The Incidence of Tuberculosis in Brazil from 2001 to 2018: Use of Polynomial Regression Combined with a Stochastic Volatility Model

Jorge Alberto Achcar¹, Ricardo Puziol de Oliveira² and Emerson Barili²*

¹Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil
²State University of Maringá, Maringá, PR, Brazil

Abstract

Monthly tuberculosis (TB) counts from January 2001 to December 2018 showed an upward trend (data from SINAN - Information System for Notifiable Diseases). The increase in the incidence of TB in general is associated with an increase in the rate of extreme poverty, an increase in AIDS cases and other factors. Combining the polynomial linear regression and stochastic volatility models, the purpose of this study was to analyze monthly count data as well as the AIDS/TB, extreme poverty/TB and urban/TB rates associated with the incidence of TB. The results were obtained using the Monte Carlo method in Markov chains under a Bayesian approach. In analyzing the data with the presence of great volatility, despite the growth of TB cases in Brazil, some rates such as AIDS/TB and extreme poverty/TB are decreasing in the same period, indicating that TB cases have increased in the last few years for all sectors of the population of Brazil. The proposed model can be of great use for other epidemiological data, in addition to modeling TB counts and rates associated with TB.

Keywords

TB count, Polynomial regression models, Stochastic volatility models, Bayesian methods, MCMC simulation

Introduction

Tuberculosis (TB) is an infectious disease, caused by Mycobacterium tuberculosis. The disease is usually presented as pulmonary TB, but it can also affect other organs (extrapulmonary TB). Tuberculosis usually progresses slowly from the latent phase (infection without active disease, LTBI-Latent Tuberculosis Infection) to active TB, except in HIV-infected patients where the progression can be rapid and fatal. Only a small proportion (5% to 10%) of people infected with M. tuberculosis progress to active TB as reported by the [1]. Despite great scientific advances, the evolution of treatment and the discovery of a vaccine have contributed to the control of TB, it is still among the infectious diseases of high lethality in the world. Global trends show that the incidence of tuberculosis (TB), prevalence and mortality rates are gradually decreasing worldwide. TB counts in all countries and regions of the world were in great decline until the early 1980’s when the first cases of patients infected with the HIV virus leading to AIDS emerged and consequently leading to a large increase in the incidence of TB in the World. After the discovery of new therapies against HIV, TB counts started to drop in practically all countries. Tuberculosis remains a poverty-related disease with a high burden of disease in low- and middle-income countries and in countries with a high incidence of HIV [2].

In Brazil, especially with the great economic crisis from the years 2014 with the increase of unemployed workers and significant increase of individuals living in extreme poverty, a new trend of increase in TB cases has been observed, regardless of the increase in AIDS cases. Many studies related to the count of individuals diagnosed with TB have been published in the literature in Brazil and in other countries.

In this direction [3], analyzed some estimates made...
by the WHO and indicators of incidence, mortality, occurrence of multidrug resistance and association with HIV for tuberculosis in Brazil and worldwide. These authors present estimates (8.7 million new cases per year in the world and 116,000 new cases per year in Brazil) or actually verified (Brazil in 2000 notified 82,249 new cases) pointing out to a serious public health situation, mainly in developing countries, which it requires energetic and effective measures for its control. In a country of high prevalence like Brazil, the actions for the discovery of new cases, associated with the measures of biosafety, are of interest to all health professionals, mainly those who work in large hospitals or emergencies.

DC Costa [4] analyzed several aspects that influence the pattern of tuberculosis mortality and morbidity: The variation in living conditions, the effects of chemotherapy and the dynamics of disease transmission which characterizes the indicators used in the measurement of tuberculosis that are built from prevalence surveys and notification data. The main indicators are: risk of infection, mortality, incidence coefficient of tuberculous meningitis, incidence coefficients of tuberculosis cases, especially pulmonary bacilliferous among others. In that study, data on the incidence of bacilliferous pulmonary cases in the Brazilian provinces of Pará and Rio Grande do Sul and mortality from tuberculosis in the capitals of Pará, Rio Grande do Sul, Pernambuco and Rio de Janeiro were analyzed, which reveal in general, a downward trend at that time, still that at a different pace.

Many other studies are presented in the literature. LMW Acosta, et al. [5] relate some social determinants and the incidence of TB in Porto Alegre, RS, Brazil. Several recent studies consider different statistical models that capture temporal or spatial effects in the analysis of TB epidemiological series [6-12].

Some studies on the incidence of TB have used ARIMA [13] time series models (autoregressive integrated moving average models) or ARCH [14] models (autoregressive conditional heteroscedasticity) in the analysis of epidemiological series. In this sense, RR Sarkar, et al. [15] consider the application of different time series models for epidemiological data with application to malaria. YL Zheng, et al. [16] use ARIMA models and a combination of ARIMA-ARCH models for predicting and monitoring monthly tuberculosis morbidity in Xinjiang, China. In this case, the ARCH model is suggested to provide tuberculosis surveillance, providing estimates on tuberculosis morbidity trends in Xinjiang, China. Other applications in epidemiology can also be seen in [17-25].

In the present study, a new formulation of statistical modeling (combination of polynomial linear regression models with stochastic volatility models) is proposed for the analysis of TB count data and associated rates using Bayesian methods of statistical inference. The article is formulated as follows: in section 2 the monthly data on TB in Brazil for the period 2001 to 2018 are presented; section 3 presents the material and methods; section 4 presents the obtained results; finally, section 5 presents conclusions and discussion of the obtained results.

TB Data in Brazil for the Period 2001-2018

This study considers TB count data in Brazil between the years 2001 and 2018 and some existing relationships given by other time series in the same period such as rates of AIDS patients among patients diagnosed with TB, rates of patients living in urban centers among patients diagnosed with TB, rates of male patients among patients diagnosed with TB, rates of patients in extreme poverty among patients diagnosed with TB among several other factors. Figure 1(a) shows a graph of the yearly time series of TB patient counts in Brazil (data from the SINAN-Information System for Notifiable Diseases-http://sinan.saude.gov.br/sinan/login/login.jsf) for the period between January 2001 and December

*Figure 1:* Yearly counts of TB patients in Brazil and yearly rates per 100,000.
2018 (data set in Appendix 1 at the end of the paper). Figure 1(b) also shows the rates (number of TB cases in the year divided by the estimates of Brazil’s population in the year, source IBGE). The data are presented in Table 1, from where it is observed an increase in TB rates between the years 2001 and 2003; from 2003 onwards there was a decline in yearly TB rates until 2016, possibly due to the efficacy of drugs to control AIDS; after that year, an increase in TB rates is observed again, perhaps as a result of the great economic crisis affecting Brazil (increase in extreme poverty, increase in the unemployed rates, lower investments in health, among many other factors). From Figure 1, it is observed two change-points: One change-point around 2003 and a second change-point close to the year 2017.

Figure 2 shows the monthly TB counts in Brazil (data from SINAN) from January 2001 to December 2018 (216 months). From the graphs presented in Figure 2, it can be seen that:

- The monthly counts of patients diagnosed with TB in the period between January 2001 and December 2018 (original scale and logarithmic scale) show seasonality for the seasons and there is clearly a yearly decrease in counts until the year 2014 where the beginning of an increase in TB cases is observed.
- It is also observed that until the mid-2014 there is a tendency to increase the rates of AIDS patients among TB patients; in that year there is a beginning of a decrease in the rates of AIDS patients among those infected with tuberculosis. This is also seen for rates of HIV-positive patients among TB patients. It is important to note that the incidence of TB was in great decline worldwide until the mid-1980s, the year of the beginning of the AIDS epidemic, a factor that led to the reappearance of TB in large numbers. From these graphs it is possible to conclude that there is a strong indication of the increase in TB in Brazil for all segments of the population regardless of the effect of HIV, as the rates of HIV-infected patients diagnosed with TB are declining, despite the increase in TB cases.
- It is also observed that the rates of urban patients among those infected with TB had an annual increase until mid-2012; as of that year, there has been a decline in the number of patients living in

| Year | Total TB | Estimated pop BR | Yearly TB rate/100,000 |
|------|----------|------------------|------------------------|
| 2001 | 87265    | 177196054        | 49.25                  |
| 2002 | 92859    | 179537520        | 51.72                  |
| 2003 | 93773    | 181809246        | 51.58                  |
| 2004 | 92980    | 184006481        | 50.53                  |
| 2005 | 92056    | 186127103        | 49.46                  |
| 2006 | 86160    | 188167356        | 45.79                  |
| 2007 | 85779    | 190130443        | 45.12                  |
| 2008 | 87622    | 192030362        | 45.73                  |
| 2009 | 87227    | 19386508         | 44.99                  |
| 2010 | 86146    | 195713635        | 44.02                  |
| 2011 | 88560    | 197514536        | 44.84                  |
| 2012 | 86905    | 199287299        | 43.61                  |
| 2013 | 86208    | 20103912         | 42.88                  |
| 2014 | 85213    | 202763739        | 42.03                  |
| 2015 | 85443    | 204471769        | 41.79                  |
| 2016 | 86026    | 206163053        | 41.73                  |
| 2017 | 90210    | 207833823        | 43.40                  |
| 2018 | 94155    | 209469323        | 44.95                  |

Table 1: Yearly counts of TB patients in Brazil and yearly TB rates.

Figure 2: Time series of counts and rates of TB patients in Brazil.
urban centers (people most susceptible to HIV infection) among TB patients. This also shows that the incidence of TB is gradually increasing for all segments of the population in Brazil, a result of great serious impact to the Brazilian health authorities. The same is observed for patients in extreme poverty (annual decrease in the rates of individuals in extreme poverty among TB patients).

Figure 2 shows, in most cases, for the data analysis and statistical data modeling, the need for polynomial linear regression models in the presence of the covariate (independent variable) month (temporal order) that can capture possible effects of linearity and that also have regression coefficients associated with quadratic and cubic effects. A quadratic term or cubic term transforms a linear regression model into a curve. Since the regression model has the covariate month squared or cube, and not the regression coefficient, the model remains a linear regression model \[26,27\]. An application of polynomial regression models with epidemiology data is given by \[28\]. The presence of a quadratic term in the model creates an inverted U- or U-shaped curve, and not the regression coefficient, the model remains a linear regression model \[26,27\]. An application of polynomial regression models with epidemiology data is given by \[28\].

The main goals of the study are:

• To model the monthly TB counts in Brazil with data on the logarithmic scale from 2001 to 2018, assuming stochastic volatility models to estimate the time series volatilities;
• To model some rates of factors associated with the incidence of TB using stochastic volatility models such as rates of patients diagnosed with TB and AIDS, rates of patients diagnosed with TB and urban dwellers and rates of patients diagnosed with TB and belonging to the socioeconomic groups of extreme poverty.
• To study the variability in the rates of patients diagnosed with TB and AIDS, rates of patients diagnosed with TB and urban dwellers and rates of patients diagnosed with TB and belonging to extreme poverty socioeconomic groups that are decreasing despite the growth of TB cases seen in Brazil in recent years.
• To model yearly TB rates in Brazil from 2001 to 2018, assuming linear regression models in the presence of change-points.

Material and Methods

Use of standard linear regression models

In a first data analysis, it is assumed a standard polynomial linear regression model given by:

\[
Y(t) = \beta_0 + \beta_1 \text{months}(t) + \beta_2 \text{[months(t)]}^2 + \beta_3 \text{[months(t)]}^3 + \epsilon(t) 
\]

for \( t = 1, 2, ..., 216 \) (months) where \( \epsilon(t) \) are noises considered independent and identically distributed random variables with a normal distribution \( N(0, \sigma^2) \). Under a classic approach, the regression parameters are usually estimated using the least squares method (EMQ). An alternative of epidemiological interest would be to model the series not only to estimate monthly averages, but also to estimate monthly variances (volatilities) that are of interest to the public health area, possibly relating these volatilities to the occurrence of non-observed factors associated with the months (season, unemployment rate, extreme poverty rate, among others). For this purpose, time series models are used in the data analysis, which simultaneously estimate the monthly average and the monthly volatility.

Use of polynomial regression models and stochastic volatility for TB data in Brazil

Stochastic volatility (SV) models have been widely used to analyze financial time series \[29\] as a powerful alternative to some existing regression models in the literature, such as the ARCH (conditional heteroscedastic autoregressive) models introduced by \[30\] and the generalized autoregressive conditional heteroscedastic models (GARCH) introduced by \[31\], but not widely used in the health field \[32-34\]. In one of these few health applications, \[35\] studied neural signals using a database of medial temporal lobe (MTL) recordings from 96 neurosurgical patients, using time series models with volatility described by a multivariate stochastic latent variable process and lagged interactions between signals in different brain regions providing new insights into the dynamics of brain function. In the financial area, these models are considered for modeling the logarithms of financial returns between current data and previous data (can be hours, days or months) without taking into account the modeling of the means depending on covariates.

The novelty of this study is the introduction of stochastic volatility models for epidemiological time series, in particular, for tuberculosis counting series in Brazil.

For the definition of the models, \( N \geq 1 \) is a fixed integer that records the amount of observed data (in our case, it will represent the monthly TB counts on the logarithmic scale and also the rates of patients in certain groups among those infected with TB).

In the presence of heteroscedasticity, that is, variances depending on time \( t \), assume that the time series \( Y(t), t = 1, 2, ..., N \) assume the combined polynomial linear regression model with the stochastic volatility model for the analysis of TB count data on the logarithmic scale and some rates related to the TB epidemic:
Y(t) = β_o + β_1 months(t) + β_2 [months(t)]^2 + β_3 [months(t)]^3 + σ(t)e(t)  \hspace{1cm} (2)

where it is assumed that e(t) are noises considered independent and identically distributed random variables with a normal distribution N(0, σ^2) and σ(t) is the square root of the variance of (1) (for simplicity, we can assume σ^2 = 1). The variance of Y(t) is modeled by σ_e^2 e^(2ζ) where h(t) depends on an unobserved latent variable.

Remark: From model (2), it is observed that:

The mean of Y(t) is given by E[Y(t)] = β_o + β_1 months(t) + β_2 [months(t)]^2 + β_3 [months(t)]^3 since E[σ(t)e(t)] = 0.

The variance of Y(t) is given by var[Y(t)] = var[σ(t)e(t)] = σ_e^2(t) since we are assuming that var[e(t)] = σ_e^2 = 1.

To analyze the data set, it is introduced a latent variable (unobserved variable) defined by an auto-regressive model AR (2), for t = 1, 2, 3, ..., N (N = 216 months).

h(1) = μ + ζ(1), t = 1,
h(2) = μ + φ_1 h(1) - μ + ζ(2) \hspace{1cm} (3)
h(t) = μ + φ_1 h(t-1) - μ + φ_2 h(t-1) - μ + ζ(t), t = 3, 4, ..., N,

where ζ(t) is a noise considered to be independent and identically distributed random variables with a normal distribution N(0, σ_ζ^2), which is associated to a latent variable h(t). The quantities σ_ζ, μ, φ_1, and φ_2 are unknown parameters that should be estimated (0 < φ_1 < 1, 0 < φ_2 < 1).

Bayesian inference procedures, based on Markov Chain Monte Carlo simulation methods (MCMC) [34,36,37] have been widely used to analyze stochastic volatility models. The main reason for using Bayesian methods is that, in general, we may have great difficulties in obtaining inferences (point and interval estimation) for the parameters of interest of the stochastic volatility model when using a standard classical inference approach. These difficulties can appear in the form of high dimensionality and likelihood function without closed form among other factors.

For a Bayesian analysis of the model (1), we assume that the prior distributions for the parameters μ, φ_1, and ζ = 1/σ_ζ^2, v = 1, 2 are respectively given by the normal distribution, Beta(b,c) distribution and a Gamma(d,e) distribution, where Beta(b, c) denotes a beta distribution with mean b/(b + c) with variance bc/[(b + c)^2(b + c + 1)] and Gamma(d, e) denotes a gamma distribution with mean d/e and variance d/e^2. The hyperparameters a, b, c, d, e are assumed to be known and specified previously. Also assume that the regression parameters β are independent normal distributions with known hyperparameters, j = 0, 1, 2, 3.

Change-points for the yearly rate series of TB in Brazil

To complete the statistical data analysis of the TB counting in Brazil for the period 2001-2018 we know consider the TB yearly counting in place of the month TB counting. Considering the yearly rates of TB in Brazil, we can see (Table 1 and Figure 1) the presence of two change-points around the years 2003 and 2016 (upward rate until 2003, downward rate between the year 2003 and year 2016, rising rate after year 2016. In this case we also can use a linear regression Bayesian model to be fitted by the data in the presence of two points of change. In this case it is assumed the linear regression model,

Y(t) = β_o [j] + β_1 [j] years(t) + e(t) \hspace{1cm} (4)

for t = 1, 2, ..., N and j = 1, 2, 3 where β_o[1] and β_1[1] are regression parameters for the model (data set in logarithmic scale) corresponding to time[t] ≤ ζ_1 (ζ_1 is the first change-point); β_o[2] and β_1[2] are regression parameters for the model (data set in logarithmic scale) corresponding to ζ_1 < time[t] ≤ ζ_2 (ζ_2 is the second change-point); β_o[3] and β_1[3] are regression parameters for the model (data set in logarithmic scale) corresponding to time[t] ≥ ζ_2 and e(t) is the error assumed to be independently and identically distributed random variables with a normal distribution N(0, σ_e^2[3]), j = 1, 2, 3 where σ_e^2[1] for time[t] ≤ ζ, σ_e^2[2] for ζ < time[t] ≤ ζ and σ_e^2[3] for time[t] ≥ ζ. In the Bayesian analysis, discrete uniform prior distributions are assumed for the change-points k1 and k2 concentrated around the years 2003 and 2016; Gamma(1,1) prior distributions for the parameters ζ = 1/σ_ζ^2[1], j = 1, 2, 3; β_o[1] ~ N(-7,6;1), β_o[2] ~ N(-7,7;1), β_o[3] ~ N(-7,7,5;1) and β_1[3] ~ N(0,0,01). We further assume prior independence.

Results

From a preliminary analysis of the data, it is observed that in addition to the count times series for the monthly number of TB in Brazil, other series are related for the data analysis of the monthly counting of TB in Brazil from January 2001 to December 2018. In this study we consider four series in the data analysis:

- Time series for the logarithm of the TB count versus time order,
- Time series for the rate AIDS/TB versus time order,
- Time series for the rate urban patients/TB versus time order,
- Time series for the rate of patients in extreme poverty/TB versus time order.

These four series are analyzed in this section. Initially, polynomial regression models are considered in the form (1) considering constant variances and classic least squares (LSE) estimation techniques. We consider four regression models denoted by,

- Model M1: The regression model (1) associated with the logarithms of TB counts versus time order,
• Model M2: Associated with rates of patients with AIDS among TB patients versus time order,
• Model M3: Associated with rates of urban patients among TB patients versus time order,
• Model M4: Associated with the rates of patients living in extreme poverty among TB patients versus temporal order.

Use of a standard polynomial regression model

Table 2 shows the LSE for the regression parameters of each proposed regression model obtained using Minitab® software version 17. The necessary assumptions (normality and constant variance of the residuals were verified from the residuals graphics).

Use of polynomial regression models and stochastic volatility

Assuming the volatility model introduced in section 2 (equations (1) and (2)) for the four series above, also denoted here by model M1 (log (TB counts)), model M2 (AIDS / TB rate), model M3 (rate (urban patients)/TB) and model M4 (rate (extreme poverty patients)/TB), let us consider the following prior distributions for the parameters of each model:

- Model M1: $\phi \sim \text{U}(0,1)$, $\zeta \sim \text{Gamma}(1,1)$; $\mu \sim N(0,1)$, $\beta_0 \sim N(9,100)$, $\beta_1 \sim N(0,1)$.
- Model M2: $\phi \sim \text{U}(0,1)$, $\zeta \sim \text{Gamma}(1,1)$; $\mu \sim N(0,1)$, $\beta_0 \sim N(0.07, 0.01)$, $\beta_1 \sim N(0, 0.01)$.
- Model M3: $\phi \sim \text{U}(0,1)$, $\zeta \sim \text{Gamma}(1,1)$; $\mu \sim N(0,1)$, $\beta_0 \sim N(0.63, 0.01)$, $\beta_1 \sim N(0, 0.01)$.
- Model M4: $\phi \sim \text{U}(0,1)$, $\zeta \sim \text{Gamma}(1,1)$; $\mu \sim N(0,1)$, $\beta_0 \sim N(0.082,0.01)$, $\beta_1 \sim N(0, 0.01)$.

Further, let us assume prior independence among the parameters. The choice of the hyperparameters for the prior distributions of the regression parameters was made based on previous knowledge obtained from the classical regression models for the regression model (1) assuming normal errors without the presence of stochastic volatility (LSE given in Table 2). In this way, empirical Bayesian methods were used [35].

In the simulation of samples for the joint posterior distribution of interest, it was used the OpenBUGS software [36], which simplifies the computational work, since this software only requires the definition of the likelihood function and the prior distributions for the parameters of the model. Thus, the conditional posterior distributions for each parameter needed for the Gibbs and Metropolis-Hastings algorithms are not presented in this study. The convergence of the sample simulation algorithm of the joint posterior distribution of interest (Gibbs/Metropolis-Hastings algorithms) was verified from time series plots of the generated Gibbs samples.

It is important to point out that from the results obtained from Table 2 and Table 3 (LSE and Bayesian estimators), it can be seen that the LSE and Bayesian estimators have close point values where in most cases related to each regression parameter, we have concordance of p-value < 0.05 and credible intervals not containing the zero value. Only for the parameter $\beta_1$ (linear effect) related to models M1 (log (TB counts) versus time), and M2 (AIDS/TB rate versus time) it is observed significance under the classic approach but no significance under the Bayesian approach and for the parameter $\beta_2$ (quadratic effect) for model M4 (rate (extreme poverty patients)/TB versus time), we have no significance under the classic approach and significance under the focus Bayesian we have different inferential results in terms of significant effects.

| Models | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
|--------|-----------|-----------|-----------|-----------|
| M1     | 8.9200    | 0.00107(+) | -0.00002(+) | < 0.00001(+) |
| M2     | 0.06678   | 0.0000052(+) | 0.0000004(+) | < 0.00001(+) |
| M3     | 0.6345    | 0.001222(+) | -0.00001(+) | < 0.00001(+) |
| M4     | 0.0817    | 0.08174(+)  | < 0.00001 | < 0.00001(+) |

(*) significative at 5% (p-value < 0.05).
Table 3: Posterior summaries for the assumed models.

| Model | \( \beta_o \) | Mean  | SD    | Lower 95% | Upper 95% |
|-------|----------------|-------|-------|-----------|-----------|
| M1    |                | 8.932 | 0.02269 | 8.8860    | 8.9740    |
|       | \( \beta_1 \) | 0.0008 | 0.00085 | -0.00073  | 0.002499  |
|       | \( \beta_2 \) | -0.000019 | 0.0000089 | -0.000037 | -0.000004 |
|       | \( \beta_3 \) | 0.00000008 | 0.000000027 | 0.00000003 | 0.00000013 |
|       | \( \mu \)     | -5.058 | 0.6944 | -5.5500 | -2.7900 |
|       | \( \phi_1 \)  | 0.1775 | 0.1411 | 0.005846 | 0.5415 |
|       | \( \phi_2 \)  | 0.3424 | 0.2518 | 0.01661 | 0.9020 |
|       | \( \zeta \)   | 4.251 | 1.575 | 1.8620 | 8.1220 |
| M2    | \( \beta_o \) | 0.06851 | 0.001675 | 0.06574 | 0.07243 |
|       | \( \beta_1 \) | -0.000002 | 0.000006 | -0.000138 | 0.000103 |
|       | \( \beta_2 \) | 0.0000043 | 0.00000061 | 0.0000032 | 0.0000057 |
|       | \( \beta_3 \) | -0.00000002 | 0.000000020 | -0.00000001 | -0.00000001 |
|       | \( \mu \)     | -4.259 | 1.036 | -6.3700 | -2.2760 |
|       | \( \phi_1 \)  | 0.6631 | 0.1781 | 0.3372 | 0.9741 |
|       | \( \phi_2 \)  | 0.3359 | 0.1778 | 0.02611 | 0.6653 |
|       | \( \zeta \)   | 2.168 | 0.8817 | 1.001 | 4.358 |
| M3    | \( \beta_o \) | 0.637 | 0.004801 | 0.6281 | 0.6477 |
|       | \( \beta_1 \) | 0.001195 | 0.000151 | 0.000857 | 0.000289 |
|       | \( \beta_2 \) | -0.0000099 | 0.00000142 | -0.000012 | -0.0000068 |
|       | \( \beta_3 \) | 0.00000002 | 0.000000004 | 0.000000014 | 0.00000003 |
|       | \( \mu \)     | -4.203 | 1.044 | -6.115 | -2.182 |
|       | \( \phi_1 \)  | 0.6546 | 0.2077 | 0.252 | 0.9783 |
|       | \( \phi_2 \)  | 0.3427 | 0.2079 | 0.0218 | 0.7514 |
|       | \( \zeta \)   | 2.324 | 0.8576 | 1.076 | 4.418 |
| M4    | \( \beta_o \) | 0.08113 | 0.00176 | 0.07796 | 0.08485 |
|       | \( \beta_1 \) | 0.000243 | 0.0000622 | 0.0000102 | 0.0000357 |
|       | \( \beta_2 \) | -0.0000036 | 0.0000006 | -0.0000046 | -0.0000022 |
|       | \( \beta_3 \) | 0.00000001 | 0.00000002 | 0.000000007 | 0.000000014 |
|       | \( \mu \)     | -4.359 | 0.9985 | -6.237 | -2.372 |
|       | \( \phi_1 \)  | 0.6784 | 0.1938 | 0.3097 | 0.9825 |
|       | \( \phi_2 \)  | 0.3219 | 0.1943 | 0.01728 | 0.6966 |
|       | \( \zeta \)   | 2.552 | 0.8985 | 1.172 | 4.655 |

Table 4: Posterior summaries for a model with two change-points.

| \( \beta_o \) | Mean | SD | Lower 95% | Upper 95% |
|--------------|------|----|-----------|-----------|
| \( \beta_o[1] \) | -7.59 | 0.7826 | -9.196 | -5.954 |
| \( \beta_o[2] \) | -7.585 | 0.235 | -8.075 | -7.125 |
| \( \beta_o[3] \) | -7.718 | 0.8682 | -9.433 | -6.032 |
| \( \beta_o[1] \) | -0.001627 | 0.09955 | -0.194 | 0.197 |
| \( \beta_o[2] \) | -0.01256 | 0.02407 | -0.06036 | 0.03539 |
| \( \beta_o[3] \) | -0.00885 | 0.05843 | -0.1162 | 0.1135 |
| \( k1 \) | 2.104 | 0.9501 | 1.0 | 4.0 |
| \( k2 \) | 16.71 | 0.976 | 15.0 | 18.0 |
| \( \zeta_1 = 1/\sigma_1^2[1] \) | 1.327 | 1.161 | 0.06158 | 4.355 |
| \( \zeta_2 = 1/\sigma_2^2[2] \) | 7.317 | 2.771 | 2.911 | 13.58 |
| \( \zeta_3 = 1/\sigma_3^2[3] \) | 1.744 | 1.375 | 0.1363 | 5.238 |
Use of a change-point model for the yearly rate series of TB in Brazil

Table 4 presents the posterior summaries obtained using the OpenBUGS software (burn-in sample = 1,000; 10,000 additional Gibbs samples). Figure 5 shows the fitted model (presence of two change-points, one change-point around the year 2003 which corresponds to the k1 estimated by 2.104 and a second change-point in the year 2016 which corresponds to the k2 estimated
Figure 5: Fitted model (presence of two change-points).

Figure 6: Graphs of the quadratic roots of the estimated volatilities (July 2001 to December 2018) for the four models.

Discussion of the Obtained Results and Concluding Remarks

From the obtained results using model (2), it is observed a good fit of the models to the four data sets denoted here by models M1, M2, M3 and M4. Regarding to the estimated variability (volatility) of the data presented in Figure 6 for the period from July 2001 to December 2018, we have:

- For the TB count series considered on the logarithmic scale (model M1), there is great variability around the year 2007; from that year on, there is a decline in variability to be increased again especially at the end of the period 2001/2018.
- For the AIDS/TB rate series (model M2), there is
great variability around the year 2007/2008 and 2016/2017, an indication that the AIDS/TB rate was in transition, indicating that the number TB cases were increasing regardless of whether a person have AIDS or not. It is a worrying result, as it indicates that TB is increasing in Brazil regardless of the increase in AIDS cases.

- For the urban patient/TB rate series (model M3), it is observed that the variability is declining; with this result and the lower average of urban patient/TB rates, there is an indication that the TB epidemic is increasing throughout Brazil regardless of being a resident of large cities.

- For the extreme poverty/TB rate series (model M4), it is observed that the variability is approximately stable regardless of the great seasonality with a slight decline, also indicating that the number of TB cases is increasing in Brazil, regardless the increasing of poverty; these results show that the increase in TB cases in Brazil is in a increasing phase regardless of the increase in extreme poverty that has been observed in Brazil, especially since 2019.

In conclusion, we can emphasize that the proposed model can be of great use for other epidemiological data in addition to modeling TB counts and rates associated with TB. Another important point of an epidemiological character obtained from the data analysis: As the number of TB cases is increasing in Brazil and some rates such as AIDS/TB and extreme poverty/TB are falling with the presence of great volatility, there is an indication that TB is increasing indiscriminately in Brazil, as AIDS and extreme poverty are two factors widely discussed in the literature that are associated with TB. This is a problem of public health concern.

It is important to point out, that in the present study we considered univariate epidemiological time series (modelM1 (log (TB counts)), model M2 (AIDS/TB rate), model M3 (rate (urban patients)/TB) and model M4 (rate (extreme poverty patients)/TB). Another possibility is to assume multivariate SV models and to study the structure of possible dependence among the series. This approach will be the goal of a future study.

Acknowledgments

The authors would like to thank the referee’s comments that led to a great improvement in the manuscript.

References

1. World Health Organization (2012) Global tuberculosis report.
2. A San Pedro, RM Oliveira (2013) Tuberculose e indicadores socioeconômicos: Revisão sistemática da literatura. Rev Panam Salud Publica 33: 294-301.
3. MA Hijar, MJPR Oliveira, GM Teixeira (2001) A tuberculose no Brasil e no mundo. Bol. Pneumol Sanit, Rio de Janeiro, 9: 9-16.
4. DC Costa (1985) Considerações sobre a tendência da tuberculose no Brasil. Cad Saúde Pública, Rio de Janeiro, 1: 313-326.
5. LMW Acosta, SL Bassanesi (2014) The Porto Alegre paradox: Social determinants and tuberculosis incidence. Rev Bras Epidemiol, São Paulo, 17: 88-101.
6. MG Almeida, DRM Barbosa, DFS Almeida (2013) Epidemiologia e distribuição espacial de casos notificados de tuberculose multirresistente (TBM) no Brasil, 2008-2012. Rev Epidemiol Control Infect 3: 117-122.
7. P Hino (2011) Padrões espaciais da tuberculose e sua associação à condição de vida no município de Ribeirão Preto. Ciênc Saúde Coletiva 16: 4795-4802.
8. DL Roza, MCGG Caccia-Bava, EZ Martinez (2012) Spatio-temporal patterns of tuberculosis incidence in Ribeirão Preto, State of São Paulo, southeast Brazil, and their relationship with social vulnerability: A Bayesian analysis. Rev Soc Bras Med Trop 45: 536-541.
9. SHF Vendramini, NSGM dos Santos, MLSG Santos, F Chiaravalloti-Neto, MAZ Ponce (2010) Análise espacial da co-infeção tuberculose/HIV: relação com níveis socioeconômicos em município do sudeste do Brasil. Rev Soc Bras Med Trop 43: 607-615.
10. R Abellana, C Ascaso, J Aponte, F Saute, D Nhalungo, et al. (2008) Spatio-seasonal modeling of the incidence rate of malaria in Mozambique. Malar J 7: 228.
11. IS Chang, SJ Fu, CH Chen, TH Wang, CA Hsiung (2009) Estimating temporal transmission parameters from infectious disease household data, with application to Taiwan SARS data. Stat Biosci 1: 80-94.
12. J Kleinischmidt, M Bagayoko, GP Clarke, M Craig, D Le Sueur (2000) A spatial statistical approach to malaria mapping, Int J Epidemiol 29: 355-361.
13. G Box, G Jenkins (1970) Time series analysis: Forecasting and control. (1st edn), Holden-Day, San Francisco.
14. RS Tsay (2002) Analysis of Financial Time Series. (1st edn), John Wiley and Sons, Canada.
15. RR Sarkar, C Chatterjee (2017) Application of different time series models on epidemiological data - Comparison and predictions for malaria prevalence. SM J Biometrics Biostat 2: 1022.
16. YL Zheng, LP Zhang, XL Zhang, K Wang, YJ Zheng (2015) Forecast model analysis for the morbidity of tuberculosis in Xinjiang, China. PLoS One 10: e0116832.
17. A Earnest, MI Chen, D Ng, LY Sin (2005) Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Serv Res 11: 5-36.
18. Q Li, NN Guo, ZY Han, YB Zhang, SX Qi (2012) Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. Am J Trop Med Hyg 87: 364-370.
19. M Gharbi, P Quenel, J Gustave, S Cassadou, G La Ruche, et al. (2011) Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. BMC Infect Dis 11: 166.
20. Q Liu, X Liu, B Jiang, W Yang (2011) Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model. BMC Infect Dis 11: 218.
21. PM Luz, BVM Mendes, CT Codeco, CJ Struchiner, AP Galvani (2008) Time series analysis of dengue incidence in Rio de Janeiro, Brazil. Am J Trop Med Hyg 79: 933-939.

22. W Wu, P Guan, JQ Guo, BS Zhou (2008) Comparison of GM(1,1) gray model and ARIMA model in forecasting the incidence of hemorrhagic fever with renal syndrome. Journal of China Medical University 37: 52-55.

23. L Yu, L Zhou, L Tan, H Jiang, Y Wang, S Wei (2014) Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. PLoS One 9: e98241.

24. NR Draper, HS Smith, HS (1981) Applied regression analysis. Wiley Online Publisher, Canada.

25. GAF Seber, AJ Lee (2003) Linear regression analysis. (2nd edn), Wiley Series in Probability and Mathematical Statistics.

26. C Chatterjee, RR Sarkar (2009) Multi-step polynomial regression method to model and forecast malaria incidence. PLoS One 4: e4726.

27. J Danielsson (1994) Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. Journal of Econometrics 61: 375-400.

28. R Engle (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50: 987-1007.

29. T Bollerslev (1986) Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31: 307-327.

30. E Ghysels (1996) Stochastic volatility, in ‘Statistical methods on finance’, North-Holland.

31. S Kim, N Shephard (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65: 361-393.

32. R Meyer, T Yu (2000) Bugs for a Bayesian analysis of stochastic volatility models. Econometrics Journal 3: 198-215.

33. AE Gelfand, AFM Smith (1990) Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85: 398-409.

34. AFM Smith, GO Roberts (1993) Bayesian computation the Gibbs sampling and related Markov Chain Monte Carlo methods. Journal of the Royal Statistical Society 55: 3-23.

35. TD Phan, JA Wachter, EA Solomon, MJ Kahana (2019) Multivariate stochastic volatility modeling of neural data. eLife 8: e42950.

36. B Carlin, TA Louis (2000) Bayes and Empirical Bayes Methods for Data Analysis. (2nd edn), Chapman & Hall, London.

37. DJ Spiegelhalter, A Thomas, N Best, D Lunn (2003) WinBUGS user manual. MRC Biostatistics Unit, Cambridge, United Kingdom.
Appendix 1: TB series in Brazil (January 2001 to December 2018).

| TB   | AIDS/TB | URB/TB | POVERTY/TB |
|------|---------|--------|------------|
| 1    | 8088    | 0.064416 | 0.634891 | 0.0817260 |
| 2    | 6542    | 0.065729 | 0.627484 | 0.0764292 |
| 3    | 8095    | 0.060531 | 0.640889 | 0.0795553 |
| 4    | 7305    | 0.061802 | 0.633676 | 0.0835044 |
| 5    | 7656    | 0.064002 | 0.643939 | 0.0858150 |
| 6    | 6805    | 0.063630 | 0.651433 | 0.0865540 |
| 7    | 6987    | 0.064692 | 0.637613 | 0.0918849 |
| 8    | 8065    | 0.067452 | 0.628518 | 0.0838190 |
| 9    | 6967    | 0.067344 | 0.626699 | 0.0895924 |
| 10   | 7524    | 0.069777 | 0.624668 | 0.0867889 |
| 11   | 6874    | 0.071720 | 0.620454 | 0.0853942 |
| 12   | 6627    | 0.068206 | 0.633016 | 0.0804285 |
| 13   | 8013    | 0.067141 | 0.660302 | 0.0842381 |
| 14   | 7346    | 0.072965 | 0.641710 | 0.0811326 |
| 15   | 7961    | 0.065444 | 0.659465 | 0.0806431 |
| 16   | 8771    | 0.064531 | 0.661612 | 0.0884734 |
| 17   | 7784    | 0.063207 | 0.674717 | 0.0908273 |
| 18   | 6713    | 0.074631 | 0.673171 | 0.0804409 |
| 19   | 7746    | 0.076039 | 0.679706 | 0.0888200 |
| 20   | 8330    | 0.073229 | 0.650660 | 0.0828331 |
| 21   | 7654    | 0.073948 | 0.662007 | 0.0837471 |
| 22   | 8200    | 0.071220 | 0.661098 | 0.0800000 |
| 23   | 7532    | 0.074615 | 0.660781 | 0.0896176 |
| 24   | 6809    | 0.078866 | 0.656925 | 0.0772507 |
| 25   | 8115    | 0.066790 | 0.659766 | 0.0853974 |
| 26   | 7985    | 0.069881 | 0.655980 | 0.0821540 |
| 27   | 7474    | 0.066899 | 0.655205 | 0.0859590 |
| 28   | 7975    | 0.067586 | 0.667712 | 0.0820063 |
| 29   | 8024    | 0.071286 | 0.685942 | 0.0844965 |
| 30   | 7082    | 0.075502 | 0.676363 | 0.0858515 |
| 31   | 7949    | 0.071833 | 0.680463 | 0.0865518 |
| 32   | 7585    | 0.075676 | 0.679944 | 0.0843771 |
| 33   | 8136    | 0.074115 | 0.673673 | 0.0886185 |
| 34   | 8521    | 0.077808 | 0.669405 | 0.0822673 |
| 35   | 7624    | 0.080273 | 0.677991 | 0.0838631 |
| 36   | 7303    | 0.076818 | 0.678078 | 0.0798302 |
| 37   | 7574    | 0.080803 | 0.691048 | 0.0764457 |
| 38   | 6743    | 0.082308 | 0.694935 | 0.0904642 |
| 39   | 8574    | 0.071495 | 0.680546 | 0.0905062 |
| 40   | 8089    | 0.071826 | 0.686086 | 0.0876499 |
| 41   | 7763    | 0.072008 | 0.684916 | 0.0888832 |
| 42   | 7314    | 0.078616 | 0.675280 | 0.0849057 |
| 43   | 7825    | 0.080985 | 0.669776 | 0.0857508 |
| 44   | 8232    | 0.078839 | 0.651361 | 0.0823615 |
| 45   | 7901    | 0.072902 | 0.655993 | 0.0898620 |
| 46   | 8016    | 0.070734 | 0.670659 | 0.0782186 |
| 47   | 7753    | 0.079040 | 0.677157 | 0.0882239 |
| 48   | 7196    | 0.075875 | 0.672457 | 0.0774041 |
|   |   |   |   |   |
|---|---|---|---|---|
| 49 | 7432 | 0.077234 | 0.674247 | 0.0805974 |
| 50 | 6748 | 0.081654 | 0.655750 | 0.0881743 |
| 51 | 8441 | 0.072859 | 0.668049 | 0.090367 |
| 52 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 53 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 54 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 55 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 56 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 57 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 58 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 59 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 60 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 61 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 62 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 63 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 64 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 65 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 66 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 67 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 68 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 69 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 70 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 71 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 72 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 73 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 74 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 75 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 76 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 77 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 78 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 79 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 80 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 81 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 82 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 83 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 84 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 85 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
| 86 | 7781 | 0.075954 | 0.668049 | 0.0900367 |
| 87 | 5282 | 0.081571 | 0.659683 | 0.0875421 |
| 88 | 7958 | 0.068233 | 0.661850 | 0.0913546 |
| 89 | 8019 | 0.068837 | 0.659683 | 0.0875421 |
99 8231 0.086988 0.681813 0.0816426 207 7893 0.093754 0.659698 0.0741163
100 7584 0.094937 0.680643 0.0901899 208 8126 0.092789 0.665272 0.0756830
101 7245 0.092478 0.695376 0.0864044 209 7878 0.097994 0.648413 0.0736212
102 6799 0.094426 0.694661 0.0897191 210 7688 0.099567 0.662585 0.0752010
103 7513 0.090643 0.701185 0.0805271 211 8085 0.094940 0.656740 0.0704210
104 7450 0.095705 0.688188 0.0783893 212 8932 0.094937 0.680643 0.0741163
105 7478 0.098288 0.687350 0.0850495 213 7663 0.096437 0.662224 0.0695550
106 7359 0.092947 0.677674 0.0852018 214 8691 0.091129 0.658842 0.0699574
107 7088 0.096924 0.681716 0.0747743 215 7682 0.091382 0.653606 0.0741994
108 6793 0.098042 0.680701 0.0791992 216 6836 0.095816 0.649210 0.0751902

**Appendix 2:** OpenBUGS code (change-points).

```plaintext
model
{
for (i in 1:N) {
  J[i] <- 1+step(i-k1)+step(i-k2)
  y[i] <- log(yearly.rate.TB[i])
  y[i] ~ dnorm(mean[i],tau[J[i]])
}
for (i in 1:N) {
  mean[i] <- beta0[J[i]]+ beta1[J[i]]*years[i]
  fitted.mean[i]<-exp(mean[i])
}
k1 ~ dcat(p1[])
k2 ~ dcat(p2[])
tau[1] ~ dgamma(1,1)
tau[2] ~ dgamma(1,1)
tau[3] ~ dgamma(1,1)
beta0[1] ~ dnorm(-7.60,1)
beta0[2] ~ dnorm(-7.70,1)
beta0[3] ~ dnorm(-7.75,1)
beta1[1] ~ dnorm(0,100)
beta1[2] ~ dnorm(0,100)
beta1[3] ~ dnorm(0,100)
}
```