

HIERARCHICAL DYNAMIC BETA MODEL

C. Q. Da-Silva ^{a,*}, H. S. Migon ^b,

^a*Department of Statistics, University of Brasília-UnB, Brasília 70910-900, Brazil*

^b*Department of Statistics, Federal University of Rio de Janeiro, Rio de Janeiro
21.945-970, Brazil*

Abstract

We develop a hierarchical dynamic Bayesian beta model for modeling a set of time series of rates or proportions. The proposed methodology enables to combine the information contained in different time series so that we can describe a common underlying system, which is though flexible enough to allow the incorporation of random deviations, related to the individual series, not only through time but also across series. That allows to fit the case in which the observed series may present some degree of level shift. Additionally, the proposed model is adaptive in the sense that it incorporates precision parameters that can be heterogeneous not only over time but also across the series. Our methodology was applied to both real and simulated data. The real data set used in this article comprise three time series of Brazilian monthly unemployment rates, observed in the cities of Recife, São Paulo and Porto Alegre, in the period from March 2002 to April 2001. The new parametrization of the precision parameter make possible the use of the same type of link function for both the mean and the precision parameters, which are then expressed in the $(0, 1)$ interval, providing more meaningful interpretation in terms of the magnitude of the scale.

Key words: Dynamic models; Beta distribution; hierarchical models; Bayesian analysis.

1 Introduction

The beta regression models, proposed by Ferrari and Cribari-Neto (2004), have attracted the attention of many researchers. Those models are useful in situ-

* Corresponding author.

Email addresses: cibeleqs@unb.br, cibeleqs@gmail.com (C. Q. Da-Silva),
migon@im.ufrj.br (H. S. Migon).

ations where the response is restricted to the standard unit interval. In this seminal work the authors developed generalized linear models (GLM) theory for dealing with the situation where only the parameter related to the mean of the beta distribution was allowed to vary. In the context of GLM's Nelder and Lee (1991) and Smyth and Verbyla (1999) describe a class of joint generalized linear models which allow both the mean and the dispersion parameters in the GLM model to vary with the response. Nelder and Lee (1991) argue that it is necessary to use two GLMs when both mean and dispersion are to be modelled, i.e., we would have the so called *mean process* and the *dispersion process*. Pregibon (1984) was the first to suggest this kind of specification. Other articles related to such perspective, in which dispersion parameter of the beta model is allowed to vary, include Cuervo-Cepeda and Gamerman (2004), Smithson and Verkuilen (2006), Espinheira (2007), Simas *et al.* (2010) and Bayer (2011). These works emphasize the need of correctly modeling the dispersion parameter of the beta regression in order to achieve efficient estimation.

Based on the class of beta regressions introduced by Ferrari and Cribari-Neto (2004), Rocha and Cribari-Neto (2009) proposed a dynamic model for continuous random variates that assume values in the standard unit interval (0,1). The proposed frequentist β ARMA model includes both autoregressive and moving average dynamics, and also includes a set of regressors. Da-Silva *et al.* (2011) proposed a dynamic Bayesian beta model for modeling and forecasting single time series of rates or proportions. In such work only the mean parameter of the beta model was allowed to vary with time.

In the present work we build upon the dynamic Bayesian beta model introduced by Da-Silva *et al.* (2011) and upon the class of conditionally Gaussian dynamic models (see Cargnoni *et al.*, 1997; Gamerman and Migon, 1993) to propose a hierarchical dynamic Bayesian beta model in which both the mean and the dispersion parameters of the beta model can vary with time. Since the proposed model is hierarchical, the parameters in the model are related both through time and hierarchically across several series, which supposedly share a common underlying trend.

We motivate our study with the problem of forecasting monthly Brazilian unemployment rates in different cities. The Brazilian Institute of Geography and Statistics (IBGE) implemented the Monthly Unemployment Survey (PME) in 1980, but since 2002 a new survey methodology has been adopted. The PME is a monthly survey about workforce and income. The most important metropolitan regions in Brazil are included in such survey: São Paulo, Rio de Janeiro, Belo Horizonte, Porto Alegre, Recife and Salvador. The data can be found at <http://www.ibge.gov.br/>. In Figure 1 we present the PME data for the cities of Recife, São Paulo and Porto Alegre. As we can observe, the three series have similar underlying trends but distinct levels and, possibly,

distinct dispersions, specially in the case of the city of Recife.

The article is organized as follows. In Section 2 we introduce the hierarchical dynamic beta. In Section 3 we describe a fully Bayesian methodology to analyze data from a hierarchical dynamic beta process. In Sections 4 to 6 we apply the methods to simulated and real data.

2 The hierarchical dynamic beta model

In this section we present a methodology for modeling a set of I time series of rates or proportions, y_{it} , $i = 1, \dots, I$, which share certain characteristics which allows us to treat them in the class of the hierarchical models.

Da-Silva *et. al.* (2001) used the parametrization of the Beta distribution given by Ferrari and Cribari-Neto (2004) to describe a dynamic beta model in which the precision parameter ζ was considered fixed. However, a more general model can be described by considering both the mean and the precision parameters varying with time. In such case, The *observation equation* of the dynamic model is given by

$$p(y_{it} | \mu_{it}, \zeta_{it}) = \frac{\Gamma(\zeta_{it})}{\Gamma(\zeta_{it}\mu_{it})\Gamma(\zeta_{it}(1-\mu_{it}))} y_{it}^{\zeta_{it}\mu_{it}-1} (1-y_{it})^{\zeta_{it}(1-\mu_{it})-1}. \quad (1)$$

We have that $E(y_{it} | \mu_{it}, \zeta_{it}) = \mu_{it}$ and $V(y_{it} | \mu_{it}, \zeta_{it}) = \mu_{it}(1-\mu_{it})/(1+\zeta_{it})$, with $0 \leq \mu_{it} \leq 1$ and $\zeta_{it} > 0$.

Another parametrization for ζ , proposed by Bayer (2011), can be used in our context, since it allows us to use link functions for the transformed ζ which are easier to interpret than, say, a log link function, whose the upper limit is unbounded. In equation (1), let $\phi_{it} = \frac{1}{1+\zeta_{it}}$ so that $\zeta_{it} = \frac{1-\phi_{it}}{\phi_{it}}$. Thus, $0 \leq \phi_{it} \leq 1$, and the *observation equation* of the model is now written as

Observation equation:

$$p(y_{it} | \mu_{it}, \phi_{it}) = \frac{y_{it}^{\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right)-1} (1-y_{it})^{(1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)-1}}{B\left(\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right), (1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right)}, \quad (2)$$

where

$$B\left(\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right), (1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right) = \frac{\Gamma\left(\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right)\Gamma\left((1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right)}{\Gamma\left(\frac{1-\phi_{it}}{\phi_{it}}\right)},$$

with $i = 1, \dots, I$, $t = 1, \dots, N$, i.e., we have I time series in study, such that $(y_{it} \mid \mu_{it}, \phi_{it})$ is independent of $(y_{jt} \mid \mu_{jt}, \phi_{jt})$ for $i \neq j$. Equation (2) incorporates heterogeneity in the precision parameter that may occur both over time or across the series.

Other components, which essential in the description of our hierarchical dynamic beta model, are given, respectively, by (i) real transformations applied to μ_{it} and ϕ_{it} , allowing the use of some simplifying Gaussian properties; (ii) the **structural equations** represented in terms of linear models relating the transformed parameters and the latent states and (iii) the **system equation** of the dynamic model.

Real valued transformations:

Take $\eta_{1it} = h_1(\mu_{it})$ and $\eta_{2it} = h_2(\phi_{it})$ with $\eta_{it} = (\eta_{1it}, \eta_{2it})'$ such that η_{it} is a real valued vector. Now, y_{it} is parametrized by η_{it} , i.e., $(y_{it} \mid \eta_{it}) \sim \text{Beta}(y_{it} \mid \eta_{it})$. Notice that functions $h_1(\cdot)$ and $h_2(\cdot)$ describe the link functions associated to, respectively, the *mean process* and the *dispersion process*.

Structural equations:

$$\eta_t = F_t \theta_t + v_t, \quad v_t \sim N(\mathbf{0}, V), \quad (3)$$

for each t , with $\eta_{it} = F_{it} \theta_t + v_{it}$, $v_{it} \sim N(\mathbf{0}, V_i)$, for each $i \in \{1, \dots, I\}$.

System Equation:

$$\theta_t = H_t \theta_{t-1} + w_t, \quad w_t \sim (\mathbf{0}, W). \quad (4)$$

We assume that the error terms v_{it} and w_t are all mutually independent. The notation used is given below.

Notation:

- θ_t is a real valued s -dimensional vector of latent states;
- $F_t = (F'_{1t}, \dots, F'_{It})'$ is the $2I \times s$ design matrix for all the I series at time t ;
- $\eta_t = (\eta'_{1t}, \dots, \eta'_{It})'$ is the $2I \times 1$ vector of structural parameters for all the I series at time t ;
- $v_t = (v'_{1t}, \dots, v'_{It})'$, is the $2I \times 1$ vector of errors for the structural equations;
- $V = \text{block-diag}(V_1, \dots, V_I)$ is a $(2I \times 2I)$ block diagonal matrix;
- H_t is a specified $s \times s$ state evolution matrix;
- $y = (y_1, \dots, y_N)$ with $y_t = (y_{1t}, \dots, y_{It})'$;
- $\theta = (\theta_0, \dots, \theta_N)$;
- $\eta = (\eta_1, \dots, \eta_N)$.
- $W = \text{block-diag}(W_1, \dots, W_k)$, i.e., we are considering k effects associated to the latent states and their respective covariance matrices.

In the prior specification for θ_0 , V and W , we assume that θ_0 , V_1, \dots, V_I and W_1, \dots, W_k are mutually independent, $\theta_0 \sim N(m_0, C_0)$, V_i , $i = 1, \dots, I$ have a common inverse Wishart prior and that W is block-diagonal with an inverse Wishart prior for each block.

Notice that equations (3) and (4) represent a standard dynamic linear model for the state vector θ_t . Additionally, θ is conditionally independent of y given η . These combined features imply in a substantial simplification in the posterior computations of the parameters η and θ , as described in Cargnoni et. al. (1997). We describe next the methodology adopted here with some adaptations to the procedure proposed by Cargnoni et. al. (1997).

3 Bayesian Analysis

It is more convenient to work with the precision matrices instead of with the covariance ones. Thus, let $\Phi_0 = \text{block-diag}(\Phi_{01}, \dots, \Phi_{0I})$ and $\Phi = \text{block-diag}(\Phi_1, \dots, \Phi_k)$ where $\Phi_{0i} = V_i^{-1}$, $i = 1, \dots, I$ and $\Phi_l = W_l^{-1}$, $l = 1, \dots, k$. Suppose that Φ_{0i} , $i = 1, \dots, I$, follow independent Wishart distributions such that $\Phi_{0i} \sim W(\nu_{0i}, S_{0i})$ and, similarly, $\Phi_l \sim W(\varsigma_l, Z_l)$, independent prior distributions for $l = 1, \dots, k$.

The joint posterior distribution is given by

$$p(\eta, \theta, \Phi_0, \Phi \mid y) \propto \left[\prod_{t=1}^N \left(\prod_{i=1}^I \text{Beta}(y_{it} \mid \eta_{it}) N(\eta_{it}; F_{it}\theta_t, \Phi_{0i}^{-1}) \right) N(\theta_t; H_t\theta_{t-1}, \Phi^{-1}) \right] \\ \times N(\theta_0; m_0, C_0) \prod_{i=1}^I W(\Phi_{0i}; \nu_{0i}, S_{0i}) \prod_{l=1}^k W(\Phi_l; \varsigma_l, Z_l). \quad (5)$$

We need to obtain samples from the full conditional posteriors: $p(\eta \mid \theta, \Phi_0, \Phi, y)$, $p(\theta \mid \eta, \Phi_0, \Phi, y)$ and $p(\Phi_0, \Phi \mid \eta, \theta, y)$.

3.1 Sampling from $p(\theta \mid \eta, \Phi_0, \Phi, y)$

As mentioned before, the equations (3) and (4) represent a standard dynamic linear model for the state vector θ_t . In such setting, the fact that θ is conditionally independent of y given η implies that $p(\theta \mid \eta, \Phi_0, \Phi, y) = p(\theta \mid \eta, \Phi_0, \Phi)$. Then, in a regular DLM, η has the same rule as y , so that in the sequential updating formulations of the DLM, y will be replaced by η .

The representation of the full conditional posterior distribution of $p(\theta \mid \eta, \Phi_0, \Phi)$, considering the conditional independence structure of the DLM as well the Bayes theorem is given by

$$\begin{aligned}
p(\theta \mid \eta, \Phi_0, \Phi) &= p(\theta_N \mid \eta, \Phi_0, \Phi) \prod_{t=0}^{N-1} p(\theta_t \mid \theta_{t+1}, \dots, \theta_N, \eta, \Phi_0, \Phi) \\
&= p(\theta_N \mid \eta, \Phi_0, \Phi) \prod_{t=0}^{N-1} p(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi) \\
&\propto p(\theta_N \mid \eta, \Phi_0, \Phi) \prod_{t=0}^{N-1} p(\theta_{t+1} \mid \theta_t, \eta, \Phi_0, \Phi) p(\theta_t \mid \eta, \Phi_0, \Phi). \quad (6)
\end{aligned}$$

Thus, all the state vectors can be sampled from $p(\theta \mid \eta, \Phi_0, \Phi)$ using the FFBS (Forward-filtering, backward-sampling) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). Conditionally on the “observed values” of η , the algorithm below is such that allows us to draw a sample $\theta_N, \theta_{N-1}, \dots, \theta_0$ from $p(\theta \mid \eta, \Phi_0, \Phi)$ as follows:

(1) **Filtering**

Using the so called Kalman filter (de Jong, 1991), compute the moments m_t and C_t of the joint posterior $p(\theta_t \mid \eta, \Phi_0, \Phi)$, $t = 1, \dots, N$, by applying the standard DLM sequential updating formulas with y replaced by η . For more details see West and Harrison (1997).

- $m_t = a_t + A_t e_t$; $C_t = R_t - A_t Q_t A_t'$,
- $A_t = R_t F_t Q_t^{-1}$; $e_t = \eta_t - f_t$,
- $a_t = H_t m_{t-1}$; $R_t = H_t C_{t-1} H_t' + \Phi^{-1}$,
- $f_t = F_t a_t$; $Q_t = F_t' R_t F_t + \Phi_0^{-1}$.

(2) **Smoothing**

At time $t = N$, sample the vector state θ_N from $p(\theta_N \mid \eta, \Phi_0, \Phi)$, i.e., sample θ_N from $(\theta_N \mid \eta, \Phi_0, \Phi) \sim N(m_N, C_N)$. For times $t = N-1, \dots, 0$, sample θ_t from $p(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi)$ conditionally on the just sampled value θ_{t+1} . That is performed by sampling θ_t from $(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi) \sim N(u_t, U_t)$, where

- $u_t = m_t + B_t(\theta_{t+1} - a_{t+1})$,
- $U_t = C_t - B_t R_{t+1} B_t'$, and
- $B_t = C_t H_t R_{t+1}^{-1}$.

3.2 Sampling from $p(\eta \mid \theta, \Phi_0, \Phi, y)$

Given θ , Φ_0 and Φ , the η_{it} 's are mutually independent over all the times t and the series i . That implies that a sample from the conditional posterior of $(\eta \mid \theta, \Phi_0, \Phi, y)$ is obtained through $I \times N$ independent samples from the respective distributions given by

$$p(\eta_{it} \mid \theta_t, \Phi_{0i}, \Phi, y_{it}) \propto p(y_{it} \mid \eta_{it}) p(\eta_{it} \mid \theta_t, \Phi_{0i}). \quad (7)$$

The second term on the right-hand side of the full conditional (7) is the normal prior $\eta_{it} \sim N(F_{it}\theta_t, \Phi_{0i})$, while the first term is given by the Beta model described by expression (2), such that $\eta_{1it} = h_1(\mu_{it})$ and $\eta_{2it} = h_2(\phi_{it})$.

Since the distribution $p(\eta_{it} | \theta_t, \Phi_{i0}, y_{it})$ does not have closed form, it is necessary to use the Metropolis-Hastings algorithm (Metropolis *et al.* 1953; Hastings, 1970) in order to draw samples from such distribution. Let m represent the m -th MCMC draw. We use the following random-walk M-H with symmetric normal proposal for η_{it} :

- (a) Draw $\eta_{1it}^* \sim q_1(\eta_{1it}^{m-1}, \eta_{1it}^*) \stackrel{d}{=} N(\eta_{1it}^{m-1}, \Phi_{1i0}^{-1})$ and $\eta_{2it}^* \sim q_2(\eta_{2it}^{m-1}, \eta_{2it}^*) \stackrel{d}{=} N(\eta_{2it}^{m-1}, \Phi_{2i0}^{-1})$.
(b) Calculate the acceptance probability $\alpha(\eta_{it}^{m-1}, \eta_{it}^*) = \min\{1, R_{\eta_{it}}\}$, where

$$R_{\eta_{it}} = \frac{\pi(\eta_{it}^* | \cdot) q(\eta_{it}^*, \eta_{it}^{m-1})}{\pi(\eta_{it}^{m-1} | \cdot) q(\eta_{it}^{m-1}, \eta_{it}^*)} = \frac{\pi(\eta_{it}^* | \cdot)}{\pi(\eta_{it}^{m-1} | \cdot)},$$

with $\pi(\eta_{it}^* | \cdot) = p(y_{it} | \eta_{it}^*) p(\eta_{it}^* | \theta_t, \Phi_{0i})$, $\pi(\eta_{it}^{m-1} | \cdot) = p(y_{it} | \eta_{it}^{m-1}) p(\eta_{it}^{m-1} | \theta_t, \Phi_{0i})$, and $q(\eta_{it}^*, \eta_{it}^{m-1}) = q_1(\eta_{1it}^{m-1}, \eta_{1it}^*) q_2(\eta_{2it}^{m-1}, \eta_{2it}^*)$.
(c) Set

$$\eta_{it}^m = \begin{cases} \eta_{it}^* & \text{with probability } \alpha(\eta_{it}^{m-1}, \eta_{it}^*), \\ \eta_{it}^{m-1} & \text{otherwise.} \end{cases}$$

3.3 Sampling from $p(\Phi_0, \Phi | \eta, \theta, y)$

Reminding that $\Phi_0 = \text{block-diag}(\Phi_{01}, \dots, \Phi_{0I})$ and $\Phi = \text{block-diag}(\Phi_1, \dots, \Phi_k)$ where $\Phi_{0i} = V_i^{-1}$, $i = 1, \dots, I$ and $\Phi_l = W_l^{-1}$, $l = 1, \dots, k$, with $\Phi_{0i} \sim W(\nu_{0i}, S_{0i})$ and $\Phi_l \sim W(\varsigma_l, Z_l)$, $l = 1, \dots, k$, the full conditional distribution of Φ_l is given by

$$\begin{aligned} p(\Phi_l | \eta, \theta, \Phi_0, y) &\propto \left[\prod_{t=1}^N \prod_{m=1}^k |\Phi_m|^{1/2} \exp \left\{ -\frac{1}{2} (\theta_t - H_t \theta_{t-1})^T \Phi (\theta_t - H_t \theta_{t-1}) \right\} \right] \\ &\quad \times |\Phi_l|^{\varsigma_l - (p_l + 1)/2} \exp \{ -tr(Z_l \Phi_l) \} \\ &\propto |\Phi_l|^{N/2 + \varsigma_l - (p_l + 1)/2} \exp \left\{ -tr \left(\frac{1}{2} \sum_{t=1}^N Z Z_{l,t} \Phi_l \right) - tr(Z_l \Phi_l) \right\} \\ &\propto |\Phi_l|^{N/2 + \varsigma_l - (p_l + 1)/2} \exp \left\{ -tr \left(\left(\frac{1}{2} Z Z_l + Z_l \right) \Phi_l \right) \right\}, \end{aligned} \quad (8)$$

with $ZZ_t = (\theta_t - H_t\theta_{t-1})(\theta_t - H_t\theta_{t-1})^T$ and $ZZ_l = \sum_{t=1}^N ZZ_{l,t}$. Thus,

$$(\Phi_l | \eta, \theta, \Phi_0, y) \sim \text{Wishart}(N/2 + \varsigma_l, \frac{1}{2}ZZ_l + Z_l), \quad l = 1, \dots, k.$$

The full conditional distribution of Φ_{i0} is given by

$$\begin{aligned} p(\Phi_{i0} | \eta, \theta, \Phi_0, y) &\propto \left[\prod_{t=1}^N N(\eta_{it}; F_{it}\theta_t, \Phi_{0i}^{-1}) \right] W(\Phi_{0i}; \nu_{0i}, S_{0i}) \\ &\propto \left[\prod_{t=1}^N |\Phi_{0i}|^{1/2} \exp \left\{ -\frac{1}{2}(\eta_{it} - F_{it}\theta_t)^T \Phi_{0i} (\eta_{it} - F_{it}\theta_t) \right\} \right] \\ &\quad \times |\Phi_{0i}|^{\nu_{0i} - (p_{0i} + 1)/2} \exp \{ -tr(S_{0i}\Phi_{0i}) \} \\ &\propto |\Phi_{0i}|^{N/2 + \nu_{0i} - (p_{0i} + 1)/2} \exp \left\{ -tr \left(\left(\frac{1}{2}SS_{\eta_i} + S_{0i} \right) \Phi_{0i} \right) \right\}, \quad (9) \end{aligned}$$

with $SS_{\eta_i} = (\eta_{it} - F_{it}\theta_t)(\eta_{it} - F_{it}\theta_t)^T$. Thus,

$$(\Phi_{0i} | \eta, \theta, \Phi, y) \sim \text{Wishart}(N/2 + \nu_{0i}, \frac{1}{2}SS_{\eta_i} + S_{0i}), \quad i = 1, \dots, I.$$

4 Application

In this section we set up the hierarchical beta model for a hypothetical case in which y_{it} represents a given rate or proportion at region i and time t , $i = 1, \dots, I$ and $t = 1, \dots, N$. We take the logit transformation of both μ_{it} and ϕ_{it} and, to η_{i1t} and η_{i2t} , we fit dynamic models considering, respectively, a second-order polynomial trend seasonal effects and a second-order polynomial trend effects. The formulation of the structural equations is given below:

$$\begin{aligned} \eta_{1it} &= \log \left(\frac{\mu_{it}}{1 - \mu_{it}} \right) = F_{i1t}\theta_t + v_{i1t}, \quad v_{i1t} \sim N(0, V_{i1}), \\ \eta_{2it} &= \log \left(\frac{\phi_{it}}{1 - \phi_{it}} \right) = F_{i2t}\theta_t + v_{i2t}, \quad v_{i2t} \sim N(0, V_{i2}), \end{aligned} \quad (10)$$

with $V_i = \text{diag}(V_{i1}, V_{i2})$.

Modeling η_{1it} : In equation (10) the term $F_{i1t}\theta_t$, on the right-hand side of η_{i1t} , is the linear predictor of the logit transformed expected value of the beta model for time t and region i . We use a second-order polynomial trend seasonal effects model with offset term in order to describe η_{i1t} , that is

$$\eta_{1it} = \beta_t + \lambda_{t0} + \gamma_{it} + v_{i1t}. \quad (11)$$

The DLM representation of the model for η_{1it} is

Second-order polynomial effects for the level with respect to μ_{it} :

$$\beta_t = \beta_{t-1} + \delta_{t-1} + w_{\beta_t} \quad (12)$$

$$\delta_t = \delta_{t-1} + w_{\delta_t}$$

Free-form Seasonal effects:

$$\lambda_{tr} = \lambda_{t-1,r+1} + w_{tr}, \quad r = 0, \dots, p-2 \quad (13)$$

$$\lambda_{t,p-1} = \lambda_{t-1,0} + w_{t,p-1}$$

First-order polynomial effects for the offset term:

$$\gamma_{it} = \gamma_{i,t-1} + w_{\gamma_{it}}, \quad (14)$$

where,

- β_t represents an underlying level at time t , with respect to $h_1(\mu_{it})$, that is common to the I series;
- δ_t is the incremental growth;
- λ_{t0} represents a seasonal effect that is common to the I series. We denote as p the size of the seasonal cycle.
- γ_{it} is an offset parameter representing deviations of the unemployment rate of region i at time t with respect to the average β_t ;
- v_{i1t} represents the region i series-specific stochastic deviation.

Modeling η_{2it} : In equation (10) the term $F_{i2t}\theta_t$, on the right-hand side of η_{2it} , is the linear predictor of the logit transformed term related to the precision of the beta model for time t and region i . We use a second-order polynomial effects model with offset term in order to describe η_{2it} , that is

$$\eta_{2it} = \psi_t + \alpha_{it} + v_{i2t}. \quad (15)$$

The DLM representation of the model for η_{2it} is

Second-order polynomial effects for the level with respect to ϕ_{it} :

$$\psi_t = \psi_{t-1} + \xi_{t-1} + w_{\psi_t} \quad (16)$$

$$\xi_t = \xi_{t-1} + w_{\xi_t}$$

First-order polynomial effects for the offset term:

$$\alpha_{it} = \alpha_{i,t-1} + w_{\alpha_{it}}, \quad (17)$$

where

- ψ_t represents an underlying level at time t , with respect to $h_2(\phi_{it})$, that is common to the I series;
- ξ_t is the incremental growth;
- α_{it} is an offset parameter representing deviations of the unemployment rate of region i at time t with respect to the average ψ_t ;
- v_{i2t} represents the region i series-specific stochastic deviation.

Identifiability restrictions:

$$\lambda_{t,p-1} = -\sum_{r=0}^{p-2} \lambda_{tr}, \quad \gamma_{It} = -\sum_{i=1}^{I-1} \gamma_{it}, \quad \alpha_{It} = -\sum_{i=1}^{I-1} \alpha_{it}.$$

In order to exemplify the construction of the model, we consider $I = 3$ regions where the rates are measured over time. Thus, the vector $(\eta_{1it}, \eta_{2it})'$ is described by

$$\begin{pmatrix} \eta_{1it} \\ \eta_{2it} \end{pmatrix} = \begin{pmatrix} \beta_t + \lambda_{t0} + \gamma_{it} \\ \psi_t + \alpha_{it} \end{pmatrix} + \begin{pmatrix} v_{i1t} \\ v_{i2t} \end{pmatrix}, \quad i = 1, 2, 3.$$

That is,

$$\eta_{it} = F_{it}\theta_t + v_{it}, \quad i = 1, 2, 3,$$

where $\gamma_{3t} = -(\gamma_{1t} + \gamma_{2t})$, $\alpha_{3t} = -(\alpha_{1t} + \alpha_{2t})$. If, for example, we are dealing with seasonal cycles of size $p = 4$ (quarters), then $\lambda_{t3} = -(\lambda_{t0} + \lambda_{t1} + \lambda_{t2})$.

Considering a cycle of generic size p , the vector θ_t is represented by

$$\theta_t = (\beta_t, \delta_t, \lambda_{t0}, \lambda_{t1}, \dots, \lambda_{t,p-2}, \psi_t, \xi_t, \gamma_{1t}, \gamma_{2t}, \alpha_{1t}, \alpha_{2t}).$$

Consider the following matrices:

$$J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad P = \begin{pmatrix} -\mathbf{1}'_{p-2} & -1 \\ \mathbf{I}_{p-2} & \mathbf{0} \end{pmatrix}.$$

The matrices J and P are essential in the description of our dynamic model. Suppose a DLM such that the observation equation is $y_t = \beta_t + \epsilon_t$ and the system equation is given by the pair of equations in expression (12). Such model is called a *linear growth model* and it includes a time-varying slope in the dynamics of β_t . If we define $\theta_t = (\beta_t, \delta_t)'$ and $F = (1, 0)'$, then the observation equation can be represented by $y_t = F'\theta_t + \epsilon_t$, while the system equation, by $\theta_t = J\theta_{t-1} + (w_{\beta_t}, w_{\delta_t})'$.

The matrix J allows us to write a *linear growth model* such The *permutation matrix* P is *p-cyclic*, so that $P^{np} = I_p$ and $P^{h+np} = P^h$, for $h = 1, \dots, p$, and any integer $n \geq 0$. For example, suppose, for simplicity, a DLM model, with $y_t = F\theta_t + \epsilon_t$ describing the observation equation and $\theta_t = \theta_{t-1} + w_t$, the system equation. Additionally, suppose a purely seasonal series and that we have quarterly data $y_t, t = 1, 2, \dots$, so that if y_{t-1} refers to the first quarter of the year, y_t refers to the second one.

Because of the restriction $\sum_{i=1}^4 \alpha_i = 0$, the series might be described by seasonal deviations from the zero. Thus assume that $y_{t-1} = \alpha_1 + \epsilon_{t-1}$, $y_t = \alpha_2 + \epsilon_t$,

and so on, so that to $(y_{t-1}, y_t, y_{t+1}, y_{t+2}, y_{t+3}, y_{t+4}, y_{t+5}, y_{t+6})$ are associated the respective seasonal deviations from zero, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$. Consider now that $\theta_{t-1} = (\alpha_1, \alpha_4, \alpha_3, \alpha_2)$ and that $F' = (1, 0, 0, 0)$. Then, the successive application of matrix P makes possible the establishment of the desired quarterly seasonal pattern.

For the models that were formulated to η_{1it} and η_{2it} , the design matrix F_t (see expression (3)), given by $F_t = F = (F'_1, F'_2, F'_3)'$, and the state evolution matrix $H_t = H$ (see expression (4)), are shown below.

$$H = \begin{pmatrix} J_{2 \times 2} & \mathbf{0}_{2 \times (p-1)} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{(p-1) \times 2} & P_{(p-1) \times (p-1)} & \mathbf{0}_{(p-1) \times 2} & \mathbf{0}_{(p-1) \times 2} & \mathbf{0}_{(p-1) \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times (p-1)} & J_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times (p-1)} & \mathbf{0}_{2 \times 2} & \mathbf{I}_2 & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times (p-1)} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{I}_2 \end{pmatrix}$$

$$F_1 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1 \times (p-2)} & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}_{1 \times (p-2)} & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad F_2 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1 \times (p-2)} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}_{1 \times (p-2)} & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{and}$$

$$F_3 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1 \times (p-2)} & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}_{1 \times (p-2)} & 1 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}.$$

4.1 Estimated proportions and forecasting

The estimated proportions are calculated using the following procedure:

- (1) The inverse transformations $\mu_{it} = \frac{\exp(\eta_{1it})}{1 + \exp(\eta_{1it})}$ and $\phi_{it} = \frac{\exp(\eta_{2it})}{1 + \exp(\eta_{2it})}$ are evaluated at the estimated values (posterior means) of η_{1it} and η_{2it} , for $i = 1, \dots, I$ and $t = \dots, N$.
- (2) For $i = 1, \dots, I$ and $t = \dots, N$ we simulate n (say, $n = 1,000$) samples from a beta distribution $\text{Beta}\left(\mu_{it} \left(\frac{1-\phi_{it}}{\phi_{it}}\right), (1 - \mu_{it}) \left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right)$ and then we take the average value of those draws.
- (3) For the confidence bands we repeat the steps (1) and (2) for the 2.5% and 97.5% percentiles of the posterior distribution of η_{it} .

The k -step-ahead forecasting are calculated using the following considerations. For $0 \leq j < k$, at each t the forecast distribution $(y_{t+k} | D_t)$ is described next.

From the system equation, $\theta_t = H_t \theta_{t-1} + w_t$; $w_t \sim (0, W_t)$. Thus, from repeated application of the system equation,

$$\theta_{t+k} = H H_{t+k}(k) \theta_t + \sum_{r=1}^k H H_{t+k}(k-r) w_{t+r},$$

where $HH_{t+k}(r) = H_{t+k}H_{t+k-1} \times \cdots \times H_{t+k-r+1}$ for all t and integer $r \leq k$, with $HH_{t+k}(0) = I$. Thus, by linearity and independence and also taking into account the Bayesian linear estimation method,

$$\theta_{t+k} \sim (a_t(k), R_t(k)),$$

with $a_t(k) = H_{t+k}a_t(k-1)$ and $R_t(k) = H_{t+k}R_{t-k}H'_{t+k} + W_{t+k}$, and $a_t(0) = m_t$ and $R_t(0) = C_t$. Thus the “future” θ_t values are then obtained by successively sampling from the system equation (see expression (3)), followed by the structural equation (see expression (4)). The forecast rates are then obtained by following steps (1) to (3) given above.

5 Application to simulated data

We applied the model described in Section 4 to simulated data in which we considered $N = 72$ time points (say, six years), $I = 3$ subpopulations and cycles of size $p = 4$. In order to obtain initial values for the MCMC procedure, we estimated the parameters involved by running separate DLM models (described by equations (3) and (4)) for each of the subpopulations. All the routines were written using the R language (<http://www.r-project.org/>). We also made extensive use of the excellent `dlm` R library by Petris (2010).

In such DLM setting the η'_{it} s have the same rule as the observed data. Thus, in order to run those initial models we estimated η_{1it} by $\log\left(\frac{y_{it}}{1-y_{it}}\right)$ and η_{2it} by $\log\left(\frac{\tilde{\sigma}_{it}^2}{1-\tilde{\sigma}_{it}^2}\right)$ with $\tilde{\sigma}_{it}^2 = \text{var}(y_i)/(y_{it}(1-y_{it}))$ (see properties of expression (1)).

For the simulated data we considered a hierarchical dynamic Beta model in which a second-order polynomial trend seasonal effects was fitted to the parameters related to the mean, μ_{it} , and a second-order polynomial effects was fitted to the parameters related to the precision, ϕ_{it} . We run chains of size 50,000 with burn-in period of 20,000. The autocorrelations could be significantly controlled by using gaps of size 30.

Figures 2 and 3 show the true values (in red) used in the simulations, the estimated values of the parameters involved in expressions (11) and (15), and the respective confidence bands for the main effects of level, growth and seasonality. Figures 4 and 5 show the individual sub-populations effects. Figure 6 shows the estimated proportions for each of the sub-populations and their corresponding confidence bands. As we can observe, except for the η_{23t} terms, all the effects and probabilities are, in general, well estimated.

6 Applying the hierarchical dynamic beta model to Brazilian unemployment rates

In this section we apply our methods to fit the three time series of Brazilian monthly unemployment rates that were described in Section 1. We analyze monthly unemployment rates (MUR) based on PME data in the period from March 2002 to April 2011 ($N=110$ observations). Forecast rates are also provided: we used MUR data for the months of May, June, July and August of 2011.

The three subpopulations involved in the analysis are Recife, São Paulo and Porto Alegre, i.e., $I = 3$. We considered a hierarchical dynamic Beta model in which a second-order polynomial trend seasonal effects (with cycles of $p = 12$ months) was fitted to the parameters related to the mean, μ_{it} , and a second-order polynomial effects was fitted to the parameters related to the precision, ϕ_{it} . We applied the same developments discussed in sections 4 and 5.

Figures 7 and 8 show the estimated values of the parameters involved in expressions (11) and (15), and the respective confidence bands for the main effects of level, growth and seasonality. Figures 9 and 10 show the individual sub-populations effects and Figure 11 shows the estimated proportions or rates for each of the sub-populations and their corresponding confidence bands. We also added the forecast rates (see Figure 11, after the dotted vertical lines). It is really reassuring how well the model is capable of describing the observed proportions for each of the subpopulations.

7 Discussion

In this article we propose an extension to the Bayesian beta dynamic model developed by Da-Silva *et al.* (2011). We develop a hierarchical dynamic Bayesian beta model for modeling a set of time series of rates or proportions. The proposed methodology enables to combine the information contained in different time series so that we can describe a common underlying system, which is though flexible enough to allow the incorporation of random deviations, related to the individual series, not only through time but also across series. That allows to fit the case in which the observed series may present some degree of level shift. Additionally, the proposed model is adaptive in the sense that it incorporates precision parameters that can be heterogeneous not only over time but also across the series. The use of two link functions, one for the *mean process* and another to the *dispersion process*, make such extension possible. Additionally, the choice of the matrices F_t and H_t allow for a multiplicity of ways of specifying the model, even allowing for the inclusion of covariates. The work of Cargnoni *et al.* (1997) was used in the development of our model.

Missing observations can be easily accommodated: if the observation at time t is missing, then $y_t = NA$ and y_t does not carry any information. Then, we set $p(\theta_t | D_t) = p(\theta_t | D_{t-1})$.

Our methodology was applied to both real and simulated data. The real data set used are three time series of Brazilian monthly unemployment rates, observed in the cities of Recife, São Paulo and Porto Alegre, in the period from March 2002 to April 2001. We used a second-order polynomial trend seasonal effects to the parameters related to the mean, μ_{it} , and a second-order polynomial effects to the parameters related to the precision, ϕ_{it} . The very good features of the proposed model can be appreciated when looking at the graphs presented. The new parametrization of the precision parameter that was proposed by Bayer (2011) was used in the model formulation. It is very convenient since both, the link function for μ_{it} and ϕ_{it} are expressed in the $(0, 1)$ interval, which gives us a more meaningful interpretation in terms of the magnitude of the scale.

For future work we envision the possibility of extending the current model to enable the inclusion of different regimes for the level of the processes of the mean, μ_{it} , and also of the precision, ϕ_{it} .

Acknowledgements

The authors are deeply indebted to the Office to Improve University Research (CAPES-Brazil), for financial support via Project PROCAD-NF 2008. Cibele Q. Da-Silva was supported by CNPq-Brazil, BPPesq. Helio S. Migon was supported by the National Research Council (CNPq-Brazil, BPPesq) and the Rio de Janeiro State Research Foundation (Faperj) together with CAPES via the Pronex project.

References

Bayer, F.B. (2011). Inferência e Modelagem em Regressão Beta. Doctoral Dissertation. Statistics Department, Federal University of Pernambuco, Recife, Brazil.

Cargnoni, C., Müller, P. and West, M. (1997). Bayesian Forecasting of Multinomial Time Series Through Conditionally Gaussian Dynamic Models. *Journal of the American Statistical Association*, **92**: 438, 640-647.

- Carter, C. K, and Kohn, R. (1994). On Gibbs Sampling for State-Space Models. *Biometrika*, **81**, 541-553.
- Cuervo-Cepeda, E., Gamerman, D. (2004). Bayesian modeling of joint regressions for the mean and covariance matrix. *Biometrical Journal*, **46**(4), 430-440.
- Da-Silva, C.Q. ; Migon, H.S. ; Correia, L. T. (2011). Dynamic Bayesian Beta Models. *Computational Statistics Data Analysis* (Print), **55**, 2074-2089.
- Espinheira, P.L. (2007). Regressão beta. Doctoral Dissertation. Statistics Department, Institute of Mathematics and Statistics - University of São Paulo (USP).
- Ferrari, S.L.P., Cribari-Neto, F., (2004). Beta regression for modeling rates and proportions. *Journal of Applied Statistics* **31**, 799-815.
- Früwirth-Schnatter, S. (1994). Data Augmentation and Dynamic Linear Models. *Journal of the Royal Statistical Society*, Ser. B, **55**, 629-642.
- Gamerman, D. and Migon, H. (1993). Dynamic hierarchical models, *Journal of the Royal Statistical Society*, Series B **55**: 629-642.
- Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57**, 97-109.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A., Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemistry and Physics*, **21**, 1087-1091.
- Nelder, J.A., and Lee, Y. (1991). Generalized linear models for the analysis of Taguchi-type experiments. *Appl. Stochastic Models Data Anal.*, **7**, 107-120.
- Petris, G. (2010). An R Package for Dynamic Linear Models. *Journal of Statistical Software*, **36**(12), 1-16. URL <http://www.jstatsoft.org/v36/i12/>.
- Pregibon, D. (1984). Review of generalized linear methods by McCullagh and Nelder. *Ann. Statist.*, **12**, 1589-1596.
- Rocha, A.V., Cribari-Neto, F. (2009). Beta autoregressive moving average models. *Test*, **18**(3), 529-545.
- Simas, A.B., Barreto-Souza, W. and Rocha, A.V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis* (Print), **54**, 348-366.

Smithson, M., Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, **11**(1), 5471.

Smyth G.K., Verbyla, A.P. (1999). Adjusted likelihood methods for modeling dispersion in generalized linear models. *Environmetrics*, **10**(6), 695709.

West, M. and Harrison, P.J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd edition, Springer, New York.

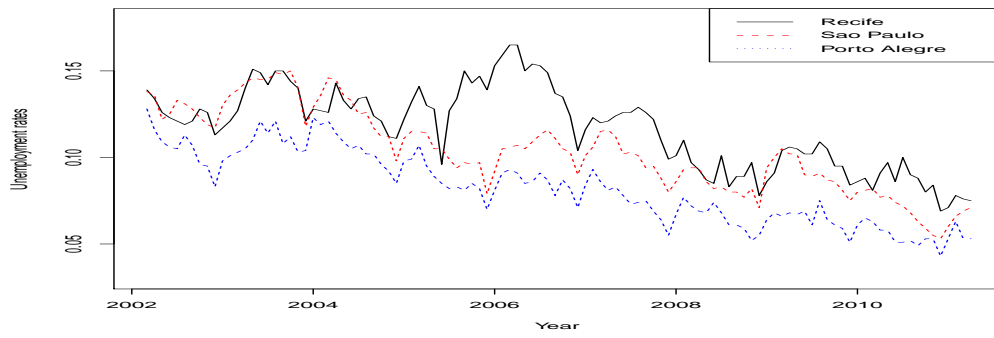
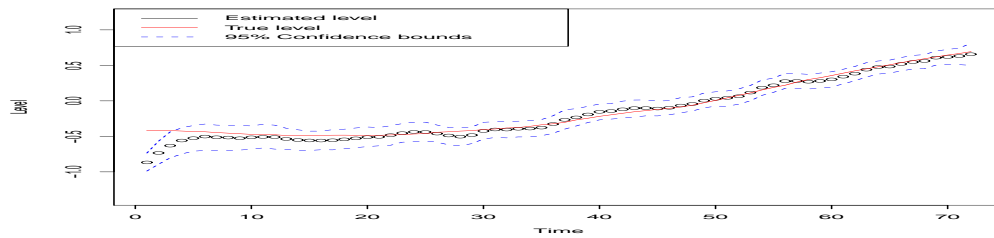
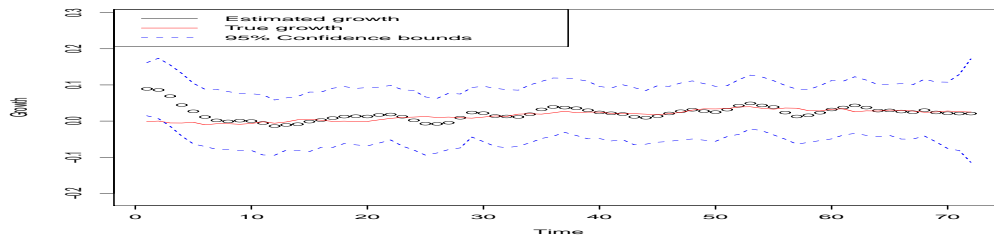


Fig. 1. Observed unemployment rates in the cities of Recife, São Paulo and Porto Alegre - Brazil.

(a)



(b)



(c)

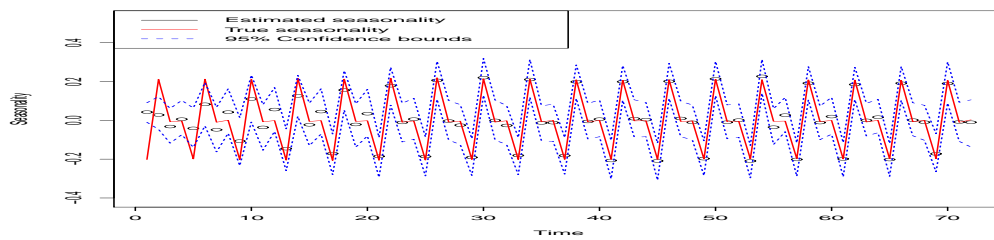
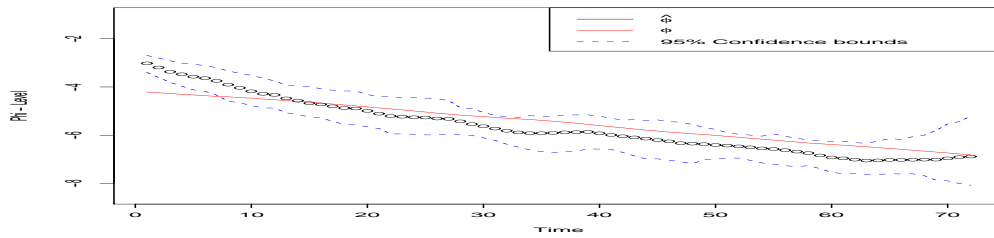


Fig. 2. Simulated data - estimated values and 95% credibility bounds for the components of η_{1it} : (a) Level (β_t), (b) Growth (δ_t), (c) Seasonality (λ_t).

(a)



(b)

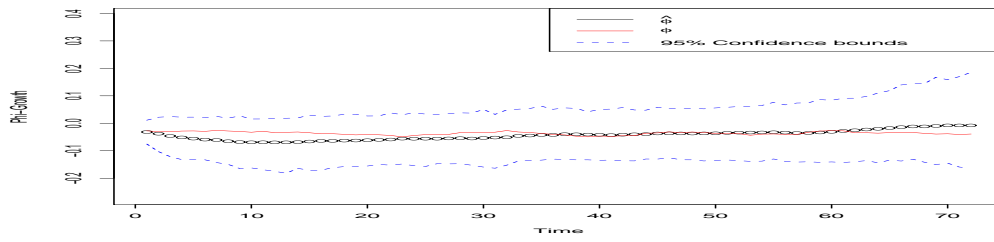
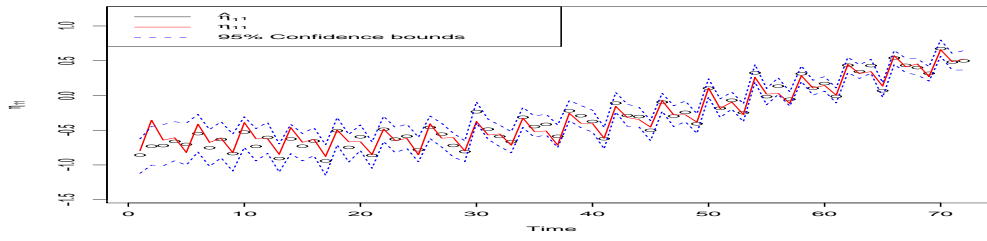
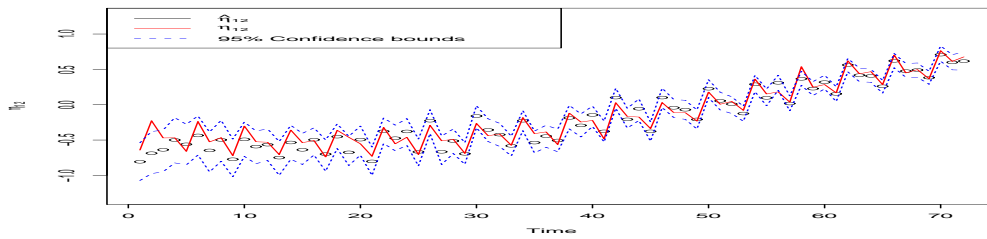


Fig. 3. Simulated data - estimated values and 95% credibility bounds for the components of η_{2it} : (a) Level (ψ_t) and (b) Growth (ξ_t).

(a)



(b)



(c)

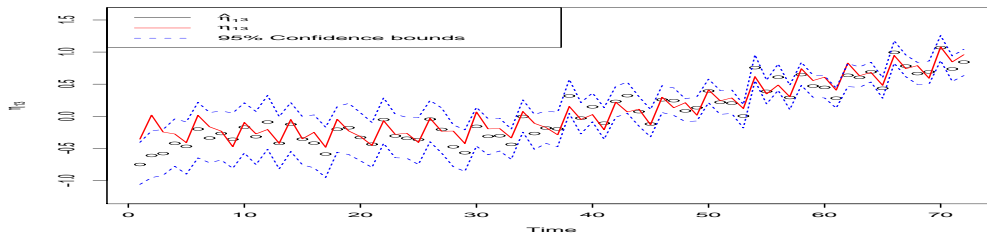
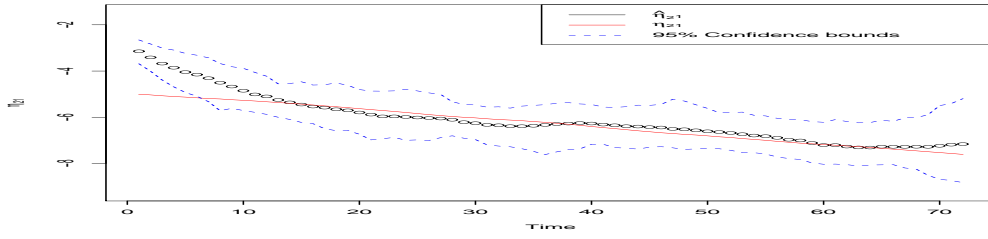
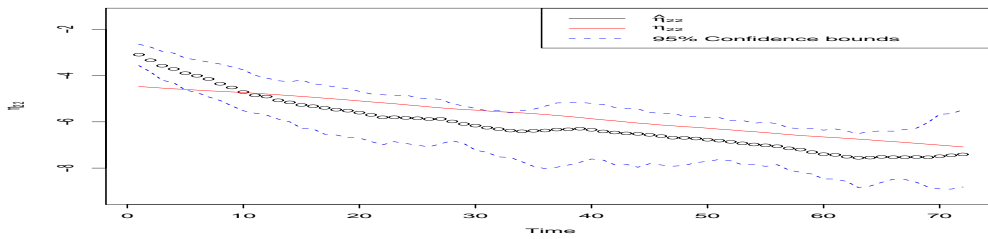


Fig. 4. Simulated data - estimated values and 95% credibility bounds for (a) η_{11t} , (b) η_{12t} , (c) η_{13t} .

(a)



(b)



(c)

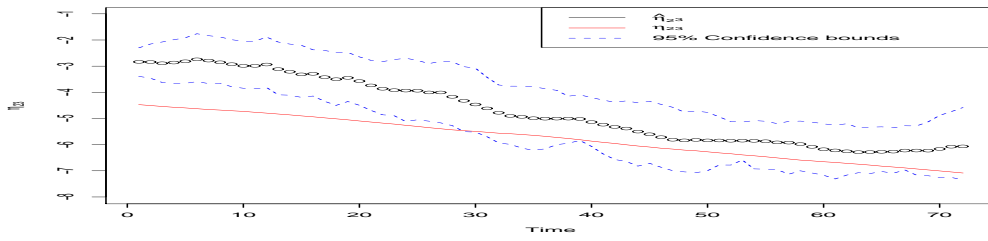
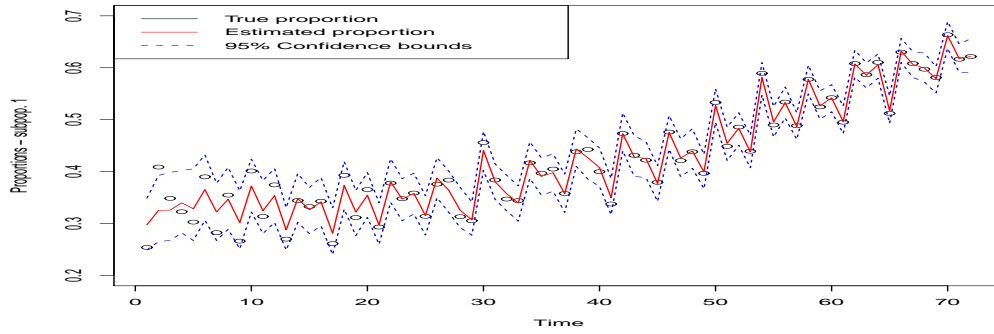
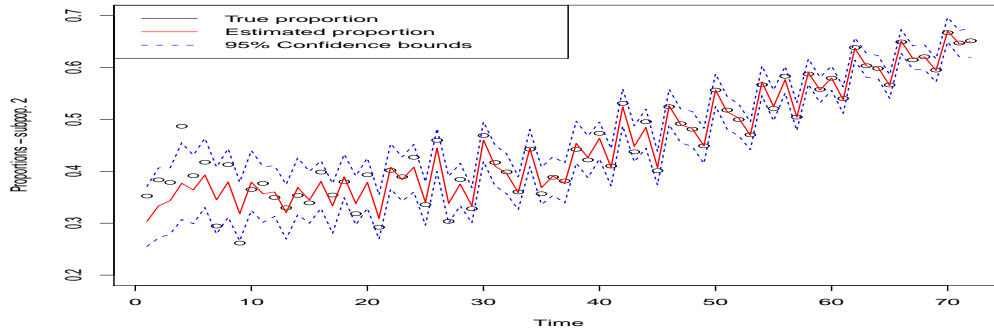


Fig. 5. Simulated data - estimated values and 95% credibility bounds for (a) η_{21t} , (b) η_{22t} and (c) η_{23t} .

(a)



(b)



(c)

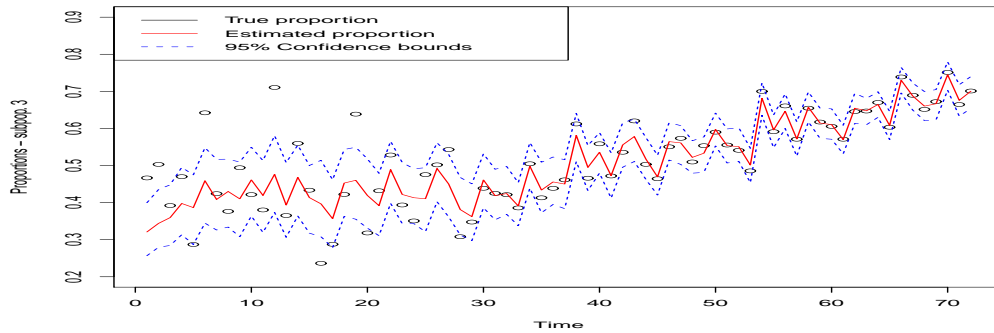


Fig. 6. Simulated data - estimated proportions and 95% credibility bounds for the three sub-populations.

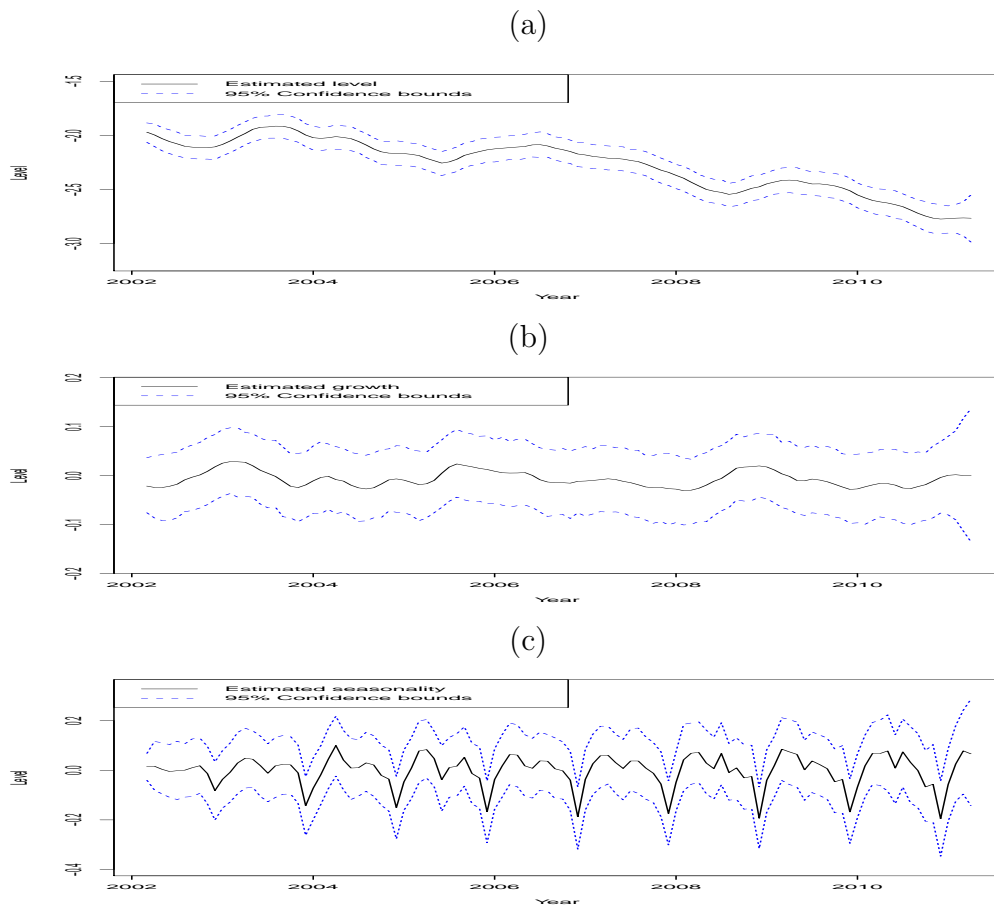
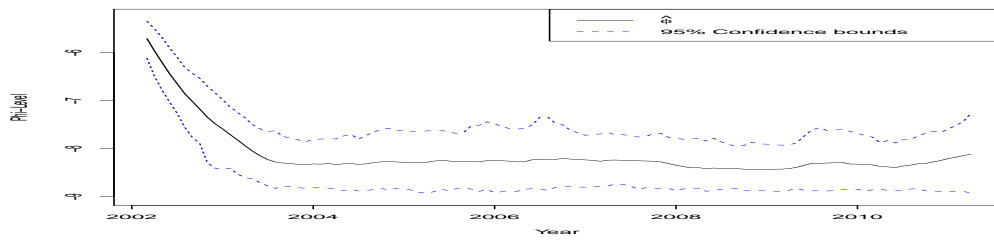


Fig. 7. MUR data - estimated values and 95% credibility bounds for the components of η_{1it} : (a) Level (β_t), (b) Growth (δ_t), (c) Seasonality (λ_t).

(a)



(b)

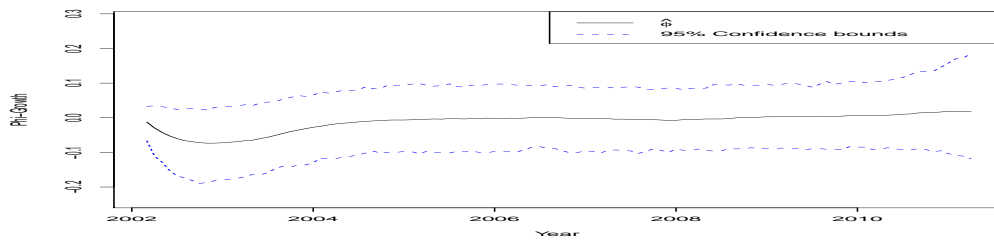


Fig. 8. MUR data - estimated values and 95% credibility bounds for the components of η_{2it} : (a) Level (ψ_t) and (b) Growth (ξ_t).

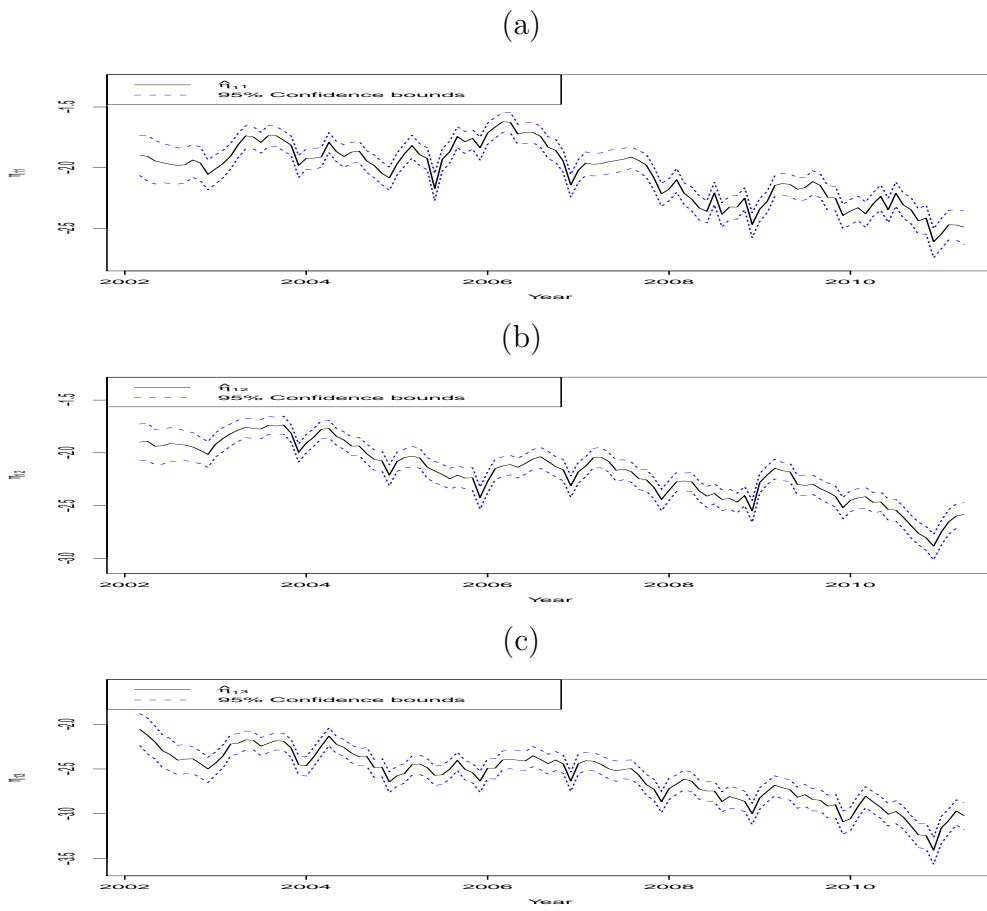
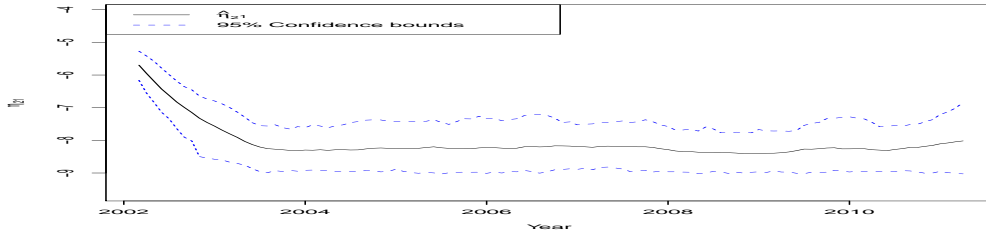
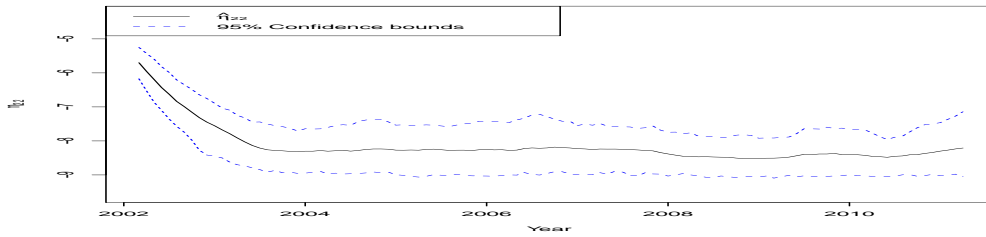


Fig. 9. MUR data - estimated values and 95% credibility bounds for (a) η_{11t} , (b) η_{12t} , (c) η_{13t} .

(a)



(b)



(c)

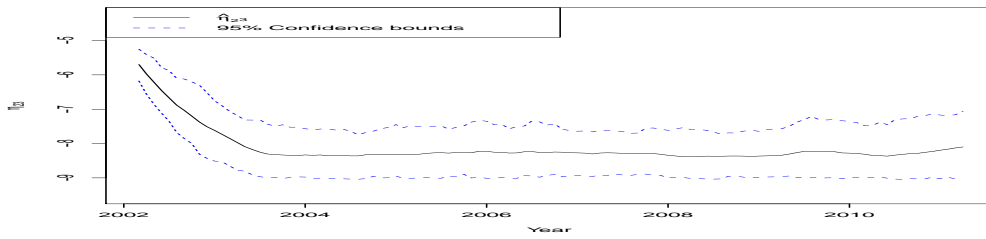
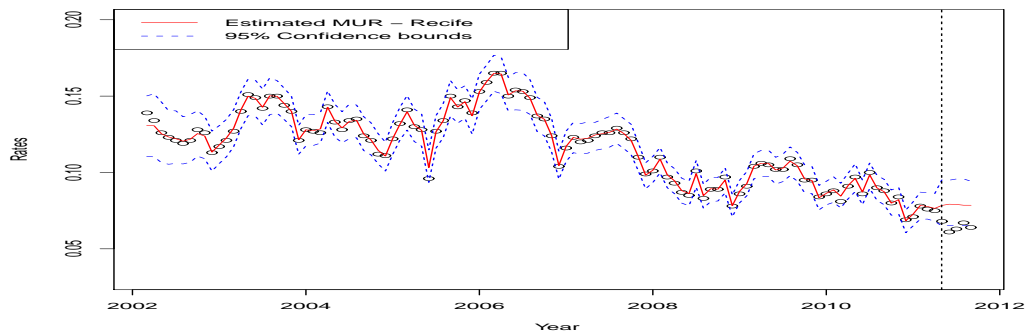
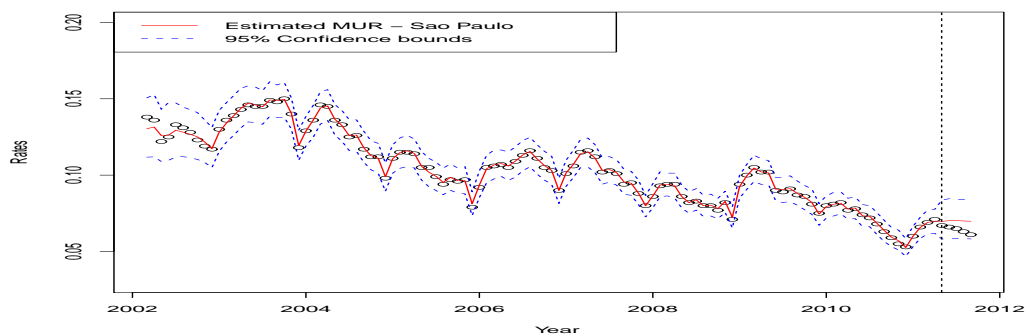


Fig. 10. MUR data - estimated values and 95% credibility bounds for (a) η_{21t} , (b) η_{22t} and (c) η_{23t} .

(a)



(b)



(c)

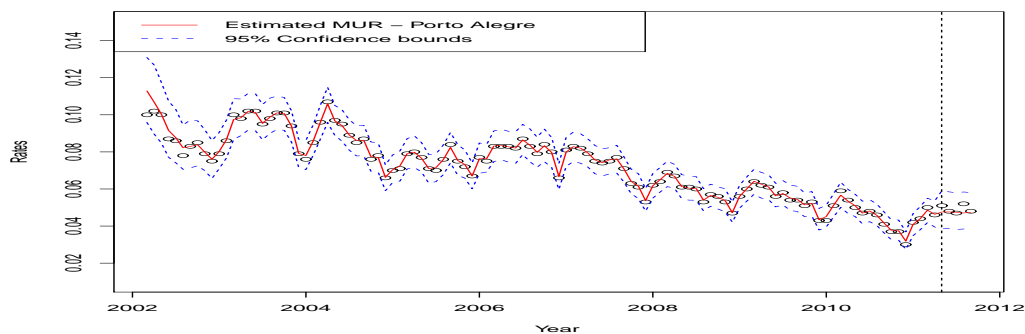


Fig. 11. MUR data - estimated proportions and 95% credibility bounds for the three sub-populations: (a) Recife, (b) São Paulo and (c) Porto Alegre. The forecast rates are presented after the dotted vertical lines.