

MATHEMATICAL MODELING OF THE EPIDEMIOLOGY OF TUBERCULOSIS IN THE ASHANTI REGION OF GHANA

ABSTRACT

In this paper, autoregressive moving average (ARMA) model is used to predict the epidemiology and incidence of tuberculosis (TB) in the Ashanti Region of Ghana. The working data from the Ashanti Health services spans from January, 2001 to March, 2013. It is evident from the analysis that tuberculosis data in the Ashanti Region of Ghana could best be modeled with ARMA (1, 0) or AR(1) and that tuberculosis epidemic in the Ashanti Region is expected to be stable between April, 2013 and September, 2013. The TB data in the Ashanti Region of Ghana depicts a Stochastic Time Series and its model is linear. The Mean Absolute Error (MAE) and the Mean Squared Error (MSE) are used to compare the in-sample forecasting performance of three selected competing models and the result shows that it is not always true that the best selected model gives the best results so far as the mean square error (MSE) is concerned. The forecasting accuracy using MAE for AR(1) is calculated as 16.3171 and the forecasting accuracy using MSE for AR(1) is also calculated as 461.3148.

Keywords: Tuberculosis, epidemiology, models, forecasting, autoregressive, TB symptoms.

1.0 INTRODUCTION

Tuberculosis (TB) is a disease of poverty affecting mainly young adults in their most productive years. The vast majority of tuberculosis-related deaths occur in the developing world. An estimated 14 million people worldwide are infected with active tuberculosis. In 2009 there were 9.4 million new reported cases of tuberculosis and 1.7 million deaths with 380,000 deaths relating to tuberculosis among people with HIV.

The latest data released by the World Health Organization (WHO) in November 2010 showed that the number of new cases continues to fall globally and in five of the six WHO regions. The exception is Southeast Asia, where incidence remains stable. In many countries tuberculosis prevalence is declining. Worldwide, deaths from tuberculosis fell by 35 percent between 1990 and 2009 (Global Fund – 2010).

TB is an infectious disease that is caused by a bacterium called *Mycobacterium tuberculosis*. TB primarily affects the lungs, but it can also affect organs in the central nervous system, lymphatic system, and circulatory system among others. The disease was called "consumption" in the past because of the way it would consume from within anyone who became infected with it (Crosta – 2009).

The TB-causing bacteria usually cause an infection in the lungs, but it can affect other parts of the body such as the kidney, spine, and brain. If not treated properly, tuberculosis can be fatal (CCOHS – 2012).

TB is a contagious lung disease that spreads through the air. When people with the disease cough, sneeze, talk or spit, they propel TB germs, known as *bacilli*, into the air. Only a small number of the bacilli need to be inhaled to cause an infection. However, not all people infected with TB bacilli feel sick. The immune system either kills the germs, or "walls off" the TB bacilli where they can lie dormant for years. Failure of the immune system to control infection with TB bacilli leads to active disease, when TB bacilli

44 multiply and cause damage in the body. Left untreated, each person with infectious TB will spread the
45 germs to about 10 to 15 people every year (WHO – 2009).

46 The signs or symptoms of active tuberculosis infection include cough that lasts three weeks or longer,
47 presence of blood or sputum in cough, pain in the chest, loss of weight, loss of energy, poor appetite, fever,
48 chills, night sweats and others (CDC – 2012).

49 TB can be recognized through medical tests such as tuberculin skin tests, chest x-ray, and sputum smear
50 tests. People infected with either latent or active TB usually show a positive tuberculin skin test and blood
51 test results. Only people with active TB produce positive sputum smear results or have an abnormal chest
52 x-ray (CCOHS - 2012).

53 Medications are the cornerstone of tuberculosis treatment. However, treating TB takes much longer than
54 treating other types of bacterial infections. With tuberculosis, the patient must take antibiotics for at least six
55 to nine months. The exact drugs and length of treatment depend on the patient's age, overall health,
56 possible drug resistance, the form of TB and the infection's location in the body (Mayo Clinic – 2012).

57 A recent study suggests that a shorter term of treatment — three months instead of nine — with combined
58 medication may be effective in keeping latent TB from becoming active TB. With the shorter course of
59 treatment, people are more likely to take all their medication and the risk of side effects is lessened. More
60 study is needed to ascertain this recent finding.

61 A TB patient may need to take just one type of TB drug. Active tuberculosis, particularly if it is a drug-
62 resistant strain, will require several drugs at once. The most common medications used to treat
63 tuberculosis include isoniazid, rifampin (Rifadin, Rimactane), Ethambutol (Myambutol) or Pyrazinamide.
64 There's some evidence that taking vitamin D during tuberculosis treatment enhances some of the effects of
65 the drugs.

66 Side effects of TB drugs are not common but can be serious when they do occur. All tuberculosis
67 medications can be highly toxic to the liver. Patients on such medications are advised to call their doctors
68 immediately if they experience any of the following: nausea or vomiting, loss of appetite, a yellowish skin
69 color (i.e. jaundice), dark urine, a fever that lasts three or more days and has no obvious cause (Mayo
70 Clinic – 2012).

71 **1.1 Objectives of the Paper**

72
73 The objectives of this paper are:

- 74
75 1. to observe the pattern of tuberculosis infections in the Ashanti Region of Ghana from January 2001
76 to March 2013.
- 77
78 2. to model the epidemiology of tuberculosis in the Ashanti Region of Ghana.
- 79
80 3. to evaluate the impact of Akaike's Information Criteria (AIC) on in-sample forecasting performance
81 of the selected ARMA models.

83 **1.2 Significance of the Reported Research**

84 The significance of the contribution in this paper can be summarized as follows:

- 85 i. This paper provides a method for assessing the effectiveness of tuberculosis in the Ashanti Region.
- 86
87 ii. This paper contributes to the research information on tuberculosis in the country, so that it can help
88 in further work in the area of research to investigate the effects of the disease on Ghanaians.

89

iii. The paper attempts to present both application and theory at a level accessible to a wide variety of students, practitioners and researchers.

From the study of the open literature no research work has been so far reported on the model of epidemiology of tuberculosis in the Ashanti Region of Ghana. This paper therefore seeks to fill the research gap by studying the epidemiology of tuberculosis cases in the Ashanti Region of Ghana using time series analysis.

The remaining part of this paper is organized as follows. Section 1.3 surveys the research work relating to the subject of the paper. Section 2 discusses the data used, how they were obtained as well as the methods applied on them to obtain the needed results. Section 3 deals with data analysis, modeling and forecasting. Section 4 contains the summary and findings, conclusions and the recommendations of the paper.

1.3 Literature Review

From the work of White and Garnett (2010), despite the infectious agent that causes tuberculosis having been discovered in 1882, many aspects of the natural history and transmission dynamics of TB were still not fully understood. This was reflected in differences in the structures of mathematical models of TB, which in turn produced differences in the predicted impacts of interventions. Gaining a greater understanding of TB transmission dynamics required further empirical laboratory and field work, mathematical modelling and interaction between them. Modelling could be used to quantify uncertainty due to different gaps in their knowledge to help identify research priorities. Fortunately, the present moment was an exciting time for TB epidemiology, with rapid progress being made in applying new mathematical modelling techniques, new tools for TB diagnosis and genetic analysis and a growing interest in developing more-effective public-health interventions.

Claude et al (2011) addressed the spread of tuberculosis through one-strain and two-strain models. They first presented a basic model that incorporated fast and slow progression, effective chemoprophylaxis and therapeutic treatment. The system exhibited the traditional behavior. They proved that if the basic reproduction ratio $R_0 = 1$, then the disease-free equilibrium is globally asymptotically stable on the nonnegative out-hunts and if $R_0 > 1$ an endemic equilibrium exists and is globally asymptotically stable. Based on the first model, the second model dealt with the problem of drug resistance as a competition between multiple types of strains of mycobacterium tuberculosis: those that were sensitive to anti-tuberculosis drugs and those that were resistant. Their objective was to characterize the role of multi-drug-resistant in the transmission of tuberculosis. The coexistence and stability of the associated equilibria were discussed.

According to Caroline et al (2006), TB was a leading cause of infectious mortality. Although anti-biotic treatment was available and there was vaccine, tuberculosis levels were rising in many areas of the world. They used mathematical models to study tuberculosis in the past and had influenced policy; the spread of HIV and the emergence of drug-resistant TB strains motivated the use of mathematical models today. Here, they reviewed and compared the mathematical models of tuberculosis dynamics in the literature. They presented two models of their own: a spatial stochastic individual-based model and a set of delay differential equations encapsulating the same biological assumptions. They compared two different assumptions about partial immunity and explored the effect of preventative treatments. They argued that seemingly subtle differences in model assumptions could have significant effects on biological conclusions.

Laura and Maria (1997) on their part evaluated the efficacy of recommended tuberculosis infection control measures. They used a deterministic mathematical model for airborne contagion. They examined the percentage of purified protein derivative conversions under various exposure conditions, environmental control strategies, and respiratory protective devices. They concluded that environmental control cannot eliminate the risk for TB transmission during high-risk procedures; respiratory protective devices, and

138 particularly high-efficiency particulate air masks, may provide nearly complete protection if used with air
139 filtration or ultraviolet irradiation. Nevertheless, the efficiency of those control measures decreased as the
140 infectivity of the source case increased. Therefore, administrative control measures were the most effective
141 because they substantially reduced the rate of infection.

142 Also deserving mention is the work of Bowong et al (2012). They studied the spread of tuberculosis
143 through a two-patch epidemiological system $SE_1 \cdots E_n$ which incorporated migrations from one patch to
144 another just by susceptible individuals. Their model was considered with bilinear incidence and migration
145 between two patches, where infected and infectious individuals cannot migrate from one patch to another
146 due to medical reasons. They discussed the existence and uniqueness of the associated endemic
147 equilibria. They used quadratic forms and Lyapunov functions to show that when the basic reproduction
148 ratio is less than one, the disease-free equilibrium (DFE) is globally asymptotically stable, and when it is
149 greater than one there exists in each case a unique endemic equilibrium which was globally asymptotically
150 stable. Numerical simulation results were provided to illustrate the theoretical results.

151 Sematimba et al (2005) formulated mathematical models to establish the conditions on the size of the area
152 occupied required minimizing and thereafter eradicating tuberculosis. Both numerical and qualitative
153 analyses of the model were done and the effect of variation in the area size and recruitment rate on the
154 different epidemiological groups was investigated. Their results of the analysis showed that there exists a
155 stable disease-free equilibrium point provided that the characteristic area was greater than the product of
156 the probability of survival from the latent stage to the infectious stage and the number of latent infections
157 produced by a typical infectious individual during his/her mean infectious period. Their study recommended
158 that the characteristic area per individual should be at least 0.25 square kilometres in order to minimize the
159 tuberculosis incidence.

160 Adetunde (2009) reviewed a mathematical model of the dynamical behaviour of tuberculosis disease in the
161 Upper East Region of the Northern part of Ghana. The equilibrium points of the model system were found
162 and their stability was investigated. His model exhibited two equilibria, namely, the disease-free equilibrium
163 and the endemic equilibrium. He used stability theory and computer simulation, and observed that
164 population determine the infection rate of tuberculosis hence the higher the population density, the greater
165 the risk of instability of the disease free equilibrium.

166 According to Gammack et al (2005) the use of different mathematical tools to study biological processes is
167 necessary to capture effects occurring at different scales. They studied as an example the immune
168 response to infection with the bacteria *Mycobacterium tuberculosis*, the causative agent of tuberculosis.
169 They showed that the immune responses were both global as well as local in nature. They used four
170 different mathematical tools to explore both the global immune response as well as the more local one and
171 compared and contrasted results obtained using those methods. Applying a range of approaches from
172 continuous deterministic models to discrete stochastic ones allowed them to make predictions and
173 suggested hypotheses about the underlying biology that might otherwise go unnoticed. The tools
174 developed and applied were also applicable in other settings such as tumor modeling.

175 The strengths and limitations of using homogeneous mixing and heterogeneous mixing epidemic models
176 were explored in the context of the transmission dynamics of tuberculosis by Aparicio and Carlos (2009).
177 Their focus was on three types of models: a standard incidence homogeneous mixing model, a non-
178 homogeneous mixing model that incorporates 'household' contacts, and an age-structured model. The
179 models were parameterized using demographic and epidemiological data and the patterns generated from
180 those models were compared. Furthermore, the effects of population growth, stochasticity, clustering of
181 contacts, and age structure on disease dynamics were explored. That framework was used to assess the
182 possible causes for the observed historical decline of tuberculosis notifications.

183 The study looked at the estimation of the economic burden and household welfare impact of tuberculosis
184 (TB) in the Western Region of Ghana. Studies into the economic burden of TB in Ghana have been limited.
185 WHO's (2002a) guidelines on cost and cost effectiveness of TB management were followed in the

estimation of cost of TB from the patient/household and health provider perspectives. Human capital method was applied in the cost estimation. Wells-Riley model and multiple regression technique were employed in the estimation of the probability of transmission within households and the household welfare impact of TB. Results established that tuberculosis causes a significant deterioration in household income and welfare. The study also found that TB imposes various catastrophic economic costs on affected households and utilize considerable resources within the public health system. It is recommended that safety nets or income insurance be establish for households affected by TB to help cope with high economic burden and help patients fully complete treatment (Blankson – 2012).

2.0 MATERIALS AND METHODS

The data used for the modeling and analysis was obtained from the Ministry of Health in the Ashanti Region of Ghana. It was a secondary data. The data on monthly bases consist of the tuberculosis cases from various hospitals in the Ashanti Region for the period of January 2001 to December 2012.

Ashanti region is used as our case study area since the region is a cosmopolitan region in Ghana. All sorts of people are living in the region. The region is the third largest of 10 administrative regions in Ghana, occupying a total land surface of 24,389 square kilometers or 10.2 per cent of the total land area of Ghana. In terms of population, however, it is the most populated region with a population of 4,780,380 in 2010 population and housing census (PHC), accounting for 19.4 per cent of Ghana's total population; however, its density (148.1 per square km) is lower than those of the Greater Accra (895.5/km²) and Central (162.2/km²) Region. (Ghana Statistical Service, 2010).

The region is centrally located in the middle belt of Ghana. It lies between longitudes 0.15W and 2.25W, and latitudes 5.50N and 7.46N. The region shares boundaries with four of the ten political regions, Brong-Ahafo Region in the north, Eastern region in the east, Central region in the south and Western region in the south west. The region is divided into 30 districts, each headed by a district chief executive (Wikimedia Commons, 2013).

Majority of the region's population are Ghanaians by birth (87.3%) with about five per cent naturalized Ghanaians. A smaller proportion (5.8%) of the population originate from outside Ghana, made up of 3.7 per cent mainly from the five English-speaking countries of ECOWAS and 2.1 per cent from other African countries. The non-African population living in the region is 1.8 per cent of the total population. Akans are the predominant ethnic group in the region, representing 77.9% of Ghanaians by birth. A high proportion (78.9%) of the Akan population is Asante. The non-Akan population in the region comprises the Mole-Dagbon (9.0%), the Ewe (3.2%), the Grusi (2.4%), the Mande-Busanga (1.8%) and the Ga- Dangme (1.4%). The other smaller ethnic groups form about 1.3 per cent of the population of the region. (Wikipedia, 2013).

The model used to analyze the collected data is the Auto-Regressive Moving Average (ARMA) which was developed by Box and Jenkins in 1970, simply because the ARMA model fits the collected data very well as compared to other models. The **R** software package has been used in addition to manual calculations to model the given data. The Internet and Koforidua Polytechnic library are among the resources exploited.

The basis of the Box-Jenkins approach to modeling time series is summarized below and consists of three phases: identification, estimation and testing, and application (Spyros et al., 1998).

2.1 Identification Stage

Transform data to stabilize variance.

2.1.1 Difference data to obtain stationary series:

For non-seasonal data, first differencing is usually sufficient to attain stationarity. The first-order difference is denoted by equation (1).

$$\nabla X_t = X_t - X_{t-1} \quad (1)$$

2.1-2 The Sample Autocorrelation Functions

Another very important diagnostic tool for examining dependence is the sample autocorrelation functions. In this section, we examine data, Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to identify potential models.

The autocorrelation function (ACF), $\rho_{t,s}$, is given by the formula in equation (2).

$$\rho_{t,s} = \text{Corr}(X_t, X_s) = \frac{\text{Cov}(X_t, X_s)}{\sqrt{\text{Var}(X_t)\text{Var}(X_s)}} \quad \forall t, s \in \{0, \pm 1, \pm 2, \pm 3, \dots\} \quad (2)$$

$$\text{where } \text{Cov}(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)] = E(X_t, X_s) - \mu_t \mu_s$$

The autocorrelation coefficient estimated from sample observations at lag k is given by equation (3).

$$\gamma_k = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad \text{for } k = 1, 2, 3, \dots \quad (3)$$

For a variety of reasons, this has become the standard definition for the sample autocorrelation function. A plot of r_k versus lag k is often called a **correlogram** (Cryer and Chan, 2008).

2.3 Autoregressive Moving Average Model (ARMA)

We now proceed with the general development of autoregressive, moving average, and mixed autoregressive moving average (ARMA), models for stationary time series. A time series $\{X_t; t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ is ARMA (p, q) if it is stationary and given by equation (4).

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_{t-p} + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t \quad (4)$$

with $\phi_p \neq 0$, $\theta_q \neq 0$ and $\rho_\omega^2 > 0$.

The parameters p and q are called the autoregressive and the moving average orders, respectively. If X_t has a non-zero mean μ , we set $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ and write the model as shown in equation (5).

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_{t-p} + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \dots + \theta_q \omega_{t-q} + \omega_t \quad (5)$$

where $\{\omega_t; t = 0, \pm 1, \pm 2, \dots\}$ is a Gaussian white noise sequence and α is the intercept (Cryer and Chan, 2008).

2.3-1 ARMA (1, 1) Model

An example of an ARMA (p, q) model is the ARMA (1, 1) model given by equation (6).

$$X_t = \alpha + \phi_1 X_{t-1} + \theta_1 \omega_{t-1} + \omega_t \quad (6)$$

281 The ARMA (1, 1) model is stationary if $-1 < \phi_1 < 1$ and it is invertible if $-1 < \theta_1 < 1$.
 282 In an ARMA (1, 1) model both ACF and the PACF trial off to zero (Cryer and Chan, 2008).
 283
 284

285 2.4 Testing Stage

286
 287 While the approximate linear decay of the sample ACF is often taken as a symptom that the underlying
 288 time series is nonstationary and requires differencing, it is also useful to quantify the evidence of
 289 nonstationary in the data-generating mechanism. This can be done via hypothesis testing.
 290 Consider the model $X_t = \phi_1 X_{t-1} + Y_t$ for $t = 1, 2, 3, \dots$, where $\{Y_t\}$ is a stationary process. The process
 291 $\{X_t\}$ is nonstationary if the coefficient $\phi_1 = 1$, but it is stationary if $|\phi_1| < 1$. So, the null hypothesis
 292 corresponds to the case where the AR characteristic polynomial has a unit root and the alternate
 293 hypothesis states that it has no unit roots (Cryer and Chan, 2008).
 294

296 2.4-1 Augmented Dickey-Fuller (ADF) Test

297 The hypotheses $H_0: X_t$ is non – Stationary against the alternate hypothesis $H_1: X_t$ is Stationary
 298 can be tested in the regression equation (7).

$$299 \Delta X_t = \beta_0 + \alpha t + \beta_1 X_{t-1} + \sum_{i=1}^p \gamma_i \Delta X_{t-i} + \varepsilon_t \quad (7)$$

300 Accept H_0 if $P - value > 0.05$, else accept H_1 .
 301
 302

303 2.4-2 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

304 An alternative approach to the ADF test is the KPSS test. A hypotheses of $H_0: X_t$ is level or trend
 305 stationary is tested against $H_1: X_t$ is non- stationary in the regression equation (8).
 306

$$307 X_t = \alpha_t + \beta t + \mu_t, \text{ where a random walk, } \alpha_t = \alpha_{t-1} + \varepsilon_t \text{ is allowed.} \quad (8)$$

308
 309 Accept H_0 if $P - value > 0.05$, else accept H_1 .
 310
 311

312 2.5 Select Best Model Using Suitable Criterion:

313 A number of other approaches to model specification have been proposed since Box and Jenkins' seminal
 314 work. One of the most studied is Akaike's Information Criterion (AIC). This criterion says to select the
 315 model that minimizes AIC (Cryer and Chan, 2008).
 316

318 2.5-1 The Akaike's Information Criteria (AIC)

319 The AIC is equal to twice the number of parameters in the model minus twice the logarithm of the likelihood
 320 function. Mathematically, AIC is calculated by equation (9).
 321

$$322 AIC(p, q) = 2k - 2 \log (\text{maximum likelihood}) \quad (9)$$

323
 324 where $k = p + q + 1$ if the model contains an intercept or constant term and $k = p + q$ otherwise.
 325

326 Given two or more competing models, the one with the smaller AIC value will be deemed more appropriate
 327 (Cryer and Chan, 2008).

The AIC is a biased estimator, and the bias can be appreciable for large parameter per data ratios. Hurvich and Tsai (1989) showed that the bias can be approximately eliminated by adding another non-stochastic penalty term to the AIC, resulting in the corrected AIC, denoted by AICc and defined by equation (10).

$$AICc = AIC + \frac{2(k+1)(k+2)}{n-k-2} \quad (10)$$

where n is the sample size and k is the total number of parameters excluding the noise variance.

2.5-2 The Schwartz's Bayesian Information Criteria (BIC)

Another approach to determining the ARMA orders is to select a model that minimizes the Schwarz Bayesian Information Criterion (BIC). It is defined mathematically by equation (11).

$$BIC(p, q) = k \times \log(n) - 2 \times \log(\text{maximum likelihood}) \quad (11)$$

If the true process follows an ARMA (p, q) model, then it is known that the orders specified by minimizing the BIC are consistent; that is, they approach the true orders as the sample size increases. However, if the true process is not a finite-order ARMA process, then minimizing AIC among an increasingly large class of ARMA models enjoys the appealing property that it will lead to an optimal ARMA model that is closest to the true process among the class of models under study (Cryer and Chan, 2008).

2.6 Testing the Model for Adequacy (Portmanteau Test)

After identifying an appropriate model for a time series data, it is very important to check that the model is adequate. Box and Jenkins (1994) provided a modified portmanteau test statistic for checking the randomness of the error terms. Their statistic is given by equation (12)

$$Q^* = n(n+2) \times \sum_{k=1}^h \left(\frac{r_k^2}{n-k} \right) \quad (12)$$

which is approximately distributed as a χ^2 with $h-p-q$ degrees of freedom, where n is the length of the time series, h is the first h autocorrelations being checked, p is the order of the AR process and q is the order of the MA process and r_k is the estimated autocorrelation coefficient of the k^{th} residual term. If the calculated value of Q^* is greater than χ^2 for $h-p-q$ degrees of freedom, then the model is considered inadequate and the model is adequate if Q^* calculated is less than χ^2 for $h-p-q$ degrees of freedom. If the model is tested inadequate, then the forecaster should select an alternative model and test for the adequacy of the model (Spyros et al., 1998).

2.7 Measuring Forecast Accuracy

We now turn to another fundamental concern – how to measure the suitability of a particular forecasting method for a given data set. In most forecasting situations, accuracy is treated as the overriding criterion for selecting a forecasting method. In many instances, the word “accuracy” refers to “goodness of fit,” which in turn refers to how the forecasting model is able to reproduce the data that are already known. To the consumer of forecasts, it is the accuracy of the future forecast that is most important (Spyros et al., 1998).

2.7-1 Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is defined mathematically by equation (13).

375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \tag{13}$$

2.7-2 Mean Squared Error (MSE)

Symbolically, the Mean Squared Error (MSE) is defined by equation (14).

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \tag{14}$$

where X_t is the actual observation for time period t , F_t is the forecast value for the same period and $e_t = X_t - F_t$ is the error term and n is the number of forecasting values.

3 DATA ANALYSIS AND RESULTS

This chapter discusses the data analysis and modeling the data collected on tuberculosis cases in the Ashanti Region’s time series. Here we applied the **R** Statistical Package in modeling the time series. That is, all the plots and numerical output displayed in this paper have been produced with the **R** software. Most of the numerical outputs have been edited for additional clarity or for

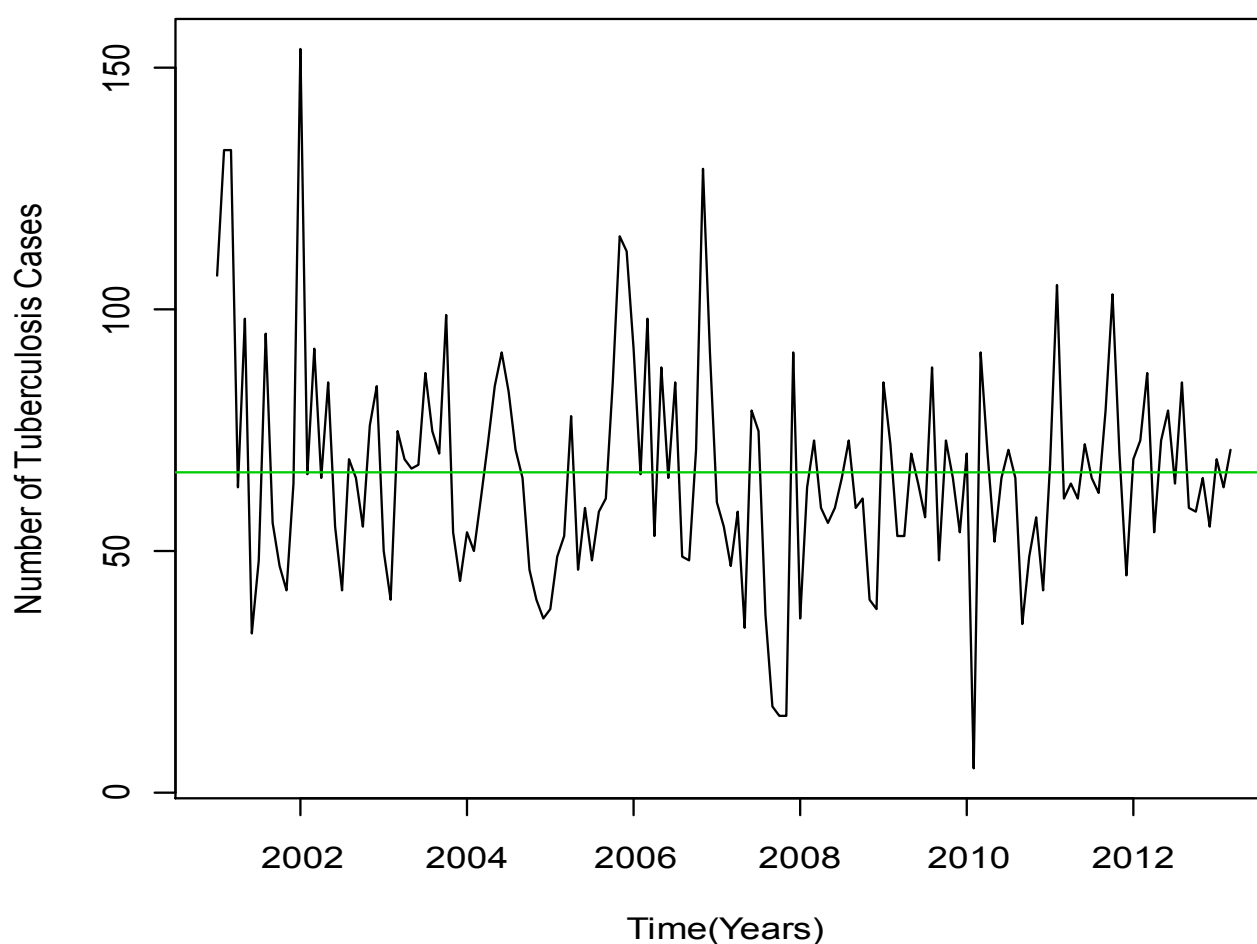
418 simplicity. Actual tuberculosis data drawn from various hospitals in the Ashanti Region of Ghana
 419 are used throughout in this chapter to illustrate the methodology explained in chapter 2.

420
 421

422 3.1 Time Plot of Prevalence of Tuberculosis (TB) in the Ashanti Region of Ghana

423

424 In general, the trend in TB prevalence in the Ashanti Region of Ghana seems to be irregular. The annual
 425 TB time plot in Figure 1 does not exhibit seasonal variation. The green line shows the mean of the series.
 426 Most of the data points are very close to the mean. This indicates that there is a clear case of stationarity in
 427 the mean. It follows that the TB series is stationary in the mean since the data fluctuate around a constant
 428 mean, independent of time, and the variance of the fluctuation remains essentially constant over time.
 429 There was not any seasonal behavior in the time plot, and hence the TB data looks to be approximately
 430 stable for further investigations.



431
 432

433 **Figure 1:** Time plot of TB cases in the Ashanti-Region of Ghana from January 2001 to March 2013.

434 3.1-1 Stationarity checks using the ACF and PACF

435 Figure 2 depicts the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the
 436 TB data. From Figure 2 (i) the spikes from 2 to 20 autocorrelation do not exceed two standard errors above
 437 zero (they are significantly near to zero). This shows that there is a stationarity in the TB data. Figure 2 (ii)
 438 exhibits the partial autocorrelation function (PACF) of the TB data. At lag zero the PACF is far away from
 439 unity (1) which confirms that the tuberculosis time series is stationary.

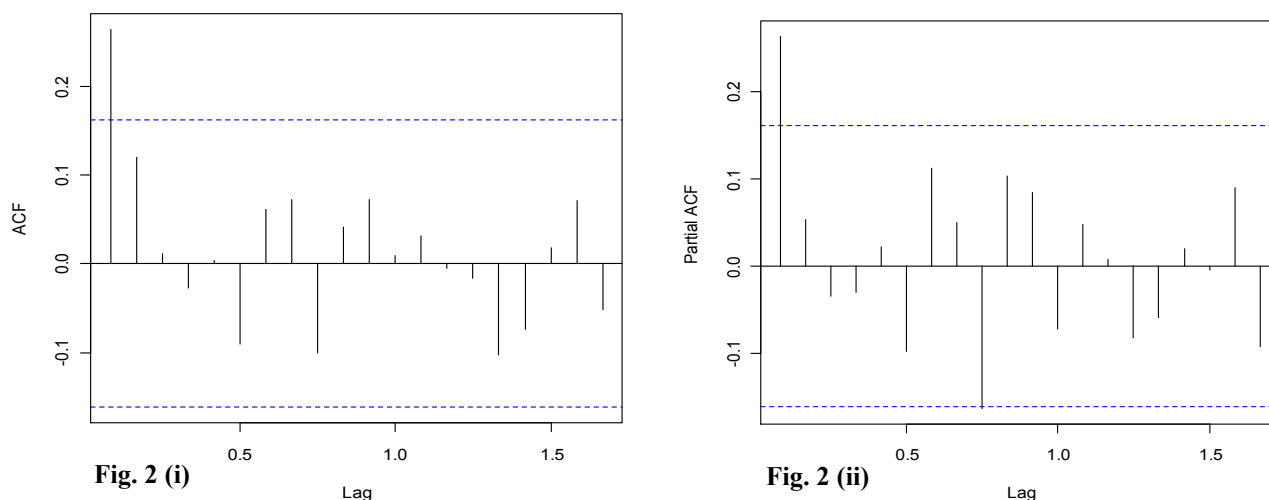


Figure 2: The Autocorrelation Function and Partial Auto-Correlation Function of prevalence of TB

3.1-2 Stationarity checks using the Augmented Dickey-Fuller (ADF) Test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

The Augmented Dickey-Fuller (ADF) Test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test for tuberculosis data are discussed below.

Table 1: ADF and KPSS Tests

Augmented Dickey-Fuller (ADF)Test		
Dickey-Fuller	Lag Order	P-Value
-5.1017	5	0.01
Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test		
KPSS Level	Lag Parameter	P-Value
0.2961	2	0.1

Table 1 shows both ADF and KPSS test results. There was stationarity in the original tuberculosis (TB) data; since p-value for ADF test was less than 0.05 and that of KPSS was greater than 0.05 using a 5% significant level. The two tests confirmed that there was stationarity in the original TB data and hence mean stationarity is achieved.

3.2 Selecting Competing Models Using ACF and PACF of the Tuberculosis (TB) Cases

The first part of Figure 2 shows the sample ACF of the TB cases. Except for marginal significance at spike 1, the model seems to have captured the essence of the dependence in the series. Inspecting the sample ACF, we see that PACF is tailing off and the ACF is cutting off at spike1 as shown in Figure 2 (i). This would suggest that the TB data follows an MA (1) model.

Figure 2 (ii) shows the sample PACF of the TB cases in the Ashanti Region of Ghana at different lags. Inspecting the sample PACF, we see that the ACF is tailing off and the PACF is cutting off at spike 1. Except for marginal significance at spike 1, the model seems to have captured the essence of the dependence in the series. This suggests an AR (1) for the TB cases.

As a preliminary analysis, we will fit both models. It follows that, in both the ACF and the PACF of the tuberculosis (TB) data in Figure 2, the following models were suggested:

1. ARMA (0, 1) or MA (1);
2. ARMA (1, 0) or AR (1)
3. ARMA (1, 1).

3.3 Estimation of Tentative Models

The estimations of the four selected tentative models with non-zero mean are discussed below.

3.3-1 Parameter estimate and diagnostics of ARMA (0, 1) model

Table 2: Parameter estimate for ARMA (0, 1) with non-zero mean

<i>Coefficient</i>	<i>Estimate</i>	<i>Standard Error</i>	<i> t – Value </i>	<i>Intercept</i>
ma1	0.2194	0.0735	2.9850	71.5757
<i>AIC</i>	<i>AICc</i>	<i>BIC</i>		<i>Constant</i>
1326.41	1326.69	1333.37		-0.0733

The coefficient of the estimated MA(1) parameter is within the causality condition bounds since its absolute t value is greater than 2 as shown in Table 2. The estimated ARMA (0, 1) model can be written as shown in equation (15).

$$X_t = 0.2194 \omega_{t-1} + 71.5024 \quad (15)$$

Table 3: Box-Ljung test of ARMA (0, 1) with non- zero mean

<i>Box - Ljung Test</i>		
X Squared	d f	p-value
14.8703	20	0.7838

Results from Table 3 showed that the model's residuals were non - significant with Ljung Box test statistic of 14.8703 and a p-value of 0.7838. Hence the model was adequate for forecasting.

3.3-2 Parameter estimate and diagnostics of ARMA (1, 0) model

Table 4: Parameter estimate for ARMA (1, 0) with non-zero mean

<i>Coefficient</i>	<i>Estimate</i>	<i>Standard Error</i>	<i> t – Value </i>	<i>Intercept</i>
--------------------	-----------------	-----------------------	--------------------	------------------

ar1	0.2557	0.0802	3.1883	71.7416
AIC	AICc	BIC	Constant	
1324.95	1325.23	1336.91	-0.0749	

The coefficient of the estimated AR (1) parameter is within the causality condition bounds since its absolute t value is greater than 2 as shown in Table 4. The estimated ARMA (1, 0) model can be written as shown in equation (16).

$$X_t = 0.2557 X_{t-1} + 71.6667 \quad (16)$$

Table 5: Box-Ljung test and Forecasts from ARMA (1, 0) with non- zero mean

<i>Box - Ljung Test</i>		
X Squared	d f	p-value
15.2002	20	0.7648

Results from Table 5 shows that the model's residuals were non-significant with Ljung Box test statistic of 15.2002 and a p-value of 0.7648. Hence the model was adequate for forecasting.

3.3-3 Parameter estimate and diagnostics of ARMA (1, 1) model

Table 6: Parameter estimate for ARMA (1, 1) with non-zero mean

<i>Coefficient</i>	<i>Estimate</i>	<i>Standard Error</i>	<i> t-value </i>	<i>Intercept</i>
ar 1	0.4184	0.2537	1.6492	71.9467
ma 1	-0.1716	0.2687	0.6386	
<i>AIC</i>	<i>AICc</i>	<i>BIC</i>	<i>Constant</i>	
1326.57	1326.99	1341.52	-0.0768	

The coefficients of the estimated ARMA (1, 1) parameters are outside the causality and invertibility condition bounds since their absolute t -values are less than 2. The estimated ARMA (1, 1) model can be written as shown in equation (17).

$$X_t = 0.4184X_{t-1} - 0.1716\omega_{t-1} + 71.8699 \quad (17)$$

Table 7: Box-Ljung test and Forecasts from ARMA (1, 1) with non- zero mean

<i>Box - Ljung Test</i>		
X Squared	d f	p-value
15.2481	20	0.7620

Results from Table 7 showed that the model's residuals were non - significant with Ljung Box test statistic of 15.2481 and a p-value of 0.7620. Thus the model was adequate for forecasting.

3.4 Forecasting from ARMA (1, 0), the best model

Table 10: Point Forecasts from ARMA (1, 0) with non-zero mean

Point Forecast From ARMA (1, 0)					
Apr-13	May -13	Jun -13	Jul -13	Aug -13	Sep -13
67.4969	66.5555	66.3024	66.2344	66.2162	66.2112

From **Figure 3**, the yellow region depicts the 95% confidence interval, the red region is the 85% confidence interval and the blue line is the forecasting points. The model was used to forecast six months ahead and showed that the tuberculosis prevalence in the Ashanti Region of Ghana will be slightly decreased from April, 2013 to September, 2013 as shown in the blue line as well as Table 10, its forecast points.

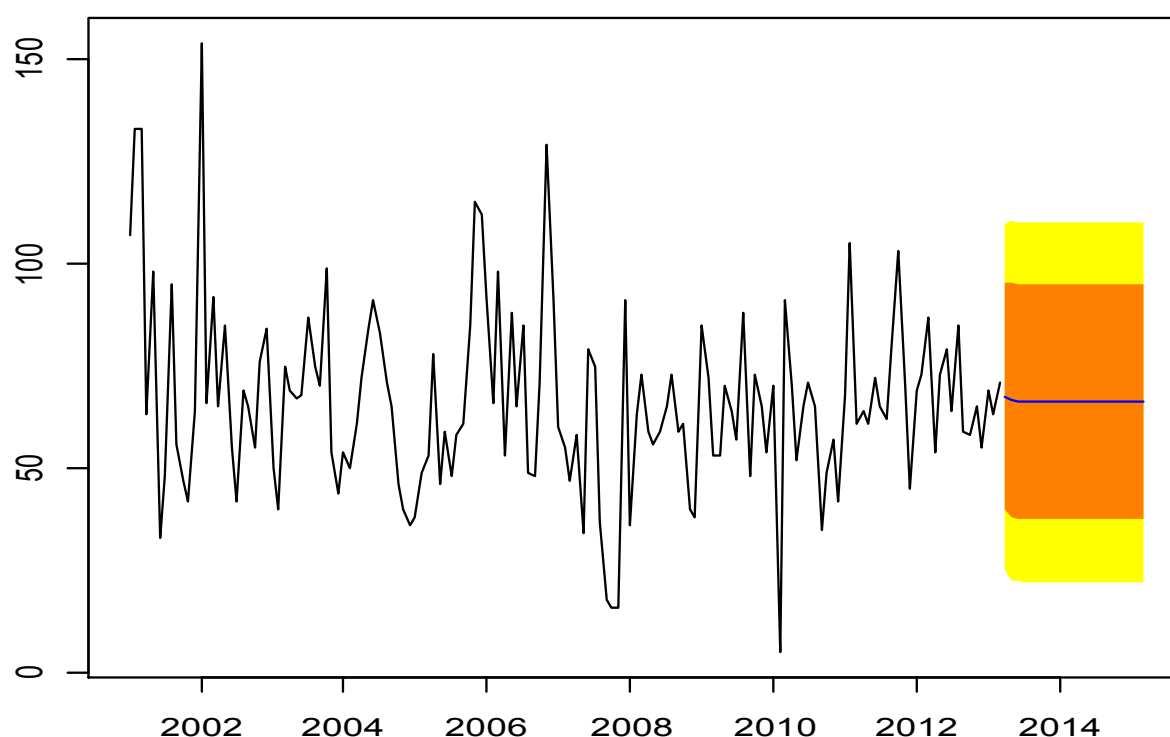


Figure 3: Forecasts from ARMA (1, 0) with non-zero mean

3.5 The Error Metrics

Forecasting accuracy based on the Mean Absolute Error (MAE) of the forecasted values was checked for each fitted model as shown in Table 12. It highly favored the forecasted value of ARMA (1, 0), the best selected model. This means that, the ARMA (1, 0) forecast error of 16.3171 out-performed all the forecast errors so far as the MAE is concerned. Hence ARMA (1, 0) was confirmed to be the best model.

Similarly, the forecasting accuracy based on the Mean Squared Error (MSE) of the forecasted values favored ARMA (1, 1), one of the competing selected models. This means that, the ARMA (1, 1) forecast error of 460.1114 out-performed all the forecast errors so far as the MSE is concerned.

573 **Table 12: The Mean Absolute Error (MAE) and the Mean Squared Error (MSE)**

(i) Mean Absolute Error (MAE)		(ii) Mean Squared Error (MSE)	
ARMA(0, 1)	16.3394	ARMA(0,1)	465.9801
ARMA(1, 0)	16.3171	ARMA(1, 0)	461.3148
ARMA(1, 1)	16.3172	ARMA(1, 1)	460.1114

581
582

4.0 CONCLUSION, FINDINGS, RECOMMENDATIONS AND SUGGESTIONS

4.1 Conclusion

586 From Table 13, ARMA (1, 0) had the lowest AIC value of 1324.96. It was the best model so far as the AIC
587 was concerned.

588

589 **Table 13: Summary of Diagnostics Test**

MODEL	AIC	AICc
ARMA (1, 0)	1324.96	1325.23
ARMA (0, 1)	1326.35	1326.69
ARMA (1, 1)	1326.57	1326.99

590
591
592
593
594
595
596

597 Hence the best selected model was $X_t = 0.2557 X_{t-1} + 71.6667$

598

599 However, the forecasting accuracy based on the Mean Absolute Error (MAE) for ARMA (1, 0), the best
600 selected model, was the lowest and It's forecast error was calculated as 16.3171.

601

602 In conclusion, the research study reported in this monograph has found that tuberculosis data in the
603 Ashanti Region of Ghana could best be modeled with ARMA (1, 0) or AR(1). The study again found out
604 that tuberculosis prevalence in the Ashanti Region is expected to stable from April, 2013 to September,
605 2013.

606

607 This model did not consider mass vaccination as one of the methods to prevent the prevalence of
608 tuberculosis in the region. The results of this paper can be used as a tool to facilitate the introduction of
tuberculosis vaccine and improve tuberculosis vaccination in the country as a whole.

4.2 Findings:

610

611 The results of this paper showed that it is not always true that the best selected model gives the best
612 results so far as the mean squared error (MSE) is concerned.

613

614 From definitions, the tuberculosis data in the Ashanti Region of Ghana from January, 2001 to March, 2013
615 depicts a *Stochastic Time Series* and its model is *linear*.

616

4.3 Recommendation

617 The following recommendation is being made to the stakeholders of Ashanti Region:

- 618 • Ghana Health Service should continue the mass tuberculosis vaccination in the region to possibly
619 eradicate the disease.

620

621

4.4. Suggestions

The following suggestions are being made to the stakeholders of Ashanti Region:

- Stakeholders including traditional authorities, politicians and opinion leaders should be actively involved in the tuberculosis campaign to encourage patients to seek early treatment.
- The participants should use film shows, public forum and drama among other interventions to educate the people about the disease.
- Proper education on TB would help reduce the huge amount of money government spends in combating and treating the disease.

REFERENCES

- Adetunde I. A.(2009) "The Mathematical Models of the Dynamical Behaviour of Tuberculosis Disease in the Upper East Region of the Northern Part of Ghana: A Case Study of Bawku", *Current Research in Tuberculosis*, 1: 15-20. **DOI:** 10.3923/crt.2009.15.20 , **URL:** <http://scialert.net/abstract/?doi=crt.2009.15.20>
- Akaike, H. (1974). "A new look at the statistical model identification." *IEEE Transactions on Automatic Control*, 19, 716 – 723.
- Aparicio Juan Pablo and Carlos Castillo-Chavez Abstract (2009), "Mathematical modelling of tuberculosis epidemics", *Mathematical biosciences and engineering MBE* (2009), Volume: 6, Issue: 2, Publisher: School of Science and Technology, Universidad Metropolitana, San Juan 00928-1150, Puerto Rico. juan.p.aparicio@gmail.com, Pages: 209-237
- "Ashanti Region" (2010) Ghanadistricts.com. <http://ghanadistricts.com/region/?r=2&sa=8>.
- "Ashanti Region" (2010) Ghanadistricts.com. <http://ghanadistricts.com/region/?r=2&sa=12>.
- Blankson Hammond Kodwo (2012), "Economic Burden of Tuberculosis in the Western Region of Ghana",
- Bowong S., Tewa J. L. and Noutchie O. S. C. (2012), "Mathematical analysis of a two-patch model of tuberculosis disease with staged progression", <http://dx.doi.org/10.1016/j.apm.2012.01.026>.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994), "Time Series Analysis: Forecasting and Control", 3rd edition, Prentice Hall, Englewood Cliffs, New Jersey.
- Canadian Centre for Occupational Health and Safety (2012), www.cchs.ca/oshanswers/disease/tubercul.html, OSH
- Caroline C., Ted C. and Megan M. (2006), "Mathematical Models of Tuberculosis: Accomplishments and Future Challenges", *Proceedings Trim Size: 9in x 6in colijn*, seis.bris.ac.uk/~encgc/colijn_brazil.pdf
- Centers for Disease Control and Prevention (2012), 1600 Clifton Rd. Atlanta, GA 30333, USA
800-CDC-INFO (800-232-4636) TTY: (888) 232-6348

- 660 Claude Jean Kamgang, Bowong Samuel, Jean Jules Tewa (2011), “Stability analysis of the transmission
661 dynamics of tuberculosis models”,
662 www.worldacademicunion.com/journal/1746.../wjmsvol07no02paper01.pdf
- 663 Crosta Peter M. A. (2009), ‘Medical News Today’, www.medicalnewstoday.com/articles/8856.php,
664 0:00 PST
- 665 Cryer D. Jonathan and Kung – Sik Chan (2008), “Time Series Analysis With Applications in R”, Second
666 Edition, U.S.A., Springer.
- 667 Gammack David, Suman Ganguli, Simeone Marino, Jose Segovia-juarez, Denise and E. Kirschner
668 (2005), “Understanding the immune response in tuberculosis using different mathematical models
669 and biological scales”
- 670 Ghana News Agency (2011), www.ghanadistricts.com/districts1on1/kma/
- 671 Ghana Statistical Service (2010). Population and Housing Census (PHC).
672
- 673 Global Fund (2010), www.theglobalfund.org/Tuberculosis.
- 674 Hurvich C. M. and Tsai C. L. (1989), “Regression and time series model selection in small samples.”
675 *Biometrika*, 76, 2, 297 – 307.
- 676 Laura Gammaitoni and Maria Clara Nucci (1997), “Using a Mathematical Model to Evaluate the Efficacy of
677 TB Control Measures”, wwwnc.cdc.gov/eid/article/3/3/97-0310_article.htm
678
- 679 Mayo Clinic (2012), www.mayoclinic.com/.../tuberculosis/.../DSECTION=treatments-and-drugs.
- 680 Sematimba, A.; Mugisha, J. Y. T.; Luboobi, L. S. (2005), “Mathematical Models for the Dynamics of
681 Tuberculosis in Density-Dependent Populations: The Case of Internally Displaced Peoples' Camps
682 (IDPCs) in Uganda”, Academic journal article from *Journal of Mathematics and Statistics*, Vol. 1,
683 No. 3
- 684 Shumway H. R. and Stoffer S. D. (2006), “Time Series Analysis and Its Applications With R Examples”,
685 Second Edition, U.S.A., Springer.
- 686 Spyros M., Wheelwright S. C. and Hyndman R. J. (1998), “Forecasting: Methods and Applications”, Third
687 Edition, U.S.A., John Wiley & Sons, Inc.
- 688 White P. J. and Garnett G. P. (2010), “Mathematical modelling of the epidemiology of tuberculosis”, *Adv*
689 *Exp Med Biol.* 2010; 673:127- 40, www.ncbi.nlm.nih.gov/pubmed/20632534
- 690 Wikimedia Commons (2013), http://commons.wikimedia.org/wiki/File:Ashanti_districts.png
- 691 Wikipedia (2013), http://en.wikipedia.org/wiki/Ashanti_Region
- 692 World Health Organization (2009), www.who.int/features/qa/o8/en/index.html
693
694