average orders, and the partial autocorrelation function, which gives direct correlation between two observations of a time series. Thus, it is useful for determining the autoregressive orders. Models for stationary series can be studied under the Autoregressive Moving Average (ARMA) models which are also called Box-Jenkins models. A Moving Average (MA) model of order q is represented as a linear combination of present and q past terms of white noise error terms. White noise error terms are independently and identically distributed random variables with constant mean and variance. Moreover, an Autoregressive (AR) model of order p is represented as a linear combination of p past values of the series itself plus a white noise error term. Lastly, an Autoregressive Moving Average (ARMA) model of order (p, q) is represented as a linear combination of present and q past terms of white noise error terms and p past values of the series itself (Sariaslan, 2010).

1.4 The Box and Jenkins Approach

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model . The Box–Jenkins methodology applies ARIMA models to find the best fit of a time series to past values of this time series, in order to make forecasts. They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity. Box and Jenkins popularized an approach that combines the moving average and autoregressive approaches (Box & Jenkins, 1970). This resulted in autoregressive moving average model (ARMA). The Box-Jenkins model assumes that the time series is stationary. Box and Jenkins recommend differencing non-stationary series one or

more times to achieve stationarity. Doing so produces an ARIMA model, with the "I" standing for 'Integrated'. The Box-Jenkins methodology is a powerful approach to the solution of many time series analysis problem (Spyros, Wheelwright, & Hyndman, 1998). The basis of the Box-Jenkins approach to modelling time series consists of three phases: identification, estimation and testing, and application.

Box and Jenkins effectively put together in a comprehensive manner the relevant information required to understand and use univariate time series ARIMA models. The theoretical underpinnings described by Box and Jenkins and later by Box, Jenkins, and Reinsel (1994) are quite sophisticated, but it is possible for the non-specialist to get a clear understanding of the essence of ARIMA methodology

The main condition to use the Box-Jenkins methodology is that the time series of interest must be stationary

Stationary series is one whose basic statistical properties (like mean and variance) have no changes over time (Hanke & Reitsch, 1991), and differencing can be used to transform a non-stationary time series to a stationary one. This methodology does not assume that time series produces particular pattern. Instead of that, uses an iterative approach to identify potentially an appropriate model from general class of models. Then the selected model checked over the historical data to be sure that model is adequate (Chatfield, 1996).

The appropriate model, whose error terms are small, will be considered randomly distributed, and independent. If the specified model is not appropriate, the process is repeated using another class of models to improve the suggested one (Bowerman, 2005).

Box-Jenkins is a relatively accurate technique and powerful forecasting tool. However, it is quite complex and require computer analysis to perform many of computations which required identifying the appropriate model, estimating parameters and checking that the model is adequate (Bowerman, 2005). However, ARIMA models are criticized for their black box approach that makes no attempt to discover the factors affecting the system of interest

1.5 RESEARCH METHODOLOGY

Methodology is the systematic, theoretical analyses applied to a field of study or analyses of the body of the methods and principle associated with a branch of knowledge. It typically encompasses concept such as theoretical model phases and qualitative or quantitative techniques. A methodology does not

set out to provide solution but rather offer the theoretical support for understanding which method, set of methods or so called best practice that can be applied to a specific case. It analyzes the principle of method rules and postulate employed by a discipline. A methodology is the design process for carrying research or the development and is not itself an instrument for doing those things. It is used in careful study or enquiry for the purpose of presenting a particular phenomenon using a set of procedure or methods of data and result presentation. The concept of research is one that represents some orderly, organized, systematic and consistence pattern of thinking that gives structures and direction to investigation, knowledge of facts and methods. This chapter of this research work discusses the model used for the data analysis and the statistical software used for the analysis is MINITAB.

1.6 SOURCE OF DATA

The data used in this research work is a secondary data for a period of 10 years (2008-2017). The data is on monthly number/ record of HIV /AIDS patients in Ado/ Odo-Ota General Hospital, Ogun state.

1.7 THEORETICAL FRAME WORK

Univariate time Series Analysis shall be employed for this research work. Time series components estimated in the study are Autoregressive process, Moving average and Autoregressive Moving average which are the family of ARIMA (Autoregressive Integrated Moving average). ARIMA is a univariate time series which is based on the theory of forecasting the future using past values. Auto Regressive (AR) models were first introduced by Yule in 1926. They were consequently supplemented by Slutsky who in 1937 presented Moving Average $X_t = \emptyset_t X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + e_t - \emptyset_1 e_{t-1} - \emptyset_2 e_{t-2} - \cdots \emptyset_q e_{t-q}$ (MA) schemes. It was Wold (1938), however, who combined both AR and MA schemes and showed that ARMA processes can be used to model all stationary time series as long as the appropriate order of p, the number of AR terms, and q, the number of MA terms, was specified. This means that any series $\mathbf{X_t}$ can be modeled as a combination of past $\mathbf{X_t}$ values and/or past $\mathbf{e_t}$ errors, ……..... (1)

Using (1) for modeling real life time series requires three steps:

First the original series X_t must be transformed to become stationary around its mean and its variance.

Second, the appropriate order of p and q must be specified.

Third, the value of the parameters $\phi_1, \phi_{2,\ldots,}\phi_p$ must be estimated using some non-linear optimization procedure that minimizes the sum of square errors or some other appropriate loss function.

❖ STANDARD TIME SERIES MODELS

i. White noise process order
ii. General linear process order
iii. Auto-regressive process of order
iv. Moving average process order
v. Auto-regressive moving average
vi. Auto-covariance and auto-correlation functions

1.7.1 WHITE NOISE PROCESS ORDER

A time series is called a white noise process or purely random process if it satisfies the following conditions.

$$\boldsymbol{E(X_t) = 0}$$

$$\boldsymbol{E\left(X_t^{\,2}\right) = \sigma^2}$$

$$\boldsymbol{E(X_t X_h) = 0}\ for\ all\ values\ of\ \boldsymbol{h \neq t}$$

A time series ($\boldsymbol{X_t}$) is said to be random walk if $\boldsymbol{X_t = X_{t-1} z E_t}$ **if 't'** assumes the following numbers

0, 1, 2 … then, when t = 0 random walk is $\boldsymbol{X_t z E_t}$

Random walk P process is non-stationary as the mean and variance depends on it

1.7.2 GENERAL LINEAR PROCESS OF ORDER

A time series($\boldsymbol{X_t}$) is said to follow at general linear process if it satisfies the difference equations.

$$X_t = E_t + \phi_1 E_{t-1} + \phi_2 E_{t-2} + \ldots$$

$$= E_t + \sum_{r=0} \phi_r E_{t-r}$$

The general linear process (G.L.P) can also be written as

$$X_t = E_t + \Pi_1 X_{t-1} + \Pi_2 X_{t-2}$$

Proof

Let $X_t = E_t + \phi_1 E_{t-1} + \phi_2 E_{t-2}$

Making E_t the subject of the formular, we have

$$E_t = X_t - \phi_1 E_{t-1} - \phi_2 E_{t-2} \quad \ldots equation\ (i)$$

$$E_{t-1} = X_{t-1} - \phi_1 E_{t-2} - \phi_2 E_{t-3} \quad \ldots equation\ (ii)$$

$$E_{t-2} = X_{t-2} - \phi_1 E_{t-3} - \phi_2 E_{t-4} \quad \ldots equation\ (iii)$$

$$E_{t-3} = X_{t-3} - \phi_1 E_{t-4} - \phi_2 E_{t-5} \quad \ldots equation\ (iv)$$

Substituting equation (ii) and (iii) in equation (i), we have

$$E_t = X_t - \phi_1[X_{t-1} - \phi_1 E_{t-2} - \phi_2 E_{t-3}] - \phi_2[X_{t-2} - \phi_1 E_{t-3} - \phi_2 E_{t-4}]$$

Make X_t the subject of the formular, we have

$$X_t = E_t + \phi_1[X_{t-1} - \phi_1 E_{t-2} - \phi_2 E_{t-3}] + \phi_2[X_{t-2} - \phi_1 E_{t-3} - \phi_2 E_{t-4}]$$

$$X_t = E_t + \phi_1 X_{t-1} - {\phi_1}^2 E_{t-2} - \phi_1\phi_2 E_{t-3} + \phi_2 X_{t-2} - \phi_2\phi_1 E_{t-3} - {\phi_2}^2 E_{t-4} + \ldots X_t$$
$$= E_t + \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + \ldots$$
$$X_t = E_t + \sum \Pi X_{t-2}$$

1.7.3 AUTO-REGRESSIVE PROCESS OF ORDER (p)

The autoregressive process of order p is denoted $AR_{(p)}$ and defined by

$$X_t = \sum_{r=1}^{p} \phi_r X_{t-r} + \epsilon_t$$

Where $\phi_1, \ldots, \phi_r$ are fixed constants and $\{\epsilon_t\}$ is a sequence of independent (or uncorrelated) random variables with mean 0 and varianceσ^2.

1.7.4 MOVING AVERAGE PROCESS OF ORDER (q)

The moving average process of order q is denoted $MA_{(q)}$ and defined by

$$X_t = \sum_{s=0}^{q} \theta_s \epsilon_{t-s}$$

Where $\theta_1, \dots, \theta_q$ are fixed constants, $\theta_0 = 1$ and $\{\epsilon_t\}$ is a sequence of independent (uncorrelated) random variables with mean 0 and variance σ^2.

Using white noise process to power that

$E(X_t) = 0$

$Var(X_t) = \sigma$

The general linear process (G.L.P) is written as

$X_t = \sum \phi_r X_{t-r}$ Taking expected value of both sides, we have

$$E(X_t) = E\left(\sum \phi_r X_{t-r}\right)$$

Using backward shift operator, we have

$$E(X_t) = \sum \phi E(X_{t-r})$$

$$E(X_t) = \sum \phi \beta_r E(X_{t-r})$$

$$since\ E(X_t) = 0$$

$$Therefore\ E(X_t) = \sum \phi E(X_{t-r}) = 0$$

$$Var(X_t) = E(X_t - \mu)(X_t - \mu)$$

$$Var(X_t) = E({X_t}^2 - 2\mu X_t + \mu^2)$$

$$var(X_t) = E({X_t}^2) - 2\mu E(X_t) + \mu^2$$

$$since\ \mu = E(X_t)$$

$$var(X_t) = E({X_t}^2) - 2E(X_t)E(X_t) + E(X_t)E(X_t)$$

$$var(X_t) = E\left({X_t}^2\right) - 2[E(X_t)]^2 + [E(X_t)]^2$$

$$var(X_t) = E\left({X_t}^2\right) - E(X_t)$$

$$since\ E(X_t) = 0$$

$$Var(X_t) = E({X_t}^2)$$

Using the white noise process, the general linear process is written

$$X_t = \sum \phi_r X_{t-r}$$

Therefore $var(X_t) = E[\sum \phi_r X_{t-r}]^2$

$$var(X_t) = E(\sum {\phi_r}^2 \sum {X_{t-r}}^2)$$

When r = 0

$$var(X_t) = E(\sum {\phi_0}^2 \sum {X_t}^2)$$

$$var(X_t) = \sum {\phi_0}^2 E(\sum {X_t}^2)$$

$$since\ E({X_t}^2) = \sigma^2 and \sum \phi_0 = 1$$

$$var(X_t) = \sigma^2$$

1.7.5 AUTO-REGRESSIVE MOVING AVERAGE [ARMA$_{(p,q)}$]

Auto-regressive moving average process is the combination of the auto-regressive process (AR) and moving average (MA). This is called the principle of PARSIMONY. This process is appropriate for situations where an original simpler series is aggravated overtime e.g. where monthly data are aggravated to give quarterly or yearly data.

A stochastic process ($\boldsymbol{X_t}$) follows an auto-regressive moving average of order **(pq)** if it satisfy the following equation.

$$X_t - [\theta_1 X_{t-1} + \ \dots\ \theta_p X_{t-p}] = X_t + \theta_1 X_{t-1} + \ \dots\ \theta_r X_{t-s}$$

$$X_t - \sum_{r=1}^{p} \theta_r X_{t-r} = \sum_{s=0}^{q} \theta_r X_{t-s}$$

1.7.6 AUTO-COVARIANCE AND CORRELATION FUNCTIONS

Auto-covariance: The covariance between X_t and X_{t+k} separated by lag_k (either to the left or right) is called auto-covariance of lag_k, denoted by γ_k and it is given by

S/N	YEAR	QTR1	QTR2	QTR3	QTR4
1.	**2008**	51	42	147	157
2.	**2009**	156	163	194	182
3.	**2010**	225	228	181	152
4.	**2011**	66	127	128	134
5.	**2012**	161	155	242	223
6.	**2013**	222	267	241	205

$$\gamma_k = cov(X_t, X_{t+k}) = \sum (X_t - \mu)(X_{t+k} - \mu) \quad k = 0, \pm 1, \pm 2, \ldots$$

Where $\mu = \sum X_t$ and is constant for all t

$$when\ \gamma_0 = \sum (X_t - \mu)(X_t - \mu)$$

Table 1: Quarterly Number Occurrence of HIV patient (2008-2017)

7.	**2014**	161	160	141	110
8.	**2015**	100	90	142	166
9.	**2016**	175	77	87	125
10.	**2017**	138	161	173	144

$$\gamma_0 = \sum (X_t - \mu)^2$$

$$\gamma_0 = var(X_t)$$

$$k = 0, \qquad \gamma_0 = var(X_0)$$

$$proof, \qquad \gamma_0 = \sum (X_t - \mu)(X_{t+0} - \mu)$$

In practice, estimation of auto-covariance is

$$\gamma_k = C_k = \frac{1}{N}\sum (X_t - \bar{X})(X_{t+k} - \bar{X})$$

Where N is the sample size and $\bar{X}$ is the sample mean.

Auto-correlation: The auto-correlation of lag_k of a time series X_t is given by

$p_k = \frac{\gamma_k}{\gamma_0}\ for\ k = 0,1,2,\dots n$. So that we can have auto-covariance (γ_k) series and auto-correlation sequence (p_k)

2.0 DATA PRESENTATION AND ANALYZINIG

DESCRIPTIVE STATISTICS FOR THE NUMBER OF HIV PATIENTS

The summary of the descriptive statistics of this research work data is presented in the table below. The skewness is an indicator of the asymmetry or deviation of variable from a normal distribution with an expected value of zero, the kurtosis defines the degree of flattening or peakedness of a distribution with an expected value of three and jarque-Bera statistic determine the normally or otherwise of a distribution.

Table 2: Descriptive statistics of HIV patients

STATISTICS	HIV PATIENTS
Mean	154.97
Median	156.65
Std. dev.	52.57
MIN	42
MAX	267
Kurtosis	-0.07
Skewness	-0.39
Trimmed	155.78
Mad	40.03
S.E	8.34
Range	225
Number	40

2.1 DATA ANALYSIS

Test for stationarity

To make statistical inference about the structure of a time series, it is prerequisite to make some assumptions about the structure. The most important assumption is the stationarity. This is to

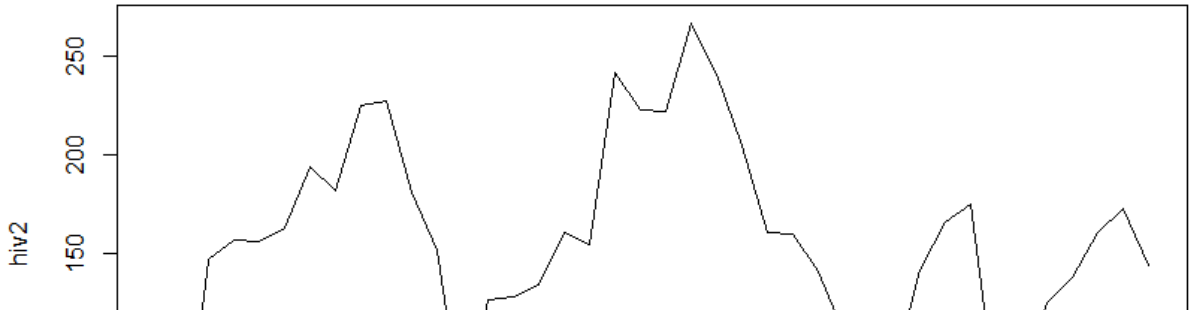

Figure1: Time plot of the HIV patients from Jan.2008 – Dec. 2017

determine that the process is in a statistical equilibrium. We therefore determine the stationarity of the time series data by plotting a time series graph of the original data. This is represented

The result in figure 1 shows a time series plot of the original data. It is clear from this figure that there is no increasing trend in mean and variance of the time series plot. There do not seem to be any seasonal pattern in the data. The original series is stationary and non-seasonal

2.2 TREND

Table 3: Model Summary and Parameter Estimates for The Linear Trend

Min	1Q	Median	3Q	Max
-107.834	-28.428	2.278	25.066	114.460

Source: R Statistical package version 3.5.1 2017 Edition, by R Foundation for statistical computing.

Model of parameter	Estimate	Std. Error	T value	Pr(>|t|)	Decision	Conclusion
(intercept)	158.2231	17.2100	9.914	3.35e-11 ***	reject Ho	Significant
Xt	- 0.1584	0.7315	-0.217	0.83	reject Ho	Significant
Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 53.05 on 38 degrees of freedom Multiple R-squared: 0.01439, Adjusted R-squared: -0.01155 F-statistic: 0.5546 on 1 and 38 DF, p-value: 0.461						

The linear trend equation is given as $Y_t = 158.2231 - 0.1584(x_t)$

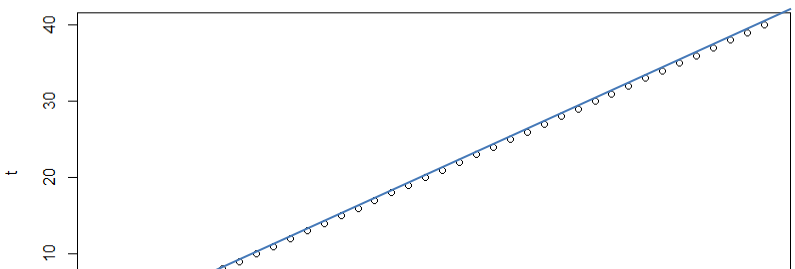

Fig 2. Ordinary Least Square Trend

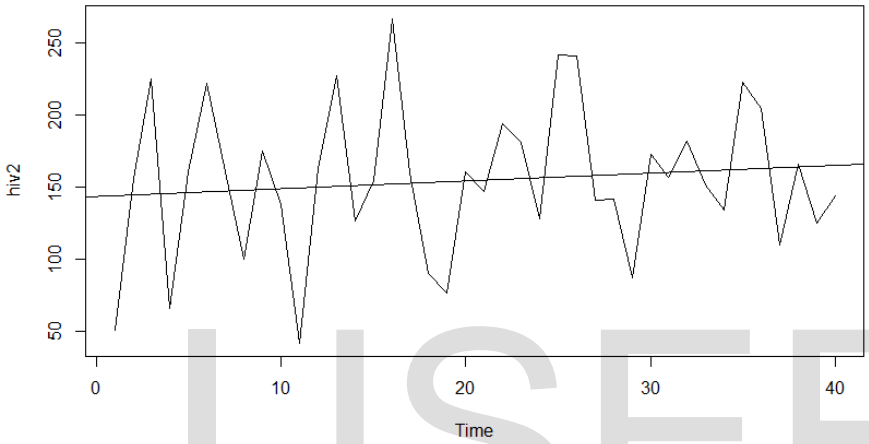

Fig 3: Observed Data and OLS Linear Trend

2.3 SEASONAL DECOMPOSITION

Seasonal decomposition

	seasonal	trend	remainder
2008 Q1	-5.413964	46.97972	9.4342417
2008 Q2	-6.704948	79.90921	-31.2042580
2008 Q3	11.575727	110.78443	24.6398419
2008 Q4	0.543137	142.87983	13.5770376
2009 Q1	-5.413964	161.77053	-0.3565689

2009 Q2 -6.704948 170.69062 -0.9856683

2009 Q3 11.575727 181.61455 0.8097206

2009 Q4 0.543137 198.40861 -16.9517514

2010 Q1 -5.413964 208.05840 22.3555598

2010 Q2 -6.704948 203.21080 31.4941515

2010 Q3 11.575727 177.85063 -8.4263523

2010 Q4 0.543137 142.09473 9.3621352

2011 Q1 -5.413964 122.21610 -50.8021350

2011 Q2 -6.704948 115.05525 18.6496967

2011 Q3 11.575727 125.71989 -9.2956141

2011 Q4 0.543137 140.39081 -6.9339443

2012 Q1 -5.413964 158.15973 8.2542381

2012 Q2 -6.704948 183.78305 -22.0781038

2012 Q3 11.575727 204.05721 26.3670629

2012 Q4 0.543137 224.81960 -2.3627385

2013 Q1 -5.413964 238.51369 -11.0997282

2013 Q2 -6.704948 238.42756 35.2773889

2013 Q3 11.575727 227.87034 1.5539329

2013 Q4 0.543137 204.25579 0.2010715

2014 Q1 -5.413964 178.84566 -12.4316970

2014 Q2 -6.704948 155.12144 11.5835037

2014 Q3 11.575727 135.16179 -5.7375132

2014 Q4 0.543137 117.99168 -8.5348137

2015 Q1 -5.413964 109.12231 -3.7083509

2015 Q2 -6.704948 115.73774 -19.0327918

2015 Q3 11.575727 133.26027 -2.8359991

2015 Q4 0.543137 145.25127 20.2055896

2016 Q1 -5.413964 136.27804 44.1359236

2016 Q2 -6.704948 118.63369 -34.9287396

2016 Q3 11.575727 107.42586 -32.0015863

2016 Q4 0.543137 116.99568 7.4611855

2017 Q1 -5.413964 139.78081 3.6331528

2017 Q2 -6.704948 148.16997 19.5349783

2017 Q3 11.575727 154.34217 7.0821022

2017 Q4 0.543137 158.74488 -15.2880125

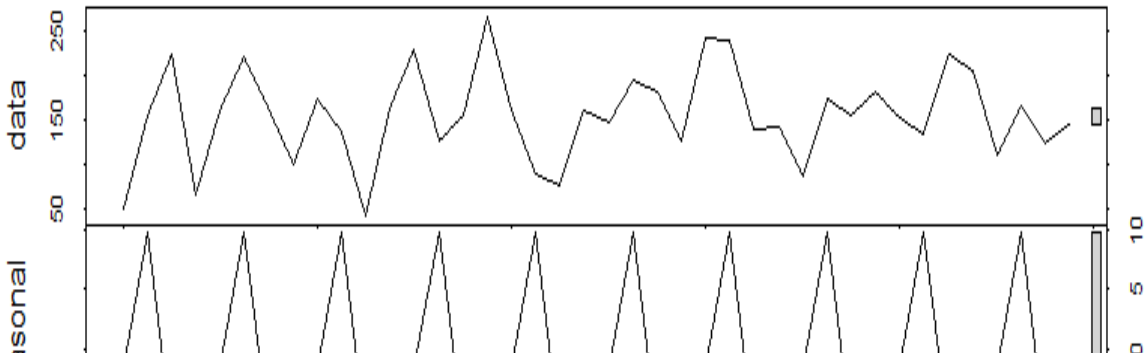

Fig. 4: Plot of Seasonal Decomposition

2.3 Seasonal decomposition (moving average trend, trend, seasonal, cyclical and irregular as remainder)

Discussion of result

According to the series time plot, the data appeared to be non-seasonal with low seasonal factor and the series time plot shows no seasonal fluctuation. To confirm the seasonality in the data, the level autocorrelation function and partial autocorrelation function ACF plot of the time series data to expose properly the seasonality in time series data.

2.4 CORRELOGRAM OF THE RESIDUALS

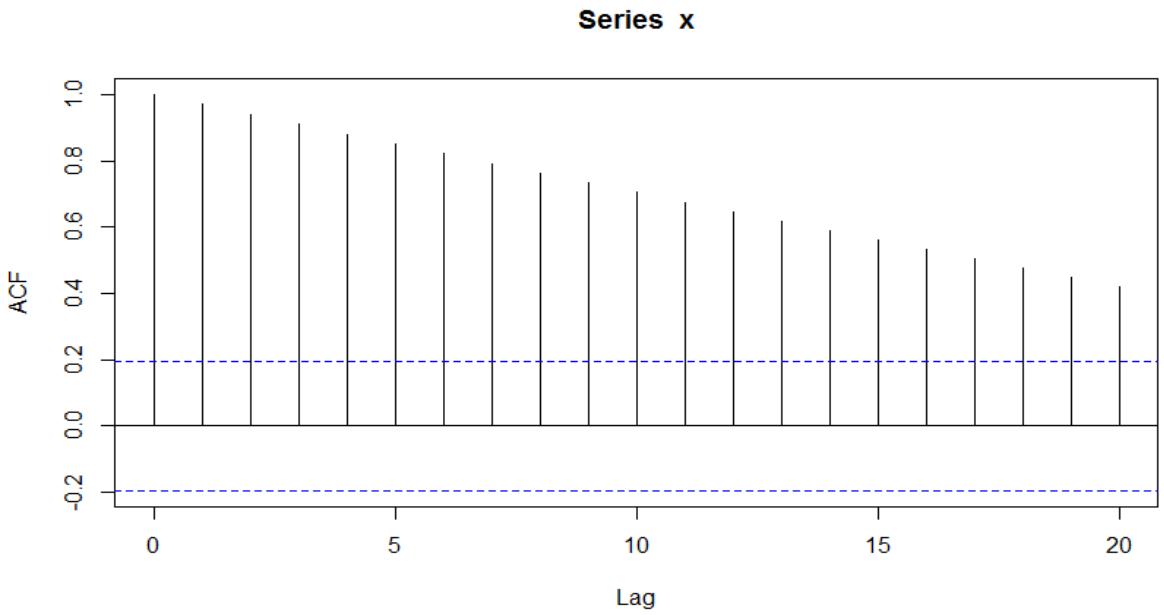

Figure 5 ACF & PACF of residuals for HIV patient

Figure 5 shows that ACF and PACF plots of the residuals are well within their two standard error limit. Hence the residuals are white noise model classified as ARIMA that there is no differencing involved, no AR & MA part since Y_t does not depend on ε_{t-1} therefore the model is efficient/ a good fit. The observed series ACF and PACF shows that the data is non-seasonal since there is no significant spike at equal interval on the plots.

2.5 TEST FOR STATIONARITY

Table 4: Augment Dickey – Fuller Test

D-F TEST Value	Lag Order	P – value	Decision	Conclusion

-3.8025	3	0.02993	Reject H_0	Stationary

H_0 : Not Stationary

H_1 : Stationary

Decision Rule: Reject H_0 if $P<0.05$

Table 5: Runs Test for Randomness

Median	**Obs, Exp. (Runs)**	**n1, n2**	**P value**	**Decision**	**Conclusion**
-1.0567	1, 1	0, 40	0.915	Do not Reject H_0	There is rand- omness in the series

H_0 : There is randomness in the series

H_1 : There is no randomness in the series

Decision Rule: Reject H_0 if $P<0.05$

2.6 Model Identification Stage

It is not always easy to determine the appropriate model to fit a time series data even after the time plot have been properly examined. It is also necessary to examine the two model identification tools which are autocorrelation Function (ACF) and partial autocorrelation function (PACF). This is represented graphically as shown in Fig 4.4.0 & 4.4.1 respectively

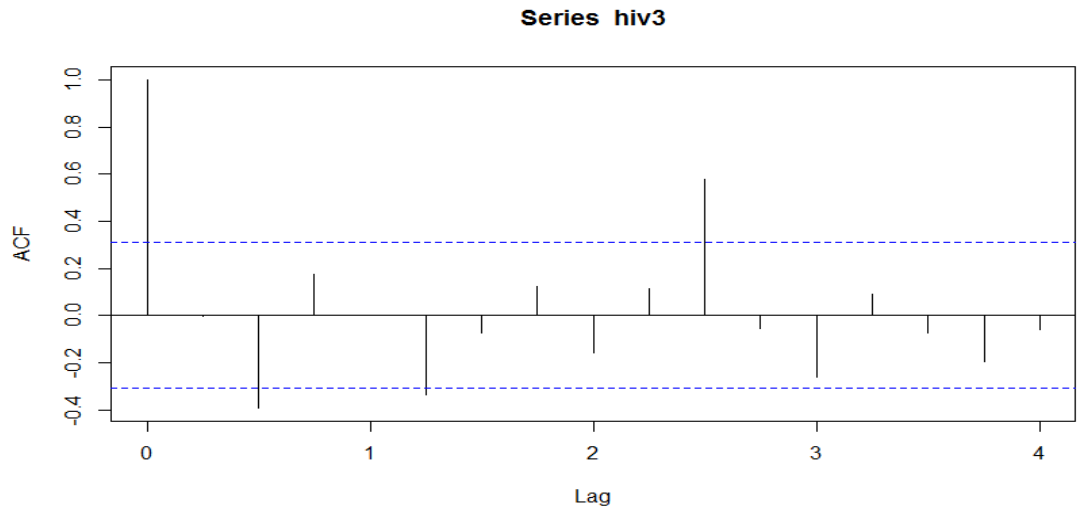

Fig 6: Autocorrelation Function of the original series

The Autocorrelation plot has 95% confidence band which is constructed based on the assumption that the process is a Moving Average (MA) process. From the graphical representation, the value of the autocorrelation function is less than one. Since the parameter of any given lag as shown in the Fig 6 is less than one, and decays slowly which confirms non – stationarity. Therefore, we check partial autocorrelation function plot to confirm this.

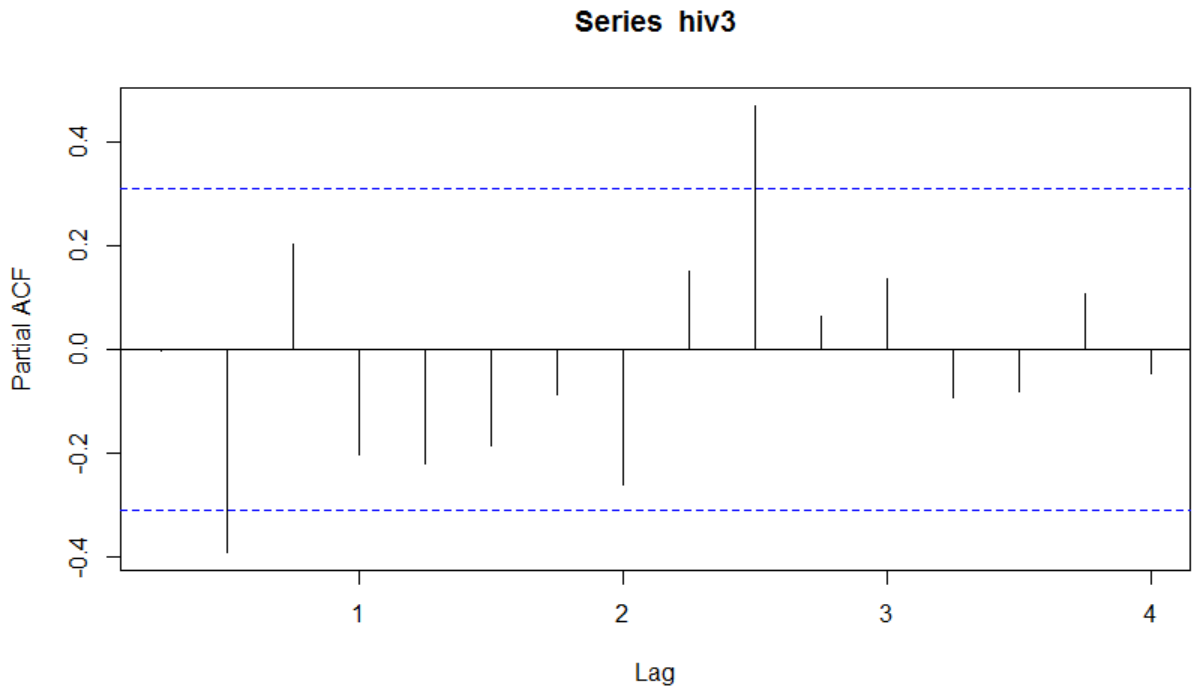

Fig 7 Partial Autocorrelation Function of the original series

The Partial Autocorrelation Function has a 95% confidence band which is constructed based on the assumption that the process is an Autoregressive (AR) Process. From the graphical representation, it can be clearly seen that the PACF has only one significant spike at lag 1 and other lags tends to zero. This confirms stationarity and hence, we try AR (1).

Discussion of result: both ACF of the simulated model and the observed data above decay exponentially, while their PACF cut off. Since the order of AR process is determine using the PACF plot, it was clear that the simulated PACF cut out at the 1stlag (order 1) and the observe data PACF cut off only at 11stlag (order 1). This shows that the appropriate model for the data should be AR1 model. The order of the model will be identified by varying the order of the AR process.

Applying Box Jenkins Methodology, the series was differenced zero to make it stationary, therefore the series is integrated of order 1 that is d =0. Thus, ARIMA (p,q,d) = ARIMA (p,q,0) model since the model is autoregressive integrated and moving average model ARIMA. We vary value of p and q then judge the model suitability based on their Akaike Information Criterion (AIC), Log likelihood and MSE (Sigma^2). The model with the lowest AIC will be the best model.

Table 6: Show the MSE, Likelihood AIC of ARIMA (p, 1, q) model

Model **ARIMA (p, 1,q)**	**MSE** **SIGMA ^2**	**LOG LIKELIHOOD**	**AIC**
(1,0,0)	1357.21	-201.3942	408.7884
(1,0,1)	133.586	-201.0479	410.0957
(1,0,2)	1315.667	-200.8037	411..2525
(2,0,0)	1321.221	-211.2915	410.583
(2,0,1)	1295.388	-200.6262	411.2525
(2,0,2)	1144.06	-199.124	410.2481

Discussion of result: ARIMA (1,0,0) has the lowest Akaike Information Criterion (AIC) =408.7884 .

Table 7: Estimation of The Parameters Of Identified

Model	**Parameter** **ARIMA(1,0,0)**	**S.E** **Std. Error**	**AIC**	**SIGMA^2**	**Log Likelihood**
ARIMA(1,0,0) **Mean**	154.9750	8.2354	410.0957	1333.586	-201.0479

2.8 DIAGNOSTIC CHECKING

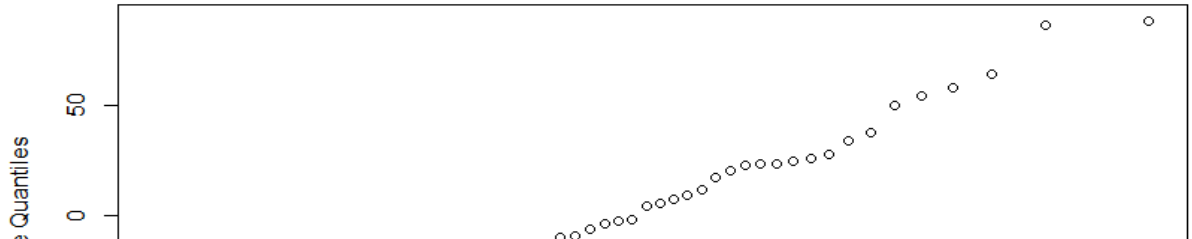

Fig 8: Normal Q-Q Plot

2.9. BOX-PIERCE TEST

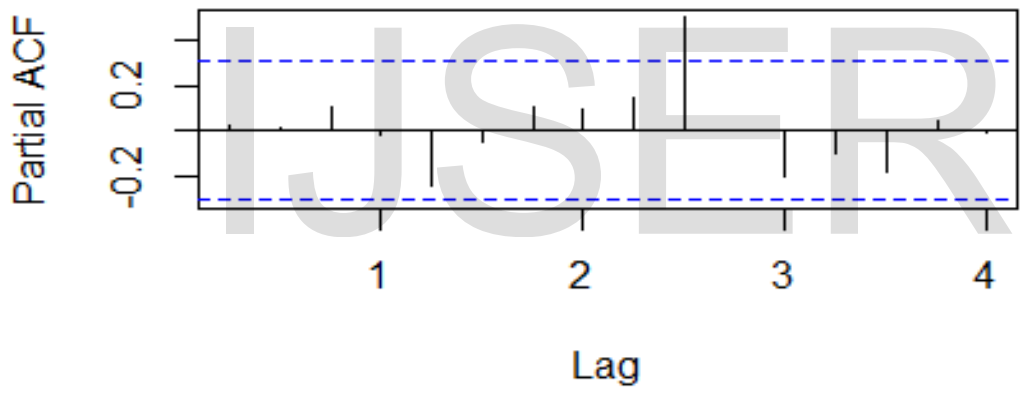

Fig 9: Series fit7&residuals

Table 8: Box-Ljung Test

X-SQUARED	D.F	P-VALUE	DECISION	CONCLUSION
0.18012	1	0.6713	Reject H_0	The model residual are independently distributed

Decision Rule: Reject H_0 if P< 0.05

Table 9: Box-Ljung test

X-SQUARED	D.F	P-VALUE	DECISION	CONCLUSION
0.19397	1	0.6596	Reject H_0	The model residuals are independently distributed

Decision Rule: Reject H_0 if P< 0.05

DISCUSSION OF RESULT: we reject H_0 in both Box – Pierce and Ljung-Box test and conclude that the model residuals are independently which shows that the model is correctly specified. The Error to misspecification of the time series model is adequately avoided**.**

2.9 MODEL FORECAST

The fitted model forecasted values for the next five years' time periods (2018-2022) are presented in the tables below.

Point Forecast Lo 80 Hi 80 Lo 95 Hi 95

2018 Q1	145.1745	97.96174	192.3873	72.96881	217.3802
2018 Q2	146.0261	87.70973	204.3425	56.83891	235.2133
2018 Q3	146.6435	83.26561	210.0214	49.71538	243.5716
2018 Q4	147.0911	81.20829	212.9740	46.33201	247.8503
2019 Q1	147.4157	80.25351	214.5779	44.69999	250.1314
2019 Q2	147.6510	79.82598	215.4761	43.92158	251.3805
2019 Q3	147.8216	79.65074	215.9925	43.56325	252.0800
2019 Q4	147.9453	79.59334	216.2973	43.40998	252.4807
2020 Q1	148.0350	79.58802	216.4820	43.35436	252.7157
2020 Q2	148.1000	79.60315	216.5969	43.34308	252.8570
2020 Q3	148.1472	79.62408	216.6703	43.35014	252.9442
2020 Q4	148.1814	79.64449	216.7183	43.36326	252.9995
2021 Q1	148.2062	79.66203	216.7503	43.37697	253.0354
2021 Q2	148.2241	79.67620	216.7721	43.38912	253.0591
2021 Q3	148.2372	79.68723	216.7871	43.39908	253.0752
2021 Q4	148.2466	79.69562	216.7976	43.40692	253.0863
2022 Q1	148.2535	79.70192	216.8050	43.41292	253.0940
2022 Q2	148.2584	79.70659	216.8102	43.41745	253.0994
2022 Q3	148.2620	79.71004	216.8140	43.42081	253.1032

2022 Q4 148.2646 79.71257 216.8167 43.42330 253.1060

Source: R statistical package version 3.2.0 2015 Editions

It is clear from this forecast that the total number of HIV patient on quarterly basis in general hospital Ado-odo ota local govt area, ogun state is expected to follow the positive trend visible in the time series plot of the original data.

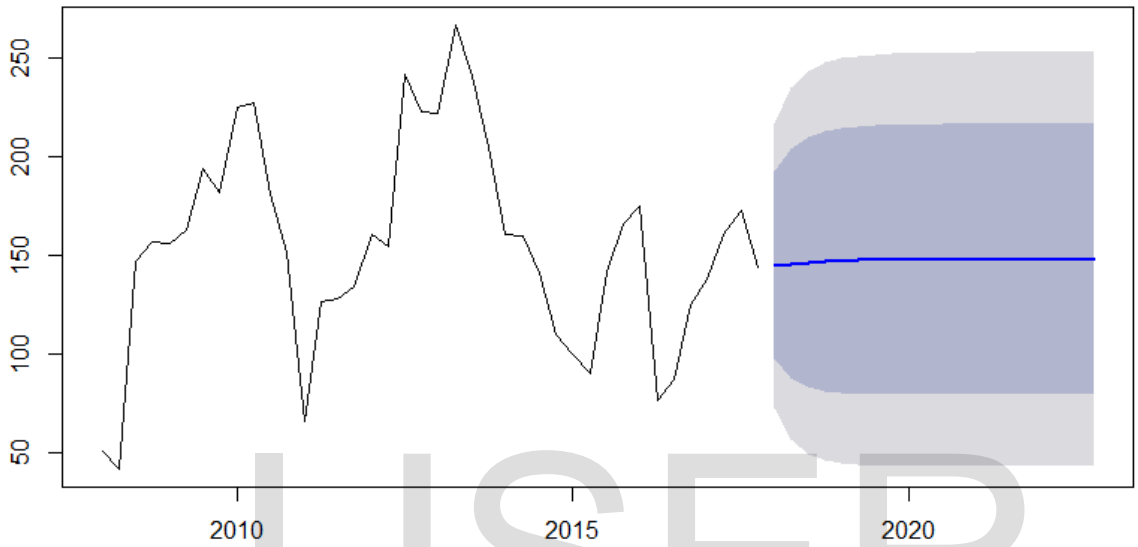

Table 10: Forecast from ARIMA (1,0,0) with non-zero mean

Source: R statistical package version 3.2.0 2015 Edition

3.0 FINDINGS

It will be essential at this junction to point out that statisticians rely on the observation of the past and present time series to forecast the future. In fact, one of the most fundamental objectives in time series is modeling and forecasting. The findings shows that there is an increase in the number of the HIV Patient for some period of time but later exhibit a decreasing pattern. This study reveals that the monthly data of total number of HIV patients was stationary, the appropriate model for the time series data collected is Autoregressive model AR(1). The statistical model for the forecasted values is given as; $\mathbf{Y_t = 158.2231 - 0.1584(x_t)}$

3.1 CONCLUSIONS

Based on the findings carried out in this study, the following conclusions can be drawn;

- The monthly data of total number of HIV patients was stationary.
- The appropriate fitted model for the data was AR(1).
- The forecasted value shows that there is an increase in the number of the HIV patients for some period of time and then approximately remain constant for the period of three years from 2020-2022.

3.2 RECOMMENDATIONS

From the result of the analysis carried out in chapter four of this research work and the outcome of the findings and conclusion, we need to sound our recommendation to concerned bodies specifically health workers, HIV care centers, governmental and non-governmental organization that;

- More effort to be put in terms of reaching and sensitizing large number of people so that there would be reduction in the spread of the disease. Also to address the ignorant of the basic facts about HIV/AIDS, aggressive media hype should be embarked upon to properly educate and sensitive the people on ways of contracting the virus and the evil of stigma and discrimination.
- Patients who are tested positive should be followed up with proper referral systems to make sure that treatment is started to stop the spread of the virus.
- People living with HIV/AIDS should be cared for, encouraged and allowed to be partaker in activities which uninfected people do in order for them not to have the mind set of spreading the disease.

REFERENCES.

[1] Akgun, B. (2003). "*Identification of Periodic Autoregressive Moving Average Models". Turkey:* Department of Statistics, Middle East Technical University.

[2] Bilali, C. (2017). *"Ways to prevent HIV/AIDS. Nigeria"* Daily Trust Newspaper.

[3] Bowerman, O'connell, Koehler. (2005), *"Forecasting, time series and regression an Applied Approach"*. (4th Edition) U.S.A: Curt Hinrichs.

[4] Box, G., & Jekins, G. (1970). *"Time Series Analysis, Forecasting and Control"*. (3rd Edition.). New Jersey: Prentice Hall, Englewood Clifs.

[5] Chatfield, C. (1996). *"The Analysis of Time Series"* (5th Edition). New York: Chapman & Hall.

[6] Delius, P. & Glaser, C. *"Sex, diseases and stigma in South Africa"*. Historical Perspective.Afr JAIDS Res.2006;4 (1), 29-36.

[7] Ezegwui HU, Ikeme AC, & Onwasigwe CN. *"Attitude of antenantal patients in Enugu towards HIV screening"*.Troop J obsetGynecol.2005; 22 (2),:156-158.

[8] Federal Ministry of Health. (2010). *"HIV/STI Integrated Biological and Behavioural Surveillance"* Survey 2010, Abuja.

[9] Gottaman, M .John.(1981), *"Time series Analysis, A Comprehensive Introduction for Social Scientists"*. Cambridge University Press, British.

[10] Hanke, J., & Reitsch, A. (1991). *"Understanding Business Statistics"*. (D. Richard, Ed.) IRWIN INC. http://hiv.gov/hiv-basics/staying-in-hiv-care/hiv-treatment/hiv-treatment-overview. (n.d.). http://dx.doi.org/10.4314/afrrev.v10i5.3

[11] Hudu, M. (2009). *"Interrupted Time Series Analysis of the Rate Inflation in Ghana"*. West Africa: Department of Mathematics, Kwame Nkrumah University of Science and Technology.

[12] Joint United Nations Programmes on HIV and AIDS. (2004). *"Epidemiological Fact Sheet on HIV/AIDS and Sexually Transmitted Infection in Nigeria"*: An update,htpp://data.unaids.org/publication/fact-sheets01/Nigeria_en.pdf (accessed 10 Febru

[13] Lim, J.Y (2002). *"An Evaluating of Alternative Forecasting Model for Natural Rubber* Economics and Finance, Curtin University of Technology.

[14] McDowel, D., R. McCleary, E.E Meidlinger, and A.H Richard Jr. (1980).

[15] Interupted *"Time Series Analysis"*. Sage Publications, Inc.,Boca Raton, FL, USA.