MODELLING TUBERCULOSIS AND HIV CO-DYNAMICS IN KENYA

BY

BERNARD KIPKOECH LANGAT

School of Mathematics

College of Biological and Physical Sciences

University of Nairobi

A project submitted in partial fulfillment of the requirements for the degree of Master of Science in Biometry

2013

Declaration

Candidate

I declare that this research project is my original work and has not been presented for a degree in any other university or for any other award.

---------- ----------

Signature Date

Bernard Kipkoech Langat I56/65810/2010

Supervisor

This research project has been submitted with my approval as university supervisor.

---------- ----------

Signature Date

Dr. Chumba Isaac Kipchirchir

School of Mathematics

University of Nairobi

P. O. Box 301197-00100

Nairobi

Acknowledgement

I acknowledged the support of my supervisor Dr. Chumba Isaac Kipchirchir, colleagues, friends and other lecturers for immense contribution and guidance.

I am indebted to the NASCOP who maintains the national HIV sentinel surveillance system and DLTLD who maintain the TB routine surveillance system for permission to use their surveillance data.

Further appreciation goes to my employer, the Ministry of Health and colleagues for allowing time for me to pursue this master's degree course and seeing it as an investment in furthering the TB control in the country.

Dedication

To Caren, Chepkemoi, Kipkorir and Chemutai

Abstract

A time series approach using autoregressive integrated moving average (ARIMA) modeling has been used in this study to model the co-dynamic relationship between HIV and Tuberculosis.

The study has showed that ARIMA(0,1,2) model provides the best fit for HIV prevalence rate and that the ARIMA(1,2,0) model provides the best fit for the TB case notification rate.

TB case notification rate and HIV prevalence rate time series demonstrated that there is a long run equilibrium relationship between HIV prevalence and TB notification rates. The current declining trends of TB cases may indicate that the efforts in HIV control could be driving down the TB epidemic. The Kenyan twin epidemic has time lag of 6 years between the trends of HIV prevalence and TB case notification rate.

The study also showed that there is Granger causal relationship between HIV and TB trends and that HIV Granger causes TB.

.

List of abbreviation

ACF	Autoregressive Correlation Function
AIDS	Acquired Immune Deficiency Syndrome
ARIMA	**A**uto**R**egressive **I**ntegrated **M**oving-**A**verage
CDR	Case Detection Rate
CNR	Case Notification Rate
DOTS	Directly Observed Treatment Short Course
DTLC	District TB and Leprosy Coordinator
EPTB	Extra-Pulmonary Tuberculosis
HAART	Highly Active Anti-retroviral Therapy
HIV	Human Immune-deficiency Virus
MDRTB	Multi-Drug Resistant Tuberculosis
PACF	Partial Autoregressive Correlation Function
PTB	Pulmonary Tuberculosis
RRv	Ross River Virus
SARS	Severe acute respiratory syndrome
STI	Sexual Transmitted Infection
TB	Tuberculosis
WHO	World Health Organization

Contents

Chapter 1: Introduction

1.1 Background information

Tuberculosis remains infectious disease of great public health importance in the world. In its 2010 Global TB control report, WHO estimates of the global burden of disease caused by TB in 2009 are as follows: 9.4 million incident cases (range, 8.9 million–9.9 million), 14 million prevalent cases (range, 12 million–16 million), 1.3 million deaths among HIV-negative people (range, 1.2 million–1.5 million) and 0.38 million deaths among HIV-positive people (range, 0.32 million–0.45 million). There were 5.8 million notified cases of TB in 2009, equivalent to a case detection rate (CDR, defined as the proportion of incident cases that were notified) of 63% (range, 60–67%), up from 61% in 2008

The burden of TB remains great with Kenya being ranked 15th among the 22 highest TB burden countries worldwide and 5th in Africa (1). In its annual report of 2010 and 2011, Kenya notified total of 102,083 TB cases (all forms of tuberculosis) to national programme in 2011. This is a reduction from total of 106,083 TB cases (all forms of tuberculosis) were reported in 2010.

The global response to threat of TB is encapsulated in the Stop TB Partnership's Global Plan to Stop TB, 2006–2015, launched in January 2006. It set out the scale at which the interventions included in the Stop TB Strategy need to be implemented to achieve the 2015 Millennium development Goals.

Directly Observed Treatment Short-course chemotherapy (DOTS) strategy remains at the heart of the Stop TB Strategy. The DOTS approach has five basic components:

1) Political commitment with increased and sustained financing for TB control by way enabling legislation, planning, human resources, management and training

2) TB Case detection through quality-assured bacteriology which involves strengthening TB laboratories, drug resistance surveillance

3) Standardized treatment with supervision and patient support

4) An effective drug supply and management system that ensures no interruption of treatment

5) Monitoring and evaluation system and impact measurement which involves national TB recording and reporting systems

In line with the global plan and National Health Sector Strategic Plan II, the Division of Leprosy, Tuberculosis and Lung Disease developed and is currently implementing its strategic plan 2011-2015.

Donald A Enarson and Nils E Billo, in their critical evaluation of the Global DOTS Expansion In Southern sub-Saharan Africa, TB and HIV are closely linked; for example, in the highest-burden settings, 75% of TB patients are also living with HIV/AIDS. Due to the link between TB and HIV, sub-Saharan Africa is projected to supersede all other regions in the burden of TB over the coming decades. The trend of rising TB case rates can only be reduced if HIV infection rates are also reduced.

The natural history of TB as described by Hans Reider; starts from the exposure of mycobacterium, TB causing bacilli from a symptomatic Pulmonary TB patient. Exposure may lead to infection, where TB bacilli establish itself in body tissues. Often times, the infection

occurs in the lung tissues. This infection may proceed to symptomatic phase in which the patient has signs and symptoms of TB disease. However in most times, the bacilli goes into latency in which bacteria becomes dormant until it is reactivated by lowered body immunity. It is estimated that globally that 1/3 of the world population is infected with TB and the estimate in sub-Saharan Africa, 2/3 of the population is estimated to be infected. In general population, the annual risk of TB disease (reactivation of latent) is estimated at 1% and 10% life time risk. HIV/AIDS increases the risk of TB disease to 10% annual risk and 50% life time risk. At an individual level, HIV infection lowers immunity and thereby increasing the risk of reactivation of latent TB. Likewise HIV/AIDS increases the risk of progression of TB infection into primary TB disease.

1.2 Statement of problem

The co-dynamic relationship between HIV and TB is majorly associated with rising burden of TB and this is attributed with the rise of TB epidemic in Kenya and other Sub-Saharan countries. The question arises whether the trends of TB can be forecasted from HIV/AIDS trends. Further to this, question arises as to what extent can the temporal information on time series data assist in the modeling and understanding of the causal relationship between TB and HIV? Demonstrating requires cross-correlations between HIV and TB temporal time series data. Furthermore, there is need to demonstrate if there is presence of “disease causality”.

1.3 Objectives

The overall objective is to demonstrate presence or absence of granger causality between TB and HIV

The specific objectives that will be considered are

1. To identify the appropriate ARIMA model for TB and HIV trends
2. To investigate Granger causality between TB and HIV

1.4 Significance of the study

The goal of TB control programme is to decrease morbidity due to TB and cut transmission of TB in the community. Understanding the co-dynamic relationship between TB and HIV is key to the control of the twin epidemic.The study will give insight into co-dynamics of TB and HIV which is crucial in the control and management of these diseases. Further, it will guide policy makers on the control and prevention and be able to prescribe proper interventions.

Chapter 2: Literature Review

Mathematical models are tools that can be used to explore epidemic interactions and disease co-infections at population. Lih-Ing W. Roeger, et al (2009) noted that the study of the joint dynamics of HIV and TB present formidable mathematical challenges due to the fact that the models of transmission are quite distinct. Furthermore, although there is overlap in the populations at risk of HIV and TB infections, the magnitude of the proportion of individuals at risk for both diseases is not known.

Elisa F. Long, et al (2008) recognizes role of mathematical models in guiding policymakers to allocate resources for the prevention and control of infectious disease epidemics. In their study, mathematical analyses of the HIV-TB co-epidemics suggest that exclusive treatment of only one disease may substantially reduce that epidemic, but may exacerbate the other epidemic; that prevention programs can have a greater effect on reducing latent disease than treatment alone; and that comprehensive treatment for HIV, latent TB, and active TB must be combined with increased prevention efforts can diminish both epidemics. Finally, when modeling two or more synergistic infectious disease epidemics, it is important to include the effects of each disease on the transmission and progression of the other disease.

Joe Suyama, et al (2003), notes that epidemiologists have utilized applied mathematics to perform analysis of data in the public health and disease surveillance areas. Using time-series analysis, many seemingly unrelated spectra, series, or systems can be compared to determine whether important relationships exist. For example, researchers were able to correlate the incidence of Ross River virus (RRv) infections in Cairns, Australia, with the presence or absence of certain weather patterns in the preceding months of an outbreak. Certain climate changes were predictive of RRv epidemics. Similar time-series analysis and cross-correlations have

assisted in the analysis of complex and voluminous data collected for public health reasons. Smallpox epidemics in England between 1600 and 1800 were studied to determine whether any exogenous factors correlated with the development of smallpox epidemics in various populations. Using CCFs, five year cycles of smallpox outbreaks were determined.

Stephen D. Lawna et al (2005) in their study of long term incidence and risk of Tuberculosis among HIV-infected patients receiving HAART in a South African cohort, they found out that long-term HAART confers a greater reduction in TB risk than previously reported and HAART may, therefore, contribute more to TB control in low-income countries than previously estimated. Further, they found out that incidence of TB continued to decrease during the first 5 years of HAART. Patients with advanced pretreatment immunodeficiency had persistently increased risk of TB during HAART; this may reflect limited capacity for immune restoration among such patients.

Saeed Akhtar et al (2008), while studying nonlinear pattern of PTB among immigrants at entry in Kuwait, employed standard time series methods to assess and model long term trends in the data. The time series model describe temporal trend in the proportions of tuberculosis cases among migrants at entry in Kuwait. The trend estimation was done by first de-seasonalizing the series using the moving average smoothing method. The goodness-of-fit of the final model was evaluated via residual analysis by plotting residuals against fitted values and also versus the time variable.

Sánchez S., et al (2009), in their mathematical modeling of Kenya's TB and HIV trends, reported incongruence in trends TB–HIV co-dynamics, a deviation from international finding of high congruence. They postulated two explanations namely that there is an unaccounted improvement in TB case detection that has occurred, or that HIV is not declining as reported in

sentinel surveillance. The study recommended the need to reevaluate trends of both diseases in Kenya, and identify the most critical epidemiological factors at play.

José Leopoldo Ferreira Antune, et al (1999), applied a time-series analysis using ARIMA model to develop epidemiological profile of tuberculosis in the city of São Paulo. The study looked at mortality rates of Tuberculosis. Time series proved to be efficient in many ways: improving the use of statistical methodology in the health sciences; bypassing the difficulties inherent to the characteristics of data values (autocorrelation, heteroskedasticity, collinearity, and non-normality of forecast error distribution); integrating quantitative analysis with the historical interpretation of the study phenomena; projecting estimates of future trends in the behavior of variables; and systematizing methodology for application in future social research.

Hong-Jen Chang, et al (2004), in their review of impact of the SARS epidemic on the utilization of medical services, applied time-series autoregressive-moving average (ARIMA) analysis to determine whether the SARS epidemic was significantly associated with changes in medical service utilization rates. Over the study period, they observed significant utilization reductions at the peak of the SARS epidemic. The model demonstrated that the fears of SARS significantly influenced people's care-seeking behavior and that this fear seriously compromised their accessibility to quality care.

Andrew Arnold, et al (2007) in their study on temporal causal modeling with graphical Granger methods, assert that the need for mining causality, beyond mere statistical correlations, for real world problems is recognized widely. The applications involve temporal data, which raises the challenge of how best to leverage the temporal information for causal modeling. The concept of "Granger causality", based on the intuition that a cause helps predict its effects in the future, has gained attention in many domains involving time series data analysis.

Chapter 3: Methodology

3.1 Data

The study uses data the annual TB case notification rate data as reported to the national TB control program (DLTLD) and HIV prevalence rates data obtained from the national HIV sentinel surveillance system from the National STI and AIDS control program (NSACOP). The data covers the years 1990 to 2010.

3.2 ARMA model

A general ARMA (p,q) model with mean μ is given by;

$$Y_t - \mu = \sum_{i=1}^{p}(\alpha_i Y_{t-i} - \mu) + \varepsilon_t + \sum_{i=1}^{q}\theta_i \varepsilon_{t-i} \qquad (3.1)$$

where $\{\varepsilon_t\}$ is a sequence with mean μ and variance σ^2 ,that is , $\varepsilon_t \sim (\mu, \delta^2)$

3.3 ARIMA model

In practice, trends exist in many data sets and hence there is need to remove these effects. Trends in time series can be removed by differencing. This differencing is integrated into the ARMA models creating the ARIMA models. The differencing is done typically once, twice or three times, until the series is at least approximately stationary exhibiting no obvious trends.

ARIMA (p,d,q) defines a model with an Autoregressive part of order p, a Moving average part of order q and having applied d order differencing and is given by

$$(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p)(1 - B)^d Y_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + B^q)\varepsilon_t \qquad (3.2)$$

where, B is a backshift operator. B^k shifts time series by k time units, that is,

$$B^k Y_t = Y_{t-k} \qquad (3.3)$$

3.4 The autocorrelation function (ACF) and partial autocorrelation function (PACF)

The principle way to determine which AR or MA model is appropriate is to look at the ACF and PACF of the time series. Table 3.1 below gives the behaviour of these functions for different ARMA models.

Table 3.1: Theoretical framework for AR, MA and ARMA

MODEL	**ACF**	**PACF**
AR(p)	Exponential decay or damped sine wave. The exact pattern depends on the signs and sizes of $\alpha_1 \dots \alpha_p$	Spikes at lags 1 to p, then zero.
MA(q)	Spikes at lags 1 to q, then zero.	Exponential decay or damped sine wave. The exact pattern depends on the sign and sizes of $\theta_1 \dots \theta_q$
ARMA(p,q)	Both decay exponentially and may contain damped oscillations	

Having determined *p*, *d*, and *q*, the coefficients of the autoregressive and moving average terms are estimated using nonlinear least square method or maximum likelihood. This is followed by diagnostic checking of the fitted model, and, if necessary, the model may be modified in terms of the values of *p* and *q* to achieve the desired level of model adequacy. The general points to be considered while fitting the model are:

(1) The model should have a relatively small residual variance

(2) The model should be parsimonious that is the number of parameters should be kept small without compromising the model adequacy

(3) The residuals should be normally distributed and independent with a mean value of 0.

There are some things to bear in mind in choosing appropriate ARIMA model. First, values of p, q or d of more than 3 are very rarely needed. Second, it is often the case that many different ARIMA models give more or less the same predictions, so there is some flexibility in the choice of p, d and q.

3.5 Which ARIMA (p,d,q) model to use?

Tentative model of the ARIMA class is identified through the analysis of historical data. This involves examining whether the time series is stationary or non-stationary. If non-stationary, it is first transformed into a stationary time series by applying a suitable degree of differencing. Where differencing is called for, then difference the data once, d = 1, and inspect the time plot of first order difference ΔY_t where Δ=1-B. If additional differencing is necessary, then one can try differencing again and inspect a time plot of second order difference $\Delta^2 Y_t$. Care must be taken not to over difference because this may introduce dependence where none exists.

Tentative values of *p* and *q* are found by examining the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the stationary time series.

Box-Jenkins procedure provides a 4 step process namely

1. If necessary, data should be transformed, such that covariance stationarity is achieved.

2. Inspect, ACF and PACF for initial guesses of p and q.

3. Estimate proposed model.

4. Check residuals (diagnostic tests) and stationarity of process.

If step 4 fails, go to step 2 and repeat. If in doubt, choose the more parsimonious model specification.

3.6 Cointegration

A variable Y_t is said to be integrated of order d that is $Y_t \sim I(d)$ if it has stationary, invertible, non-deterministic ARMA representation after differencing d times. Variables, all of which achieve stationarity after differencing, may have linear combinations which are stationary without differencing and this equilibrium relationship is term as cointegration between time series. Cointegration can be interpreted economically as the presence of a long-run equilibrium, the relationship between the variables being stable.

Given two time series $\{X_t\}$ and $\{Y_t\}$ which are both I(d) (they have compatible long-run properties), any linear combination of $\{X_t\}$ and $\{Y_t\}$ will be also I(d). In particular, variables X_t and Y_t describe TB case notification rate and HIV prevalence rate respectively.

$$\Delta Y_t = \beta \Delta X_t + \varepsilon_t \qquad (3.4)$$

where $\{\varepsilon_t\}$ is a sequence with mean μ and variance σ^2 that is $\varepsilon_t \sim (\mu, \delta^2)$

But this collapses to nothing in the long run, that is, $Y_t = Y_{t-1} = Y$, $X_t = X_{t-1} = X$ hence all the difference terms will be zero that is $\Delta Y_t = 0$, $\Delta X_t = 0$

One way to get around this problem is to use both first difference and level terms, that is,

$$\Delta Y_t = \beta_1 \Delta X_t + \beta_2 (Y_{t-1} - \gamma X_{t-1}) + \varepsilon_t \quad (3.5)$$

where $(Y_{t-1} - \gamma X_{t-1})$ is known as the error correction term.

Provided that $\{X_t\}$ and $\{Y_t\}$ are cointegrated with cointegrating coefficient γ, then $(Y_{t-1} - \gamma X_{t-1})$ will be I(0) even though the variables are each I(1). This is Granger representation theorem which states that any cointegrating relationship can be expressed as an equilibrium correction model.

3.7 Cointegration: Engle-Granger Test

This is a three step process, namely;

1. Pre-test the variables for the presence of unit roots and check if they are integrated of the same order
2. Regress the long run equilibrium model
3. Test whether the residuals are I(0).

3.7.1 Phillips and Perron unit root test

Phillips and Perron (PP) Test is non-parametric correction based on estimated long-run variance of ΔY_t Phillips-Perron tests assess the existence of a unit root in a univariate time series$\{\Delta Y_t\}$.

Consider model

$$Y_t = c + \alpha Y_{t-1} + \varepsilon_t \ , \quad t = 1,2,3, \dots, T \quad (3.6)$$

Where $\varepsilon_t \sim$ is serially correlated and c is a constant.

The constant c is included to capture the nonzero mean under the alternative hypothesis.

Phillips-Perron test equation is given by

$$\Delta Y_t = c + \delta Y_{t-1} + \varepsilon_t \, , \varepsilon_t \sim I(0) \tag{3.7}$$

Where $\delta = \alpha - 1$

$H_0 : \delta = 0$ (there is a unit root)

versus

$H_1 : |\delta| \neq 0$ (there is no unit root)

In order to remove the dependence of asymptotic distributions on nuisance parameter σ^2 and σ_ε^2 , Philips-Perron test uses modified statistics denoted Z_t and Z_β given by

$$Z_t = \left(\frac{s_\varepsilon}{s}\right)^{1/2} . t_\alpha - \left(\frac{1}{2}\right)(s^2 - s_\varepsilon^2).\left(s^2 T^{-2} \sum_{t=1}^{T} Y_{t-1}^2\right)^{-1/2} \tag{3.8}$$

and

$$Z_\beta = T(\hat{\alpha} - 1) - (s^2 - s_\varepsilon^2)\left(2T^{-2} \sum_{t=1}^{T} Y_{t-1}^2\right)^{-1} \tag{3.9}$$

where s^2 and s_ε^2 are consistent estimates of σ^2 and of σ_ε^2 respectively.

Computed Z-statistics in absolute value is compared to critical value. If the computed Z-statistic is less than the critical value, we fail to reject H_0, which implies that there exists a unit root and conclude that time series is integrated of order one

3.7.2 Testing for cointegration

This is based on Augmented Dickey-Fuller (ADF) test. The critical values to be used here are approximated critical values. The residuals are not the actual error terms, but estimated values from the long run equilibrium equation.

Equation 3.5 can be expressed as

$$\Delta Y_t = \beta \Delta X_t + \beta_2 \hat{\varepsilon}_{t-1} + \varepsilon_t \qquad (3.10)$$

where $\hat{\varepsilon}_{t-1} = (Y_{t-1} - \gamma X_{t-1})$

ε_t should be I(0) if the variables X_t and Y_t are cointegrated.

We can test the residuals of equation (3.10) to see if they are non-stationary or stationary. So we have the regression equation

$$\hat{\varepsilon}_t = Y_t + \hat{\beta} X_t \qquad (3.11)$$

This simply checks if a series is stationarity. The error correction process is a constructed series from estimated parameters with different distributions. ADF test is given by

$$\Delta \hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + v_t,\ v_t \sim \text{iid} \qquad (3.12)$$

Sets of critical values tabulated by Engle and Granger (1987) are used to test for cointegration.

ADF test examines the hypothesis that two time series are cointegrated. The null hypothesis is non-cointegration against the alternative of cointegration

Hypothesis

$H_0: \rho = 1, \quad \text{no cointegration}$

versus

$H_1: \rho < 1, \quad cointegration$

A large test statistic rejects the null of non-cointegration that is, a large test statistic 'accepts' cointegration

3.8 The Granger Causality

"Granger causality" tests are statistical tests of "causality" in the sense of determining whether lagged observations of another variable have incremental forecasting power when added to a univariate autoregressive representation of a variable. A series $\{X_t\}$ may be said to cause a series $\{Y_t\}$ if and only if the expectation of X_t given the history of $Y_t$is different from the unconditional expectation of X_t.

Consider two random variables X_t and Y_t

Forecast for Y_t, s period ahead is given by

$$Y_t(s)^{(1)} = E(Y_{t+s}]Y_t, Y_{t-1}, \dots), \; Y_t(s)^{(2)} = E(Y_{t+s}]Y_t, Y_{t-1}, \dots X_t, X_{t-1}, \dots) \qquad (3.13)$$

and

$$\mathrm{MSE}\left(\mathrm{Y_t(s)}\right) = \mathrm{E}\left(\mathrm{Y_{t+s}} - \mathrm{Y_t(s)}\right)^2 \qquad (3.14)$$

This is the mean squared error of $\mathrm{Y_t(s)}$

If $\mathrm{MSE}\left(\mathrm{Y_t(s)}^{(1)}\right) = \mathrm{MSE}\left(\mathrm{Y_t}\ (s)^{(2)}\right)$ then $\mathrm{X_t}$ does not Granger − cause $\mathrm{Y_t} \forall s >$ 0 which implies that $\mathrm{X_t}$ is not linearly informative to forecast $\mathrm{Y_t}$

3.8.1 To test for Granger-causality

Assume a lag of p

$$\mathrm{X_t} = c_1 + \gamma_1 \mathrm{X_{t-1}} + \gamma_2 \mathrm{X_{t-2}} \cdots \gamma_p \mathrm{X_{t-p}} + \beta_1 \mathrm{Y_{t-1}} + \beta_2 \mathrm{Y_{t-2}} + \cdots \beta_p \mathrm{Y_{t-p}} + \epsilon_t \qquad (3.15)$$

Estimate the parameter by ordinary least square (OLS) and test the hypothesis

$$\mathrm{H_0}: \beta_1 = \beta_2 = \cdots = \beta_p = 0\ (\mathrm{Y_t\ does\ not\ Granger - cause\ X_t}\ )$$

versus

$$\mathrm{H_1}: \beta_i \neq 0 \text{ for some i}$$

3.9 Analysis software

Statistical analysis of time-series is performed using the R-Gui software.

Chapter 4: Data Analysis and Results

In this chapter data analysis and results of the time series data of TB CNR and HIV prevalence are presented.

4.1 Fitting an ARIMA model to HIV prevalence time series data

4.1.1 Model Identification

The time plot of HIV prevalence rates showed an increasing a trend that peaked in 1999 and begun to decline in the past decade as shown in figure 1.

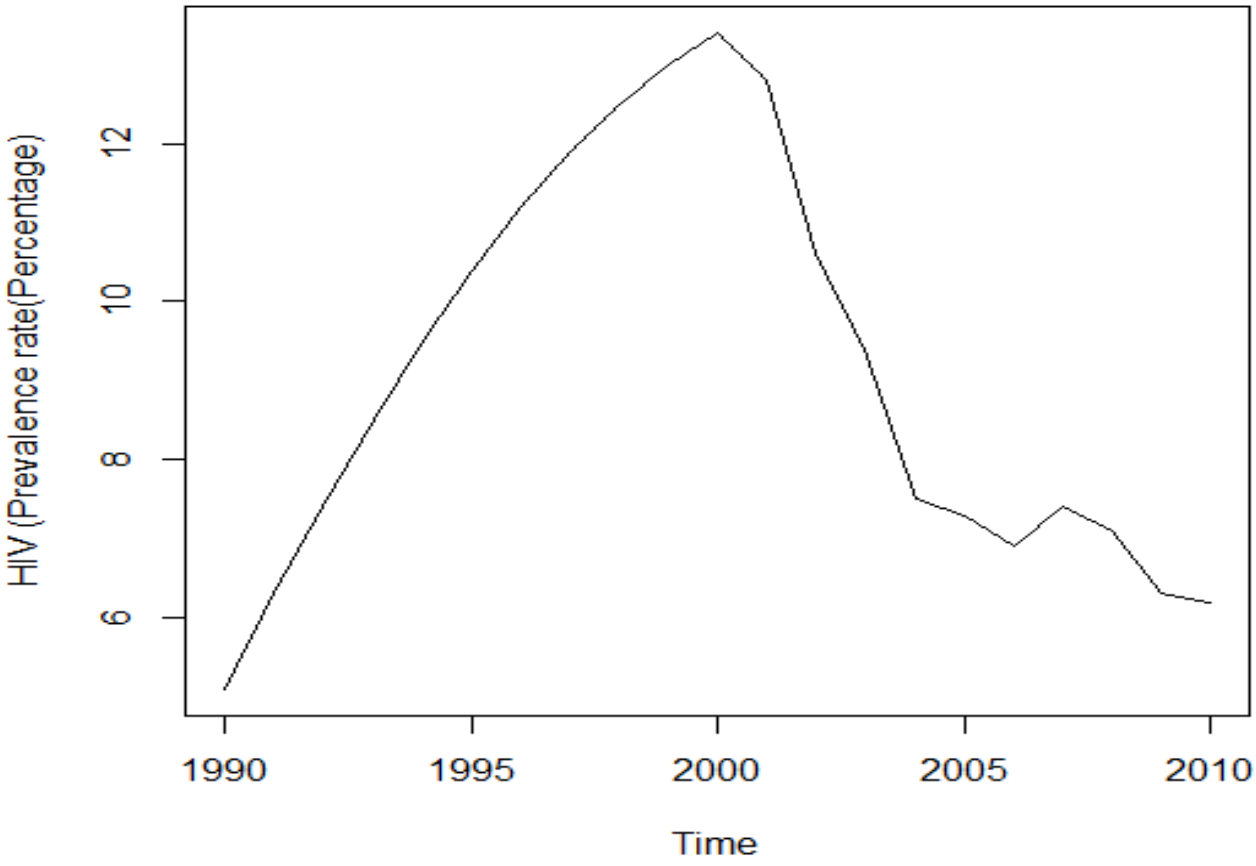

Figure 1: Time plot of HIV prevalence rate.

To identify AR and MA terms autocorrelation and partial autocorrelation functions were plotted for HIV prevalence rates at various lags as shown in figure 2.

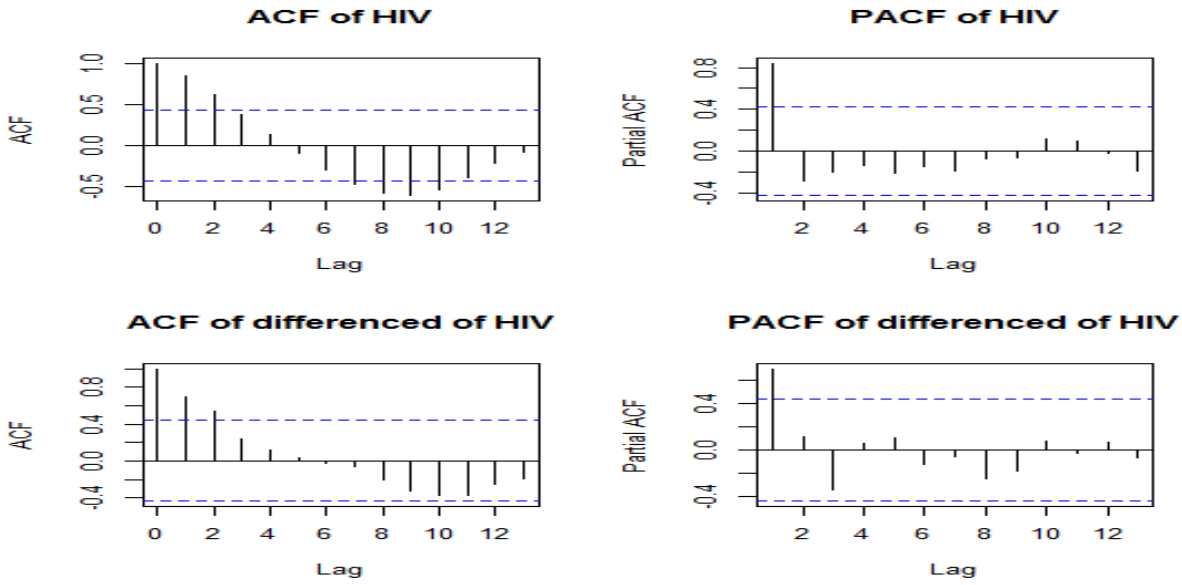

Figure 2: ACF and PACFplot HIV prevalence rate

The differenced HIV fits an MA(2) as shown in figure 2 with lower AIC=45.95 compared to MA(1) which has AIC=52.64.

4.1.2 Model diagnostics

Residual diagnostic tests and AIC are used here to determine the goodness-of-fit of the selectted ARIMA model to the original time series

HIV prevalence rates fitted ARIMA(0, 1, 2)

Coefficients:

	ma1	ma2
	0.7148	0.6531
s.e.	0.2579	0.1943

sigma^2 estimated as 0.4041: log likelihood = -19.98, aic = 45.95

The ACF of the residuals is also used as a diagnostic tool. Here we see that the ACF values except one which can be treated as outlier are all within the 95% zero-bound as shown in figure 3 below. This indicate that there is no correlation amongst the residuals hence an indicator of the independence of the residual terms.

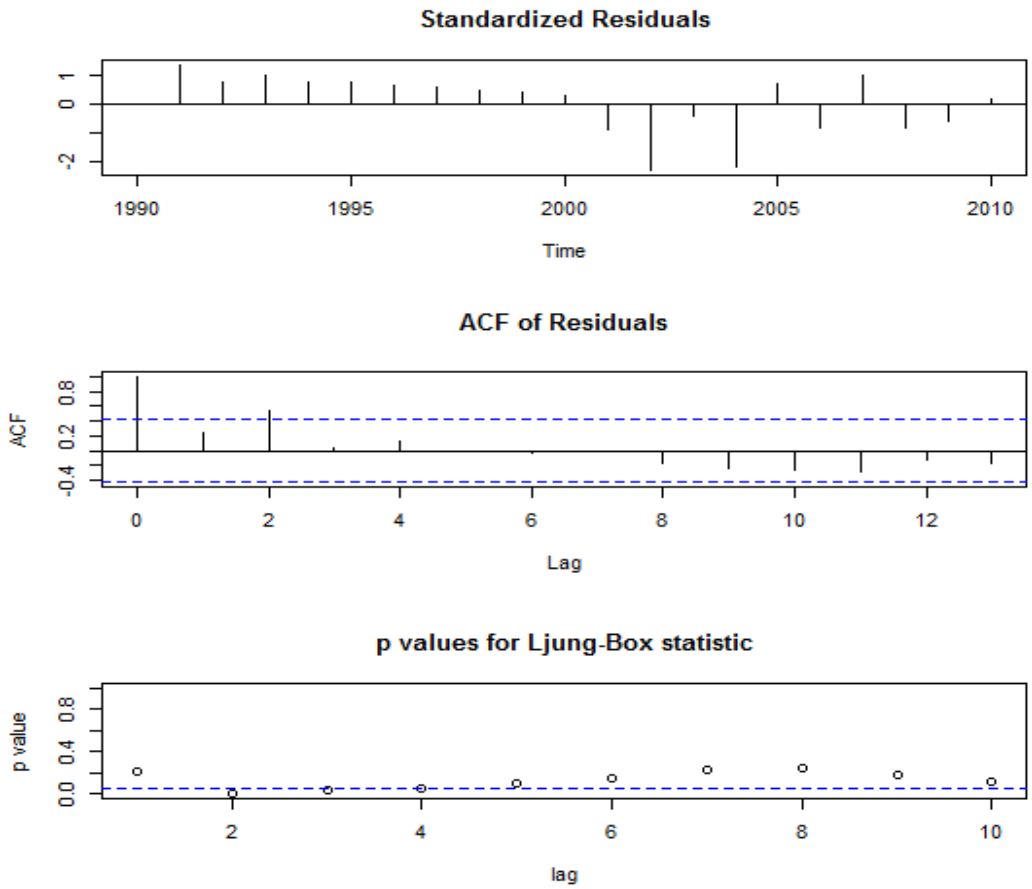

Figure 3: Residual diagnostics plots of ARIMA (0,1,2) model for HIV prevalence rate

QQ-plot is used to test for Normality. Here we can see that the QQ-plot approximately follows the QQ-line visible on the plot as shown in figure 4. This is good indicator of near normal residuals.

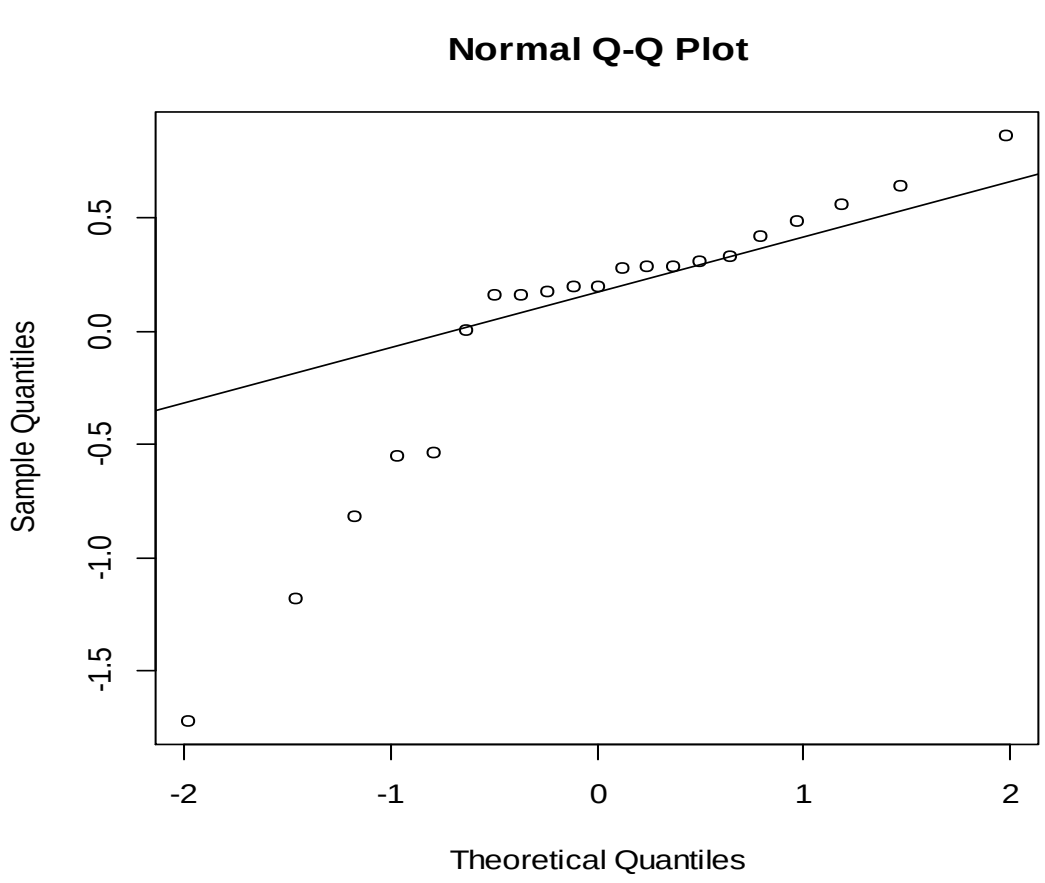

Figure 4: QQ-plot of ARIMA (0,1,2) model for HIV prevalence rate

4.1.3 Fitted Model

We therefore conclude that the ARIMA(0,1,2) the best- fit model for HIV prevalence rate. The fitted model is of the following form:

$$Y_t - Y_{t-1} = \varepsilon_t + 0.7148\varepsilon_{t-1} + 0.6531\varepsilon_{t-2} \quad (4.1)$$

4.2 Fitting an ARIMA model to TB case notification rate time series data

4.2.1 Model Identification

The time plot of TB case notification rates showed increasing trend that peaked in 2005 and begun to decline as shown in figure 5. The TB trends peak was observed 6 years later than that of HIV. The TB seems to be stable prior to the year 1990 and the year 1993 herald the rise in the

TB case notification. Comparison to HIV prevalence rates cannot be made prior to 1990 since the levels of HIV prevalence were unknown due lack of national HIV surveillance system.

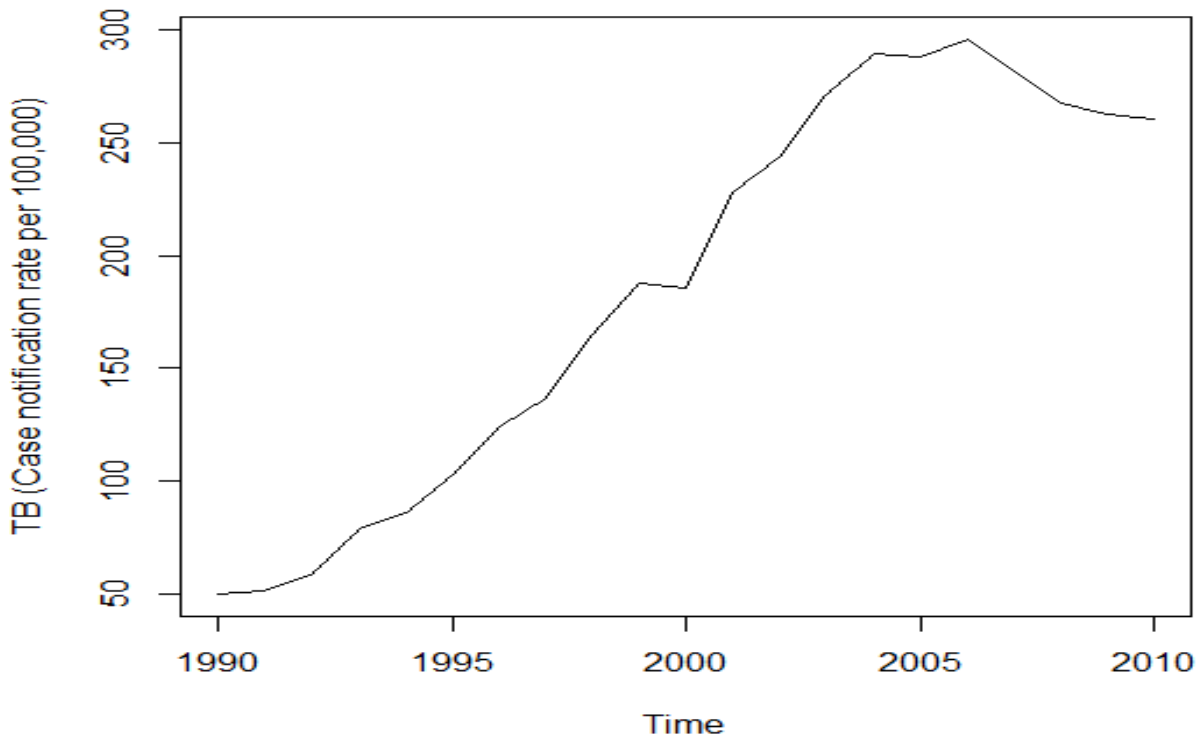

Figure 5: Time plot of TB case notification rate.

To identify AR and MA terms, autocorrelation and partial autocorrelation functions were plotted for TB case notification rates at various lags. The second order differenced fits an AR (1) model (figure 6) with low AIC=157.74 as compared with first order difference with AIC=168.32.

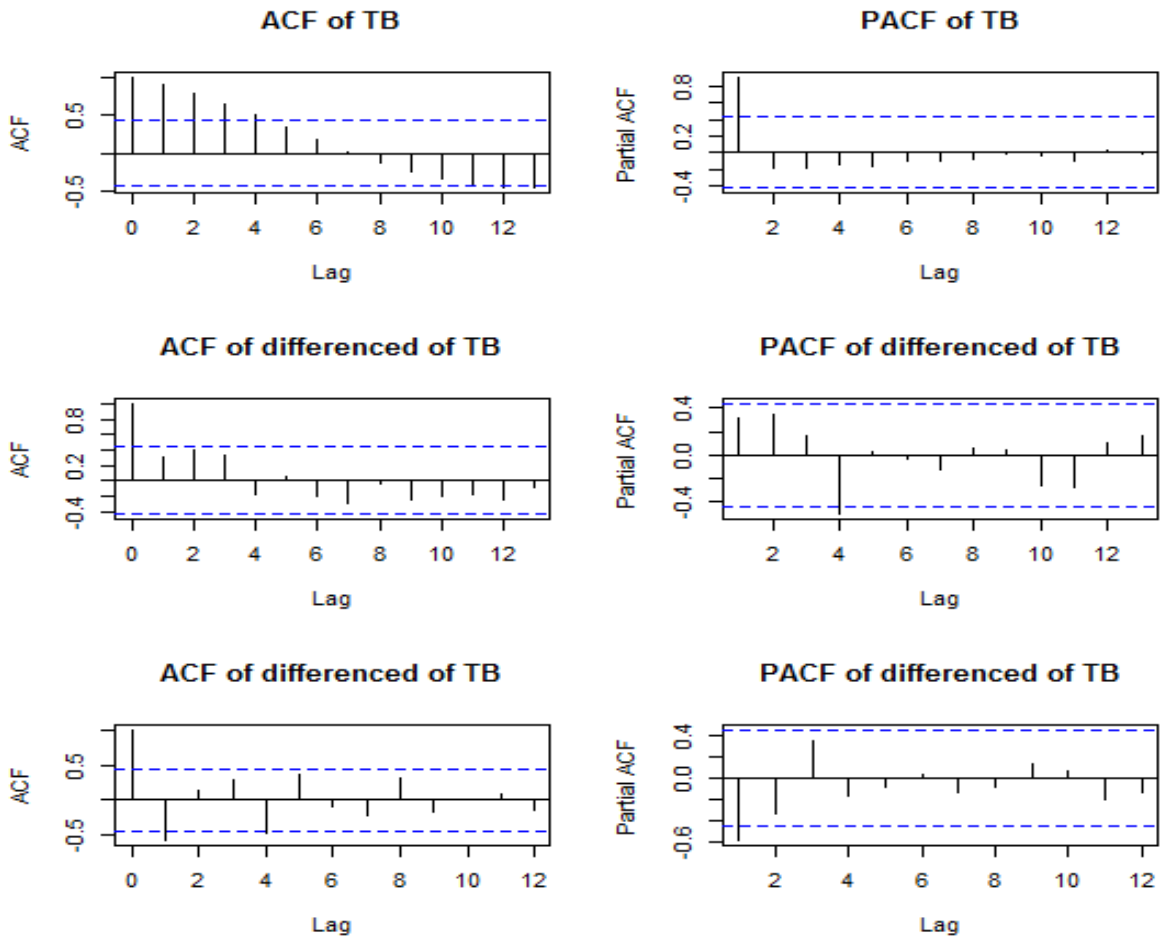

Figure 6: TB case notification rate ACF and PACF plots for first and second difference

4.2.2 Model diagnostics

Residual diagnostic tests are used here to determine the goodness-of-fit of the selectted ARIMA model to theTB notification rate data. TB case notification rate fitted an ARIMA (1, 2, 0)

Coefficients:

ar1

-0.5633

s.e. 0.1780

sigma^2 estimated as 185.5: log likelihood = -76.77, aic = 157.54

The ACF of the residuals is also used as a diagnostic tool. Here we see that the ACF values are within the 95% zero-bound as shown in figure 7 below. This indicate that there is no correlation amongst the residuals hence an indicator of the independence of the residual terms.

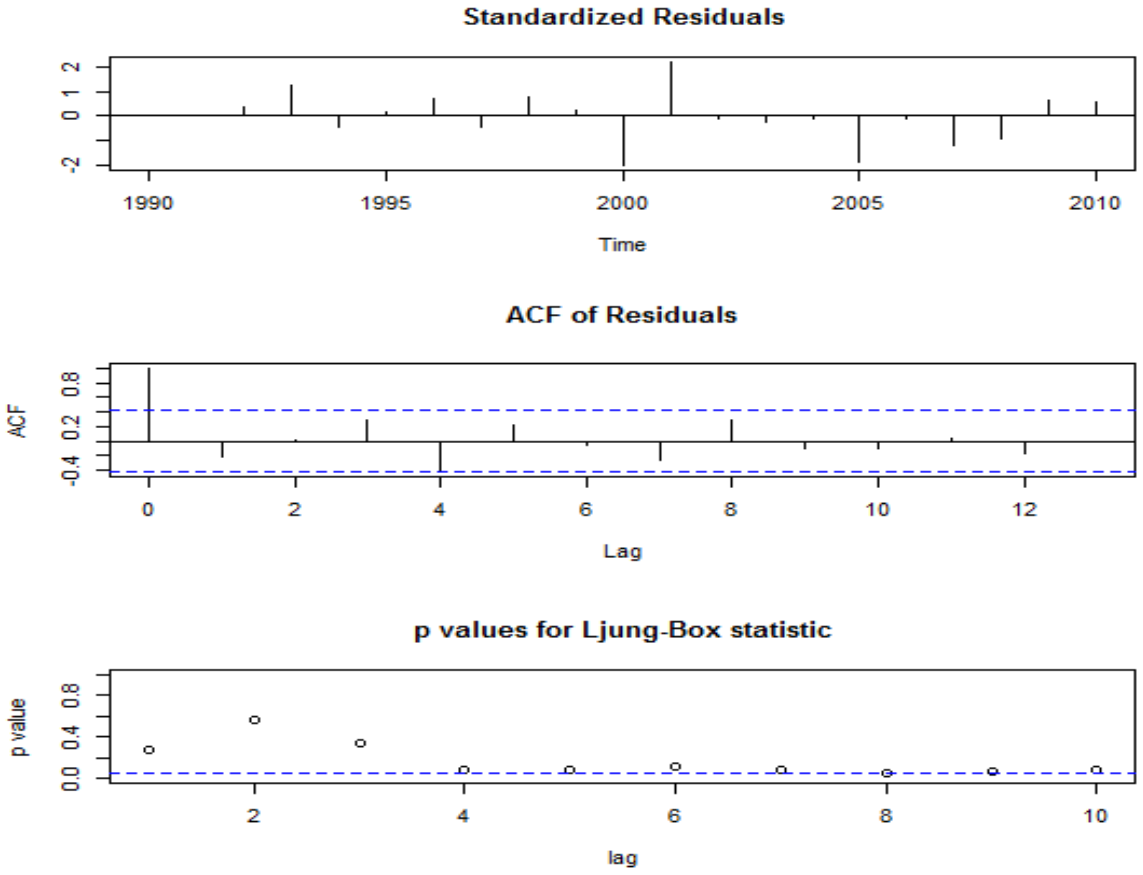

Figure 7: Residual diagnostics plots of ARIMA (1,2,0) model for TB case notification rate

The test for normality is carried out using Q-Q plot as shown in figure 8 below

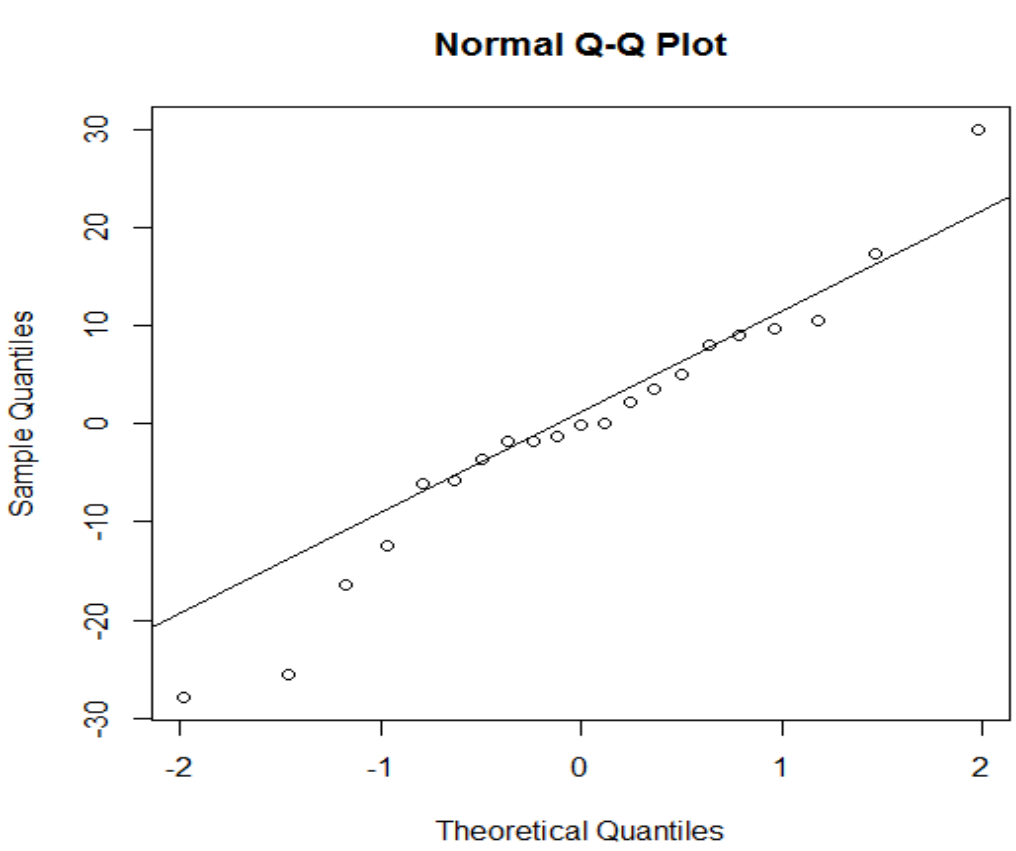

Figure 7: Q-Q plot of ARIMA (1,2,0) fit of TB case notification rate

Here we can see that the QQ-plot approximately follows the QQ-line visible on the plot. This is a good indicator of near normal residuals.

4.2.3 Fitted Model

We therefore conclude that the ARIMA(1,2,0) model is the best- fit ARIMA model for the original TB case notification. The final model is of the following form:

$$\Delta^2 X_t = \varepsilon_t - 0.5633\Delta^2 X_{t-1} \qquad (4.2)$$

4.3 Test for autoregressive unit root test

To test for autoregressive unit roots Phillips-Perron Unit Root Test is applied to regression model of TB case notification rate to its lags.

Table 4.1: Phillips-Perron Unit Root Test result for TB case notification rate

Critical values for Z statistics		1%	5%	10%
Z-tau-mu	-0.2454	-4.50005	-3.659125	-3.26775
Z-tau-beta	-0.4538	-4.50005	-3.659125	-3.26775

Table 4.1 shows that computed test-statistics for the Phillips-Perron Unit Root Test is less in absolute value than the critical value at 5% level of significance and conclude that the TB case notification have unit root, that is, difference stationary. This can be observed in the residual plot of Phillips-Perron regression for TB case notification rate trends as shown in figure 8 (Faint plot line representing fitted residuals and bolder plot line the actual data).

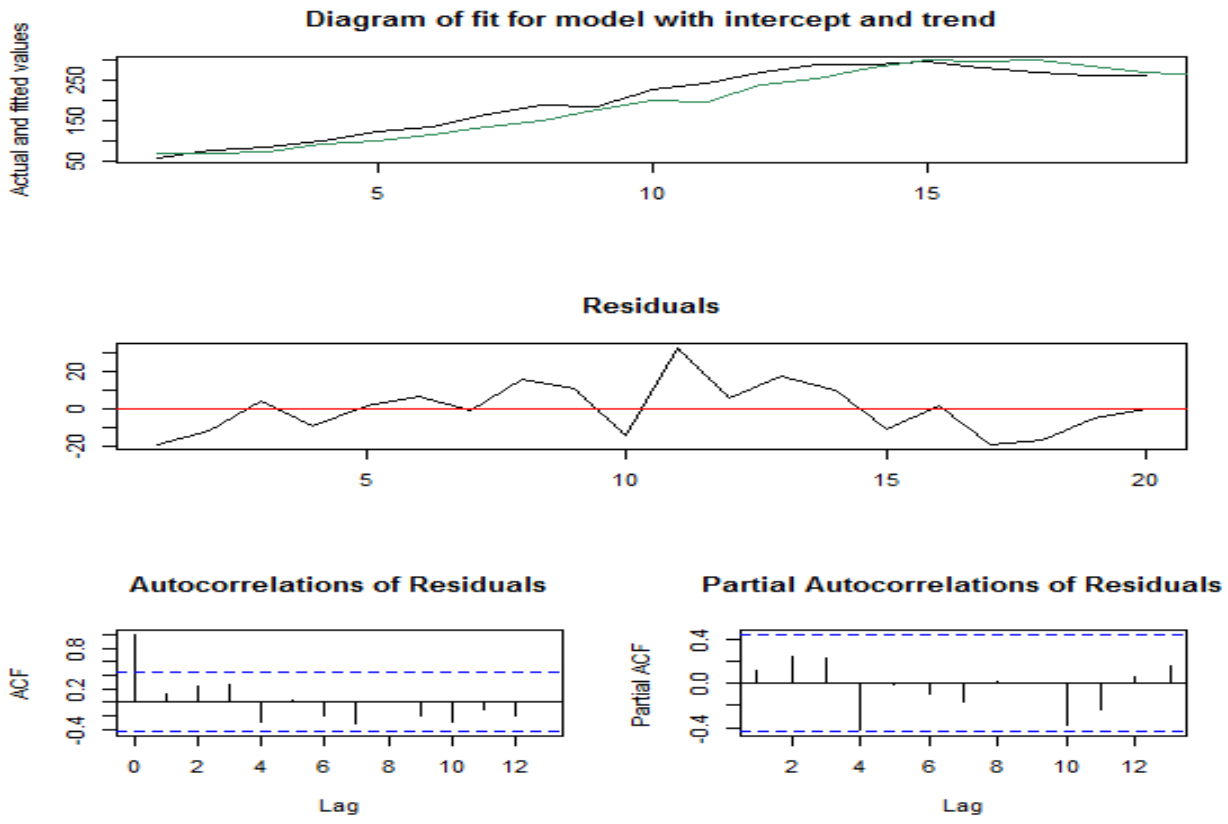

Figure 8: The residual plot of Phillips-Perron for TB case notification rate

The Phillips-Perron Unit Root Test of HIV prevalence rate regression with intercept and trend are shown in Table 4.2. The computed Z-statistics for the Phillips-Perron Unit Root Test is less in absolute value than the critical value at 5% level of significance and showing existence of the unit root (integrated of order one).

Table 4.2: Phillips-Perron Unit Root Test result for HIV prevalence rate

Critical values for Z statistics		1%	5%	10%
Z-tau-mu	1.1643	-4.50005	-3.659125	-3.26775
Z-tau-beta	-2.9706	-4.50005	-3.659125	-3.26775

This can be observed in the residual plot of Phillips-Perron regression for HIV prevalence rate as shown in figure 9 with faint plot line representing the fitted residuals and bold plot line representing the actual HIV prevalence rate data

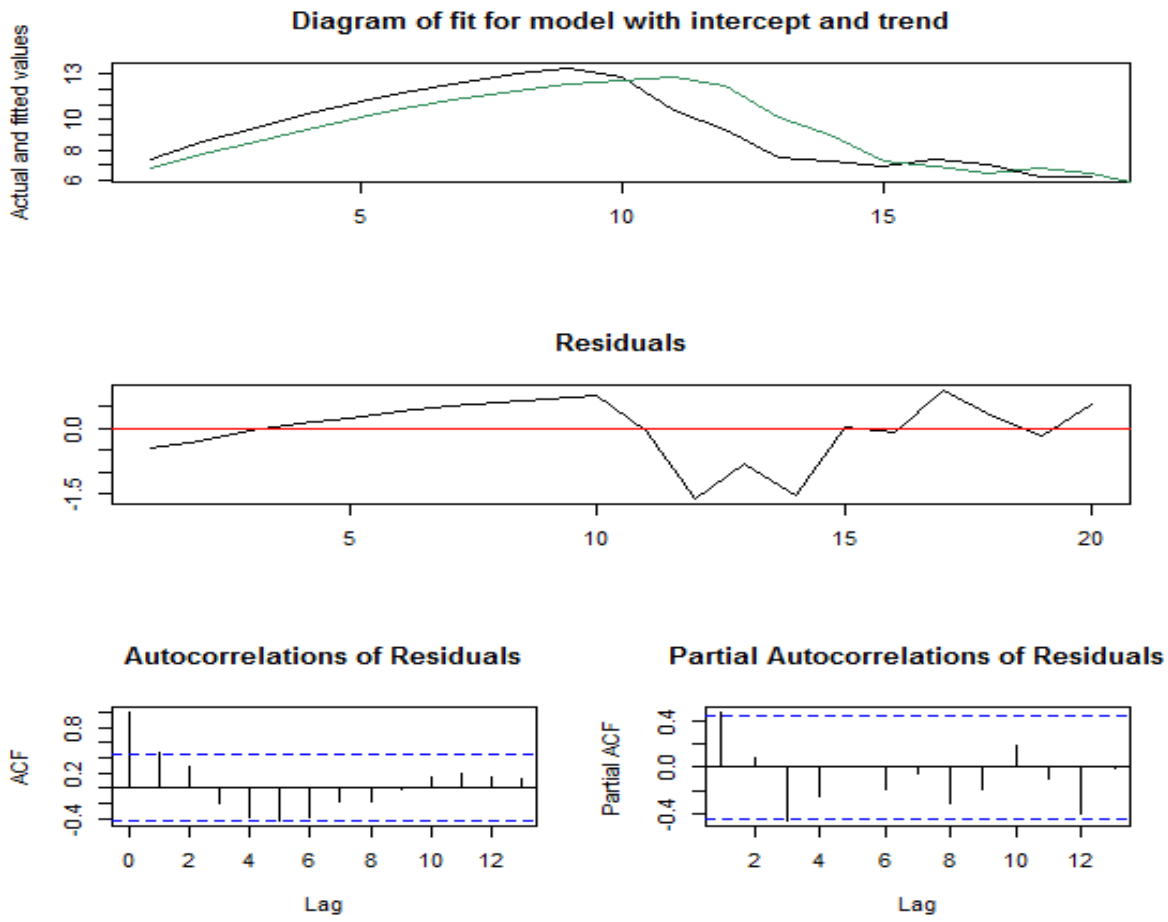

Figure 9: The residual plot of Phillips-Perron regression for HIV prevalence rate trends

4.4 Test for Cointegration

To conduct a cointegration test, the study utilized the ADF test. The test was performed to answer the question whether there is cointegrating relationship between the HIV prevalence and TB case notification rate.

This test whether the residuals are I(0). If they are $I(0)$ such that $Y + \beta X \sim I(0)$ then it is cointegrated. If this combination is $I(1)$, then the variables are said to be not-cointegrated.

Regression model of TB case notification on HIV prevalence rate was fitted and Augmented Dickey-Fuller Test Unit Root Test conducted and the results are shown in table 4.3.

Table 4.3: Table ADF unit root test for cointegration

ADF	Estimate	Std. Error	t value	Pr(>|t|)
z.lag.1	-0.06211	0.04066	-1.528	0.1450
z.diff.lag	0.53652	0.18677	2.873	0.0106 *

Residual standard error: 14.66 on 17 degrees of freedom; Multiple R-squared: 0.425, Adjusted R-squared: 0.3574. F-statistic: 6.283 on 2 and 17 DF, p-value: 0.009059

Table 4.4: ADF test z-statistics

Critical values for test statistics:

ADF	Test statistic	1%	5%	10%	Remark
z. tau1	-1.5275	-2.66	-1.95	-1.6	

The computed test statistics for the ADF test is less in absolute value than the critical value at 5% level of significance as shown in table 4.4 and hence conclude that TB and HIV are cointegrated.

The ADF cointegration residuals are plotted in figure 10 and the plot of residuals appear to vary about a fixed level thereby confirming the existence cointegration.

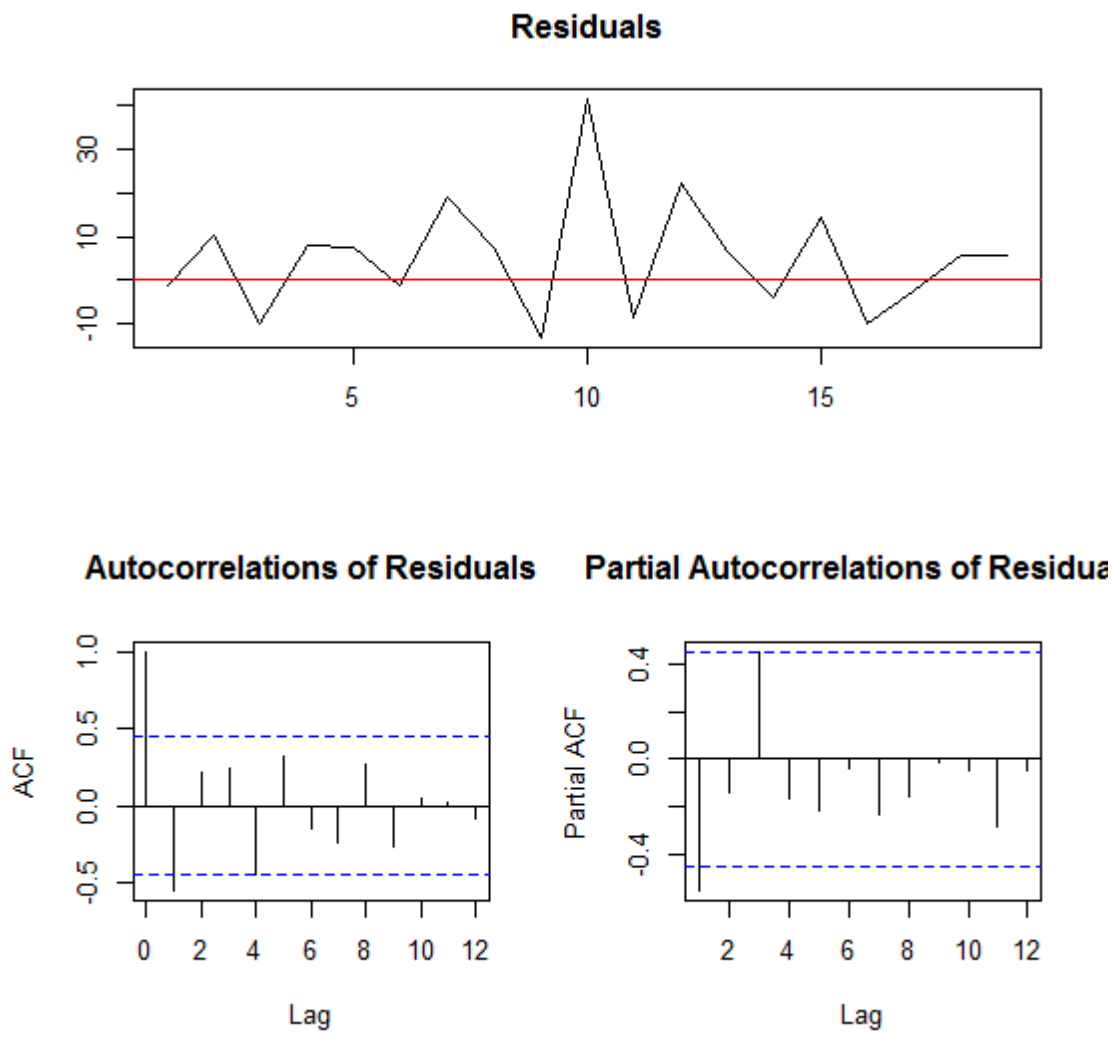

Figure 10: Plot of ADF cointegration residual

4.5 Granger causality

The test for Granger causality was conducted using Granger causality test.

The first test was whether HIV does not Granger cause TB for a model. Table 4.5 below shows the Granger causality test that HIV does not Granger cause TB

Table4.5: Granger causality test that HIV does not Granger cause TB

Model 1: TBCNR ~ Lags(TBCNR, 1:2) + Lags(HIV, 1:2)

Model 2: TBCNR ~ Lags(TBCNR, 1:2)

	Diff	Df	F	Pr(>F)
Complete model	14			
Reduced model	16	-2	9.817	0.00214 **

The p-value of 0.00214 means the null hypothesis is rejected hence conclude that HIV Granger cause TB

The second test is whether TB does not Granger cause HIV for model. The test results is shown in table 4.6 below

Table4.6: Granger causality test that TB does not Granger cause HIV

Model 1: HIV ~ Lags(HIV, 1:2) + Lags(TBCNR, 1:2)

Model 2: HIV ~ Lags(HIV, 1:2)

	Diff	Df	F	Pr(>F)
Complete model	14			
Reduced model	16	-2	1.2575	0.3146

The p-value of 0.3146 means the null hypothesis that TB does not cause HIV is not rejected hence conclude that TB do not Granger cause HIV

Chapter 5: Conclusions and Recommendations

The aim of this study was to determine the appropriate ARIMA model for TB case notification rates and HIV prevalence and to investigate presence or absence of granger causality relationship between TB and HIV.

5.1 Conclusions

The TB case notification rates showed increasing trend that peaked in 2005 and is presently declining. In comparison, the HIV prevalence rates showed an increasing a trend that peaked in 1999 and begun to decline in the past decade. The Kenyan twin epidemic show has time lag of 6 years between the trends of HIV prevalence and TB case notification rate.

The study has showed that ARIMA(0,1,2) model provides the best fit for HIV prevalence rate and that the ARIMA(1,2,0) model provides the best fit for the TB case notification rate.

The study demonstrated that there is a long run equilibrium relationship between HIV prevalence and TB notification rates. The current declining trends of TB cases may indicate that the efforts in HIV control could be driving down the TB epidemic

The study has demonstrated that there exist Granger causality relationship between TB and HIV. HIV Granger causes TB.

5.2 Recommendations

Kenya through NASCOP is moving towards universal ARV coverage for person living with HIV/AIDS (PLHIV) and the expect impact is the stagnation and possible rise of HIV prevalence as PLHIV live longer. TB case notification rate is expected to continue declining since ARVs is protective from TB disease. A further investigation on this co-dynamic relationship is recommended in view of the expected changed in pattern of trends.

References

Andrew Arnold, Yan Liu, Naoki Abe, 2007, Temporal Causal Modeling with Graphical Granger Methods, KDD'07 August 12–15

Division of Leprosy, TB and Lung Disease, 2011, TB Recording and Reporting system, DLTLD publication

Division of Leprosy, TB and Lung Disease, 2010, Annual Report, DLTLD publication

Division of Leprosy, TB and Lung Disease, DLTLD Strategic Plan 2011-2015, DLTLD publication 2010

Donald A Enarson and Nils E Billo,2007, Critical evaluation of the Global DOTS Expansion Plan, Bulletin of the World Health Organization 2007;85:395-403

Elisa F. Long, Naveen K. Vaidya, Margaret L. Brandeau, November–December 2008, Controlling Co-Epidemics: Analysis of HIV and Tuberculosis Infection Dynamics, Operations Research Vol. 56, No. 6, pp. 1366–1381

Hans L. Rieder, 1999, Epidemiologic Basis of Tuberculosis Control, First edition, IUATLD

Hong-Jen Chang, Nicole Huang, Cheng-Hua Lee, Yea-Jen Hsu, Chi-Jeng Hsieh, Yiing-Jenq Chou, April 2004The Impact of the SARS Epidemic on the Utilization of Medical Services: SARS and the Fear of SARS American Journal of Public Health, Vol 94, No. 4

Ion Dobre and Adriana AnaMaria Alendru, 2008 "Modelling unemployment rate using Box-Jenkins procedure", Journal of applied quantitative methods, Vol 3 No. 2

Joe Suyama, Matthew Sztajnkrycer, Christopher Lindsell, Edward J. Otten, Judith M. Daniels, Amy B. Kressel, July 2003, Surveillance of Infectious Disease Occurrences in the Community: An Analysis of Symptom Presentation in the Emergency Department, ACAD EMERG MED d, Vol. 10, No. 7 d www.aemj.org

José Leopoldo Ferreira Antune, Eliseu Alves Waldman, July/Sept. 1999, Tuberculosis in the twentieth century: time-series mortality in São Paulo, Brazil, 1900-97, Cad. Saúde Pública vol.15 n.3 Rio de Janeiro

Lih-Ing W. Roeger, Zhilan Feng, Carlos Castillo-Chavez, October 2009, MATHEMATICAL Biosciences and Engineering Volume 6, Number 4,

Robert Alan Yaffee, Journal of Statistical Software December 2007, Volume 23, Software Review 1.

Sánchez S, James O. Lloyd-Smith, Brian G. Williams, Travis C. Porco, Sadie J. Ryan, Martien W. Borgdorff, John Mansoer, Christopher Dye, Wayne M. Getz, 2009, Incongruent HIV and tuberculosis co-dynamics in Kenya: Interacting epidemics monitor each other, Epidemics 1,pg 14–20.

Saeed Akhtar, Hameed GHH Mohammad, 2008, Nonlinear pattern of pulmonary tuberculosis among migrants at entry in Kuwait: 1997–2006, BMC Public Health, 8:264

Spyros Makridakis and Michèle Hibon, ARMA models and the Box Jenkins methodology, INSEAD

Stephen D. Lawna, Motasim Badria and Robin Wooda, 2005, Tuberculosis among HIV-infected patients receiving HAART: long term incidence and risk factors in a South African cohort, AIDS, 19:2109–2116

World Health Organization, 2010, Global TB control report, WHO publication

World Health Organization, 2006, The Global Plan to Stop TB, 2006–2015: Actions for life towards a world free of tuberculosis, WHO/HTM/STB/2006.35